THE UNIVERSITY OF SHEFFIELD

PhD Mechanical Engineering

Transfer learning for population-based structural health monitoring

by

Jack Poole

May 2025

Keith Worden, Nikolaos Dervilis

Thesis submitted to the School of Mechanical, Aerospace and Civil Engineering, University of Sheffield for the degree of Doctor of Philosophy

Acknowledgements

First and foremost, I would like to thank my supervisors. Thank you Prof. Keith Worden for teaching me so much and providing valuable guidance throughout the project. I would also like to thank Prof. Nikoloas Dervilis, who has kept me on track, made sure I always think about telling a story, and often provided an alternative opinion to Keith in supervisor meetings. You make a great supervisor team, and your constant support has made the PhD enjoyable and as stress-free as a PhD can be. Also, a special thanks to Dr Paul Gardner and Dr Aidan J. Hughes for all the long meetings to help me develop the ideas in this thesis. I would also like to give a special thanks to Robin Mills for the support in designing the experiments in this thesis.

I feel lucky to have been surrounded by such a friendly group of open-minded people in the last few years; thank you to all of the DRG! I've learnt so much from you all. To the members of RC02a – Tristan, Chris, and Josie – thank you for putting up with the interruptions. You've always helped me develop my ideas, often as they sprang into my head! I would also like to thank Valentina Giglioni for her help with the design and data collection for modular bridge experiments in this thesis. I also can't forget the wider team on ROSEHIPS (and adjacent ROSEHIPS members), it's been a pleasure working with you all!

I have also been surrounded by an incredibly supportive network of family and friends – thank you all. I would particularly like to thank my parents for helping me stay on track with my education and for the constant support throughout my time at university. I also want to give a special thanks to Sarah K for helping me improve my writing over the years. Also, thanks to Falcone, Sam L, Matt T and Fiona for their support during the stressful times.

Most of all, I need to thank Laura for regularly dealing with me saying "I'll make dinner in 5", and then continuing to code for 1-3 hours, and of course, for her constant support. Now you are at the Max Plank Institute, do you have better access to rats to grow that extra ear and perhaps some wings (obviously, with no rats harmed)? If not, I'll have to put out a call to the wider scientific community, so please let me know.

I couldn't have done this without you guys!

This thesis is dedicated to Mary Windmill.

Abstract

Structural health monitoring (SHM) systems aim to proactively identify damage and provide diagnostic information to support maintenance decisions in mechanical, aerospace, and civil infrastructure. A critical challenge for the application of SHM systems – particularly those that provide contextual information – is the feasibility and cost of acquiring comprehensive data. Population-based SHM (PBSHM) presents a potential solution by leveraging data from related structures. However, differences between structures often prevent conventional machine learning models from generalising across domains. This issue motivates the use of transfer learning, which seeks to improve predictive performance in a target domain using data from a related source domain.

In PBSHM, target structures will often only have data for a limited range of health states. Therefore, to enable transfer when target labels are sparse, this thesis presents novel statistic alignment (SA) methods that require only undamaged target data. These methods are shown to facilitate the generalisation of models learnt using only labelled source data.

Quantifying similarity between structures and their features is essential to ensure that transfer learning will yield positive results. This thesis investigates using physics knowledge to address limitations with data-based similarity measures in sparse-data scenarios. This approach is incorporated into a feature-selection criterion to identify transferable, damage-sensitive features. Subsequently, it is used within a regression framework to predict the quality of predictions when transferring between a specific source/target pair, supporting decisions about when transfer is appropriate.

Previous work has not considered how to incorporate transfer learning into an online framework that updates as labels are collected during a monitoring campaign. Thus, a Bayesian model is proposed that uses the SA methods to define mappings early in the monitoring campaign and updates sequentially as labels are obtained. This model is integrated into an active-sampling strategy that guides inspections by selecting the most informative observations to label.

Contents

Publications

Journal Publications

J. Poole, K. Worden, N. Dervilis, V. Giglioni, R. S. Mills, P. Gardner, A. J. Hughes, Active transfer learning for structural health monitoring, Submitted to Mechanical Systems and Signal Processing

J. Poole, P. Gardner, A. J. Hughes, N. Dervilis, T. Dardeno, R. S. Mills, Physics-informed transfer learning for SHM via feature selection, *Mechanical Systems and Signal Processing*, vol. 237, Art. 113013, 2025

J. Poole, P. Gardner, N. Dervilis, L. Bull, K. Worden, On statistic alignment for domain adaptation in structural health monitoring, *Structural Health Monitoring*, pp. 1581–1600, 2022.

Conference Publications

J. Poole, A. J. Hughes, N. Dervilis, P. Gardner, L. A. Bull, V. Giglioni, R. S. Mills, K. Worden. Active transfer learning for SHM of bridges under changing environmental conditions, In *Proceedings of the European Workshop on Structural Health Monitoring (EWSHM)*, 2024.

J. Poole, P. Gardner, N. Dervilis, J. H. Mclean, T. J. Rogers, K. Worden, Towards physics-based metrics for transfer learning in dynamics, *Society for Experimental Mechanics Annual Conference and Exposition*, 2023.

J. Poole, P. Gardner, A. J. Hughes, R. S. Mills, T. A. Dardeno, N. Dervilis, Physics-informed transfer learning in PBSHM: A case study on experimental helicopter blades, *Structural Health Monitoring*, 2023.

J. Poole, P. Gardner, N. Dervilis, L. Bull, K. Worden, On the application of partial domain adaptation for PBSHM, *Proceedings of the European Workshop on Structural Health Monitoring (EWSHM)*, 2022.

J. Poole, P. Gardner, N. Dervilis, L. Bull, K. Worden, On normalisation for domain adaptation in population-based structural health monitoring, *Structural Health Monitoring*, 2021.

Glossary

SHM - structural health monitoring
PBSHM - population-based structural health monitoring
SA - statistic alignment
MAC - modal assurance criterion
DA - domain adaptation
GP - Gaussian process
EoVs - environmental and operational variations
AG - attributed graph
IE model - irreducible element model
DA - domain adaptation
MMD - maximum mean discrepancy
KL divergence - Kullback-Leibler divergence
DNN - deep neural network
PCA - principal component analysis
TCA - transfer component analysis
JDA - joint-distribution adaptation
BDA - balanced distribution adaptation
PAD - proxy-A distance
RKHS - Reproducing Kernel Hilbert Space
JMMD - joint maximum mean discrepancy
GAN - generative adversarial network
DANN - domain adversarial neural network
CDAN - conditional domain adversarial neural network
GMM - Gaussian mixture model
ARTL - adaptation regularization transfer learning
CORAL - correlation alignment
A-standardisation - independent standardisation of the source and target domains
N-standardisation - standardisation by pooling source and target data
NCA - normal condition alignment

NCORAL - normal correlation alignment

GFK - geodesic flow kernel

kNN - k-nearest neigbours

TFC - transfer feature criterion

KBTL - kernelised Bayesian transfer learning

MES - maximum-entropy sampling

RVM - relevance vector machine

DA-RVM - domain adaptation relevance vector machine

SVM - support vector machine

OMA - output-only modal analysis

SSI - stochastic-subspace identification

Chapter 1

Introduction

Structural health monitoring (SHM) is a field of engineering that aims to provide diagnostic information about damage in real-time for mechanical, aerospace, and civil infrastructure [1]. SHM has the potential to automate the inspection process, reducing maintenance costs. In addition, it could improve diagnostic capabilities, both in terms of timely detection and sensitivity to damage, increasing the safety of structures and potentially de-risking design. By providing a rationale for less conservative designs, SHM may also help reduce redundancies in structures, ultimately lowering costs and CO_2 emissions from construction.

These systems typically derive insights from measured sensor data via physics-based or data-based approaches. Data-based approaches have received significant attention in recent years [1], as these methods can be used to learn relationships between measurable sensor data and damage in scenarios where physical understanding is insufficient to accurately model complex damage mechanisms. However, obtaining data to learn data-based models can be prohibitively expensive or infeasible for expensive and/or safety-critical assets. This issue becomes particularly restrictive for SHM systems capable of providing more in-depth diagnostics about the location, type and severity of damage, as these systems require contextual information relating to a range of different damages and environmental conditions. Thus, developing methods to reduce the data requirement for training models is a critical area of research in SHM [1–4].

The recent emergence of the field of population-based SHM (PBSHM) presents a potential solution to the issue of data scarcity by developing a framework for leveraging data from across a population of related structures [5–7]. However, data from different structures will have discrepancies, meaning that naively applying standard data-based models learnt using data from one structure to another structure would result in poor

performance, and potentially misinform asset managers. Thus, this thesis aims to develop transfer-learning methods for PBSHM, which aim to account for discrepancies in data acquired from different structures, allowing information-rich datasets obtained from one structure be used to train models that can make predictions about incipient damage in structures with limited data.

1.1 Structural health monitoring

In the context of SHM, damage refers to changes in the material properties or geometry of a structure that compromises its performance or safety, either currently or in the future [1]. SHM systems aim to provide insight into emerging damage in structures of interest. Therefore, SHM systems can reduce maintenance costs by alerting operators to structural changes and incipient damage, indicating when, and potentially to some extent what, maintenance is necessary. Additionally, these systems can help extend the design life of structures via more efficient maintenance and informing operators about structural degradation. These systems also have a pivotal role in safety-critical applications, where real-time damage diagnostics can help inform emergency interventions.

The effort required to develop an SHM system depends on the diagnostic information it is capable of providing. The level of information a SHM system could provide can be defined as a hierarchy, as proposed by Rytter [8]. This framework is presented below [8]:

1. Detection – is damage present?
2. Localisation – what is the location of damage?
3. Classification – what type of damage is it?
4. Quantification – what is the extent of the damage?
5. Prognosis – what are the outcomes of the damage and what is the remaining useful life of the structure?

It is generally considered that each proceeding level of the hierarchy would require additional information. Therefore, as the usefulness to decision support improves, the efforts to derive this information also increase. In addition, prognosis is distinguished from the other levels as it is generally considered to be only achievable with physics knowledge [1]. Consequently, the design of an SHM system should carefully consider the utility of each level with respect to safety and/or economical benefits.

1.2 SHM approaches

Damage is impossible to measure directly. Therefore, SHM systems rely on models to extract damage information from indirect measurements. Generally, these models can be classified as either physics-based or data-based models [1]. Physics-based models, often referred to as white-box models, seek to use understood laws of physics to describe the observed behaviours of the system. Measured data are then used to update model parameters and validate the model. On the other hand, data-based methods (also called black-box models), are derived by learning relationships between previously measured data and damage-states of interest.

Each of these models has its respective advantages. Physics-based models can be used to generate a range of operational and environmental conditions, and generally offer the benefit of interpretability as they are based on well-understood laws of physics [9]. However, they can be computationally intensive and may require detailed knowledge of the system, and the most prominent damage mechanisms, which is not always available or easily obtainable. Additionally, they must be regularly validated to ensure they sufficiently reflect the asset of interest [1, 10]. Data-based models, in contrast, can model complex behaviours without constructing a model from first physical principles. However, these methods require large datasets that capture the range of physical behaviours of interest. In addition, they also often lack the level of interpretability provided by physics-based models. Fundamentally, data availability can limit both approaches, as physics-based models require data for validation, while data-based models require data for both training and validation.

Models used for SHM, and engineering in general, could leverage both approaches, as valuable insights can be obtained from both methodologies. Consequently, models that incorporate physics into data-based models (grey-box models) [11] and methods that combine both simulated and measured data [12] are active areas of research. However, these hybrid approaches may still require expensive-varied datasets to develop models informative of all health-states of interest; thus, this thesis aims to reduce data requirements by developing data-based models capable of sharing information between datasets collected from different structures. The following section provides a brief overview of the data-based approach to SHM to motivate the application of transfer learning.

1.2.1 Data-based SHM

The data-driven approach to SHM can be considered as a pattern-recognition problem, where a statistical model is tuned using a set of training data previously obtained from

the structure and is used to make predictions about new in-service data. These methods typically require large quantities of measurements, and to provide more in-depth diagnostics, they also need contextual information, often encoded in the form of labels that describe the state of the structure. The process of building a data-based SHM system can be summarised in the following steps:

1. **Sensing and data acquisition** – damage-sensitive quantities are measured via sensors permanently installed on the structure. These quantities may be related to measurements directly from the structure, such as strain, accelerations, displacements, or correspond to environmental or operational conditions (EoVs), such as temperature, traffic loading, wave height, wind speed, or wind direction.

2. **Data-processing and feature extraction** – raw sensor data should be processed to extract a set of quantities that are sensitive to damage, known as damage-sensitive features. Standard procedures include domain transformation, as well as data fusion, filtering to reduce noise, and normalisation to remove confounding effects. Additionally, it may involve dimensionality reduction, as well as feature scaling and selection, to enhance the effectiveness of subsequent modelling.

3. **Machine Learning** – an appropriate statistical model is defined and parameters are tuned using historic data. This model can then be used to make predictions about future data.

4. **Decision** – predictions on data acquired online are generated using the machine learning model, which are then used in a decision process to inform potential inspection and/or interventions.

Several significant research challenges remain for data-based SHM; these challenges could be categorised as the following three research areas. Firstly, the acquisition of data and the extraction of damage-sensitive features is currently a bespoke procedure, requiring specialist knowledge to design sensor systems and process data. Moreover, often extraction of damage-sensitive features requires manual processes, such as modal analysis. Furthermore, obtaining a set of features sensitive to a variety of damage locations, types, and extents remains a challenging problem.

Secondly, confounding influences, such as varying EoVs, often mask changes in data relating to damage; therefore, these must be taken into account to ensure damage can be effectively identified, which is often achieved via methods that attempt to remove these effects [1].

Third, data-based models require large and varied historic datasets to train machine-learning models. This issue particularly impedes the application of models that provide in-depth diagnostics relating to damage location, type, and extent. These methods generally require *supervised* learning methods, which need observations from all damage-states and corresponding contextual information encoding damage information (i.e. location, extent, type). There are several methods for obtaining suitable datasets for training SHM systems; three examples are summarised as follows:

1. Data corresponding to damage could be obtained by directly damaging the structure; however, this approach is clearly prohibitively expensive in most cases, and often raises safety concerns.

2. Damage data could be acquired as it naturally occurs during the monitoring campaign, and corresponding contextual information could be obtained via inspections. This approach is more feasible, although it limits the capability of SHM systems to making predictions about previously observed damage-states and still requires operators to repeat labelling efforts for each structure, which is itself requires costly inspections. Thus, contextual information can typically only be obtained for a subset of all measurements because of budget constraints.

3. Surrogate data sources, such as data generated from a finite element (FE) model [13], could be used to train an initial model. While this option has the potential to generate a wide range of training data at a low cost, any surrogate data source will inevitably differ from the structure of interest, which invalidates a fundamental assumption made by conventional machine-learning methods – that training data follow the same distribution as the testing data.

The cost and limited availability of data have driven significant research efforts to reduce the data requirements for training machine-learning models in SHM. A few notable approaches include incorporating information from unlabelled data (partially-supervised learning) [14] and physics (physics-informed machine learning) [11]. While these approaches have proven highly effective in enhancing data-based models with sparse measurements, they typically still require representative labelled data to make prediction about a specific health-state of interest.

1.2.2 Population-based SHM

The emerging field of population-based SHM (PBSHM), presents a potential solution to the issue of data sparsity by leveraging data from across a group, or *population*, of

related structures, consisting of multiple related structures [6, 15, 16]. In this paradigm, when informative health-state data are obtained for one structure, they could be used to update predictive models across the population, extending the value of these data. In addition, data from historical monitoring campaigns or datasets generated using FE models containing information relating to various health-states could be leveraged to learn models capable of providing diagnostic information about damage location, type, and extent when insufficient data are available in the target structure to learn such models. Thus, PBSHM presents a framework for reducing costs and improving the diagnostic capabilities of SHM systems.

There are numerous examples of populations of structures where PBSHM could transform asset management. For instance, the UK government anticipates £50 billion in investment in offshore wind by 2030 [17], driven by the urgency of the climate crisis and the need for sustainable energy generation. The operation and maintenance of these wind farms pose significant challenges because of the sheer number and size of these structures, as well as issues related to accessibility. Nevertheless, wind farms typically consist of many nominally-identical structures, offering the potential for substantial data to facilitate PBSHM. Similarly, in 2020, National Highways, the UK's main highways agency, managed 9,392 bridges [18], highlighting another context where PBSHM could enhance the management of large populations of structures. Monitoring of bridges is conventionally carried out via periodic visual inspections, presenting a significant cost. Furthermore, even following visual inspections, damage may remain undetectable; for example, the Malahide Viaduct collapse in 2008 occurred three days after an inspection [19].

1.2.2.1 The role of transfer learning in PBSHM

The damage response of any two structures will differ, either because of manufacturing tolerances and variation in boundary conditions, or differences in design. Thus, a fundamental assumption made by conventional machine-learning models will be invalidated – that training and testing data were generated by the same generative process – likely leading to insufficient predictive performance [20]. Methods to account for discrepancies between data derived from different structures are therefore a central component of PBSHM. This issue motivates the research presented in this thesis, which focusses on one such technology – transfer learning – a branch of machine learning that aims to improve the performance in a target structure with sparse data, using data from related source structures with more abundant data [21].

1.3 Thesis contributions

This thesis aims to develop and validate methods for transfer learning for PBSHM to address limitations related to data availability. The methods presented in the subsequent chapters address key challenges related to "when to transfer?", "what to transfer?" and, "how to transfer?", with the objective of moving towards transfer learning pipelines suitable for practical PBSHM applications. The core contributions of this thesis are summarised as follows:

1. Transfer learning (statistic alignment – SA) methods that are robust in sparse data scenarios are adapted to extend their application to scenarios where data are imbalanced, which is typical in SHM. The resulting methods allow for labelled data from a source structure to be used to learn predictive models applicable to a target structure using only data from the undamaged target structure.

2. A novel similarity measure is proposed using the modal assurance criterion (MAC), providing a method for quantifying joint distribution divergence – the primary indicator transfer is feasible – using only data from the undamaged target structure. This similarity measure is shown to address limitations with prevalent unsupervised data-based measures.

3. The MAC-based similarity measure was formulated into a transfer feature-selection criterion. The feature-selection criterion presents the first physics-based method for selecting transferable features, with results showing it to provide consistent improvements when using transfer learning to learn a damage classifier.

4. A regression framework is proposed to address issues relating to the validation of predictive models learnt through transfer learning based on domain similarity. The framework presents a tool that can support decisions relating to when transfer is appropriate. Furthermore, the MAC-based similarity measure was shown to be an effective measure for this application.

5. An online framework incorporating transfer learning and active learning is presented. To this end, a novel Bayesian model is introduced that allows for mapping parameters to be defined using unsupervised domain adaptation.

6. This thesis presents two novel datasets, collected via vibration testing of composite and metal helicopter blades, and laboratory-scale beam and slab bridges. Using these datasets, this thesis presents some of the first examples validating the use of transfer learning to generalise a damage classifier to a target structure learnt using only source data between data collected from different blades and bridges in a lab.

1.4 Chapter Summary

The remainder of this thesis is as follows:

- **Chapter 2** – The necessary background on probability theory and machine learning is provided.
- **Chapter 3** – The fundamental elements of PBSHM are discussed, the general research challenges in transfer learning are introduced, and a literature review of transfer learning for PBSHM is presented.
- **Chapter 4** – SA methods are introduced as a form of transfer learning for sparse data scenarios. These methods are adapted to facilitate their application where target data are not representative of all the health states in the source dataset. The proposed methods are demonstrated using numerical data and data from two real bridges to facilitate the sharing of a damage classifier or the joint clustering of data from multiple structures. These applications are shown for situations where structural discrepancies arise due to differences in design or changes to the structure from repair. In addition, a demonstrative case study is used to discuss the importance of SA methods as a preprocessing step before the application of further transfer learning.
- **Chapter 5** – Limitations related to data-based similarity measures are investigated with a view towards developing methods to select suitable source domains and their corresponding features. A similarity measure leveraging the MAC between the undamaged structures is proposed to provide an indication of joint distribution similarity without target labels. The MAC is then incorporated into a transfer feature selection criterion. Results investigating the possibility of transferring a damage classifier are presented for a numerical population and a population of different helicopter blades. In addition, the methods presented in Chapter 4 and several prominent methods from the literature are compared.
- **Chapter 6** – A regression framework is presented to predict the outcomes of transfer, motivated by the challenge of validating predictive models without comprehensive labelled target data. Specifically, a generalised Gaussian process (GP) model, using a beta likelihood, is proposed as a suitable model to constrain the prediction of classification rates, using a similarity measure as an indicator of the success of transfer learning. In this chapter, the MAC-based measure is used to demonstrate the possibility of predicting accuracy resulting from transfer learning. It is also discussed how this tool could be used to help decide when transfer is appropriate.

- **Chapter 7** – An online transfer-learning framework is presented, incorporating an active-learning strategy to guide the labelling process by selecting more informative labels. To this end, a novel Bayesian model is proposed to update mappings learnt using SA methods with label information as it is acquired throughout the monitoring campaign of the target structure. Results are presented for the transfer of a damage classifier between data from laboratory-scale bridges with varying span lengths.

- **Chapter 8** – Conclusions and future work are presented.

Chapter 2

Structural health monitoring and machine learning

Machine learning is primarily concerned with estimating statistical models from data. In engineering applications, such models are particularly valuable in scenarios where the governing physics are either unknown or only valid under strict assumptions. In scenarios where sufficient data can be acquired, machine learning provides a powerful framework capable of addressing various SHM tasks. This chapter introduces key machine learning concepts relevant to SHM and provides a brief overview of probability theory, as it underpins many of the core concepts in machine learning.

2.1 Fundamental probability theory

Probability theory provides a framework for explaining scenarios where precise signals cannot be assigned to certain values or categories, which is often inevitable when using measured data in engineering applications. A brief overview of the relevant probability theory is provided here; for a more in-depth understanding, the interested reader may refer to [22].

In the context of SHM, an observation of a measured quantity x can be considered as a realisation of a (typically continuous) random variable X that can take values in a domain $\mathcal{X}$, i.e. $x \in \mathcal{X}$. The domain $\mathcal{X}$ is a space that allows a probability measure to be defined. For continuous variables, $\mathcal{X}$ is typically the set, or a subset, of real numbers, $\mathcal{X} \subseteq \mathbb{R}$; for discrete variables, $\mathcal{X}$ is a finite set of discrete outcomes. Generally speaking, the probability of a specific outcome describes the likelihood it will occur relative to all other possible outcomes, where $P(X) = 0$ indicates the outcome will definitely not occur

and $P(X) = 1$ that it is certain to occur. Therefore, it is bounded within the interval $[0, 1]$, and the sum over all possible outcomes must equal one.

A continuous random vector is associated with a probability density function (p.d.f) $p(x)$, which describes the probability that X will take a value x in $a < X < b$ as follows,

$$P(a < X < b) = \int_a^b p(x)dx \quad \text{s.t.} \quad \int_{\mathcal{X}} p(x)dx = 1, \quad 0 \leqslant p(x) \leqslant 1 \tag{2.1}$$

Often, the objective of SHM is to assign observations to discrete categories y that describe the state of the structure, which can take one of K classes in a finite set $y \in \mathcal{Y} = \{1, 2, \ldots, K\}$ – this is called a *label space* in machine learning. For a random variable Y, the probability mass function (p.m.f) $P(Y)$ describes the probability for each of the possible outcomes; a valid p.m.f follows,

$$P(Y = y) \quad \text{s.t.} \quad \sum_{\mathcal{Y}} P(Y = y) = 1, \quad 0 \leqslant P(Y = y) \leqslant 1 \tag{2.2}$$

For brevity, probabilities such as $P(Y = y)$ will be denoted $P(y)$, and $p(\cdot)$ will refer to either a p.d.f. or p.m.f., depending on the context. In addition, multivariate distributions will be indicated by vector inputs. For example, $p(\mathbf{x})$ would represent the multivariate distribution describing the probability of observing a vector $\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^d$ for a d-dimensional random vector.

Given two random variables X and Y, it is often useful to describe the probability that they each take a certain value simultaneously, which is described by the joint probability distribution $p(x, y)$. The fundamental product rule states that the joint distribution can be decomposed into,

$$p(x, y) = p(y|x)p(x) \tag{2.3}$$

where $p(y|x)$ is the conditional probability distribution, which describes the probability of a particular state y given that x has taken a certain value. The conditional distribution is often an object of interest in machine learning, and it is desirable to measure values that are dependent on the quantity of interest y, such that predictions can be made about the current structural state with high certainty. If the random variables are independent, the joint distribution can be decomposed as follows,

$$p(y, x) = p(y)p(x) \tag{2.4}$$

Independence between y and x would imply that knowledge of one value would not restrict the range of values the other would be expected to take.

Often in machine learning, it is useful to estimate the probability $p(y|x)$, given knowledge of $p(x|y)$. Bayes' rule can be derived using the product rule, and it is given by,

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \tag{2.5}$$

where $p(y|x)$ is the *posterior probability*, representing the probability of y after observing x, $p(x|y)$ is the *likelihood*, representing the probability of observing x given y, $p(y)$ is the *prior probability* of y, and $p(x)$ is the *marginal likelihood* or the evidence, which normalises the posterior such that it is a valid probability distribution.

Generally, data generated from a real system can be considered to be samples from a feature space $\mathcal{X}$, and label space $\mathcal{Y}$. Each observation of the response of the system consists of a d-dimensional feature vector, $\mathbf{x}_i \in \mathcal{X}$, and can be assumed to have been generated under a discrete system state, denoted by a label $y_i \in \mathcal{Y}$. Data can be considered as being generated according to an underlying joint probability distribution $p(y, \mathbf{x})$. If this distribution could be accurately modelled, it would enable explicit statements about the likelihood that observations were generated by the system and what the corresponding system states the data were likely generated under, which could then be used to inform the management of the system. Clearly, constructing a perfect model of this underlying distribution would be extremely complex and require the observation of high quantities of data, many variables, and would require the specification of numerous states [23]. However, it is generally useful to think of the modelling problem as an attempt to capture this underlying distribution.

2.2 A brief overview of machine learning for SHM

Machine-learning approaches use a *training dataset* consisting of previously observed data to tune the parameters of an adaptive model [24]. These models can then be used to make inferences about new (test) data. Broadly, these models can be allocated to two categories – *unsupervised* or *supervised* learning.

2.2.1 Unsupervised machine learning

In unsupervised machine learning, it is assumed that only input feature data are available, i.e. the training dataset is $\{\mathbf{x}_i\}_{i=1}^{n_u}$, with n_u unlabelled samples. It is generally used

to find groups in data (clustering), to perform dimensionality reduction for visualisation or to aid downstream learning tasks, or for density estimation [20].

Unsupervised methods are often used in SHM to achieve damage detection [1, 25]. Specifically, using only measurements of damage-sensitive features corresponding to the undamaged structure, density estimation can be used to establish a baseline representing the distribution $p(\mathbf{x})$, which represents the likely range of values for the measurements of the structure under "normal" conditions. During testing, if data deviates from the baseline, the algorithm will indicate novelty – this process is often referred to as *anomaly* or *novelty detection*. These methods are generally considered sufficient to perform damage detection; in some cases, they can also be used to attempt damage localisation [1]. Assuming confounding factors are negligible or have been removed, these methods can provide a strong indication of damage [25–28].

2.2.2 Supervised machine learning

Supervised machine-learning methods aim to learn the relationship between a set of input features and a target variable t_i by training a model using a dataset consisting of paired input features and corresponding target variables $\{\mathbf{x}_i, t_i\}_{i=1}^{n_l}$, with n_l labelled data. Target variables can either be continuous values, in which case the task is *regression*, or *discrete values* representing specific categories, in which case the task is classification, where the target variables are often referred to as labels, which will be denoted by y_i.

In the context of SHM, supervised methods are typically required to achieve more in-depth diagnostics. For example, classification algorithms can be trained to assign data to specific labels, encoding damage information such as location, type, and/or extent [1]. As a regression problem, in some cases more precise predictions can be made by predicting damage location or extent as a continuous variable. While the methods developed in this thesis may also be applicable to some regression tasks, in this thesis predicting the current health state of a structure is primarily treated as a classification problem so contextual information will be referred to as labels for the remainder of the thesis. As a classification problem, models for SHM may address tasks from Rytter's hierarchy independently or encode damage into categories that integrate information from multiple levels, such as damage location, type, and/or extent.

From a probabilistic perspective, supervised-learning methods can be viewed as aiming to model the conditional distribution $p(y|\mathbf{x})$ (discriminative approaches) or the joint distribution $p(y, \mathbf{x})$ (generative approaches) [20]. Previous studies have demonstrated

the application of supervised machine learning for various SHM tasks [4, 29, 30, 30–33]; these methods have mainly been validated using data derived from FE models and controlled experiments.

Following training of a model, it is important to perform validation to ensure the model *generalises* to data that was not used in the training process. For this purpose, datasets are typically split into training and testing data. The quality of a predictive model can be quantified in multiple ways, such as via assessing the likelihood of predictions or via quality measures. Two quality measures that are used for assessing the quality of classification throughout this thesis are accuracy and the F1-score [20]. Accuracy provides the rate of correct classifications; it is given by,

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.6}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. In cases where the quantity of data in each class varies – referred to as class imbalance – accuracy will be disproportionately influenced by predictions in the classes where data are more abundant. The F1-score provides an alternative measure that considers both precision and recall, and is defined as,

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.7}$$

where precision and recall are given by,

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \tag{2.8}$$

The F1-score can also be extended to find an average across multiple classes; a common example is the macro F1-score. It is given by the mean of the F1-scores for each class,

$$\text{Macro F1-score} = \frac{1}{C} \sum_{c=1}^{C} \text{F1-Score}_c, \tag{2.9}$$

where C is the total number of classes, and F1-Score_c represents the F1-score for class c.

2.2.3 Partially supervised learning

Obtaining fully labelled datasets is often infeasible in SHM; however, it is possible to obtain labels for a subset of measurements via periodic inspections throughout the monitoring campaign of a structure. As such, SHM datasets may contain large quantities of unlabelled data and a few data with associated labels. This paradigm is well-suited to partially supervised learning methods [34], which aim to develop regression or classification models by integrating information from both labelled and unlabelled data. By incorporating unlabelled data into the learning process, partially-supervised methods can reduce the requirement for labelled data to train predictive models.

Generally, partially-supervised learning can be categorised as either *semi-supervised* or *active* learning [2]. Semi-supervised learning leverages labelled data, $\mathcal{D}_l = \{\mathbf{x}_i, y_i\}_{i=1}^{n_l}$, and unlabelled data, $\mathcal{D}_u = \{\mathbf{x}_i\}_{i=1}^{n_u}$. These models are trained using a unified scheme, $\mathcal{D} = \mathcal{D}_l \cup \mathcal{D}_u$, which incorporates both types of data. On the other hand, active learning does not explicitly use unlabelled data in the training process. Instead, it uses a predictive model, trained on labelled data, to estimate the information content of unlabelled samples and guide the labelling procedure to obtain more informative labelled datasets.

2.2.4 The problems with conventional machine learning in SHM

A fundamental assumption made when applying a machine learning model to new data is that the testing data were well represented in the training dataset, i.e. $p_{train}(\mathbf{x}) = p_{test}(\mathbf{x})$[1] for density estimation, $p_{train}(y|\mathbf{x}) = p_{test}(y|\mathbf{x})$ for discriminative classifiers or $p_{train}(y, \mathbf{x}) = p_{test}(y, \mathbf{x})$ for generative models, where $p_{train}(\cdot)$ and $p_{test}(\cdot)$ represent the training and testing distributions respectively. Thus, training robust novelty detectors requires observations representative of the structural response across a wide range of EoVs. The requirement for supervised methods is significantly more demanding, as training data must include observations and corresponding labels for each health state that may be encountered during testing. As previously discussed, acquiring these datasets directly from the structure of interest is often prohibitively expensive and/or unsafe. Furthermore, because machine-learning models typically fail to generalise to data generated by different underlying distributions, data sourced from FE models, or different structures cannot reliably be used to train models via conventional machine-learning methods.

[1] While in novelty detection the desired behaviour of the model is to indicate when $p_{train}(\mathbf{x}) \neq p_{test}(\mathbf{x})$, the assumption is that for the normal condition $p_{train}(\mathbf{x}) = p_{test}(\mathbf{x})$, such that data corresponding to abnormal operating conditions can be identified.

Obtaining unlabelled datasets of the normal condition is often feasible in practical scenarios, facilitating the application to damage detection [25–27, 35], which is highly-valuable for safety-critical applications. However, it is generally not feasible to obtain complete labelled datasets from a single structure, impeding the practical implementation of SHM systems that provide more detailed diagnostic information, even though many studies have demonstrated supervised methods are capable of achieving damage localisation [29, 30], classification [31] and quantification [32, 33] using experimental data.

Consequently, substantial efforts have been made to reduce the need for extensive labelling by leveraging unlabelled data [2, 3, 36, 37], incorporating physics-based knowledge [11, 38, 39], or using surrogate data sources such as FE models [13, 40]. These methods can use information more efficiently than many conventional supervised methods, leading to more robust models when data are sparse. However, they still rely on the availability of labelled data from the target structure of interest being representative of all health states of interest. As a result, they cannot assess specific damage before that particular damage state occurs, and some labels are acquired via an inspection.

Chapter 3

Population-based structural health monitoring and transfer learning

As previously discussed, data relating to health states of interest for decision makers such as damage or abnormal EoVs are often costly and/or unfeasible to obtain. The cost and availability of data motivates the development of PBSHM systems. However, PBSHM introduces several unique challenges – requiring methods to identify what conditions PBSHM is possible and how machine learning methods can leverage related information. This chapter aims to give a brief overview of these problems. First, the idea of homogeneous and heterogeneous populations is introduced, along with current approaches to quantify structural similarity. An overview of transfer learning is then presented, highlighting the key research challenges in the context of PBSHM.

3.1 Population-based SHM

To address the issues related to data scarcity, the emerging field of PBSHM aims to develop a framework to leverage information between groups of related structures [6, 15, 16], i.e. populations. When considering all of the data across a population of structures, it is more likely that data corresponding to a relatively complete damage-label set will become available. Furthermore, sharing unlabelled data across similar structures may lead to more robust novelty detection [41–43].

Considering datasets from a population of structures introduces a number of unique challenges. These challenges can be divided into two essential elements of the PBSHM

framework. First, structures must be "similar" enough that data obtained from one structure can reliably improve the predictive capabilities of another. The structural similarity measures within the context of PBSHM should guide two core transfer learning questions – "when to transfer?" and "what to transfer?" [21]. Second, appropriate methods must be developed to account for the variations between data from different structures, such that predictive models can effectively generalise across the population. The proceeding sections aim to provide an overview of the PBSHM framework.

3.1.1 Homogeneous and heterogeneous populations

The first stage in designing a PBSHM system involves assessing the similarity of structures that are available for data acquisition. The suitability and feasibility of a PBSHM approach, as well as the choice of transfer method, will depend on the degree of similarity between the population members and the amount of information available in each of their corresponding datasets.

The method for transfer may vary significantly depending on structural similarity. To this end, it is useful to make a distinction between two types of population. The most internally-similar population is a *homogeneous* population, which is defined as follows,

Definition 1: A population can be considered as *homogeneous* if each member is *structurally equivalent*, and the geometric, material and physical parameters $\boldsymbol{\theta}$ can be considered as random draws from an underlying distribution $p(\boldsymbol{\theta})$ [16].

An example of a homogeneous population would be an off-shore wind farm, which would contain many wind turbines built to the same specifications. In this case, variation will mostly be attributed to manufacturing tolerances and defects, as well as differences in the boundary conditions, i.e. the connection to the seabed may vary. In a homogeneous population, it would be expected that the response of structures to damage would be similar, and in certain cases, a general model – or a *population form* – may be able to represent the behaviour of all members of the population.

In many instances, structures are bespoke to suit a set of specific requirements. For example, highway bridges often have varying boundary conditions, lengths, support locations, and sometimes have different numbers or types of supports; these parameters are adapted to suit each specific site. However, these bridges generally adhere to a similar beam-and-slab design and are comparable in scale, as they are constructed to handle similar traffic loads and span highways, which are typically of similar width. Thus, it is intuitive that it may also be possible to share information between these structures.

A population of highway bridges is an example of a *heterogeneous population*, which is formally defined as,

Definition 2: A population can be considered as *heterogeneous* if each member differs in design; there may be differences in the structures' geometry, connections, or material properties [6].

There are numerous examples of heterogeneous populations; these include sets of radio masts, buildings, nuclear power plants, and bridges. In addition, even in cases where a large homogeneous population is being monitored, it may still be useful to consider transfer from previous monitoring campaigns of structures following a previous design. For example, at the start of the monitoring campaign for a new wind farm there may be more data available from a historic monitoring campaign of a wind farm of a similar design. Thus, this thesis largely focuses on developing methods capable of transferring label information between these heterogeneous structures.

3.1.2 Structural similarity quantification in heterogeneous populations

Intuitively, it may be apparent to an engineer that a population of heterogeneous structures may share some degree of similarity. However, merely having a sense that structures are similar is not sufficient to decide whether transfer is possible, or to inform what transfer methods should be used, as in the worst-case-scenario sharing information can lead to worse predictive performance compared to only using target data. In addition, the complexity of engineering structures makes it challenging to assess their similarity and evaluate whether their responses to damage will be related.

To perform similarity assessment in a principled manner, Gosliga *et al.* proposed formulating abstract representations of structures as attributed graphs (AGs) [6], by first reducing structural components to *irreducible element* (IE) models. Representing structures as AGs has several core advantages for PBSHM. By representing structures as graphs, similarity measures from graph theory, such as the maximum common subgraph or the Jaccard index [6], can be utilised, or similarity measures can be learned via methods such as graph neural networks [44].

Several examples of IE models and their corresponding AGs are presented in Figure 3.1; this figure was first presented in [45]. Illustrations of a wind turbine (a), passenger jet (b), turboprop (c), and two-, three-, and four-span bridges indicate regions of the structures that correspond to their irreducible elements, and below each illustration is their corresponding AG representation.

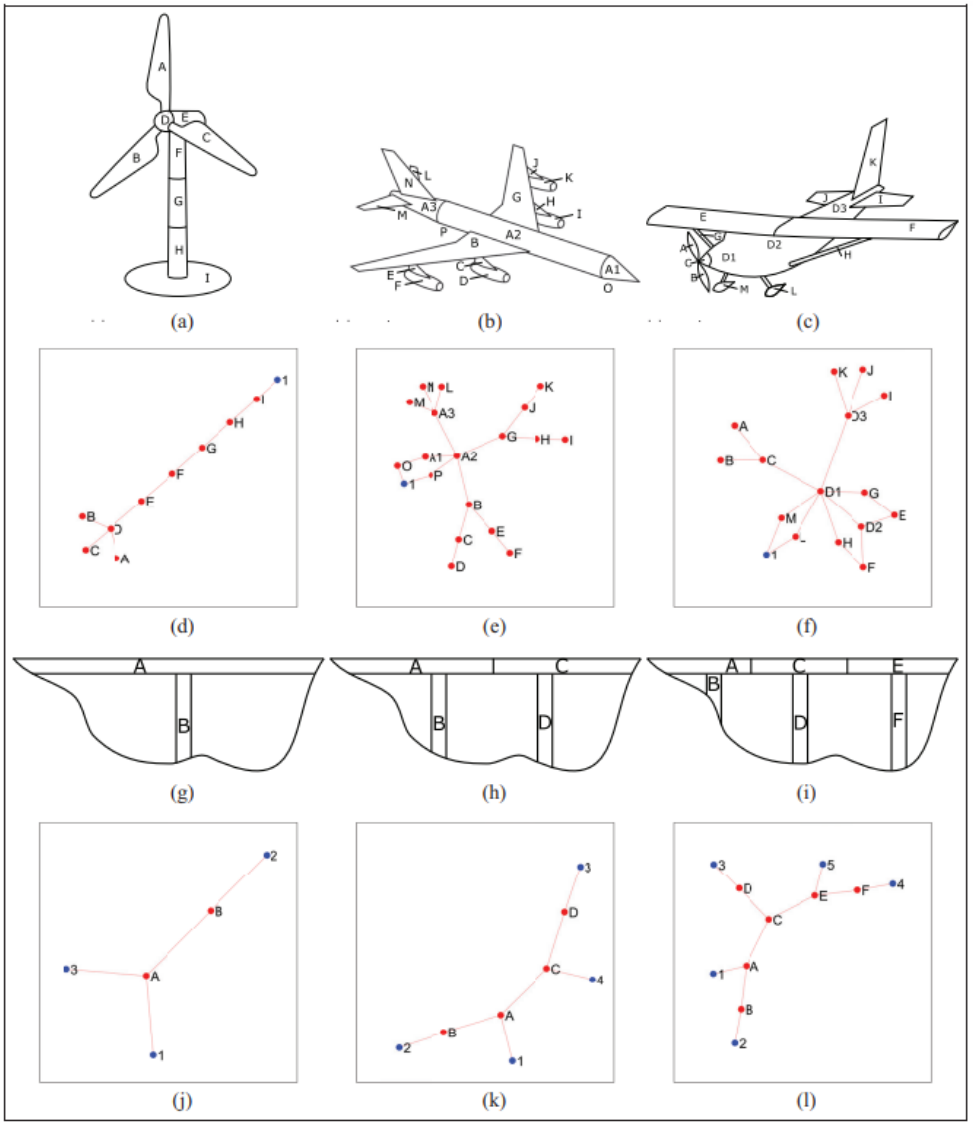


FIGURE 3.1: An example of reducing structures to irreducible elements and their corresponding attributed graph representations; this figure was originally presented in [45]. Examples are presented for a wind turbine (a), a passenger jet (b), a turboprop (c), and two- (g), three- (h) and four-span bridges (i).

Assuming the process of abstracting a structure as an AG is standardised, comparisons via AGs may facilitate principled structural comparison, meaning similarity scores can be meaningfully interpreted. In addition, they may allow for automatic similarity quantification between many structures, meaning potential candidates for transfer could be selected from large databases containing many structures.

The process of building IE models is briefly summarised here – for a more in depth description of IE models the interested reader may refer to [6, 46]. First, IE models are formulated to simplify the representation of structures by breaking them down into fundamental components, such as beams, plates, or shells. The idea is to capture the

essential characteristics of a structure that are most relevant to the response of the structure and potential damage states. For standardised comparison between structures, the process of constructing IE models must be principled, such that an IE model made by any two engineers will be identical; methods for ensuring IE models are standardised is currently an important research topic [47]. AGs are then created by converting IE models into a graphical representation where the topology between nodes describes the connections between the components of the structure, and attributes at each node describe the structural components' geometry and material properties; as such, they highlight three main sources of variation between structures:

- *Geometry* can be described as several geometry classes and each with a corresponding shape, where the shape is further defined by its dimensions, i.e. length, width, and thickness. The geometric information is thereby organised in a hierarchical manner, which may be advantageous when reasoning about PBSHM, since some tasks may only require high-level similarity for certain components or regions of a structure.
- *Material* differences are also described hierarchically, at the coarsest level, it is described by its material class – for example, metal or ceramic etc. At a finer level, the specific description of the material (i.e. the type of steel) and the material properties (i.e. elastic modulus) may also be given.
- *Topology* is determined by the physical connections between elements in a graph. Topology may also be important for determining labels in damage localisation.

There may also be additional sources of variation in the data themselves, as a result of data acquisition. For example, sensor placement, type, or sampling rate may vary. Thus, the overall framework should select data sets to transfer from, based on both structural and data similarity [7].

3.1.3 Knowledge transfer for PBSHM

Once a population of related structures has been identified, the second essential element of a PBSHM framework involves developing methods to train predictive models that can generalise to each structure in the population. As previously discussed, conventional machine learning methods are not appropriate when considering different structures. However, several sub-fields of machine learning aim to improve predictive performance by learning from multiple related datasets, where the features and/or target variables may be different. Several potential approaches for addressing discrepancies in data

include transfer learning [21, 48, 49], multi-task learning [50], domain generalisation [51] and multi-view learning [52]. The methodologies in these sub-fields are often similar, and they largely differ in their objectives. A brief comparison with the related technologies is given to motivate the application of transfer learning for PBSHM.

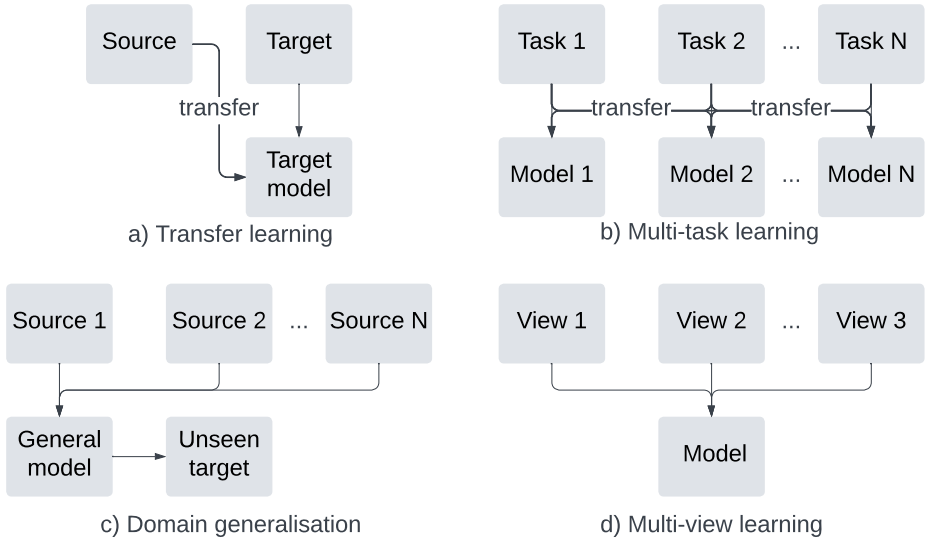


FIGURE 3.2: Illustration of different fields of machine learning that aim to learn from multiple domains/tasks.

The general flow of information for each of these methods is shown in Figure 3.2. The focus of transfer learning is to leverage information contained in a more information-rich dataset (a source domain) to improve the performance of a domain where data are sparse (the target domain). Transfer learning is closely related to *multi-task learning*, which aims to use information from multiple related tasks to improve the performance in each task [50], where each task may share the same input data or have independent input datasets. Multi-task learning differs from transfer learning as the objective is to improve performance in each task equally, whereas transfer learning is often only concerned with improving performance of the target task; although, multi-task learning is sometimes considered a form of transfer learning [21].

In practice, methodologies for performing transfer learning and multi-task learning often use similar approaches [48, 50]. Both approaches have their merits for PBSHM, and both transfer learning [12, 41, 53, 54], and multi-task learning [42, 43, 55] have been demonstrated to be beneficial in several PBSHM applications; however, transfer learning – i.e. prioritising the performance of a specific structure – is more applicable for several important applications. For example, transfer learning would be more relevant where one (or more) structure has data relating to damage, and the objective is

to improve the predictive performance of a damage classifier for structures that have not yet experienced damage. In this case, structures for which damage-state data are available could be considered source structures, and transfer learning could be applied to a target structure that has not observed damage (or damage data are much more limited). Transfer learning would also be more appropriate in scenarios where the source represents a historic monitoring campaign that already contains large amounts of data, and the objective is to improve predictive performance for a target structure where the monitoring campaign has just started.

Another related field is *domain generalisation* [51]. Domain generalisation aims to develop models that can perform well in one or more unseen target domains by learning features that are invariant to the specific characteristics of individual domains [51]. Transfer learning differs, as it assumes that some target data are available, either labelled or unlabelled, and it aims to use these data to reduce discrepancies between a specific source domain (or multiple source domains) and a target domain. While domain generalisation may be applicable to PBSHM, in most cases some data in the structure of interest will become available close to the start of a monitoring campaign, where damage is unlikely, and will continue to become more abundant and varied throughout the monitoring campaign. As such, it would be more efficient to use these data either via transfer learning or multi-task learning.

In addition, transfer learning shares some similarities with *multi-view learning* [52]. Multi-view learning aims to leverage several distinct feature sets, called "views", to improve overall performance in a single prediction task. While in transfer learning, distinct datasets are sourced from related systems, multi-view learning typically uses multiple representations of the same data source. For example, multi-view learning would be applicable if the objective was to improve damage classification using acceleration data from different locations on a structure, or using both acceleration and strain measurements, under the assumption that each feature set (view) includes additional information[1]. While both multi-view and transfer learning may attempt to account for differences between several feature sets, they differ more significantly in their objectives compared to the other approaches, as multi-view learning generally only uses measurements from a single system.

In SHM, scenarios where damage-state data are unavailable or sparse in a structure of interest are prevalent. Transfer learning is one of the most appropriate technologies to leverage information from different datasets that contain the required damage-state information when there is limited data in the domain of interest, although in some applications both transfer learning and multi-task learning approaches would be related and

[1]It should be noted that multi-view learning is also similar to data fusion [1].

may produce similar results. Thus, this thesis aims to develop transfer learning methodologies for PBSHM. The following section presents an overview of transfer learning and the challenges for its application in PBHSM.

3.2 Transfer learning

A major limitation of standard machine learning models is that they can only be expected to produce accurate predictions on testing data that are well represented in the training dataset. This assumption restricts the application of machine learning to scenarios where sufficient training data can be acquired for the specific system of interest. As such, for each task, an independent model must be learnt. However, this learning paradigm differs from how biological intelligence typically learns. Consider the task of learning the acoustic guitar. It would be natural to assume that a musician proficient in another string instrument, such as a bass or electric guitar, would be able to use their previous experience to quickly learn the acoustic guitar. It is also likely that such individuals would be able to use their knowledge of music to learn non-string instruments more efficiently. Essentially, most new learning tasks are aided to some extent by an innate ability to identify similarities between previous experiences and identify useful knowledge. Thus, it is intuitive that some mechanism for exploiting related information between similar domains/tasks could be incorporated into artificial learning algorithms; in a general sense, this is the motivation of transfer learning. This section aims to provide a general overview of transfer learning to motivate its application to PBSHM.

3.2.1 Transfer learning definitions

There are various categories of transfer learning motivated by different learning scenarios. To provide a general definition of transfer learning and its various subfields, two core objects must first be defined – a *domain* and a *task*; their definitions are as follows:

Definition 3: A *domain* $\Omega = \{\mathcal{X}, p(\mathbf{x})\}$, is defined by a feature space $\mathcal{X}$ and a marginal probability distribution $p(\mathbf{x})$ on that space, where $\mathbf{x} \in \mathcal{X}$.

Definition 4: A *task* for a given domain is defined by $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$, where $\mathcal{Y}$ is the label space and $f(\cdot)$ is a predictive function learnt from a finite sample $\{\mathbf{x}_i, y_i\}_{i=1}^{n_l}$, where $\mathbf{x}_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$. In a probabilistic setting, the predictive function can also be viewed as modelling the conditional distribution $p(y|\mathbf{x})$.

The most common paradigm studied in transfer learning considers using a single source and target domain. For simplicity, the following definitions will make this assumption;

however, transfer learning may also be applied to scenarios where there are multiple source domains – called multi-source transfer learning [56]. Following these definitions, transfer learning can be defined as follows:

Definition 5: In *transfer learning*, given a target domain Ω_t and corresponding task $\mathcal{T}_t$., the objective is to use information contained within a source domain Ω_s and the corresponding task $\mathcal{T}_s$ to improve a target predictive function $f_t(\cdot)$.

Typically, it is assumed that the source domain contains sufficient labelled data to learn a predictive function, while the target domain has significantly fewer data, often with few or no labelled examples. The approach to transfer learning often varies depending on the label setting. Pan *et al.* classify the main forms of transfer learning as follows [56]:

In *unsupervised transfer learning*, the source domain consists of labelled data, $\mathcal{D}_{s,l} = \{\mathbf{x}_{s,i}, y_{s,i}\}_{i=1}^{n_{s,l}}$, where $n_{s,l}$ is the number of labelled source instances, and $\mathbf{x}_{s,i}$ and $y_{s,i}$ represent the features and labels, respectively, which are assumed to be generated following the underlying source joint distribution $p_s(y, \mathbf{x})$[2]. The target domain consists of unlabelled data, $\mathcal{D}_{t,u} = \{\mathbf{x}_{t,j}\}_{j=1}^{n_{t,u}}$, where $n_{t,u}$ is the number of unlabelled target instances, which are assumed to be generated from the underlying target joint distribution $p_t(y, \mathbf{x})$.

In *supervised transfer learning*, both the source dataset, $\mathcal{D}_{s,l}$, and the target dataset, $\mathcal{D}_{t,l} = \{\mathbf{x}_{t,j}, y_{t,j}\}_{j=1}^{n_{t,l}}$, have labelled data, where $n_{t,l}$ denotes the number of labelled target data. Similar to conventional machine learning, this setting can also be extended to partially-/semi-supervised transfer learning by incorporating both labelled and unlabelled data into the learning process [56].

A less studied area is where the source dataset, $\mathcal{D}_{s,u} = \{\mathbf{x}_{s,i}\}_{i=1}^{n_{s,u}}$, including $n_{s,u}$ unlabelled instances, and the target dataset, $\mathcal{D}_{t,u}$, both consist of only unlabelled observations. In this setting, transfer learning aims to enhance the performance of unsupervised learning tasks in the target domain, such as clustering, density estimation, and dimensionality reduction [21]. In [56], this paradigm is not explicitly defined; however, in [21, 49] this setting is defined as unsupervised transfer learning. It is important to note that there are a few inconsistencies in the definitions of different transfer-learning paradigms in the literature [21, 48, 49, 56]; however, the remainder of this thesis will use the definitions provided above, and the setting where no data are labelled in both domains will not be considered. In addition, for the remainder of the thesis subscripts "u" and "l" will only be used when it would otherwise be unclear whether notation refers to labelled or unlabelled data notation refers.

[2]The source and target distributions denoted by $p_s(\cdot)$ and $p_t(\cdot)$, respectively take the corresponding measurements and labels $\mathbf{x}_s$ and/or y_s and $\mathbf{x}_t$ and/or y_t, respectively. However, in this thesis the subscript for features and labels will be dropped for brevity.

In addition, several transfer scenarios arise depending on the differences between the features and label spaces of each domain. These are categorised as homogeneous and heterogeneous transfer learning, as defined in [56]. In *homogeneous transfer learning*, it is assumed that the feature and label spaces are the same, i.e. $\mathcal{X}_s = \mathcal{X}_t$ and $\mathcal{Y}_s = \mathcal{Y}_t$, but there are differences in the marginal $p_s(\mathbf{x}) \neq p_t(\mathbf{x})$ and/or conditional distributions $p_s(y|\mathbf{x}) \neq p_t(y|\mathbf{x})$. The objective is then to account for the differences in the distributions to improve a target predictive function $f_t(\cdot)$. Whereas, in *heterogeneous transfer learning*, it is assumed there are differences in the feature and label spaces, i.e. $\mathcal{X}_s \neq \mathcal{X}_t$ and $\mathcal{Y}_s \neq \mathcal{Y}_t$. This scenario leads to a more complex objective, where available source and target data need to be used to find a shared feature space where data distributions are similar, such that source data can be used to improve the performance of a target predictive function $f_t(\cdot)$.

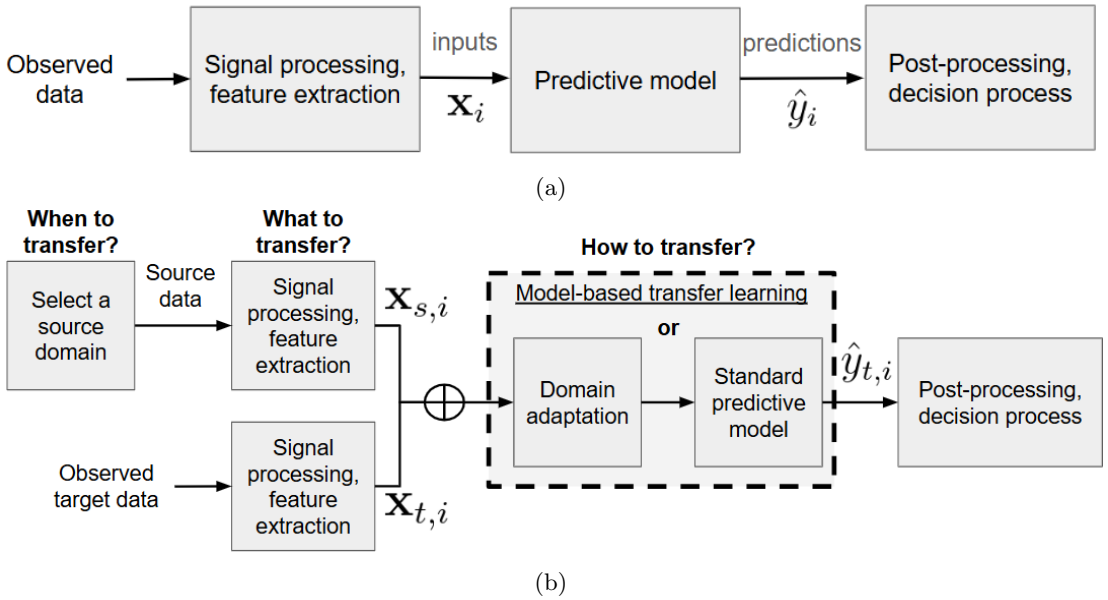


FIGURE 3.3: Frameworks of (a) a conventional SHM process and (b) a transfer learning-based SHM framework. The dashed line in (b) indicates that transfer can be performed in two stages using DA or in one stage with model-based transfer learning.

When developing a transfer learning strategy, it is useful to frame the problem with regard to three questions [21]: **"when to transfer?"**, **"what to transfer?"**, and **"how to transfer?"**, each of which presents important research issues. A comparison of a standard SHM framework and a transfer learning-based SHM framework are presented in Figure 3.3(a) and Figure 3.3(b), respectively. It can be seen that in comparison to conventional SHM (Figure 3.3(a)), transfer learning introduces additional considerations relating to each of these core questions (Figure 3.3(b)). First, evaluation of available source domains/datasets must be carried out to decide whether transfer is possible. Second, signal processing and feature extraction must not only focus on extracting damage sensitive features, but also features that can be used for transfer learning. Finally, the

modelling procedure requires an additional mechanism to share information, either by introducing a domain adaptation (DA) step or using end-to-end transfer learning models (model-based transfer learning). The current approaches and core challenges relating to these questions are introduced in the context of PBSHM in the following sections.

3.2.2 When to transfer? – avoiding negative transfer

"When to transfer?" seeks to assess in which situations transfer learning should be applied. The objective of transfer learning is to enhance the predictive performance of a target predictive function by leveraging information from a related source domain. However, transfer learning is not always beneficial. In certain scenarios where the problem is ill-posed or the domains are not sufficiently related, transfer learning may result in worse performance than only using target data. This phenomenon is known as *negative transfer* – which may have catastrophic consequences in SHM, where poor generalisation of a source model can lead to unnecessary costly inspections or severe damage caused by missing critical maintenance.

A visual representation of negative transfer in the context of DA is presented in Figure 3.4. Figure 3.4(a) presents the desired outcome from DA, where post-alignment data from corresponding classes (indicated in blue and red) occupy the same region of the feature space in both the source and target domains; thus, a classifier trained using source data would be expected to generalise well to the target data. In contrast, Figure 3.4(b) illustrates a scenario where alignment led to data corresponding to different labels in the source and target domains to be aligned, which in this case would lead to incorrect predictions in the target domain, a worse result than even a random guess; thus, it represents negative transfer.

To determine whether transfer learning has led to negative transfer, one approach would be to compare the *risk* associated with different predictive models. For a predictive function $f(\cdot)$, output by an algorithm $f = \mathcal{A}(\mathcal{D}_s, \mathcal{D}_t)$, risk on the target domain is defined as follows,

$$R(f) = \mathbb{E}_{(x_t, y_t) \sim \mathcal{D}_t}[\ell(f(\mathbf{x}_t), y_t)], \tag{3.1}$$

where $R(f)$ represents the risk of a predictive function $f(\cdot)$ applied to a target dataset $\mathcal{D}_t$, and $\ell(\cdot, \cdot)$ is a loss function. In the context of transfer learning, negative transfer can be defined as follows,

Definition 6: *Negative transfer* occurs when the risk of the predictive function, learnt via transfer learning, $R(\mathcal{A}(\mathcal{D}_s, \mathcal{D}_t))$, is greater than the risk of a predictive function

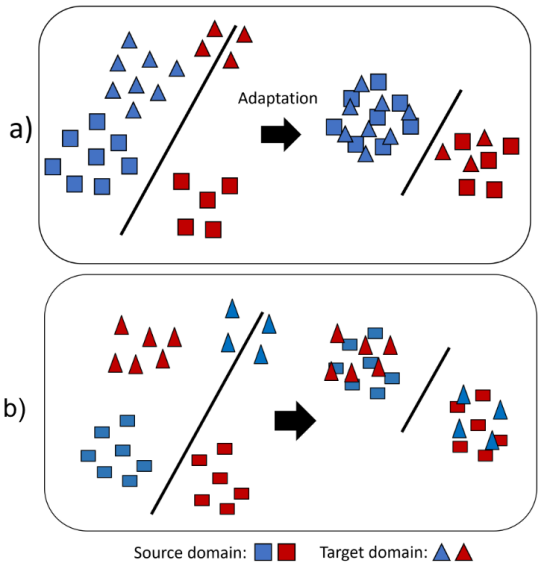


FIGURE 3.4: Example of an ill-posed transfer learning problem where adaptation leads to the data relating to different labels in the source and target domains to be aligned.

learnt using only the target data, $R(\mathcal{A}(\varnothing, \mathcal{D}_t))$, i.e. $R(\mathcal{A}(\mathcal{D}_s, \mathcal{D}_t)) > R(\mathcal{A}(\varnothing, \mathcal{D}_t))$, where $\varnothing$ represents an empty set [57].

"When to transfer?" can be framed with respect to avoiding negative transfer, which is dependent on two main factors 1) the similarity between the source and target domains, and 2) the quantity of information available to learn regularities between domains [57].

The first factor, requires the joint distributions $p_s(y, \mathbf{x})$ and $p_t(y, \mathbf{x})$ to be related. Quantifying similarity between the joint distributions is challenging, and often data intensive [58]. Limited target data, particularly labelled data, often impedes the direct application of similarity metrics. In many transfer learning studies, joint distribution similarity is implicitly assumed [21, 48, 49], and only a few studies provide principled justification for transfer; for example, by using marginal distribution distance metrics [59–61]. In PBSHM, engineering expertise can also be leveraged to quantify domain (structural) similarity, as discussed in previous sections; however, identifying important characteristics that indicate the possibility of transfer is still an ongoing topic of research [46].

The likelihood of negative transfer is also heavily dependent on the availability of data in the target domain. On one hand, the likelihood of negative transfer is related to the performance of the target-only model. For example, consider a supervised learning task. If labelled target data are unavailable ($n_{t,l} = 0$), a target-only model would be a weak random model, meaning prediction probability would be uniform across all classes; thus, negative transfer is less likely to occur. As some labels become available, semi-supervised methods become more appropriate to learn a target-only model [20], which are likely to perform better than a weak random model. At the other end of the spectrum, if labelled

target data are abundant, if the source joint distributions differs even slightly from the target, transfer may harm generalisation.

On the other hand, data availability also influences the practicality of accounting for distribution shift. When labelled target data are unavailable ($n_{t,l} = 0$), transfer-learning algorithms often rely on minimising the discrepancy between the marginal distributions of the observed data $p_s(\mathbf{x})$ and $p_t(\mathbf{x})$, which requires a strong assumption to be made about the similarity of the joint distributions [60]. Conversely, if some labels are available in both domains ($n_{t,l} > 0$ and $n_{s,l} > 0$), a wider variety of methods become applicable to learn shared regularities between the joint distributions [48].

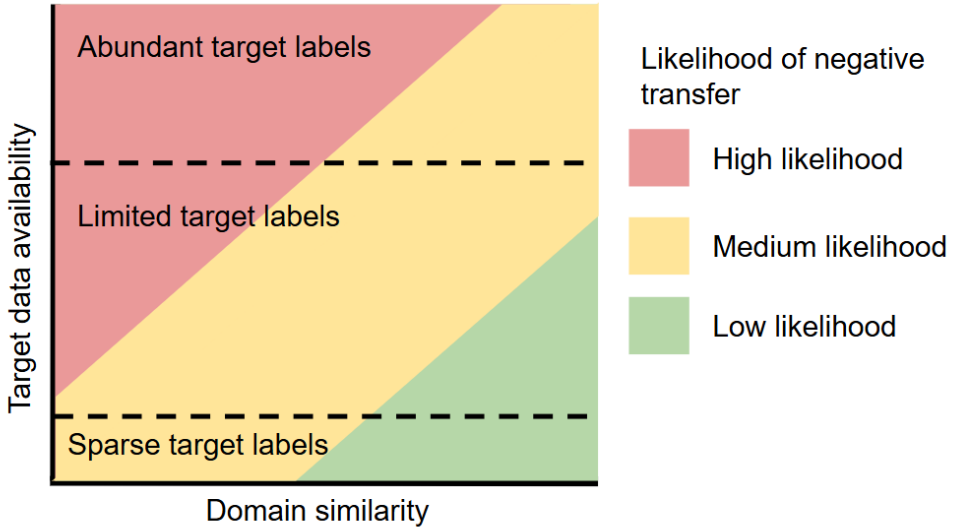


FIGURE 3.5: A purely demonstrative illustration of the likelihood of negative transfer relative to target data availability and source domain similarity.

Figure 3.5 illustrates this balance between target labelled-data quantity and domain similarity, highlighting how these factors may influence the likelihood of negative transfer[3]. While it may be intuitive that transfer learning is most appropriate where data are "insufficient" to learn a robust target-only model and there is a "similar" enough source domain to transfer from – i.e. transfer is most appropriate when the task falls in the green region in Figure 3.5 – these concepts currently lack rigorous methods to assess the suitability of transfer in specific PBSHM scenarios.

The most direct approach for deciding when to transfer would be cross-validation, where target labels are used to evaluate the empirical risk. However, if data are sparse or unrepresentative of the testing data, this method would be unreliable [62]. Alternatively, several bounds rooted in statistical learning theory exist that aim to provide an upper bound on a predictive function learnt via transfer learning [60]. These bounds suggest the

[3]This illustration is purely speculative and is for demonstration purposed only.

performance of a source model is dependent on three factors – marginal and conditional distribution divergence, and the source risk of the predictive function. In practice, obtaining valid measures on the marginal and conditional distributions is challenging, preventing the direct application of these bounds.

Another approach would be to attempt to quantify the joint-distribution divergence. However, this approach presents two distinct challenges. First, in the absence of sufficient labels to measure the similarity between the joint distributions, similarity quantification is limited to unsupervised metrics that quantify the distance between the marginal distributions [48, 60, 63]. Thus, using these methods, there would still be a reliance on domain expertise to ensure the conditional distributions are similar. Second, a *distribution divergence* $D(\cdot)$, will only indicate the distributions are the same if $D(\mathcal{D}_{s,u}, \mathcal{D}_{t,u}) = 0$. If distributions are related, but not identical, it is challenging to interpret $D(\mathcal{D}_{s,u}, \mathcal{D}_{t,u}) > 0$. Thus, developing methods to inform "when to transfer?" to ensure PBSHM systems are robust is a critical research area.

3.2.3 What to transfer? – identifying related features and relationships

"What to transfer?" focuses on identifying specific information that can be shared between domains; thus, guiding **"how to transfer?"**, which aims to develop algorithms that can improve generalisation across domains. For example, "what to transfer?" may seek to identify related features, or similar relationships between measurements and predicted values.

In homogeneous transfer learning, the feature and label spaces are assumed equivalent; thus, transfer learning aims to address distribution shift, where the joint distributions between the two domains differ. "What to transfer?" is largely driven by the type of shift between features; it is important that features with a suitable type of distribution shift are chosen for a given transfer-learning method. Two scenarios that transfer learning typically attempts to address are:

1. $p_s(\mathbf{x}) \neq p_t(\mathbf{x})$ and $p_s(y|\mathbf{x}) = p_t(y|\mathbf{x})$. The conditional distributions may be the same between domains, $p_s(y|\mathbf{x}) = p_t(y|\mathbf{x})$ but the marginals differ, a phenomenon known as *covariate shift* [64]. An illustration of this situation is shown in the top panel of Figure 3.6, where it can be seen that the regions the source and target data are likely to take overlap, but the proportion between the class distributions differs, leading to a different marginal distributions[4].

[4]Note, the shifts in the marginal distributions here are a direct result of the prior distributions differing between domains $p_s(y) \neq p_t(y)$

An example of covariate shift in SHM would be data collected from two identical structures during different seasons. For example, if one dataset contained more measurements taken during freezing temperatures, it would include a greater proportion of higher natural frequencies resulting from the increased stiffness of the structure. However, the probability of observing a natural frequency under freezing temperatures, given a specific range of natural frequency values, would remain identical for both structures. Covariate shift is particularly important to address for density estimation and generative models. In the context of classification, covariate shift, as presented in Figure 3.6, may lead to a change in the optimal boundary if classes are not separable. For example, it can be see in Figure 3.6, the value at which the class 0 becomes more likely than class 1 differs between domains, which is indicated by the black vertical lines. In addition, it may cause issues in supervised learning caused by sample bias, meaning the model may over-fit to patterns that are over-represented in the training set.

2. $p_s(y|\mathbf{x}) \neq p_t(y|\mathbf{x})$ and/or $p_s(\mathbf{x}) \neq p_t(\mathbf{x})$. A more general scenario is where the conditional distributions also differ, named conditional distribution-, or concept shift [56]. Conditional-distribution shift will likely lead to poor generalisation of a predictive model, as this is the relationship in the data that supervised learning typically aims to capture. This assumption captures a broad spectrum of scenarios, ranging from conditional distribution shifts that can be reduced by only minimising marginal distribution discrepancy, to deviations that mean transfer learning can only provide limited (or no) improvements in the target domain [48].

   An example of this scenario is shown in the bottom panel of Figure 3.6, where it can be seen that the regions of high density do not overlap between domains. This lack of overlap in the marginal distributions distinguishes this situation from covariate shift, as the absolute position of any classifier boundary could differ significantly. For example, the black lines centred at zero and ten in Figure 3.6 represent the classification boundaries for the source and target domains, respectively. This discrepancy makes directly leveraging the data in their original feature space more challenging. However, the domains share a related class structure, where Class 1 represents an increase in the value of the measured quantity. Thus, it may be intuitive that applying a mapping that removes the mean and scaling could align these domains. Importantly, it may be possible to learn such a mapping without target labels. Alternatively, this relation could be leveraged to learn a suitable inductive bias for a classifier, reducing labelled target data requirements.

As with the previous question – "when to transfer?" – identifying what type of distribution shift exists for a given pair of domains, and associated features, has previously

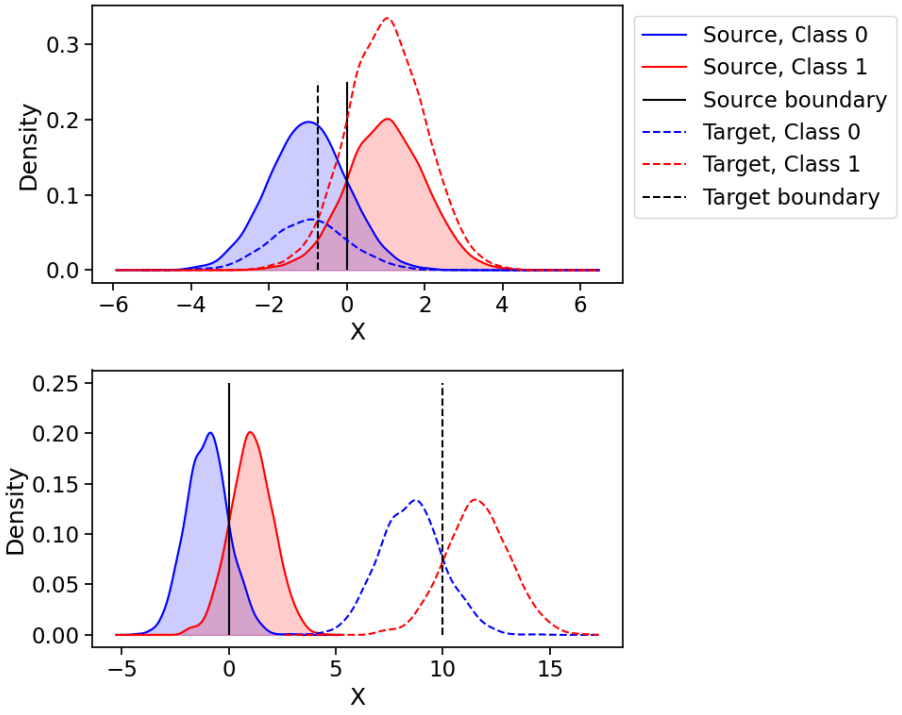


FIGURE 3.6: Examples of two types of distribution shift that transfer learning aims to address. The top panel represents covariate shift, and the bottom panel presents label or conditional-distribution shift.

largely relied on domain expertise [21, 48, 49]. Moreover, there are few well-defined methods for testing the requirements of certain distribution-shift assumptions between a set of source and target features. For example, in the second scenario, quantifying the relationship between conditional distributions that occupy distinct regions of the feature space may be challenging. An interesting direction for PBSHM would be to instil physical understanding or engineering expertise into principled methods to attempt to infer what features would have a similar response to damage and how the response to damage is related. This is one of the objective of developing metrics based-on IE models [6].

In PBSHM, both types of distribution shift may arise; however, this thesis focuses on developing methods to address the second scenario. To motivate this setting for PBSHM, consider the task of modelling the distribution of the natural frequencies $p(\boldsymbol{\omega}_n)$, for each member from a population of uniform cantilever beams. A closed-form expression for a uniform cantilever beam exists, defining a deterministic value for natural frequency given the beam geometry and boundary conditions. In practice, it is likely that observed natural frequency values $\hat{\omega}$, will also be contaminated with observation noise, which could be considered to follow a Gaussian distribution with variance σ^2 (and

some parameters will have a dependence on various EoVs). The generative process for the natural frequency observations $\hat{\omega}_n^{(j)}$ could therefore be given by,

$$\omega_n^{(j)} = \left(\frac{\beta_n^{(j)}}{L}\right)^2 \sqrt{\frac{EI}{\rho A}} \tag{3.2}$$

$$p(\hat{\omega}_n^{(j)}) = \mathcal{N}(\omega_n^{(j)},\ \sigma^2) \tag{3.3}$$

where $\omega_n^{(j)}$ represents the natural frequency of the j-th mode, $\beta_n^{(j)}$ is the j-th root of the characteristic equation for a cantilever beam, which is determined by the boundary conditions. The natural frequency is dependent on the beam geometry and material properties – the length of the beam L, the Young's modulus of the material E, the second moment of area I , the density of the material ρ, and A is the cross-sectional area of the beam. Assuming a heterogeneous population, these properties may differ significantly, meaning the mean value of $\omega_n^{(j)}$, given by equation (3.2), will change significantly, potentially resulting in little or no overlap in the regions of high density of the marginal distributions. However, this mean function is the result of scaling the fundamental natural frequencies $\beta_n^{(j)}$, which are invariant to geometry and material properties, suggesting that even though the absolute values of $\omega_n^{(j)}$ differ, there is a common factor between all of these structures which may allow for information sharing.

In heterogeneous transfer learning, both the feature and/or label spaces will differ, meaning that models cannot be directly applied between domains without first mapping data into shared spaces. For example, in SHM, feature spaces may differ because of changes in the sampling rate of acceleration measurements, meaning associated frequency-domain features would have a different dimension [53]. In this case, transfer-learning methods must assume that there is a shared latent subspace where information can be shared between domains.

3.2.4 How to transfer? – methods for transfer

Motivated by the potential of useful information that could be shared between domains, **"how to transfer?"** aims to specify an appropriate form of transfer learning algorithm. Transfer learning can facilitate information sharing where there are discrepancies between domains; it generally achieves this by either regularising/adapting a source model using target data or by manipulating the data themselves to align the distributions, thereby allowing standard machine learning methods to generalise effectively. A brief

overview of the methods used in each approach is presented as follows. There is extensive research in transfer learning for a range of applications; as such, the following summary provides a general overview and is not an exhaustive literature review. For a more in-depth review, the interested reader may refer to [21, 48, 49, 56].

3.2.5 Model-based transfer learning

Model-based algorithms generally aim to learn inductive biases using source data by finding better model initialisation [65, 66], sharing parameters [67, 68], or regularising parameters across domains [69–75]. A popular example is fine-tuning, where some or all of the weights of a neural network trained using source data are updated using limited labelled target data [65]. Fine-tuning has been shown to significantly reduce the data requirement for training large neural networks [65], leading to wide adoption for applications such as computer vision [76–80] and natural language processing [81, 82], where complex nonlinear functions are required.

A current promising area of research relating to the idea of fine-tuning is that of a "foundation model" [83–85]. A foundation model generally consists of a large neural network that has been trained on a large quantity of data from many sources with the motivation of learning a general representation useful for downstream tasks. Thus, these models can then be updated with relatively small quantities of data via fine-tuning. Furthermore, in some cases pre-training can be conducted using tasks where data are cheap to annotate; for example, next word prediction in natural language processing does not require expensive human annotators. Learning a set of bases in this way effectively restricts the model complexity and encourages learnt functions to follow a form that performs well across related learning tasks. This approach is particularly powerful for applications like natural language processing [83] and computer vision [85], where complex nonlinear functions are required to model input-output relationships; however, there are no known functions that explain these specific relationships, motivating the application of high-variance models that capture a wide range of possible functions. Nevertheless, even if these methods can represent the exact underlying functional form that describes a specific phenomenon within these bases, data relating to the domain/task of interest would likely still be required to tune parameters to a target application.

Fine-tuning essentially biases the target model to have similar weights to the source model. There are numerous approaches that attempt to directly bias the target model to have similar weights to the source model, many of which fall into the category of multi-task learning. Hierarchical Bayesian modelling presents one example, which assumes that the model parameters are drawn from a shared hyperprior [69]; thus, sharing

information by learning an informative population-level prior. This hyperprior can be seen as a population-level model, representing the variation across each domain/task. Mixed-effects models also follow a similar approach, but directly share certain parameters across multiple domains [42].

Another approach uses shared regularisation terms across multiple tasks; thus, these methods encourage the use of more informative features by inferring which features are beneficial to multiple prediction tasks [50]. These regularisation schemes have also been modified to explicitly prioritise a target domain by including regularisation terms that either enforce the source and target predictions [70, 71], or the source and target weights to be similar [73–75]. A related approach is to assume a set of shared weights and allow the target model to vary from the source model by including additional target-specific weights [68].

There also exist a few model-based algorithms that do not require target labels. These typically use unlabelled target data in a regularisation term, to ensure that the entropy in the target is low, encouraging confident predictions in the target [72], which is also a common use of unlabelled data in semi-supervised learning [20]. A few studies also use an ensemble of source models and attempt to find a consensus that is consistent for a given target sample [86, 87]. However, these methods require significant overlap between the source and target marginal distributions. For example, they would not be appropriate in scenarios such as Figure 3.6(b), as an initial source classifier would likely confidently misclassify all target data as Class 1.

3.2.6 Domain adaptation

Domain adaptation (DA) aims to reduce the differences in data distributions by re-weighting instances, known as instance-based algorithms, or by finding a new feature representation, referred to as feature-based algorithms [48]. Following successful domain adaptation, any predictive function learnt on the source data should then generalise to target data.

Instance-based algorithms focus on re-weighting individual training samples, such that specific samples are given more or less importance depending on the likelihood that they are relevant to the target task [88]. These methods typically involve modifying the loss function by adding a weight w_i to modify the empirical risk as follows,

$$\hat{R}(f) = \sum_{i=1}^{n_s} w_i \ell(f(\mathbf{x}_{s,i}), y_{s,i}) \tag{3.4}$$

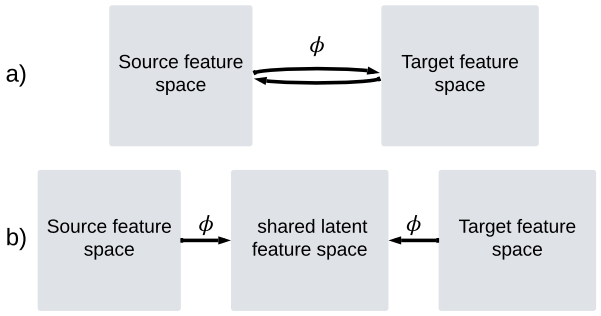


FIGURE 3.7: A comparison of the two mapping approaches for DA – asymmetric mappings (a), which project one domain to the other, and symmetric mappings (b), which project the data into a shared latent feature space.

Generally, w_i aims to estimate the ratio between the joint probabilities $\frac{p_t(y_i,\mathbf{x}_i)}{p_s(y_i,\mathbf{x}_i)}$, such that the re-weighted source instances will more closely match the target joint distributions [21]. In unsupervised transfer learning, instance-weighting is generally used to address covariate shift (or sample-bias), as such differences in the joint distributions can be reduced by finding instance weights that estimate the ratio between the marginal probabilities $w_i = \frac{p_t(\mathbf{x}_i)}{p_s(\mathbf{x}_i)}$, since $p_s(y|\mathbf{x}_i) = p_t(y|\mathbf{x}_i)$ [88–90]. These methods have been used for learning classifiers that are more representative of the target data [89–91]; in addition, they have also been applied to reduce bias in cross-validation [88].

In some applications, only a subset of the input space may follow the same conditional distribution in both the source and target domains; as such, several methods use limited target labels to improve transfer. For example, some approaches iteratively down weight instances from the source domain that are less relevant to the target task using a small set of labelled target data [92–94], while others remove samples that are less relevant to the target and use instance weighting to account for covariate shift [95].

A major limitation of instance-based methods is they require significant overlap between the marginal distributions, i.e. if the marginal distributions do not overlap $\frac{p_t(\mathbf{x}_i)}{p_s(\mathbf{x}_i)} = 0$, meaning no instances can be directly applied to the target. This assumption is often too restrictive, motivating feature-based approaches to DA.

Feature-based approaches generally aim to find a shared feature representation between domains [48]. A prevalent approach aim to learn a mapping to project data into a shared space; thus, these methods assume $p_s(\mathbf{x}) \neq p_t(\mathbf{x})$ and $p_s(y|\mathbf{x}) \neq p_t(y|\mathbf{x})$, and a mapping $\phi(\cdot)$ exists, that projects the domains into a shared feature space, where $p_s(\phi(\mathbf{x})) = p_t(\phi(\mathbf{x}))$ and $p_s(y|\phi(\mathbf{x})) \approx p_t(y|\phi(\mathbf{x}))$ [48, 96]. There are two main mapping approaches – asymmetric and symmetric mappings [48]; these are demonstrated in Figure 3.7(a) and Figure 3.7(b), respectively. An asymmetric mapping aims to project data from one domain to the other and often are only applied to one domain, allowing for

models trained in one domain to be applied directly to the other. On the other hand, symmetric mappings applied to the same mapping to both domains and aim to project data into a shared latent feature space, assuming a shared latent space exists where the distributions of both domains are invariant.

In unsupervised DA, directly estimating joint distribution discrepancy is challenging, as such a prominent approach is to assume a mapping $\phi(\cdot)$ can be learnt by minimising a divergence measure on the unlabelled data, which quantifies how different two probability distributions are; in general, the marginal-distribution divergence are used [48, 96]. A few examples include the maximum mean discrepancy (MMD) [96, 97], Kullback-Leibler (KL) Divergence [93], Jensen-Shannon Divergence [98], Bregman Divergence [99], or Hilbert-Schmidt Independence Criterion [100, 101]. These methods generally aim to learn a mapping that minimises one of these divergence measures $D(\cdot)$ between the sets of source and target feature vectors, given by $\mathbf{X}_s \in \mathbb{R}^{n_s \times d}$ and $\mathbf{X}_t \in \mathbb{R}^{n_t \times d}$ respectively, using the following objective function,

$$\phi = \underset{\phi}{\operatorname{argmin}} D(\phi(\mathbf{X}_s), \phi(\mathbf{X}_t)) \tag{3.5}$$

Within these approaches, mappings are generally found either via a kernel mapping or a feature extractor using deep neural networks (DNNs) [48]. For example, transfer component analysis (TCA), can be seen as an extension to kernelised principal component analysis (PCA) that aims to find a latent space that maximises variance and minimises the MMD [96]. Long *et al.* proposed joint-distribution adaptation (JDA) to extend TCA by also attempting to minimise the MMD between the conditional distributions [97]. Since in unsupervised DA there are no labels in the target, JDA uses pseudo-labels found via predictions of a classifier trained on the source domain to assign pseudo-labels to the target samples and then finds the distance between the class-conditional distributions $p_s(\mathbf{x}|y)$ and $p_t(\mathbf{x}|\hat{y})$, where $\hat{y}$ are label predictions. Wang *et al.* proposed a modification of JDA, balanced distribution adaptation (BDA) [102], by introducing a parameter to control the importance of the marginal and conditional distributions. In addition, a few approaches propose learning a source classifier and kernel simultaneously [103, 104].

In recent years, a significant amount of literature has focused on implementing similar objectives in DNNs [105, 106]. For example, analogous methods exist for TCA – the domain adaptation network [107], JDA – the joint adaptation network [108] and BDA – dynamic weighted learning [109]. Other neural network-based methods learn a feature representation to confuse a domain discriminator, which aims to distinguish between the source and target [110–113]; thus, these algorithms minimise the proxy-A distance

(PAD) (see the proceeding section for more details). These methods have been largely motivated by image classification [105] and natural language processing [106], and are advantageous where data are high-dimensional and complex nonlinear functions are required for mapping and predicting data. However, they will typically require more data to prevent overfitting and often do not produce low-dimensional feature representations like the kernel methods, making visualisation of features more challenging.

Aside from minimising nonparametric distribution distance metrics, a few approaches align domains by directly matching statistical moments [91, 114–116] – called statistic alignment [48]. These have the advantage of requiring less data and use linear transformations, which retain the interpretability of the original feature space. Similarly, some approaches perform DA via batch normalisation approaches [117–119], which align the means and standard deviations of the *activations* in DNNs.

Finally, several methods attempt to align domains via manifold alignment. Similar to statistical-alignment approaches, Procrustes analysis can be used to align domains via translation, scaling, and rotation operations [120, 121]; the main difference is that these methods use isotropic transformations, which maintain the local geometric properties of the data. These methods may also leverage geometric properties; for example, Riemannian Procrustes analysis extends this concept to spaces with intrinsic curvature [122]. Other manifold methods compute shared subspaces by interpolating between the manifold of subspaces (Grassmannian) [123, 124]. Similarly, gradual domain adaptation leverages ideas from this geometric perceptive to transfer via interpolating domains, as discussed by Wang *et al.* [125].

Within unsupervised domain adaptation, a few approaches exist that do not learn a mapping. For example, a few studies find shared features via feature selection. Both unsupervised metrics and supervised distribution distance metrics have been incorporated into selection criteria, using exhaustive selection schemes [126–128], and heuristic search methods [129, 130]. Furthermore, a few approaches aim to augment the original feature space [67].

While advantageous when labelled data are expensive and/or unfeasible to obtain, the application of unsupervised domain adaptation is generally restricted to pairs of domains with high structural similarity and varied unlabelled data. In the supervised transfer learning setting, mappings can also be learnt to directly optimise the performance of a predictive model by learning both a predictive function and a mapping, a general formulation for the loss function then becomes,

$$\phi, f = \arg\min_{\phi, f} \sum_{D \in \{D_s, D_t\}} \sum_{(\mathbf{x}_i, y_i) \in D} \ell\big(f(\phi(\mathbf{x}_i)), y_i\big) \tag{3.6}$$

In this paradigm, examples exist of approaches that learn asymmetric [131], and symmetric mappings [132–134]. To learn a single predictive function that optimises the loss in both domains, the mapping must reduce the differences between conditional distributions [60].

3.2.7 Distribution distance measures and unsupervised domain adaptation

As briefly discussed in Section 3.2.3, data from two different structures will often not occupy the same region of the feature space in heterogeneous populations, motivating the application of mapping-based DA. As discussed in the previous section, there are a variety of objectives used to learn DA mappings, although minimising a distribution distance measure is perhaps the most common, and previous applications to unsupervised DA in PBSHM have focused on this approach [7, 41, 135]. In addition, these similarity metrics are also used throughout this thesis for quantifying the quality of a DA mapping, as they allow for comparison of distribution divergence in the transformed feature space.

This section introduces two prominent similarity metrics used in this thesis: the MMD and the PAD. Their relevance to learning a DA mapping is demonstrated via the introduction of TCA. Two modifications of TCA are also presented that aim to address a core limitation of unsupervised DA.

3.2.7.1 Distribution divergence measures

Determining whether two datasets were drawn from the same underlying distribution or identifying the degree of divergence between their distributions is important for detecting changes in training and testing data – a critical consideration when deploying machine learning models [59–61, 136]. This problem is also known as a two-sample test [61, 136]. If the data follow a known parametric distribution, several powerful statistical tests can be leveraged [20]. However, in practice, data rarely follow a known parametric distribution, motivating the development of nonparametric metrics, such as the MMD [136] or the PAD [59].

Maximum mean discrepancy – The MMD is a nonparametric metric between two distributions, where $D(\mathcal{D}_{s,u}, \mathcal{D}_{t,u})$=0 if, and only if, $p_s(\mathbf{x}) = p_t(\mathbf{x})$ [61, 136]. It maps data into a Reproducing Kernel Hilbert Space (RKHS) and, given a rich family of kernel functions, the MMD provides a statistical measure on the marginal distributions of the

data. The empirical estimate of the (squared) MMD can be defined by,

$$\text{MMD}(\mathcal{D}_{s,u}, \mathcal{D}_{t,u}) = \| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(\mathbf{x}_{s,i}) - \frac{1}{n_t} \sum_{i=1}^{n_t} \phi(\mathbf{x}_{t,i}) \|^2 \tag{3.7}$$

where ϕ is a feature map induced using the "kernel trick" [20]. To extend the MMD to measure the distance between both the marginal and conditional distributions, labels can be used to measure the MMD between data from a given class [107]. This extension of the MMD is referred to as the joint MMD (JMMD), which is given by,

$$\text{JMMD}(\mathcal{D}_{s,l}, \mathcal{D}_{t,l}) = \| \frac{1}{n_s^{(c)}} \sum_{x_{s,i} \in \mathcal{X}_s^{(c)}} \phi(\mathbf{x}_{s,i}) - \frac{1}{n_t^{(c)}} \sum_{x_{t,j} \in \mathcal{X}_t^{(c)}} \phi(\mathbf{x}_{t,i}) \|^2 \tag{3.8}$$

where $\mathcal{X}_s^{(c)}$ and $\mathcal{X}_t^{(c)}$ denote the samples from class c, with $c \in 1, ..., C$, for the source and target domain respectively, $n_s^{(c)}$ and $n_t^{(c)}$ are the samples for class c in the source and target, and $c = 0$ represents the data from all classes, giving the MMD between the marginal distributions.

Considering that many applications of TL either have limited or no label data in the target, it is challenging to directly apply the JMMD in an unsupervised setting. However, it provides an informative measure for quantifying the quality of DA in experimental settings. For a more in-depth discussion on the MMD, the interested reader is referred to [61].

The MMD has been incorporated in the objective function of many DA algorithms, including methods that use kernel functions or DNNs to perform feature extraction [48, 49].

Proxy-A distance – The PAD presents an alternative measure [59], so-called as it provides an estimate for the H-divergence [59]. The PAD was originally proposed to identify changing distributions in streamed test data. It is given by,

$$D_{\mathcal{A}}(\mathcal{D}_{s,u}, \mathcal{D}_{t,u}) = 2(1 - 2\hat{\epsilon}) \tag{3.9}$$

where $\hat{\epsilon} \in [0, 1]$ is the test classification error rate from a binary classifier that discriminates between the domains and $D_{\mathcal{A}} \in [0, 2]$. The minimum PAD is achieved when the domains are completely harmonised, such that the classifier can only guess the domain of the instances – giving a classification accuracy of 50%. For a more comprehensive understanding of the PAD, the reader is referred to [59].

The PAD has been adopted widely in machine learning, in part because of its simplicity to incorporate with popular algorithms such as neural networks. For example, it is the core of generative adversarial networks (GAN) [137], and has been used in many neural network-based DA architectures, notable architectures include the domain adversarial neural network (DANN) [110] and the conditional domain adversarial neural network (CDAN) [112].

3.2.7.2 Transfer component analysis

These distance metrics form the basis of many unsupervised DA algorithms; thus, these algorithms all implicitly make the assumption that an appropriate mapping ϕ that will result in a feature space $p_s(y|\phi(\mathbf{x})) \approx p_t(y|\phi(\mathbf{x}))$ can be learnt by minimising the marginal distribution discrepancy between the available data. Thus, these methods mainly vary based on the machinery used to perform feature extraction and their regularisation approach. A prominent approach that uses a kernel function to learn a nonlinear mapping is TCA [96]. Using a kernel mapping, TCA is generally able to learn efficiently from sparse data, motivating its application to several SHM scenarios [12, 41, 54]. Therefore, TCA is also investigated in this thesis, primarily as a benchmark for the proposed DA methods.

First, TCA reformulates the MMD into matrix form. Given $K = \phi(\mathbf{X})^T\phi(\mathbf{X})$ where $\mathbf{X} = \mathbf{X}_s \cup \mathbf{X}_t \in \mathbb{R}^{(n_s+n_t)\times d}$ and d is the feature dimension, the MMD can be formulated as,

$$MMD(\mathcal{D}_{s,u}, \mathcal{D}_{t,u}) = tr(\mathbf{KX}) \tag{3.10}$$

where $tr()$ denotes the trace of the matrix and $\mathbf{X}$ denotes the MMD matrix, which is defined by,

$$\mathbf{M}(i,j) = \begin{cases} \frac{1}{n_s^2} & \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}_s \\ \frac{1}{n_t^2} & \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}_t \\ \frac{-1}{n_s n_t} & \text{otherwise} \end{cases} \tag{3.11}$$

with the low-rank empirical kernel embedding $\tilde{\mathbf{K}} = \mathbf{KAA}^T\mathbf{K}$ [138], MMD can be rewritten in terms of a transform $\mathbf{A} \in \mathbb{R}^{(n_s+n_t)\times m}$,

$$MMD(\mathcal{D}_{s,u}, \mathcal{D}_{t,u}) = tr(\mathbf{A}^T\mathbf{KMKA}) \tag{3.12}$$

where $\mathbf{A}$ transforms a generic kernel matrix $\mathbf{K}$ such that it is reduced rank and MMD is minimised. The problem can then be formulated as an optimisation problem,

$$\underset{\mathbf{A}^T\mathbf{KHKA}=\mathbf{I}}{min} = tr(\mathbf{A}^T\mathbf{KMKA}) + \lambda tr(\mathbf{A}^T\mathbf{A}) \tag{3.13}$$

H is a centring matrix $\mathbf{H} = \mathbf{I} - \frac{1}{n_s+n_t}\mathbf{1}$, and $\mathbf{1}$ is a matrix of ones. Optimisation is carried out with the regularisation restraint $tr(\mathbf{A}^T\mathbf{A})$, controlling the complexity of $\mathbf{A}$, with the degree of regularisation controlled by λ. It is also subject to a kernel PCA constraint, $\mathbf{A}^T\mathbf{KHKA} = \mathbf{I}$, to preserve discriminative information by maximising variance, which also prevents the trivial solution where $\mathbf{A}$ is found to be zero. This objective can be optimised using the Lagrangian approach; A can then be found by solving the following eigenvalue problem,

$$(\mathbf{KMK} + \lambda\mathbf{I})\mathbf{A} = \mathbf{KHKA}\boldsymbol{\nu} \tag{3.14}$$

where $\mathbf{A}$ is a matrix of eigenvectors with $\boldsymbol{\nu}$ representing the corresponding eigenvalues. The transformed feature space can be found by $\mathbf{Z} = \mathbf{KA} \in \mathbb{R}^{n_s+n_t \times m}$, with which a predictive model could be trained using the source labels and generalised to target test data.

3.2.7.3 Joint distribution adaptation

A major limitation of TCA and DA algorithms with related objectives, such as the DAN or the DANN, is that they assume that minimising marginal distribution divergence alone is sufficient to learn a shared feature space. A modification to the TCA objective can be made to also incorporate the distance between the class-conditional distributions $p_s(y|\mathbf{x})$ and $p_t(y|\mathbf{x})$ – two examples of such approaches include JDA [97] and BDA [102].

The formulation of JDA closely follows that of TCA, presented in the previous section. It aims to incorporate the JMMD into the learning objective by replacing equation (3.12) in TCA, with the following expression,

$$(\mathbf{K}\sum_{c=0}^{C}\mathbf{M}_c\mathbf{K} + \lambda\mathbf{I})\mathbf{A} = \mathbf{KHKA}\boldsymbol{\nu} \tag{3.15}$$

where $\mathbf{M}_c$ refers to the MMD matrix between the conditional distributions, $c \in \{0, 1, ..., C\}$, with $c = 0$ corresponding to the MMD as in TCA; it is defined as,

$$\mathbf{M}(i,j) = \begin{cases} \frac{1}{n_s^{(c)2}} & \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}_s^{(c)} \\ \frac{1}{n_t^{(c)2}} & \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}_t^{(c)} \\ \frac{-1}{n_s^{(c)} n_t^{(c)}} & \begin{cases} \mathbf{x}_i \in \mathbf{X}_s^{(c)} \ \mathbf{x}_j \in \mathbf{X}_t^{(c)} \\ \mathbf{x}_j \in \mathbf{X}_s^{(c)} \ \mathbf{x}_i \in \mathbf{X}_t^{(c)} \end{cases} \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

where $n_s^{(c)}$ and $n_t^{(c)}$ represent the number of source and target samples belonging to class c. Typically, in unsupervised transfer learning, the lack of target labels prevents the direct computation of equation (3.15). Hence, JDA leverages pseudo-labels, which are derived from label predictions using a classifier trained with source data. To find accurate target labels, the JDA mapping is learnt via the following iterative procedure:

1. A TCA mapping is inferred via solving equation (3.14), and an initial source classifier is trained.
2. The initial source classifier is used to obtain pseudo-labels, and a JDA mapping is estimated via equation (3.15).
3. A new source classifier is learnt using the JDA features, and the JDA mapping is updated with the new label predictions; this process is repeated until convergence.

3.2.7.4 Balanced distribution adaptation

Balanced distribution adaptation (BDA) follows a similar approach to JDA, modifying the TCA objective to include the MMD between the class-conditional distributions [102]. It differs from JDA in two ways. First, it introduces a *balance factor* μ that dictates the importance between the marginal and conditional distributions. Thus, the TCA objective equation (3.12) becomes,

$$MMD(\mathcal{D}_{s,u}, \mathcal{D}_{t,u}) = tr\big(\mathbf{A}^T \mathbf{K} \big((1-\mu)\mathbf{M}_0 + \mu \sum_{c=1}^{C} \mathbf{M}_c\big) \mathbf{K} \mathbf{A}\big) \tag{3.17}$$

where $\mathbf{M}_c$ follows equation (3.16). This balance factor is the core reason BDA is implemented for benchmarking in this thesis, as by setting $\mu = 1$, it allows experiments to test whether the pseudo-label approach addresses the limitation of unsupervised methods that only minimise the marginal-distribution discrepancy.

The second difference with JDA is that BDA estimates the pseudo-labels using the untransformed features (it does not include an initial TCA step). As such, it may produce better results where TCA would lead to negative transfer, but relies on the initial feature spaces being sufficiently related that initial label predictions are reasonable.

3.2.8 Mapping transfer learning to PBSHM: current approaches and challenges

In SHM there are various scenarios where the quantity and type of data available limits the application of data-based models. There are also multiple different applicable approaches to transfer; "how to transfer?" for a given scenario will depend on what is being transferred and the data available to apply transfer-learning algorithms. Thus, this section aims to discuss previous research and outstanding challenges related to three core questions for transfer learning in the context of PBSHM.

Previously, it was discussed that for each level of Rytter's hierarchy the data requirements become more demanding. However, the data requirements in a new target structure may be much lower in cases where suitable data are available from related structures within a population. Thus, it is useful to revisit Rytter's hierarchy in the context of transfer learning:

1. **Detection** is typically performed using unsupervised learning by removing the effects of confounding influences, such as EoVs. Thus, to develop a robust damage detector, a target-only dataset would require data across various environmental effects. Otherwise, benign changes to the structure may prompt many unnecessary manual assessments of the data, or even inspections, increasing costs. Therefore, the main opportunity for transfer learning for damage detection is to leverage source domains with data from spurious EoVs, or temperature ranges yet to be observed in the target structure.

2. **Localisation** can be performed by learning regression models to predict the exact location of damage, or classifiers, where damage location is associated to discrete regions or components of a structure. In both cases, labelled data are required across all conditions under which the SHM system is required to make predictions. As such, transfer has significant potential to reduce target dataset requirements, assuming a sufficiently related source dataset (or set of datasets), with varied labelled data.

3. **Classification** also generally requires labels to train classifiers that predict a damage state from a set of possible candidates. Hence, it would also be beneficial in

scenarios where suitable source data are available to either supplement or replace labelled target data.

4. Data for damage **quantification** are generally even more sparse, as they depend on all of the previously required information to identify the type and location of damage, but also labelled observations for how this specific damage state degrades over time. Thus, leveraging these data from a source domain may increase the applicability of SHM systems that can predict damage extent.

5. **Prognosis** is typically performed with physics knowledge, informed using data-based approaches; thus, while transfer could aid prognosis tasks by assisting in achieving the lower levels of the hierarchy, the direct application of transfer learning is less obvious.

Most previous investigations of transfer learning have focused on investigating "how to transfer?", with "when to transfer?" and "what to transfer?" often addressed heuristically, using engineering knowledge. For damage detection, several previous studies have focused on multi-task learning in homogeneous populations, where the dynamic response is likely to only exhibit small variations. For example, Dardeno *et al.* used hierarchical Bayesian models to improve predictions outside of previously observed temperature ranges in a population of four helicopter blades [43]. In [42], mixed effects models were used in a hierarchical Bayesian framework to improve the prediction of power output in wind farms, demonstrating an improvement in performance at the start of operation for less-common operational conditions (near max-power output). In addition, Brealy *et al.* demonstrated the use of hierarchical models for scour detection in wind turbines [139]. These hierarchical Bayesian models improve learning in sparse data scenarios when the variation in the population-level prior has smaller variance (is more informative) than a prior specified by other means. However, in heterogeneous populations, where the structural response may vary significantly in the original feature space, variation in parameters may be high when considering the original input features, potentially limiting the effectiveness of these methods for these scenarios.

A few approaches have also investigated domain adaptation to improve damage detection in heterogeneous populations. Bull *et al.* used a population of six experimental tailplanes to demonstrate transferring a damage detector using TCA [41] and [140] used a similar approach demonstrating an application to bridges, while in [141], TCA was used to learn a shared damage detector for pedestrian bridges. These studies show that a shared model can be used to identify damage with sparse data; however, they do not specifically investigate using source domains with data related to various EoVs that were previously unobserved in the target structure. In [142], it was shown that a DA-based

auto-encoder can be used to improve anomaly detection when the target does not include a wide range of operating conditions; this paper also highlights challenges relating to class imbalance in DA, where *class imbalance* where certain classes have fewer samples. Gardner *et al.* demonstrated a domain-adapted Gaussian mixture model (GMM), showing an application of transfer between datasets from a highway and railway bridge – the Z24 Bridge and KW51 Bridge – aligning data so that joint density estimation could model undamaged data from various environmental conditions in both domains [143]. Moreover, TCA has been demonstrated to address changes caused by retrofitting bridges to facilitate a shared damage detector [144].

The potential value of transfer learning increases at higher levels of Rytter's hierarchy. For example, damage location and classification typically require many expensive-to-acquire labels; thus, both unsupervised and supervised transfer learning approaches could facilitate cheaper SHM systems by reducing the requirement for labels in the target domain. Unsupervised transfer learning particularly has drawn significant attention, as the prospect of learning SHM classifiers without any labels in the target domains clearly offers the largest possible cost reduction, with regard to obtaining a sufficient target dataset. In addition, these studies largely investigate transfer in heterogeneous populations. Since variation between the responses of these structures is often large, the data from different structures often occupy distinct regions of the feature space, motivating a mapping-based approach to DA.

In a PBSHM setting, Gardner *et al.* have shown that DA can be used to transfer localisation labels between numerical and experimental structures [12], two heterogeneous aircraft wings [135], and between pre- and post-repair states in aircraft wings [145] using kernel-based DA, including TCA, JDA, BDA, and adaptation regularization transfer learning (ARTL). In [146], TCA was used for classification of several bolt-loosening states in a population of experimental beams.

A large portion of the unsupervised DA literature for damage classification focuses on DNN-based DA. Notably Xu *et al.* use a physics-informed deep DA approach to align the frequency response of multiple structures to perform damage detection and quantification [147]. Zhang *et al*, presented a method for using DA to automate surface defect classification for steel plates [148] and Narazaki *et al.* demonstrated the application of DA for image-based classification of defects in bridges [149]. In [150], the DANN was used to perform damage localisation between an FE model and an experimental beam. In [151], a modification of the DANN was presented to transfer a damage classifier between two metal plates representing laboratory-scale bridges. Meanwhile, Peng *et al.* presented an architecture similar to the DANN, using the focal loss to account for unbalanced data for health-state classification in wind turbines [152]. There have also been

a number of attempts to apply DNN-based DA to perform fault diagnosis in machines under changing loading conditions and rotation speeds [142, 153–157].

There are several limitations with these previous applications of unsupervised DA. First, these methods typically learn nonlinear mappings and use nonparametric distance metrics in their objective functions. Thus, these methods may require large quantities of data to prevent overfitting. This issue may be more severe in the DNN-based approaches, since these methods typically require tuning of many parameters and use high-dimensional inputs, such as images or the frequency domain of the dynamic response. While in SHM it may be possible to collect large quantities of data from the undamaged structure, in many scenarios only few data can be collected relating to damage or spurious EoVs. In addition, these nonlinear mappings into a latent feature space will not necessarily preserve interpretability of the original feature space, posing additional challenges for interpreting the outcomes of transfer learning.

Furthermore, previous approaches assume that the target domain is unlabelled, but there are data available for each damage class of interest. In practice, near the start of the monitoring campaign, before any damage has been observed in the structure of interest, only data relating to the undamaged target structure would be available to learn a DA mapping. Moreover, while data relating to more varied health states may be acquired throughout the operation of a structure, it is unlikely that data for all health states of interest would be obtained.

A more realistic paradigm would assume that the source dataset includes data pertaining to all the damage states of interest, while the target contains data corresponding to only a subset of the entire label space, i.e. the target label space is a subset of the source label space $\mathcal{Y}_t \subset \mathcal{Y}_s$. This setting is a partial-DA problem, where conventional DA algorithms are often prone to negative transfer [158], posing a critical challenge to the practical application of DA. Typically, this scenario is addressed using instance-weighting [48], and a few studies have investigated approaches for the detection of faults in machines [159, 160]. As far as the author is aware, Wang *et al.* present the only example of partial DA in SHM, implementing a DANN using label predictions to calculate instance weights [161]. This architecture was demonstrated for transfer of a damage classifier capable of classifying location and extent of damage between a numerical and experimental beam, using a high-dimensional frequency spectra as the input. While these instance weighting methods can mitigate negative transfer in some cases, they potentially share the same limitation as instance-based transfer learning – the untransformed data must occupy a similar region of the feature space so that instance weights can be estimated. Furthermore, all of these methods leverage large neural networks with high-dimensional feature spaces, meaning that they may require large quantities of data to prevent overfitting.

A more general treatment of this problem would be to assume that there are unshared damage states in both the source and the target domain, and a DA mapping must be learnt using a shared subset of both datasets. This setting, often called universal DA [48], would also mitigate negative transfer in scenarios where a novel health state is observed in the target domain. To the author's knowledge, the only example of this setting in the context of SHM is [162], where a DNN was used to transfer a damage classifier for the same structure at different temperatures. Generally, this more general setting is addressed using the same methods as partial-DA, but may require weighting both source and target data to learn a mapping from data relating to a shared subset of labels.

During the monitoring campaign of a structure, it may be possible to obtain a few labels via periodic or guided inspections, motivating the application of supervised transfer learning methods. Previous methods that have considered supervised transfer learning in SHM have largely focused on fine-tuning DNNs. For example, fine-tuning has been demonstrated for image-based crack detection [163, 164] and for utilising data generated from FE models [165, 166]. However, these approaches focus on repurposing expensive-to-train neural networks and still require target data representative of all health states of interest. On the other hand, in PBSHM it is likely that labelled data would only correspond to a few rare health states as discussed previously. As far as the authors are aware, the only example of supervised domain adaptation is presented in [53]. In this paper, kernelised Bayesian transfer learning (KBTL) was applied to a numerical case study, where in some structures no data for a given class were available, showing the potential for mapping-based DA to learn a classifier that can predict classes yet to be observed in the target domain. A limitation of this case study is that it only investigated settings where data for most classes were available, whereas in PBSHM often the target dataset may only contain a small subset of the entire label space.

In addition, these labelled data would likely be obtained sequentially (online), throughout the monitoring campaign, i.e. the labelled target data will still only represent a subset of the label space ($\mathcal{Y}_t \subset \mathcal{Y}_s$). As such, the motivation for TL in this scenario is similar to the unsupervised setting – by transferring label information from a source domain, a classifier could be trained that is able to make predictions about yet to be seen classes – but there may also be labels available to infer mappings. This issue motivates online transfer learning methods robust to scenarios where target data are sparse; however, as far as the authors are aware, online transfer learning has not been investigated for SHM, and only a few examples of online transfer learning exist in the machine learning literature [167–170].

A final consideration for "how to transfer?" is that the feature or label spaces may be heterogeneous, requiring the application of heterogeneous transfer learning. For example, the feature spaces may differ if data were acquired at different sampling rates, or from different types of sensors, such as accelerometers and strain gauges. In some cases, these issues could be solved using physical knowledge; for example, in both of the presented situations, natural frequencies could be derived via modal analysis. However, it may be desirable to expedite manual efforts; in these scenarios, DA techniques could be used to find a mapping into a shared feature space. For example, in [53], KBTL was used to account for different sampling rates using frequency-based features, demonstrating improved generalisation in a numerical case study.

The label space may also differ between domains. For example, in heterogeneous populations, two structures may have different components, making it unclear how to transfer labels corresponding to components that are not shared between structures. Similar to the treatment of heterogeneous feature spaces, label space heterogeneity may be resolved by finding a shared label space. It is the opinion of the author that in SHM, it is desirable to use engineering expertise or physical knowledge to ensure labels maintain a physically-meaningful interpretation for use in decision processes.

In some settings, TL methods could be used to automatically select a shared label space. For example, Gardner *et al.* attempted to transfer labels between numerical N-storey structures with varying numbers of floors [7]. In a damage-classification problem, attempting to localise damage to a specific floor, the label space from the source structure does not directly apply to the target structure. To address this, the paper proposed an algorithm that automatically selects a subset of labels using a GMM. However, the paper also discusses the difficulties in interpreting labels in this heterogeneous label space context.

The other two core research questions for transfer learning – "when to transfer?" and "what to transfer?" – have received somewhat less attention in SHM. In the context of PBSHM, "when to transfer?" could be informed by similarity metrics derived from IE models; previous examples have used established graph-based similarity measures [6] and learnt graph-based similarity measures [44, 45, 47]. However, the development of IE models remains an active area of research, with outstanding challenges, such as standardising representations [46], increasing expressiveness [171], and incorporating attributes into similarity measures [172], still being actively investigated. Currently, Gardner *et al.* [135] has presented the only example of incorporating IE models into a full transfer-learning strategy; to the author's knowledge this paper is also the only example investigating the use of data-based similarity measures to inform transfer learning in SHM.

Furthermore, to the author's knowledge "what to transfer?" has not been directly investigated. However, a meta-analysis of the investigations of "how to transfer?" could reveal interesting insights into the features and patterns that can be successfully shared using transfer learning. For example, the various applications of multi-task learning in homogeneous populations suggest that model parameters could be shared in some cases [42, 43], while several DA studies demonstrate successful transfer using natural frequencies as features in heterogeneous populations [7, 12]. A more detailed analysis of the previous research in relation to "what to transfer?" would be an interesting area of research in itself and is left for future work.

Chapter 4

Statistic alignment for transfer with sparse target data

Domain adaptation offers the opportunity to leverage data-based models in scenarios where labelled data are sparse or unavailable in the target domain. However, previous applications of DA for SHM have focused on methods that may be prone to performance degradation under class imbalance; particularly, when target data is not representative of all the health states represented in the source dataset – a scenario known as partial DA. To facilitate DA in these sparse target data scenarios, this chapter introduces statistic alignment (SA) approaches and proposes a method to adapt these methods to the partial DA setting for SHM.

4.1 Introduction

By projecting data into a shared feature space, domain adaptation can allow for a source predictive function to generalise to the target domain, reducing the costs and increasing the diagnostic systems of SHM systems by reducing the data requirements for the target dataset. As such, DA could be applied to various SHM tasks, and has been investigated for damage detection [41, 140, 142, 144], damage classification [53, 135, 143, 146, 149, 152], and damage quantification [147, 157]. The challenge of acquiring target labels has motivated many of these studies to focus on unsupervised DA [56].

Unsupervised DA can enhance the value of costly labels by allowing them to be reused in learning new SHM systems. However, previous applications of DA to SHM have generally assumed the availability of varied (unlabelled) target datasets, including data for each class of interest [53, 135, 143, 146, 149, 152]. A critical limitation of these

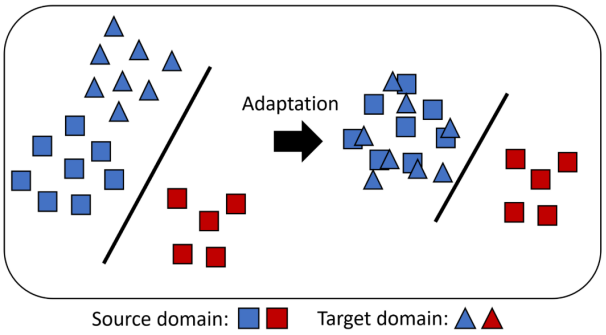


FIGURE 4.1: Illustration of a partial domain adaptation scenario.

methods is that in scenarios where the target dataset does not include observations representative of all states in the source dataset, the problem becomes a partial-DA problem, where conventional DA methods are typically prone to negative transfer [158]. Thus, conventional DA may not be robust in many PBSHM scenarios. An example of a partial-DA problem is shown in Figure 4.1.

As discussed in the previous chapter, in practice, the target dataset may only include data relating to a small subset of the states represented in the source dataset. For example, for many structures, SHM data would be obtained sequentially (online), and for extended periods of the monitoring campaign, only data relating to the undamaged structure would be available. In such scenarios, it would be beneficial to leverage a source dataset that captures the common damage states experienced by the target structure. Using such a source dataset would allow for SHM systems to provide contextual information, such as location, type and extent of damage at the first instance of damage in a target structure. However, this extreme case of class imbalance would require the estimation of DA mapping using only target data relating to a small subset of states represented in the source dataset, which would be a particularly challenging partial-DA problem [158, 173].

To facilitate learning predictive models capable of predicting yet to be observed classes in the target structure, DA methods that are robust in the partial-DA setting must be developed [48, 158]. Generally, this issue requires a subset of the source data to be selected such that source and target data relating to the same health states are used to learn a mapping. However, selecting a subset of data are challenging, particularly without labels. Typical approaches to partial-DA attempt to estimate instance weightings to increase the importance of relevant source data [158, 174–176]. However, unsupervised instance-weighting approaches typically require significant overlap between source and target data in the untransformed feature space, potentially restricting their application to scenarios where populations are strongly related.

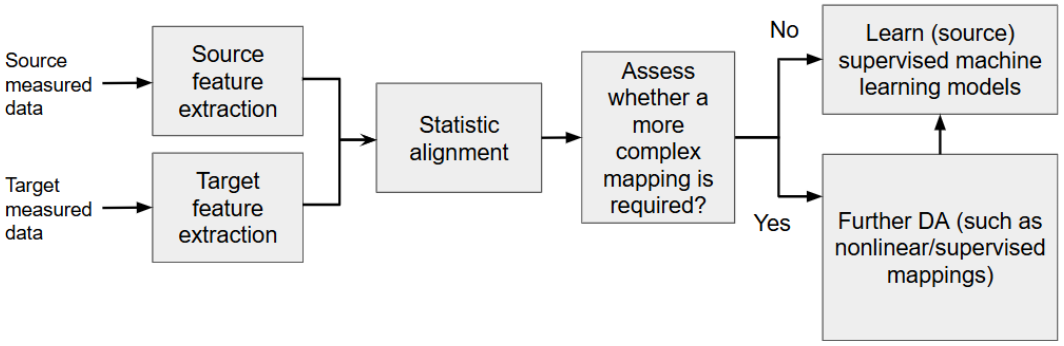


FIGURE 4.2: The proposed flow for using statistic alignment as a stand-alone DA method, as well as part of a DA pipeline for PBSHM.

Motivated by the requirement for robust methods for estimating mappings with sparse target data, this chapter proposes the application of *statistic alignment* (SA) approaches to DA for SHM [48]. Statistic alignment only requires the estimation of lower-order statistics, allowing for its application in sparse-data scenarios [116, 177]. In addition, since these methods project target data into the source feature space via a linear mapping, they can allow for source predictive functions to be reused directly. Furthermore, as they maintain the physical interpretability of the original feature space, they can also facilitate joint visualisation of the source and target data.

To allow for the application of SA in the partial-DA setting, it is proposed that SA can be performed by selecting a shared subset of the source and target data. Specifically, a mapping can be estimated using only data from the start of the operating period (or directly after an inspection), where it can be assumed that data were generated by the undamaged structure, a common assumption made for damage detection [1, 25][1]. Thus, proposed method addresses a major limitation of previously applied TL approaches for SHM [12, 41, 135, 140, 143, 144, 146, 147] and provides the first method that does not require fully representative target datasets, while also addressing issues with previous approaches presented in the DA literature, as it does not rely on distance-based instance weighting [48].

Since the resulting SA methods only require "unlabelled" target data generated by the undamaged structure, these methods could potentially be applied to learn supervised machine learning models prior to the observation of any damage-state data in the target domain, assuming the availability of a suitably similar source structure. Furthermore, in scenarios where SA alone is not sufficient, this chapter discusses the potential for its use as a preprocessing step to improve kernel- and DNN-based DA in SHM.

[1]It is also important to note that auxiliary data (i.e. temperature) could be used to ensure data were generated under similar conditions.

The outline of this chapter is as follows. In Section 4.2, the necessary background is given on DA and SA. The disadvantages of conventional DA are discussed with regard to data availability, class imbalance and partial DA, and the proposed approaches that align the domains using only normal condition data are introduced. Section 4.3 demonstrates that SA can transfer label information between numerical structures in both conventional DA and partial-DA settings, where previously investigated DA methods fail, and it is shown that the proposed methods are particularly robust for partial DA when compared to previous methods. In Section 4.4, the application of the proposed methods is demonstrated on data from a real population of bridges, the Z24 [178] and KW51 Bridges [179]. This population presents two partial-DA scenarios, which include three domains pertaining to the two heterogeneous bridges, as well as a pre- and post-repair state in the KW51 Bridge dataset. Section 4.5 discusses the limitations of only aligning the lower-order statistics and introduces SA as a pre-processing step for other prominent DA algorithms, with another numerical case study suggesting that SA may be an essential pre-processing step for the application of many DA algorithms. Finally, Section 4.6 presents a discussion on SA and discusses future work.

4.2 Partial DA and statistic alignment

This chapter aims to develop methods for *partial DA* problems, where the available target data pertains to fewer classes than the source, i.e. the target label space is a subset of the source $\mathcal{Y}_t \subset \mathcal{Y}_s$ [48]. Thus, the objective of DA is to estimate a mapping using only data relating to a subset of the health states available in the source. This scenario presents an additional challenge to DA, as the mapping should be learnt using data generated by similar underlying processes; however, in unsupervised DA, often the health states data were generated under is unknown in unsupervised DA. This issue is motivated in Figure 4.3, which shows that if the available data represent different subsets of the underlying distributions, while aligning these subsets may lead to low distribution shifts between the available (unlabelled data), it may lead to a poor alignment of the underlying distributions.

A branch of DA, *statistic alignment* (SA), provides an approach to estimate linear mappings by directly matching the lower-order statistics [48]. These transformations are restricted to affine transforms. Although these methods assume that domains can be aligned using linear mappings, they remain applicable in scenarios with sparse target data and can enable effective generalisation when such mappings are adequate to address

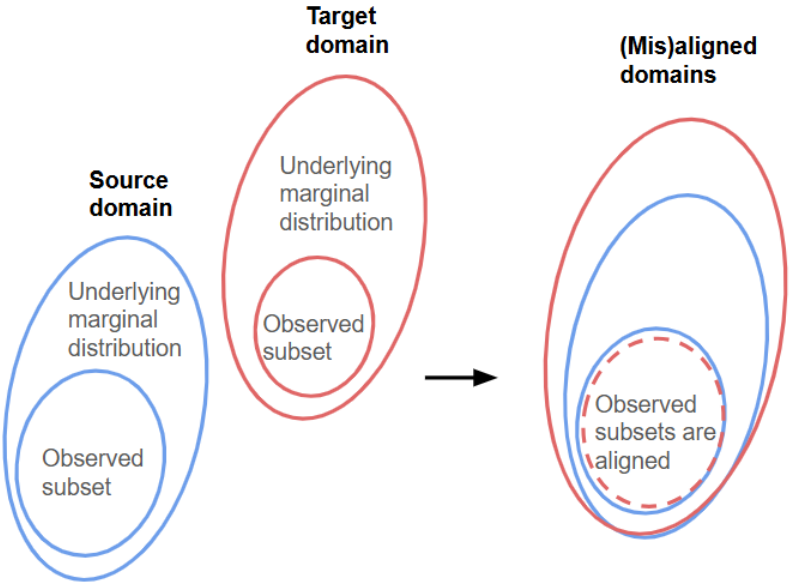


FIGURE 4.3: Demonstration of negative transfer by naively aligning domains by minimising marginal distribution distance between subsets of data that are representative of different generative processes, which is the case in partial DA and in scenarios where there is class imbalance. In the source domain (shown in blue), a wider range of all possible processes has been observed in comparison to the target domain (in red), shown by the inner ovals. Finding a mapping to align the available subsets would result in poor alignment of the underlying distributions (shown on the left), meaning predictive models learnt using source data would still not generalise to the target domain.

distribution shifts. In comparison, methods that minimise measures on the marginal distributions, such as TCA, BDA or the DANN, may require more data to accurately measure divergence [180], and may generalise less well in some scenarios since they generally learn flexible nonlinear mappings, which may overfit when data are sparse, particularly for damage-state data of interest [48].

In addition to being less data intensive, SA could also facilitate better visualisation. Many prominent DA methods [48, 49] project the data into a latent space via a nonlinear mapping. In comparison, SA maintains the structure of the original feature space, as it is restricted to affine transformations, which can be useful for physically-interpretable features, which are common in SHM. For example, features often correspond to some physical process, i.e. an increase in a natural frequency can often be interpreted as a stiffening effect.

Several statistic-alignment methods have been developed in the transfer-learning literature. A prominent SA method is *correlation alignment* (CORAL) [91], which aligns the source correlation with the target. The multiple outlook mapping algorithm (MOMAP) is a similar approach that aligns the principal components [114]. He *et al.* proposed Euclidean alignment (EA), which aligns the means of the covariance matrices for 2D

electroencephalogram (EEG) features, and demonstrated that SA methods can be generalised to cases with multiple sources [116].

A related branch of DA concerns batch normalisation approaches [117–119], which align the means and standard deviations of the *activations* in deep neural networks. These methods differ to SA, as the statistics of the original data are not directly aligned.

4.2.1 Standardisation as statistic alignment

Standardisation, a form of normalisation, is commonly used in conventional machine learning to give each feature equal treatment [24]. In DA, it also has the potential to align the means $\boldsymbol{\mu}$ and standard deviations $\boldsymbol{\sigma}$ of the marginal distributions $p_s(\mathbf{x})$ and $p_t(\mathbf{x})$ in an unsupervised manner, with the following transformations,

$$\mathbf{z}_{s,i} = \frac{\mathbf{x}_{s,i} - \boldsymbol{\mu}_s}{\boldsymbol{\sigma}_s} \tag{4.1}$$

$$\mathbf{z}_{t,i} = \frac{\mathbf{x}_{t,i} - \boldsymbol{\mu}_t}{\boldsymbol{\sigma}_t} \tag{4.2}$$

where $\mathbf{z}_{s,i}$ and $\mathbf{z}_{t,i}$ are the transformed source and target samples respectively, $\boldsymbol{\mu}_s$ and $\boldsymbol{\mu}_t$ are the means of the source and target; $\boldsymbol{\sigma}_s$ and $\boldsymbol{\sigma}_t$ are the respective standard deviations. In this chapter, this form of standardisation will be referred to as *A-standardisation*. On the other hand, standard practices for machine learning would suggest that the statistics applied to all data should be the same; for example, if some data from the source and target domain are considered training data the statistics would be calculated from $\mathbf{X} = \mathbf{X}_s \cup \mathbf{X}_t$, which may remove the effect of measurement scale without changing relative mean distance or scale between the domains. To demonstrate why this practice may lead to negative transfer in DA, this method is demonstrated in Section 4.2, and will be called *N-standardisation* throughout this chapter.

4.2.2 Correlation alignment

A-standardisation aligns the domains, ignoring the correlation between features. Correlation Alignment (CORAL) [91], extends this method to also align the covariance. This is achieved by transforming the source domain via a linear transformation matrix A, such that,

$$\mathbf{A} = \min_{\mathbf{A}} ||\mathbf{A}^T \mathbf{C}_s \mathbf{A} - \mathbf{C}_t||_F^2 \tag{4.3}$$

where $\mathbf{C}_s$ and $\mathbf{C}_t$ denote the covariance matrices of the source and target, respectively, and $||\cdot||_F$ is the Frobenius norm.

A drawback of CORAL in comparison to standardisation-based approaches is that estimating covariance suffers from the curse of dimensionality. If the number of observations in $\mathbf{X}_s$ or $\mathbf{X}_t$ are smaller than the number of dimensions d ($n < (d+1)$), the covariance matrix will be singular [27]. This can be a serious problem since vibration data are often high-dimensional. In damage detection, the curse of dimensionality has motivated the use of ensemble methods to robustly estimate covariance [26, 27].

4.2.3 Normal condition alignment

When collecting SHM data during the operation of a structure, datasets commonly exhibit class imbalance, as some health states are naturally more common than others. Thus, current SA methods may not be robust in many SHM applications. For example, given two bridge datasets, it is unlikely that both bridges will have the same quantity of available data from every damage state and environmental condition. It is also unrealistic to assume that the target will contain some data from each health state in the source – motivating the application of partial DA. As such, the sufficient statistics of both datasets would summarise different behaviours and aligning the domains based on these moments will not align the underlying distributions, leading to negative transfer, as illustrated in Figure 4.3. Similarly, prominent DA methods typically only minimise the marginal-distribution-divergence measures between all the available data [48].

To address this issue, *normal condition alignment* (NCA) is proposed, which aims to reduce the risk of aligning data generated under different structural states by utilising the assumption that data gathered at the start of a structure's operation were generated by the "normal condition" – a common assumption made for novelty detection [181]. In NCA, the source domain is first standardised via equation (4.1) to centre the data and give features equal treatment. The normal condition of the target domain is then aligned with that of the source by,

$$\boldsymbol{z}_{t,i} = \left(\frac{\boldsymbol{x}_{t,i} - \boldsymbol{\mu}_{t,n}}{\boldsymbol{\sigma}_{t,n}} \right) \boldsymbol{\sigma}_{s,n} + \boldsymbol{\mu}_{s,n} \tag{4.4}$$

where $\boldsymbol{\mu}_{s,n}$, $\boldsymbol{\mu}_{t,n}$ and $\boldsymbol{\sigma}_{s,n}$, $\boldsymbol{\sigma}_{t,n}$ are the means and standard deviations of the normal condition data for the source and target respectively.

To motivate the use of this method to predict previously-unobserved target classes, consider the scenario where variation between the datasets is assumed to be limited to a scale vector $\mathbf{a}$, and translation vector $\mathbf{b}$. In this case, the differences between domains

can be expressed by,

$$\mathbf{X}_s = \mathbf{a}\mathbf{X}_t + \mathbf{b} \tag{4.5}$$

Given that the set of feature vectors can be expressed as $\mathbf{X} = \mathbf{X}_n \cup \mathbf{X}_d$, where $\mathbf{X}_n$ is normal condition data and $\mathbf{X}_d$ is damage-state data, and the $\mathbf{a}$ and $\mathbf{b}$ define an affine transformation, it follows that the transformations for the entire domain can be learnt from the subsets $\mathbf{X}_n$, suggesting it may be possible to learn a mapping that can generalise to unseen classes using only a subset of available target data. It is noted that using only a subset of the data reduces the available data to learn the statistics, but the lower-order statistics should be able to be estimated with a limited sample size.

The advantages of aligning the domains using a mapping based on the normal condition are demonstrated in Figure 4.4, which presents a toy problem comparing the various standardisation approaches discussed. The toy problem consists of a source domain with three Gaussian clusters and a target with two classes. Hence, the problem is a partial-DA scenario, and there are also differences in the class imbalance between the available classes, with 20 samples in Class 0 (shown in red) for both domains, but with 8 and 4 samples in source and target, respectively, for Class 1 (shown in blue). Figure 4.4(b) presents N-standardisation, showing that this method does not reduce (relative) distribution shift. In Figure 4.4(c), conventional statistic alignment in the form of A-standardisation has aligned the standard deviation and mean of the two classes in the target to the three in the source, leading to poor alignment. On the other hand, NCA is shown to address this issue by only considering data from the normal condition, aligning the correct classes because the red and blue classes have a similar structure in both domains, shown in Figure 4.4d.

Selecting a subset of data using engineering knowledge in this way also addresses a critical limitation with previous partial-DA algorithms. Previous partial-DA methods presented in the machine learning literature [48, 155, 158] typically use instance weighting to select a shared subset of data. Thus, these methods may have similar limitations as instance-based approaches to DA, i.e. finding instance weights is challenging when $\frac{p_t(\mathbf{x})}{p_s(\mathbf{x})} = 0$. Meanwhile, explicitly aligning the marginal distributions of data generated by the normal condition presents a method of selecting data corresponding to a shared health state, and it is still applicable when the original data occupy distinct regions of the feature space.

An important consideration is that a finite sample of data collected from the undamaged structure may still not be representative of the same underlying structural states, as variations related to EoVs will still influence the response of the structure [182]. In some scenarios, additional measurements relating to these EoVs could be measured, allowing for further sample selection such that the normal conditions of each domain contain

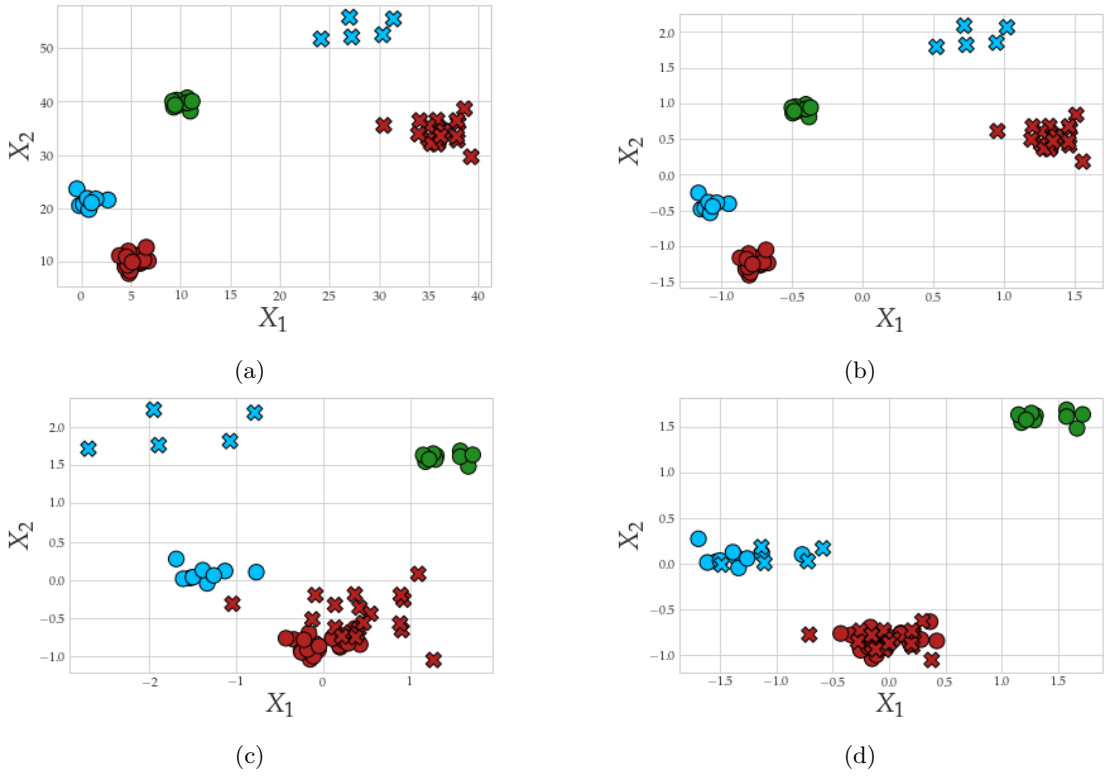


FIGURE 4.4: Demonstration of aligning a toy example, panel (a) presents a partial-DA problem, consisting of three Gaussian clusters in the source, and two in the target. Panel (b) gives the result of N-standardisation, (c) A-standardisation and (d) NCA. The source and target data are represented by (○) and (×) respectively; Classes 0,1, and 2 are depicted in red, blue and green, respectively.

data relating to similar EoVs[2]. In this chapter, samples are selected corresponding to similar EoVs (ambient temperature), and the remaining samples are excluded during the estimation of mapping. However, future work could consider additional methods to select datasets representative of similar conditions; for example, using instance weighting based on auxiliary measurements, such as temperature, wind speeds and loading conditions.

4.2.4 Normal-correlation alignment

CORAL may provide improvements where there are shifts in the correlations between features, but it may also be prone to negative transfer under class imbalance, as accurately estimating the underlying correlation would be challenging. A modification of CORAL could exploit information in the correlation between the normal condition data – Normal-Correlation Alignment (NCORAL). The first step of NCORAL is to apply

[2]It should be noted that these measurements will give an indication of what data are affected by certain EoVs, but they will not guarantee that the data were generated by a given process.

NCA; correlation alignment is then given by,

$$\mathbf{A} = \min_{\mathbf{A}} ||\mathbf{A}^T \mathbf{C}_{s,n} \mathbf{A} - \mathbf{C}_{t,n}||_F^2 \tag{4.6}$$

where $\mathbf{C}_{s,n}$ and $\mathbf{C}_{t,n}$ are the correlations of the normal condition data in the source and target, respectively. NCORAL extends the assumption that the domains differ by a scale and translation made by NCA by also considering a rotation. It is noted that NCORAL learns the correlation from a subset of the entire data, so it may have additional issues relating to the curse of dimensionality; this issue may be addressed by ensemble methods for high-dimensional features [181].

While parallels can be drawn between SA and data normalisation in conventional machine learning, it is important to emphasise that SA can facilitate DA in sparse data scenarios without any additional DA. As such, SA is first investigated as an independent DA method. Following it's investigation as an independent DA method, it is also shown to be an important form of pre-processing to aid transfer via conventional DA, such as kernel- or DNN-based DA methods.

4.3 Case study: numerical three-storey shear structure population

This section presents a numerical case study, demonstrating the ability of SA to transfer label information to facilitate damage localisation via a classification approach, with a limited quantity of data and no labels in the target domain. A number of SA methods are applied – A-standardisation, CORAL, NCA and NCORAL – and these are benchmarked against N-standardisation (showing the result of applying traditional SHM methods to a population). In addition, a range of DA methods that encompass the general DA approaches used in previous research of transfer learning in PBSHM are implemented for comparison – TCA [96], BDA [102], the geodesic flow kernel (GFK)[124], and the DANN [110].

4.3.1 Data simulation: numerical three-storey shear structure population

The numerical population used in this chapter consists of two shear structures modelled as 3DoF lumped–mass models (following the approach in [12]). The masses of each DoF were assumed to be a rectangular volume, representing a floor, parameterised by a length l_m, width w_m, thickness t_m, and density ρ, with the density sampled from a

Gaussian distribution to represent manufacturing variation. The masses were assumed connected by four cantilever beams in parallel, so stiffness is given by $k = 4k_b$, where the stiffness of each beam was found as the tip stiffness of a cantilever beam, $k_b = \frac{3EI}{l_b^3}$. The elastic modulus E was also drawn from a Gaussian distribution for each sample to introduce variability. Damping c was not derived from a physical model; instead, it was drawn from a Gamma distribution directly.

Damage at a given storey was modelled as an open crack on one of the four beams, located at the midpoint of the beam. It was modelled as a reduction in stiffness as in [183]; thus, $k = k_d + 3k_b$, where k_d is the stiffness of a damaged cantilever beam.

Having obtained the parameters of the model, the damped natural frequencies ω_d were calculated by solving the eigenvalue problem; the first three were used as features, $\mathbf{X} \in \mathbb{R}^{n\times 3}$.

	Unit	Source	Target
Beam geometry $\{l_b, w_b, t_b\}$	mm	$\{300, 40, 8\}$	$\{160, 25, 6\}$
Mass geometry $\{l_m, w_m, t_m\}$	mm	$\{400, 400, 40\}$	$\{300, 250, 25\}$
Crack geometry $\{l_{cr}, l_{loc}\}$	mm	$\{20.0, 150\}$	$\{12.5, 80\}$
Elastic modulus E	GPa	$\mathcal{N}(210, 1 \times 10^{-9})$	$\mathcal{N}(71, 1 \times 10^{-10})$
Density ρ	kg/m^3	$\mathcal{N}(7800, 50)$	$\mathcal{N}(2700, 10)$
Damping coefficient c	Ns/m	$\mathcal{G}(8, 0.8)$	$\mathcal{G}(50, 0.8)$

TABLE 4.1: Properties for the numerical 3-storey shear structure case study.

Material properties and geometry for each structure are detailed in Table 4.1. Data were simulated for the normal condition and three damage classes, representing damage located at each spring DoF. For the source structure, 200 samples were collected for each class and labels were assumed known $\{\mathbf{x}_{s,i}, y_{s,i}\}_{i=1}^{n_s}$, where $n_s = 800$. In the target structure, 100 samples were collected for each class $\{\mathbf{x}_{t,j}\}_{j=1}^{n_t}$, where $n_t = 400$. The target labels were assumed to be unknown for all classes apart from the normal condition. In addition, a separate test target dataset was generated via the same procedure as the training set, and also included 400 samples.

4.3.2 Benchmarking procedure: numerical three-storey shear structure population

To demonstrate that SA can robustly transfer label information between different structures, the labels for the three discrete damage locations, corresponding to each DoF, and the normal condition in the source were transferred to the target to facilitate damage localisation for the same health states, without using target labels.

Four SA methods were considered; two of which do not attempt to account for bias in the data A-Standardisation given by equations (4.1) and (4.2), and CORAL, as well as the proposed approaches – NCA and NCORAL. In addition, to motivate the requirement for transfer learning in this case, N-Standardisation (i.e. calculating the statistics from $\mathbf{X} = \mathbf{X}_s \cup \mathbf{X}_t$) was applied, which can be viewed as the "no DA" case.

Furthermore, SA was compared to several prominent DA approaches, including TCA, BDA, the DANN, and the GFK. These were chosen as TCA provides a comparison with a kernel-based method that aims to minimise the marginal-distribution divergence (via the MMD), while the DANN indicates the performance of methods minimising the proxy-A distance via deep neural networks. Thus, TCA and the DANN give examples of methods using one of the most prominent objectives for DA – a marginal-distribution divergence – with the main forms of feature extraction used in DA – kernels and DNNs. In addition, BDA was implemented to investigate the use of pseudo-labels to attempt to minimise the MMD between the class-conditional distributions $p_s(\mathbf{x}_s|y_s)$ and $p_t(\mathbf{x}_t|\hat{y}_t)$. Finally, the GFK provides a comparison with methods that instead take a geometric approach to DA [124]. The features were standardised for each of these methods using N-standardisation, i.e. no initial methods to account for distribution shift were applied.

After alignment, a k-nearest neighbours classifier (kNN), with one neighbour, was learnt on a source training set and used to classify data in a target test set, although any appropriate classifier could be applied following SA. A kNN is used here (and throughout this thesis), because if the source and target distributions are well aligned, data should be close in Euclidean distance.

Hyperparameter selection via cross-validation is challenging in unsupervised DA, because labels in the target are assumed unavailable; thus, it was assumed that parameters could be selected using engineering knowledge. The MMD-based methods (TCA and BDA), utilise an RBF kernel with the length scale chosen as the median of the pairwise distances [61]; the dimension of the feature space was reduced by 1 and the Frobenius-norm regularisation parameter was arbitrarily chosen as 0.1 following [12]; results were found to be largely insensitive to this value in [96]. BDA includes a "balance factor" to control the contribution of the MMD between the marginal and conditional distributions; this was chosen to be 0.5 as suggested in [102] for the unsupervised setting. The dimension of the GFK subspace must be 1 in this case study, since there is a requirement that it is less than half the original dimension [124]. The architecture of the DANN was chosen from a similar case, given in [184]. The DANN is sensitive to the random initial weights, so 100 repeats were run and the mean and one standard deviation of the results are given.

Results are given for a test dataset in the target domain, where the evaluation metric used is the macro-F1 score (see Section 2.2.2).

4.3.3 Results: numerical three-storey population

The unnormalised natural frequencies of the source and target can be found in Figures 4.6 and 4.7 respectively. Prior to DA, the differences between the domains are large, with the absolute values in the target data being about a factor of two larger than the source frequencies. Estimations of the class data distributions are given by kernel density estimation (KDE) (see [20] for more details), shown on the diagonal of each figure.

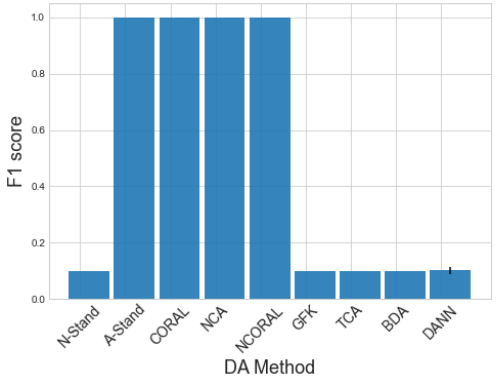


FIGURE 4.5: Classification performance of a kNN on the target domain after DA on the numerical three-storey population. The result of the DANN is given as the mean of 100 repeats with one standard deviation shown by a black line.

Results comparing the F1 scores obtained using each DA method can be found in Figure 4.5. As expected, the kNN trained on the N-standardised (unadapted) features, which can be considered as naively applying a classifier trained via a traditional SHM approach to another structure, led to poor generalisation of the source classifier. Following any of the SA methods tested, perfect classification could be achieved. Furthermore, the conventional DA algorithms did not improve classification performance upon N-standardisation. These methods should be able to align both the lower- and higher-order statistics, so this result may suggest that large differences in scale and mean may make learning challenging in conventional DA algorithms. This issue may be related to the quantity of data. It is noted that, as the mean and standard deviation are relatively simple to calculate from the data directly, SA could be used to reduce the mean and scale discrepancies before these algorithms are applied, motivating the idea of SA as a pre-processing tool, which is discussed in Section 4.4.

The features given by A-standardisation can be found in Figure 4.8. It can be seen that even though the two structures have significant structural differences, the domains

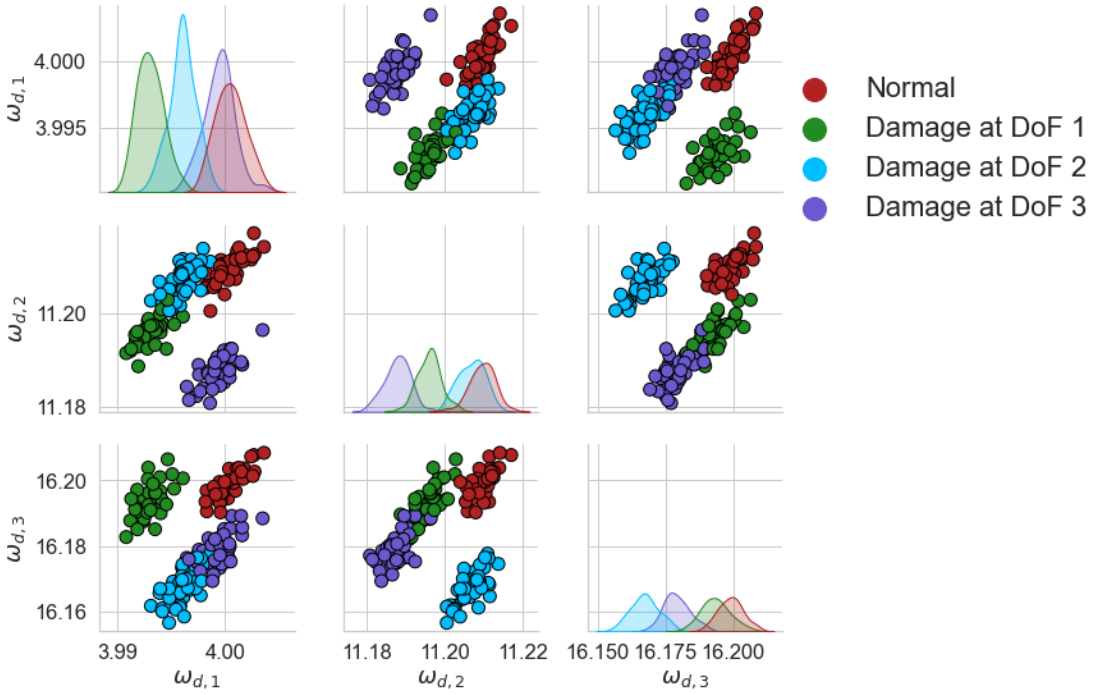


FIGURE 4.6: Unnormalised damped natural frequencies of the source structure of the numerical population of three-storey structures, in Hz. A random subset of 20% of the size of the dataset was used for visualisation.

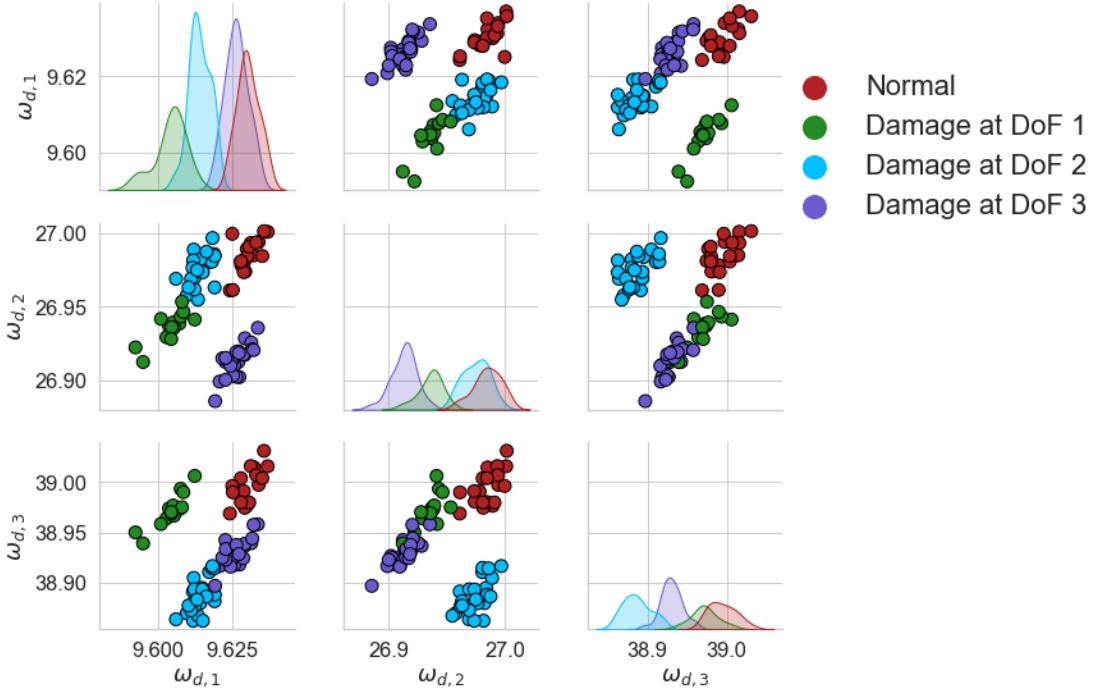


FIGURE 4.7: Unnormalised damped natural frequencies of the target structure of the numerical population of three-storey structures, in Hz. A random subset of 20% of the size of the dataset was used for visualisation.

appear to be well aligned, with data in both domains occupying the same regions of the feature space. This result perhaps suggests differences between structures caused by size

and material properties may only lead to differences in mean and scale – assuming a linear response – motivating the application of methods that estimate linear mappings for DA in SHM.

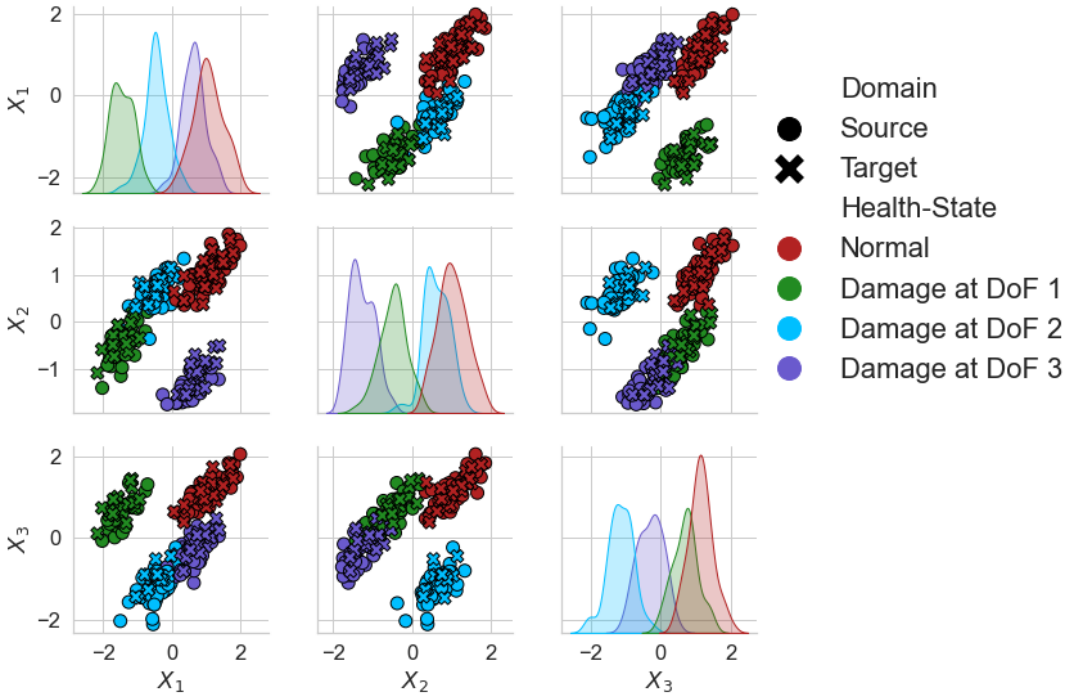


FIGURE 4.8: A-standardised features of the numerical three-storey shear structure population of structures. The source and target are depicted by (○) and (×) respectively. A random subset of 20% of the size of the dataset was used for visualisation.

4.3.4 Results: partial domain adaption with the numerical three-storey population

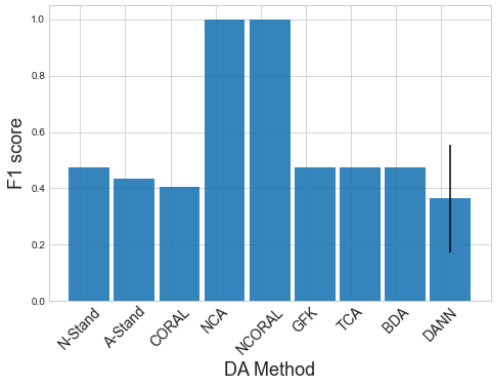


FIGURE 4.9: Classification performance of a kNN on the target domain after DA on the numerical three-storey population in a partial-DA scenario. The result of the DANN is given as the mean of 100 repeats, with one standard deviation shown by a black line.

To investigate the robustness of previously applied methods to a more realistic partial-DA setting, the target domain was downsampled to only include 10 samples from one damage state, corresponding to damage on the third storey; all other target data were excluded from this case study. Thus, this case study presents a partial-DA problem as the target datasets only have data relating to two classes, which are a subset of the four classes in available in the source dataset.

The F1 scores obtained using each method for this partial DA case study are given in Figure 4.9. As with the previous case, all the conventional DA methods failed to improve generalisation[3]. In addition, A-standardisation and CORAL caused negative transfer.

Figure 4.10 shows that aligning the global statistics using A-standardisation in this scenario has caused the two available target classes to be spread across the four classes in the source. On the other hand, Figure 4.11 shows that by selecting only normal condition data to estimate a DA mapping better alignment could be achieved, with both the undamaged and damaged data occupying a similar region of the feature space in both domains. This result could be achieved without any available damage-state data in the target, allowing damage diagnostics in real-time using contextual information from a source domain.

While this case study demonstrates the promise of leveraging varied source datasets using these partial DA techniques, it is important to note that obtaining these varied source datasets may also be a challenge. In some scenarios, such as cases where the source domain is a numerical model or lab structure, these datasets may be feasibly obtained. In other cases, such as where the source dataset is obtained from a structure in operation, it is unlikely that the source dataset will contain a large quantity of varied labelled data. Thus, in some cases, it may also be appropriate to consider technologies such as multi-source DA, which can leverage information from multiple sources, to aggregate information and increase the feasibility of obtaining diverse training datasets. Investigating methods to obtain and leverage diverse source training datasets is an important direction of future work and is discussed further in Chapter 8.

4.4 Case Study: partial domain adaptation with the Z24 and KW51 Bridges

In a population of real structures, data may be influenced by a wide range of additional effects, such as EoVs and nonlinearities. Thus, this section investigates the use

[3]The macro-F1 score for N-standardisation is higher than the previous case because there are only two classes in the target domain, so classifying all classes as one class results in a macro-F1 score of 0.5.

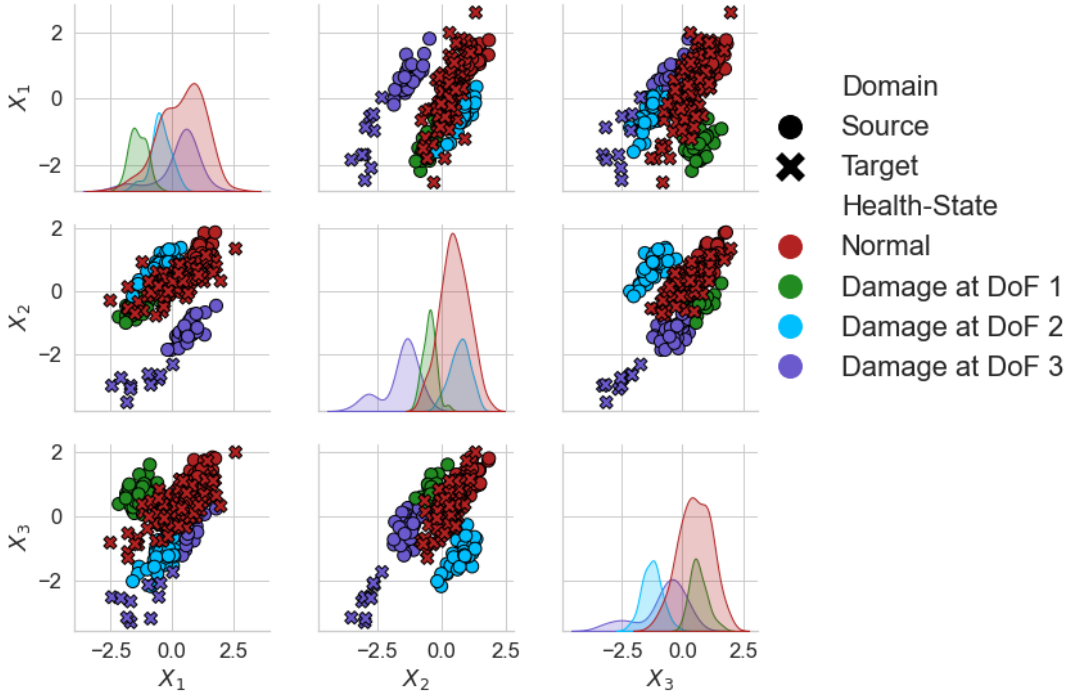


FIGURE 4.10: A-standardised features of the numerical three-storey shear structure population of structures in a partial-DA scenario, with data only from a subset of the source classes in the target. The source and target are depicted by (○) and (×) respectively. A random subset of 20% of the size of the dataset was used for visualisation.

of NCA/NCORAL using a real population of structures consisting of the Z24 [178], and KW51 Bridges [179]. This population consists of two partial-DA problems that are challenging to solve using previously investigated DA approaches, as will be discussed in the following section.

4.4.1 The Z24 Bridge and KW51 Bridge datasets

The Z24 Bridge dataset is well-studied, with data-based approaches being able to identify key events during the monitoring campaign [2, 181, 185–191]. The Z24 Bridge was a concrete highway bridge in Bern, Switzerland, which, as part of the SIMCES project, was used for an SHM campaign before its demolition in 1998 [178]. The first four natural frequencies were found via operational modal analysis (OMA), from the collected acceleration responses. Small-scale damage was introduced by lowering the pier incrementally on the 10^{th} of August 1998, before more severe damage occurred, starting with the failure of a concrete hinge on the 31^{st} of August 1998. For a more in-depth overview of this dataset, see [185].

The KW51 Bridge is a steel bowstring railway bridge in Leuven, Belgium. A monitoring campaign was carried out between 2018 and 2019 for 15 months. The acceleration

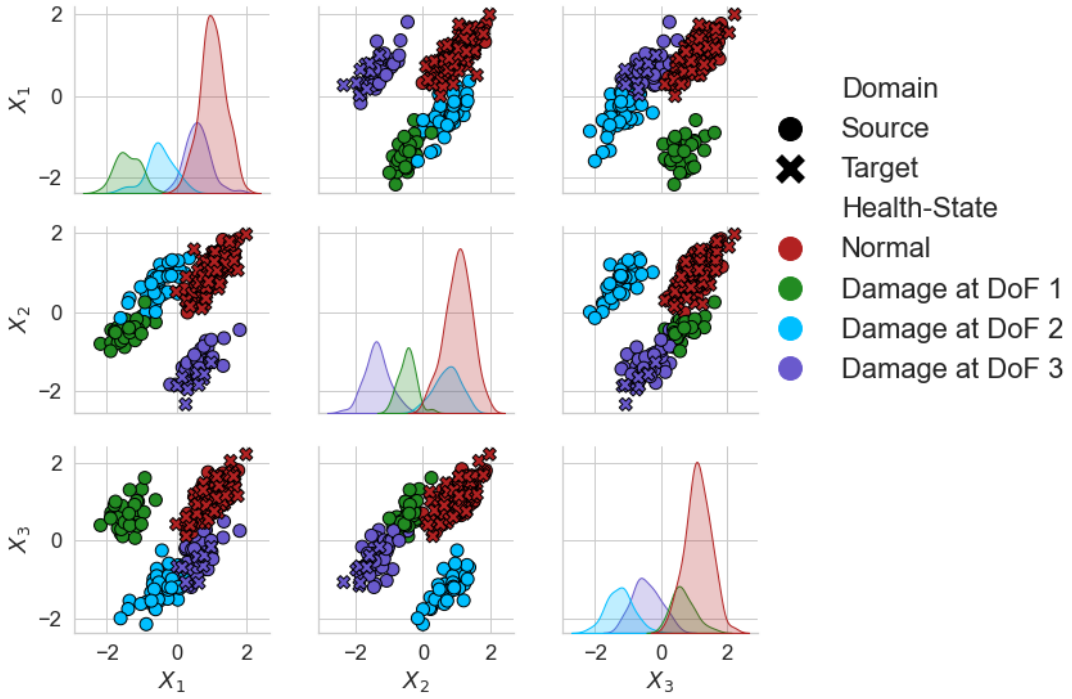


FIGURE 4.11: NCA features of the numerical three-storey population of shear structures in a partial-DA scenario, with data only from a subset of the source classes in the target. The source and target are depicted by (○) and (×) respectively. A random subset of 20% of the size of the dataset was used for visualisation.

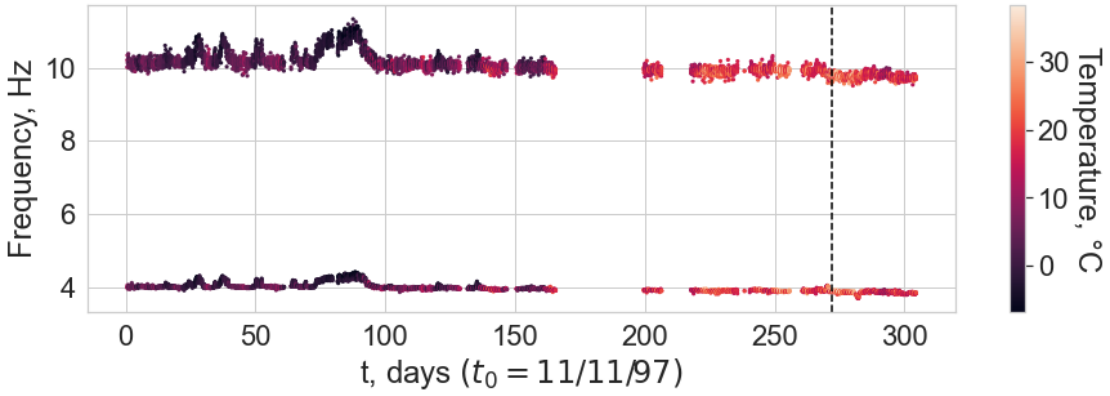


FIGURE 4.12: The first (bottom) and third (top) natural frequencies of the Z24 Bridge dataset. The first instance of damage, commencing on the 10^{th} of August, is indicated by the black line.

responses were used to obtain the first 14 natural frequencies via OMA. During the monitoring campaign, each diagonal member was retrofitted with a steel box to strengthen the design of the bridge, with the retrofit beginning on the 15^{th} of May 2019 and completed on the 27^{th} of September 2019. Novelty detection of the retrofit has been successfully demonstrated using robust PCA and linear regression trained on the pre-retrofit data [192]. For a full description of the dataset, the reader is referred to [179].

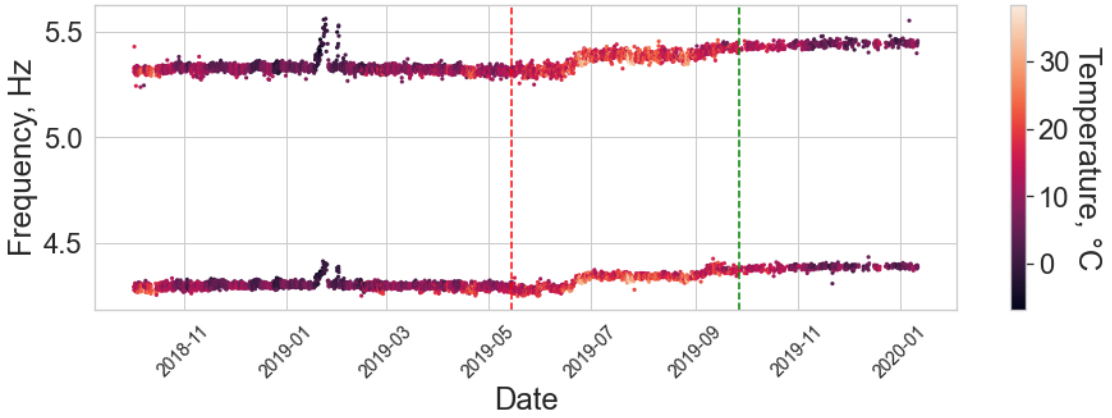


FIGURE 4.13: The tenth (bottom) and twelfth (top) natural frequencies of the KW51 Bridge dataset. The red vertical line indicates the start of the retrofit, and the green is the end.

Even though the two bridges differ significantly by design, a subset of the natural frequencies can be chosen where there are similarities in the modal response of the bridges. Specifically, the first and third natural frequencies of the Z24 Bridge and the tenth and twelfth natural frequencies of the KW51 Bridge correspond to vertical bending modes of the deck and have a similar nodal pattern (for visualisation of the mode shapes see, [178] and [179] for the Z24 Bridge and KW51 Bridge respectively) – a further discussion on feature selection approach for selecting related natural frequencies is presented in the following chapter.

While these natural frequencies were chosen as they correspond to similar modes, it is important to note that there may be a loss of discriminative information by discarding a number of less related natural frequencies from each bridge, highlighting a potential limitation of transfer learning. This issue is also further discussed in the following chapters. Another important limitation to highlight is that each bridge has different sensor locations; thus, while in this case natural frequencies can be extracted, which are not affected by sensor location, assuming they can be identified, some features may not be directly comparable when sensor networks differ, and direct comparison of mode shapes via methods such as the modal assurance criterion (MAC) is challenging.

The corresponding natural frequencies are visualised in Figures 4.12 and 4.13; it can be seen that both bridges experience stiffening effects because of below-freezing conditions, and the Z24 Bridge dataset contains additional information corresponding to damage. For estimating mappings and visualisation of shared effects, normal condition data were split into two classes – corresponding to ambient temperatures ($T > 0°C$) and freezing temperatures ($T < 0°C$). In addition, it is clear that the KW51 Bridge was stiffened by the repair, shown by the increase in frequencies in comparison to data at similar

temperatures prior to repair, meaning that the pre- and post-repair states should be considered as two domains with different underlying joint distributions.

To summarise, within the two datasets there are three domains that need aligning via DA. Clearly, the responses of the Z24 Bridge and KW51 Bridge are different and require some form of DA to facilitate the implementation of a shared predictive model. In addition, following repair, the response of the KW51 Bridge changed. This phenomenon has previously been investigated; it was found that for pre-repair data to be used to predict the health state for the post-repair structure, it must be realigned via DA [145]. Thus, to perform future predictions on the KW51 Bridge post-repair using the Z24 Bridge and KW51 Bridge pre-repair data, two partial-DA problems must be addressed:

1. Align the pre- and post-repair data so that the KW51 Bridge data can be treated as one domain. This problem is defined as one of partial DA, because there are data from the ambient and low temperature normal conditions in the pre-repair state, but only the ambient normal condition in the post-repair state.
2. Align the KW51 Bridge and Z24 Bridge data. This problem is one of partial DA, because there is an additional class in the Z24 Bridge dataset relating to damage on the deck.

In a sense, as this case study aligns three domains, this case study shows the potential for using NCA/NCORAL to align multiple domains, either in a multi-source or multi-task setting. Conversely, it may be challenging to find a shared space across multiple domains using previously-studied DA methods. For example, the methods that find a shared low-dimensional latent space (TCA, BDA, the GFK etc.) will change the information content and dimension of the original features, meaning alignment of multiple domains by sequentially applying these methods would be challenging for two reasons. First, since the transformed and untransformed features have different feature dimensions, this would become a heterogeneous DA problem, which requires specialised methods and is generally considered more challenging [49]. In addition, there is no longer necessarily a direct correspondence between the untransformed features (natural frequencies), and the transformed features (latent features), which may make transfer more challenging.

Meanwhile, the DNN-based methods shared the former problem, and also often require labels for a classification task in the source. This classification task maintains the discriminative information in the domains. Although temperature data are available for each bridge, there are no ground-truth labels that indicate the stiffening effects caused by low temperatures. Therefore, there are only noisy labels to maintain discriminative information between the ambient and low-temperature normal conditions. Even if these

labels are used, there are only 110 samples corresponding to low temperatures in the KW51 Bridge dataset, so if an unbiased subset of data are chosen, there would be 220 samples in the pre-repair source domain, which is likely too small to train a DNN. As such, this section demonstrates that NCA and NCORAL can be used to align the pre- and post-repair states of the KW51 before NCORAL is used to align all the data of the KW51 to the Z24 Bridge datasets.

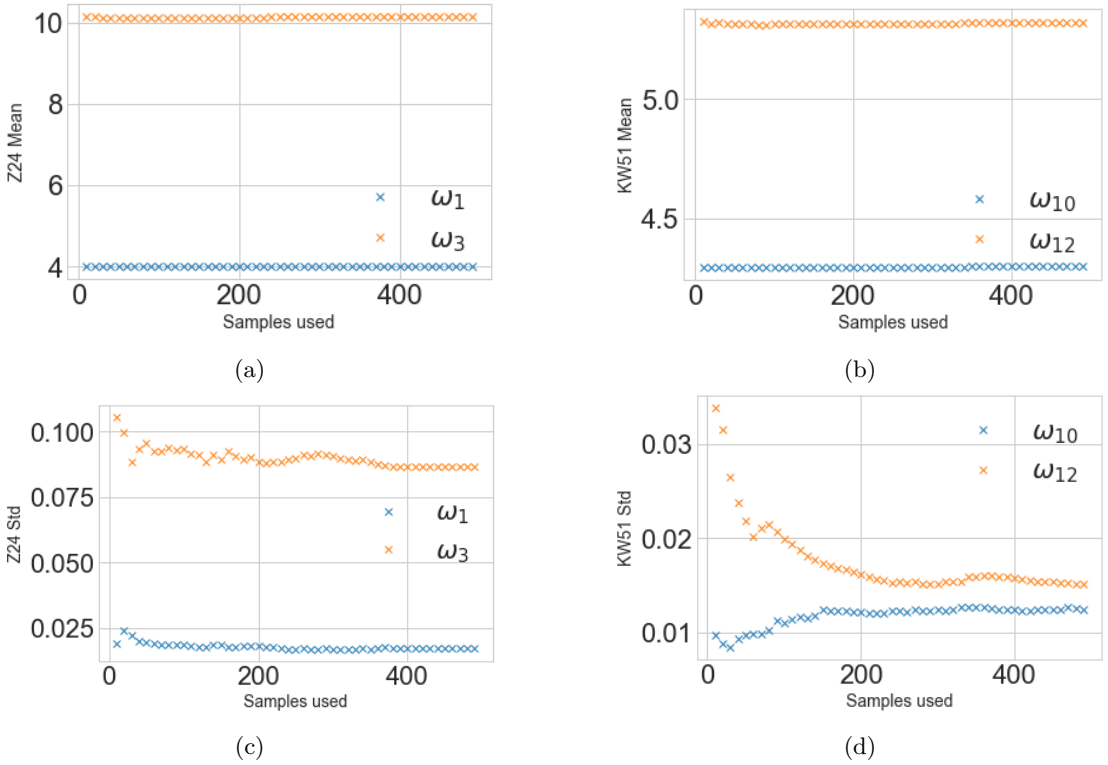
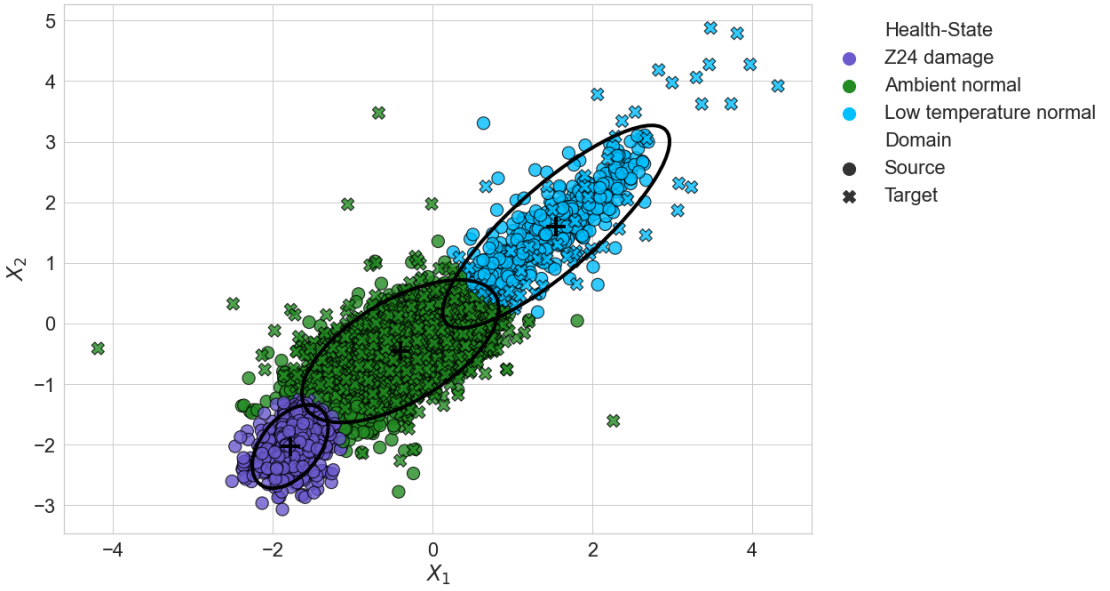


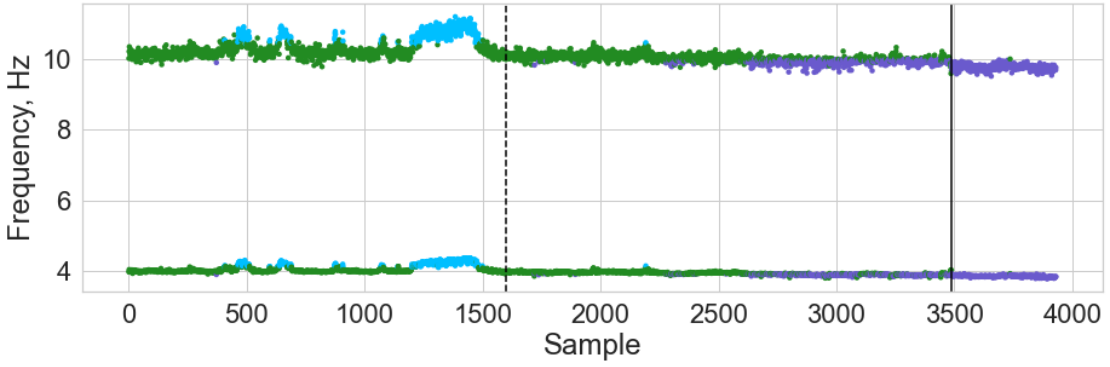
FIGURE 4.14: Sensitivity analysis for calculating the means and standard deviations for the Z24 and KW51 Bridges for a varying sample size. Panels (a) and (b) present the means for the Z24 Bridge and KW51 Bridge; panels (c) and (d) give the standard deviation.

4.4.2 Domain adaptation and clustering

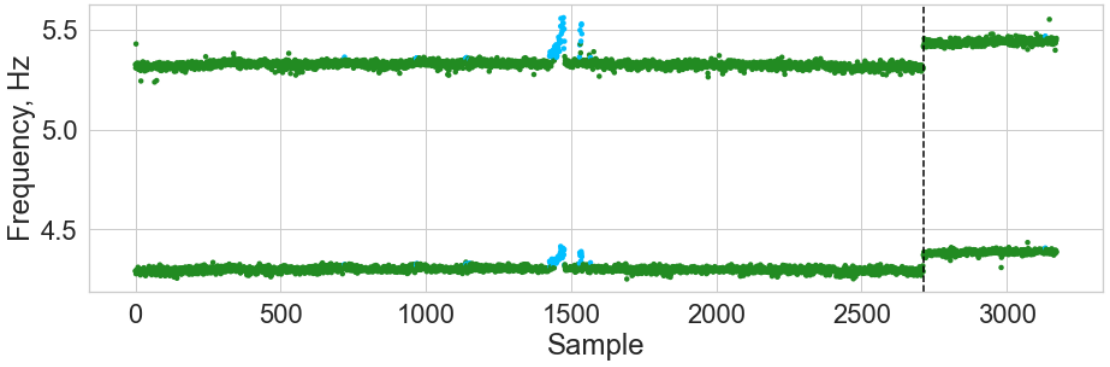
Initially, a sensitivity analysis was conducted to evaluate the quantity of data required in each domain for SA. The mean and standard deviations were calculated for each structure with varying sample sizes, starting at 10, increasing to 500 samples, with a 10-sample step size. The results of the sensitivity analysis are given in Figure 4.14. It can be seen that the mean can be accurately estimated with very limited data, suggesting that transfer could be possible between real structures if the differences are mostly summarised by the mean. As expected, the standard deviations required more data to be accurately estimated, particularly for the KW51 Bridge.



(a)



(b)



(c)

FIGURE 4.15: Unsupervised GMM predictions on the Z24 and KW51 Bridge datasets. Panel (a) gives the comparison of the two features after alignment using NCORAL, showing the four Gaussian components identified (μ (+) and $2\sum$ (—-)); the Z24 Bridge samples are denoted by (○) and the KW51 by (×). Panels (b) and (c) gives the predicted classes on the unadapted Z24 Bridge (ω_1 and ω_3) and KW51 Bridge (ω_{10} and ω_{12}) natural frequencies against sample point.

The first step was to adapt the pre- and post-repair KW51 Bridge data. Thus, NCORAL was applied, learning the statistics using the first 200 data from the pre- and post-repair state, which correspond to similar temperatures. Given the sensitivity analysis (Figure 4.14d), this sample size should correspond to a robust estimation of the statistics whilst only selecting data from a short period after inspection to increase the likelihood that it was generated by the normal condition.

To align the Z24 Bridge and KW51 Bridge datasets, NCORAL was applied. As can be seen in Figures 4.12 and 4.13, both bridges experience stiffening effects because of freezing temperatures, so a prior assumption is that, within both datasets, normal condition data can be split into two classes, pertaining to ambient ($T > 0°C$) and low temperatures ($T < 0°C$). Labels for these effects are not explicitly known, but temperature data gives an indication as to which data correspond to these effects. As such, data where $T > 0°C$ were considered "ambient normal condition" and $T < 0°C$ as "low temperature normal condition". Since there are significantly more ambient normal condition data in both domains and the Z24 Bridge dataset includes more varied low-temperature data, these temperature measurements were used for sample selection. For the Z24 Bridge, based on the sensitivity analysis (Figure 4.14c), 400 samples were selected from each normal condition class, with the low temperature normal condition being randomly selected from the subset where $T < 0°C$. In the KW51 dataset, there are only 110 samples corresponding to $T < 0°C$; so that the dataset is unbiased, 110 samples are used from each normal condition. Based on the sensitivity analysis (Figure 4.14), this sample size only underestimates the tenth natural frequency by approximately 4% and overestimates the twelfth natural frequency by 7%. The KW51 Bridge was adapted by only considering a subset of data from the first 72 days of its monitoring campaign, which can safely be assumed to be operating in its normal condition.

To demonstrate that information can be shared between the bridges after alignment, an unsupervised GMM is learned on the aligned features. Here, an unsupervised model is utilised, since ground-truth labels are unknown, but any appropriate model could be applied after SA. The prior assumption is that there are three groups within the datasets: the ambient and low-temperature normal conditions, and the Z24 Bridge damage, so a three-component model was implemented. Since ambient normal data are more abundant, to reflect the prior assumption that normal condition data are split between ambient and low temperature conditions, the temperature data were used to downsample data corresponding to ambient conditions ($T > 0°C$) [4].

The aligned features and the groups assigned by the GMM can be found in Figure 4.15. It can be seen in Figures 4.15b and c, that the ambient normal conditions of the

[4]This assumption could also be enforced with Bayesian priors.

Z24 Bridge, and pre- and post-repair KW51 Bridge are well aligned, as well as, the low temperature normal conditions of the Z24 Bridge and pre-repair KW51 Bridge, indicated by the GMMs ability to find shared groups across the domains.

The feature space has also maintained physical interpretability (Figure 4.15(a)), an aspect of SA that could be useful for joint visualisation and mitigating the risk of negative transfer. For example, it can be seen that the stiffening effect caused by low temperatures causes an increase in each feature (in blue, where $T < 0^{\circ}C$). In addition, the stiffening reduction caused by damage in the Z24 Bridge can be seen by a reduction in each feature (in purple).

One of the main advantages of aligning the population of bridges is that damage-state data from a source dataset could be used to further inform a damage detector for the target. In Figure 4.15, it can be seen that the normal condition data for the KW51 Bridge lies on the boundary with damage in the Z24 Bridge dataset, but no data are misclassified as showing damage[5]. This result motivates the idea of using damage to inform a novelty detector, which may provide a method for defining thresholds, and may provide a method to provide additional insight into the structural health by providing direct comparisons with previously observed damage from related bridges. Unfortunately, there is no damage in the KW51 Bridge dataset to further investigate this possibility.

This case study also illustrates a trade-off between selecting features that are transferable and damage sensitive, shown by Figure 4.15(b), where some normal condition data of the Z24 Bridge dataset are clustered with damage. It can be seen in Figure 4.15(a) that this is largely because damage is masked by ambient Z24 Bridge data. However, in previous studies, it has been demonstrated that using all four available frequencies allows for damage to be discriminated [181], but the most damage-sensitive natural frequency in the Z24 Bridge dataset (the second natural frequency) was not used since the mode shapes indicate that there is low physical similarity with any of the modes in the KW51 Bridge.

This trade-off may be influenced by the similarity of the population. In this population, a concrete box-girder highway bridge is used to transfer information to a steel bow-string railway bridge, with the two structures having differences in material properties, geometries and connectivity. Thus, since the aim of DA here is to find a feature space where future health-state data could be shared, the features should share physical similarity, such that it is believed that the same physical phenomena in each structure would correspond to similar parts of the feature space. This objective justifies the approach

[5]Note that the Z24-bridge damage labels were not used to learn this GMM, as the main aim is to show the domains are well aligned by SA. A supervised or semi-supervised model could be used to better define this boundary.

of selecting only frequencies with strong modal correspondence, but the dissimilarity between the bridges means that only two frequencies were deemed to be sufficiently similar. In a population where structures are more similar, it is reasonable to expect that a larger proportion of modes would be similar, reducing the severity of this trade-off; for example, two nominally-identical wind turbines would be expected to have a larger proportion of similar modes.

Despite the masking effects, the aim of this case study was to demonstrate that NCA and NCORAL can align the feature spaces of two structures in a partial-DA scenario, as well as address changes to the structural response caused by repair by only using limited normal condition data gathered in a short period after inspection. This objective has been achieved using a small quantity of inexpensive data, i.e. normal condition response and temperature data. The trade-off between transferable and damage-sensitive features is a topic for further research.

4.5 Case study: statistic alignment as pre-processing

The previous case studies have demonstrated that exclusively aligning the lower-order statistics via SA can facilitate knowledge transfer. In fact, if the underlying data distributions are similar enough, SA was shown as a suitable method to facilitate label sharing without any other form of transfer learning. In this section, a case study is investigated where it is assumed further (nonlinear) DA is required; it is proposed that SA should be used as a pre-processing step in such scenarios, aligning the lower-order statistics, such that the nonlinear mapping found by further DA is simplified.

Compared to SA, specifically NCA and NCORAL, prominent DA algorithms may have a number of additional requirements. Data available in each domain should be representative of their respective underlying distributions and abundant enough for the application of nonparametric distribution-divergence measures to be robust. Furthermore, prominent DA methods are prone to negative transfer in partial DA [158], and if the target only represents of small subset of the source classes, the risk of negative transfer will be higher [193]. Therefore, further DA should be used when there is reason to assume that there are significant nonlinear distribution shifts between domains and when available information is sufficient in each domain to mitigate the associated risk of negative transfer.

4.5.1 Data simulation: numerical three- to seven-storey population

To demonstrate SA as a pre-processing method, another numerical case study is presented. The simulation procedure follows the previous numerical case study, with this case forming a heterogeneous population consisting of 3DoF source and 7DoF target structures, giving a more complex transfer problem; the material properties can be found in Table 4.2. Damage was simulated for the first and third storeys in each structure. For the source domain, 400 samples were simulated for the normal condition and 150 samples for each of the damage states, $n_s = 700$. In the target, 200 samples were simulated for the normal condition and 75 samples for each of the damage states, $n_t = 350$. Class imbalance between the normal condition and damage states was introduced in this way to emulate practical scenarios, where normal condition data will be more abundant. The transfer-learning problem was to transfer the label information from the normal condition, first damage location and third damage location in the source structure to the same location in the target structure. Thus, this case study considers the first three storeys of the 7DoF target structure as a "sub-structure", which provides a homogeneous transfer-learning problem; this approach was first discussed in [7]. Note that this case does not investigate partial DA, as prominent DA methods are prone to negative transfer in this scenario, so identification of robust partial DA methods to use in conjunction with SA is left for future research.

	Unit	Source	Target
Beam geometry $\{l_b, w_b, t_b\}$	mm	$\{300, 40, 8\}$	$\{300, 40, 8\}$
Mass geometry $\{l_m, w_m, t_m\}$	mm	$\{400, 400, 40\}$	$\{400, 400, 40\}$
Crack geometry $\{l_{cr}, l_{loc}\}$	mm	$\{20.0, 150\}$	$\{20.0, 150\}$
Elastic modulus E	GPa	$\mathcal{N}(210, 1 \times 10^{-9})$	$\mathcal{N}(210, 1 \times 10^{-9})$
Density ρ	kg/m^3	$\mathcal{N}(7800, 50)$	$\mathcal{N}(7800, 50)$
Damping coefficient c	Ns/m	$\mathcal{G}(8, 0.8)$	$\mathcal{G}(8, 0.8)$
Storeys		3	7

TABLE 4.2: Properties for the three- to seven-storey numerical case study

The standardisation method that does not perform DA, N-standardisation, as well as the methods that perform DA A-standardisation, CORAL, NCA and NCORAL, were each used as a normalisation procedure. Following this step, each of the DA methods that were applied in the previous sections was implemented. An additional benchmark showing the effects of applying conventional machine-learning methods was presented, applying a kNN to the N-standardised space, with one neighbour.

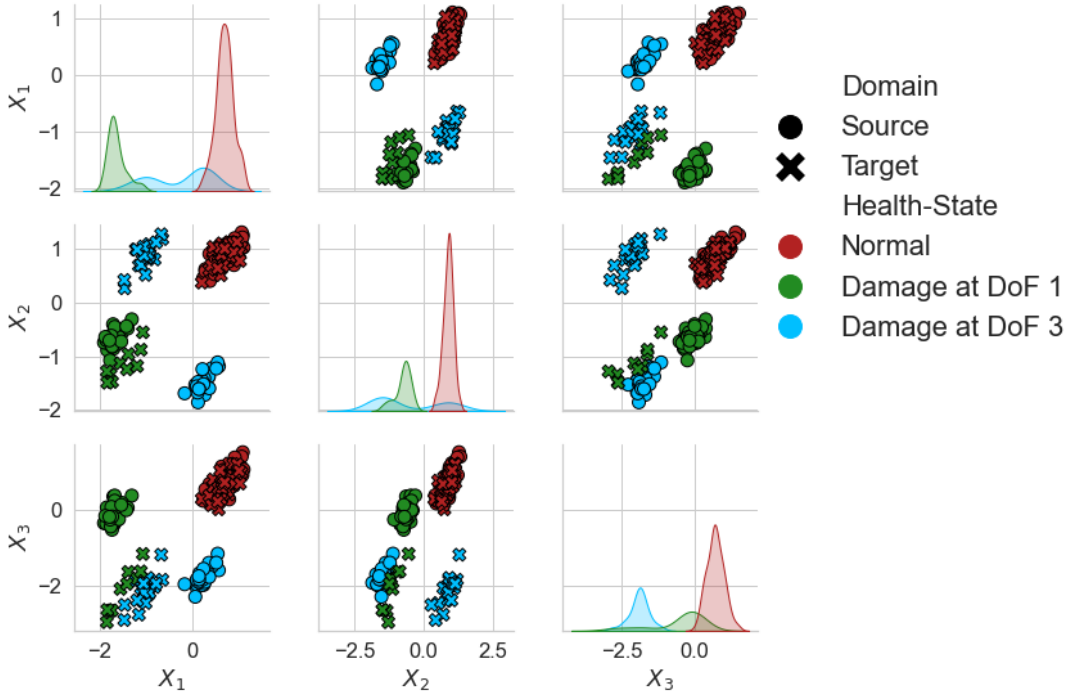


FIGURE 4.16: NCA features of the numerical three- to seven-storey population of structures. The source and target are depicted by (○) and (×) respectively. A random subset of 20% of the size of the datasets is plotted for visualisation.

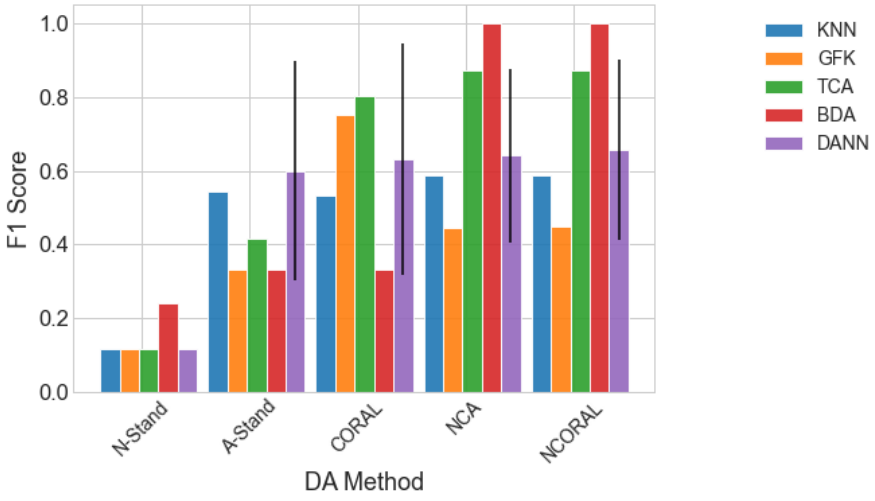


FIGURE 4.17: Classification performance of a kNN on the target domain after various SA methods, then DA on the numerical three- to seven-storey population. The result of the DANN is given as the mean of 100 repeats with one standard deviation shown by a black line.

4.5.2 Results: numerical three- to seven-storey population

The NCA features are visualised in Figure 4.16. Significant differences in the domains can be seen, including the order of Classes 1 and 2 being flipped between domains in

the second mode. Classification results can be found in Figure 4.17. Without first performing statistic alignment (shown by N-Stand in Figure 4.17), none of the nonlinear DA methods achieved adequate adaptation for knowledge transfer. Each of the SA methods alone improved upon the unaligned kNN.

Following NCA and NCORAL, excluding the GFK, all DA methods were able to improve upon SA alone (shown by NCA and NCORAL in Figure 4.17). In this case, NCORAL gave the same result as NCA, since the correlation of the normal conditions is the same in both domains. The GFK did result in negative transfer after NCA and NCORAL, but this method aligns PCA subspaces with dimension k, with the condition it must be under half the dimension of the original space d, i.e. $k < d/2$. Thus, in this case $k = 1$, which may not have been sufficient to encode enough discriminative information. These results suggest that even though without SA, the nonlinear DA methods fail to achieve knowledge transfer, they can still provide additional performance gains if SA is applied first, highlighting that it is crucial to consider SA for transfer in SHM.

While in this case, additional DA provided improved generalisation to the target data, in contexts where the datasets are similar enough for SA to provide sufficient adaptation, applying further adaptation may add an additional risk of negative transfer and lead to a loss of the physical interpretability of the feature space. Thus, whether SA would constitute a suitable transfer strategy alone is an important topic for future research, which should be considered in relation to structural and data similarity.

The presented case study shows a case where a substructure relating to three storeys in a seven-storey structure could be considered as a target domain for transfer from a three-storey structure. In this case study, it was assumed that the three storeys would exhibit a related response to damage, facilitating transfer learning. However, more principled methods for justifying such sub-structuring approaches are required for robust transfer between such structures. An important consideration is that the effects of damage will likely have different effects on these structures in practice; for example, damage on the third storey of a seven-storey structure would likely have different consequences compared to damage on the third storey of the three-storey structure. This issue motivates further research into finding equivalent label spaces in heterogeneous populations, as discussed in [7].

4.6 Discussion and conclusions

Unsupervised DA is a particularly appealing branch of transfer learning for engineering applications, as it has the potential to circumvent expensive target labelling procedures

by reusing labelled source data. However, in SHM, target information may be sparse and may not be representative of a wide range of health states; thus, information to learn DA mappings will often be limited. This limited data availability presents a challenge to the practical application of DA for SHM, as many conventional DA methods are prone to negative transfer in partial-DA scenarios [158].

To facilitate DA in scenarios where only limited normal condition target data are available, this chapter proposed statistic alignment methods to learn a linear mapping that projects target data into the source feature space. These methods only require data from the undamaged target structure; therefore, they could be applied early in a monitoring campaign. If the damage response of structures is sufficiently similar, these methods could allow for informative source-predictive models to be reused in a new target domain, potentially allowing for classification of damage in the target structure at the first instance of damage. Furthermore, these statistic alignment methods were shown to be a critical preprocessing step for several popular DA algorithms.

Three case studies were presented to evaluate the applicability of SA. The first presented a heterogeneous numerical population with differences in size and material properties. First, a conventional DA scenario was investigated. It was found that for this problem, SA methods could effectively find a shared space, allowing for a source classifier to generalise to target data. It was also found that the benchmark algorithms, TCA, BDA, the GFK and the DANN, were not able to account for this large shift in the marginal distributions, potentially because of the high initial mean shift between domains.

To investigate the partial-DA scenario, the target dataset was downsampled to only retain the undamaged class and one damage class. As expected, previous SA methods led to poor alignment because of the biased statistic estimates. However, by only using data from an assumed known class, the normal condition, NCA and NCORAL were shown to still facilitate generalisation of a source classifier.

The second case study presented a real heterogeneous population of two bridges – the Z24 and KW51 Bridges. The objective in this case study was to find a shared feature space between three distinct domains: the pre- and post-repair states of the KW51 Bridge, and the Z24 Bridge. NCORAL was applied to find a mapping to align the KW51 Bridge pre- and post-repair data with the Z24 Bridge data. To account for imbalanced data caused by differences in the EoVs represented in each dataset, data were aligned only using data acquired when ambient temperatures were above freezing. To demonstrate a shared predictive function in this space, a GMM was learnt; it was shown that the undamaged data from ambient and freezing temperatures could be jointly clustered. Moreover, the mapping estimated via NCORAL maintained the interpretability of the original natural frequencies, i.e. showing a decrease in value for damage effects and an

increase for cold-temperature effects, allowing for joint visualisation of data from the three domains.

In some cases, a linear transform learnt using only normal-condition data will not be sufficient to account for differences between the source and target domains. In the final case study, it was demonstrated that in these scenarios NCA may act as an important preprocessing step, improving the performance of nonlinear DA methods. In this demonstrative case study, it was found that several popular DA algorithms failed to improve the generalisation of a source classifier. This result is potentially because the mean shift was large, making it challenging to find a shared feature space. Applying NCA, this shift was reduced, improving generalisation of a source classifier; furthermore, following NCA the benchmark DA algorithms were able to further improve upon these results.

This chapter presents a promising approach for transfer with sparse target information; however, there is significant further work required for the reliable application of these methods. One limitation of these methods is that they are only suitable when a linear mapping is appropriate. A potential solution to this issue would be to apply further DA; in this chapter, a numerical example was presented to show that NCA can also be used to improve the performance of nonlinear DA methods. In the following chapter, these findings are extended, presenting additional case studies verifying this result, with more in-depth considerations with regards to "what to transfer?". Furthermore, several further studies have been conducted since NCA/NCORAL was published in [194], demonstrating a methodology using NCA and JDA for sharing damage labels between different damage extents in a mast structure [195], between lab-scale bridges [196], between an FE model and real bridge [197], as well as two real bridges (the S101 Bridge and the Z24 Bridge) [198], and lab representations of aircraft (Garter structures) [199].

The methods proposed in this chapter have also been demonstrated to provide sufficient alignment in several additional applications. For example, Giglioni *et al.* showed that it could be used to facilitate transfer between the S101 and Z24 Bridges [198]. Meanwhile, Wickramarachi *et al.* demonstrated the use of NCA for accounting for differences induced by repair in a mast structure to perform damage detection online with a Dirichlet process [200]. Dardeno *et al.* further extended the application of NCA to more disparate structures by using interpolating structures [201], showing that transfer can be achieved using a series of linear transforms found via NCA, in contrast to resorting to nonlinear transforms. In addition, this chapter only leverages data acquired directly after inspections. Further work could investigate how to improve these mappings using additional data as they are acquired throughout the monitoring campaign of a target structure; online transfer learning for PBSHM is further discussed in Chapter 7.

An important consideration made when estimating SA mappings in this chapter is that data could be selected in both domains such that it corresponds to similar EoVs; in this chapter, temperature measurements were used to this end. In practice, confounding influences, such as those caused by EoVs, may be challenging to account for without a wide range of measurements corresponding to these factors. As such, selecting datasets corresponding to similar generative processes could also be challenging; in addition, if one structure has a large quantity of data from an unshared health state, it may lead to negative transfer. As such, methods to account for these changes should be a focus of future work; interesting solutions may include removing the effect of benign changes [28, 182] or developing further methods to further refine the datasets used for DA based on auxiliary measurements.

Domain adaptation relies on the assumption that domains are related, such that a mapping can be learnt using the available (in this case, unlabelled) data. In practice, selecting suitable "similar" source domains, and corresponding features, is not trivial in complex engineering systems. Thus, methods to guide domain and feature selection must be developed, as well as criteria to suggest when/what transfer learning is appropriate. To this end, transfer should be guided by similarity measures, which could either be data-based or use physical similarity. Similarity measures will be discussed in the following chapters, with a physics-based measure proposed, with applications shown for feature selection ("what to transfer?"), and prediction of transfer outcomes ("when to transfer?").

Chapter 5

Physics-informed transfer learning via feature selection

A prerequisite for the application of typical unsupervised transfer learning methods is to identify suitable source structures (domains), and a set of features, for which the joint distributions are related to the target domain. Generally, the selection of domains and features has previously been reliant on domain expertise; however, for complex mechanisms, such as the influence of damage on the dynamic response of a structure, this task is not trivial. In this chapter, the modal assurance criterion (MAC) is used to quantify the correspondence between the modes of healthy structures. The MAC is shown to have high correspondence with a supervised metric that measures joint-distribution similarity, which is ultimately the primary indicator of whether a classifier will generalise between domains. Thus, the MAC is proposed as a physics-informed measure for selecting a set of features that behave consistently across domains when subjected to damage.

5.1 Introduction

In transfer learning for PBSHM, the fundamental assumption is that the response to damage of the source structure is sufficiently related to that in the target domain (another structure), so knowledge transfer can improve the performance in the target domain. Furthermore, if this assumption does not hold, transfer learning may result in negative transfer. Negative transfer has particularly critical consequences in SHM, where misclassification of damage may misinform downstream decision processes, potentially leading to unnecessary inspections or missing serious damage. This issue motivates the

development of principled methods to select domains, and corresponding features, that can be transferred.

Many previous studies of transfer learning in SHM have assumed that suitable domain and feature selection can be achieved via domain expertise [48, 49, 202]; however, in complex engineering systems this task is not trivial. This issue is one of the core motivations of physical similarity quantification by representing structures as graphs by using IE models[6]. However, these methods are yet to incorporate real data and do not directly indicate which features will have a similar response to damage.

To measure the similarity of pairs of domains for a given feature set, data-based divergence measures can be used [59, 61], as discussed in Section 3.2.3 and Section 3.2.7. However, in PBSHM target labels will often be sparse, meaning that measuring differences in the joint distributions directly is challenging; thus, unsupervised DA settings are limited to measuring marginal-distribution divergence [59, 61]. However, typically for regression or classification tasks the main requirement for a source model to generalise is that the conditional distributions are similar. Since unsupervised data-based measures can generally only measure marginal-distribution discrepancy, they may not reliably indicate whether the source and target features are sufficiently related for transfer [203].

To address these challenges, this chapter proposes using the modal assurance criterion (MAC) between a source and target structure [204], only utilising data from the undamaged state. In this way, the measure is informative of local discrepancies in the modal displacement between the structures; thus, it is sensitive to conditional distribution shift relating to damage location, addressing a key limitation of previous similarity measures, which are typically data-based [59, 60, 136]. Moreover, by incorporating the MAC into a feature-selection criterion, an unsupervised transfer learning approach based on physics is shown to select features that satisfy the assumptions made by unsupervised DA algorithms – that distribution shifts can be minimised by minimising a marginal-distribution divergence. Thus, this chapter also presents the first principled approach for addressing "what to transfer?".

To demonstrate the applicability of this methodology across different types of structures, results are presented from two distinct case studies: one involving the transfer of damage classifiers between structures within a large numerical population, and the other transferring a damage classifier between a population of two real helicopter blades, demonstrating a methodology for selecting features and performing DA using only normal-condition data from the target structure. Furthermore, it is shown that popular transfer learning methods can be applied to further improve generalisation given a sufficiently related set of features via the proposed feature-selection criterion.

The chapter is structured as follows. Section 5.2 presents a discussion on transfer learning and negative transfer in the context of PBSHM. Section 5.3 explores the idea of using mode shapes as a means to quantify data similarity, and Section 5.4 investigates these ideas on a motivating example. Section 5.5 introduces the feature-selection criterion for physics-informed transfer, with results on a numerical population, suggesting that this approach can significantly improve transfer when only subsets of the features have related conditional distributions. Furthermore, Section 5.6 demonstrates the methodology's ability to transfer label information between experimental data from two heterogeneous helicopter blades. Finally, a discussion on physics-based similarity for transfer learning in PBSHM is presented, and conclusions are drawn.

5.2 Transfer learning and the problem of negative transfer

This chapter primarily investigates unsupervised transfer learning to allow for knowledge transfer without labels about rare health states [56], i.e. damage, spurious environmental conditions etc. [1]. In unsupervised DA, the lack of labels makes estimating (and directly minimising) the discrepancy between the conditional distributions challenging; thus, unsupervised measures that measure marginal-distribution discrepancy are leveraged to learn ϕ [48]. As such, DA requires a strong correspondence between the conditional distributions in the source and target domains.

Determining when minimising the distance between the marginal distributions alone is sufficient is not trivial; if this assumption is not valid, DA will result in performance degradation – referred to as negative transfer [57] – which may have critical consequences in SHM, where poor generalisation of a source classifier can lead to unnecessary inspections or to severe damage by missing critical maintenance. Furthermore, the lack of labels makes it difficult to assess performance using traditional validation techniques, such as cross-validation, and certain issues, such as label switching, mean that instances of negative transfer may be indiscernible from successful transfer. To illustrate this issue, Figure 5.1 shows a toy example of a shift in the distributions, which can be reduced using DA, as shown in the right panel. However, without labels or prior knowledge, it is challenging to distinguish between cases where the conditional distributions are related, meaning a classifier would generalise well after DA, as shown in Figure 5.1(a), and the case where the labels are flipped, as shown in Figure 5.1(b). Minimising the marginal distribution distance in the latter case would result in a model that misclassifies the target.

[1] It should be noted that this chapter is still applicable to supervised transfer learning, as similarity of a given feature space is a core assumption of all transfer learning.

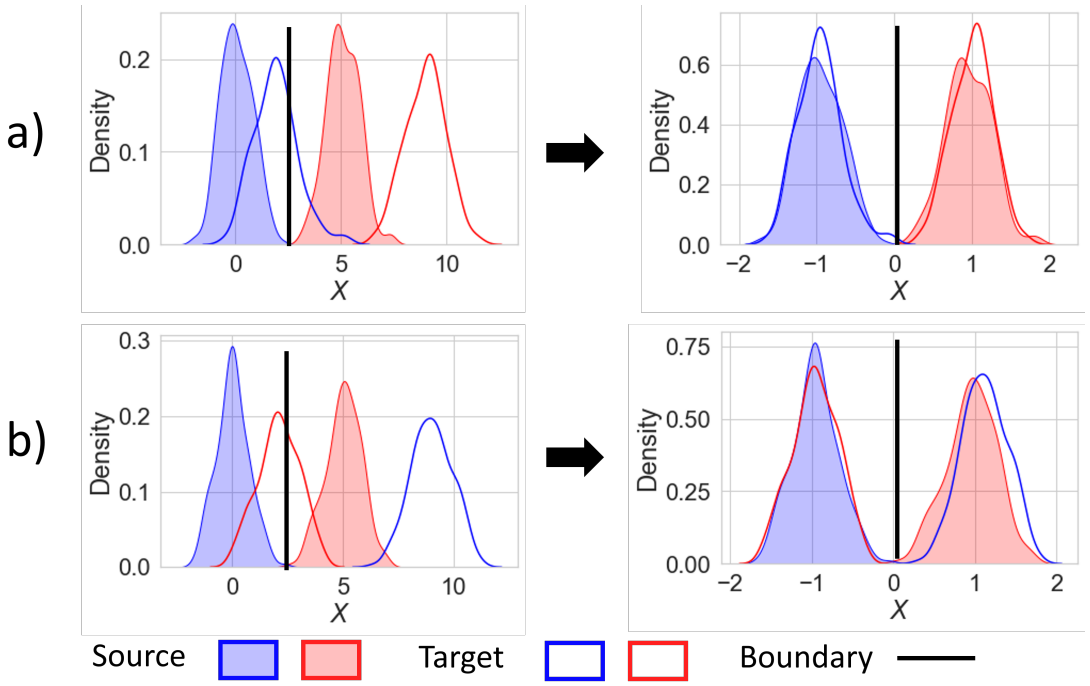


FIGURE 5.1: A demonstration of positive transfer by transferring a binary classifier when domains have similar conditional distributions, presented in (a), and the potential of negative transfer when conditional distributions are unrelated is shown in (b).

This problem highlights the need for reliable methods to select features (and domains), such that unsupervised DA is unlikely to result in negative transfer. In unsupervised DA, there are two main approaches to select domains and corresponding features – using unsupervised distribution divergence measures or domain expertise. As previously discussed, whilst domain expertise is valuable, selecting structures and sets of features that satisfy transfer conditions using domain expertise alone is not trivial. The previous transfer learning literature has also utilised several marginal-distribution distance measures [93, 98–100]. Recent DA research has largely focused on two measures [48], the MMD [136] and the PAD [59]; however, these measures have several limitations for PBSHM applications:

- In an unsupervised transfer-learning scenario, these measures are limited to estimating marginal-distribution divergence; this will not provide a robust measure of similarity if there are differences in the conditional distributions. While there are some attempts to estimate the conditional-distribution discrepancy by using pseudo-labels [104], the pseudo-labels themselves will only be accurate if the initial conditional-distribution discrepancy is small.
- These measures may require large datasets; for example, the MMD relies on data being sufficient to estimate a mean in high-dimensional feature spaces, and the

PAD may require sufficient data to train a classifier with many parameters, such that it has sufficient complexity to sufficiently measure divergence.

- The available data in the target domain may only represent a small subset of the possible classes; for example, at the start of a monitoring campaign of a target structure, only normal-condition data would be available, whereas the source might have a range of health states to transfer. In such cases, the source and target label space would be a subset of the source label space ($\mathcal{Y}_t \subset \mathcal{Y}_s$), meaning unsupervised measures would not indicate whether the underlying distributions differ, but rather that the available subset of the distributions differs.

The issue of negative transfer and the challenges in validating models motivate the need for robust methods for selecting domains, and suitable sets of features within these domains. This chapter focuses on developing a methodology to address the latter problem, while the following chapter discusses its use for deciding "when to transfer?". To this end, the following section presents a physics-based measure to address the aforementioned challenges encountered when only considering unsupervised data-based similarity measures. The core motivation is that for a subset of features that relate to similar physical mechanisms, unsupervised DA should lead to improved predictive performance of a target model, as discussed in [205].

5.3 Physics-based similarity

In the context of SHM, additional insight can be gained by exploiting physical knowledge between structures. As outlined in Rytter's hierarchy [8], assuming damage detection can be achieved using unsupervised approaches, the next task an SHM system should attempt is damage localisation. To perform damage localisation by leveraging information from a source domain, the influence of damage in a specific location l, on the response of the structures should be similar. Specifically, using only the limited available target data, it must be possible to find a mapping such that $p_s(l|\phi(\mathbf{x})) = p_t(l|\phi(\mathbf{x}))$. For brevity, the remainder of the chapter will refer to damage location as a discrete location, denoted by a label y.

Assuming linear behaviour, it is well established that the modal parameters completely characterise the dynamic response of a structure [1]. These properties have been shown to be sensitive to local changes in stiffness caused by damage [206–209]. This chapter leverages this relationship with local stiffness in a different way, using the mode shapes to assess structural similarity in the context of transfer. Specifically, it is proposed that the similarity of mode shapes be used to determine which frequency-based features are

suitable for transfer learning. In this chapter, similarity is assessed using the MAC, a widely-used tool for comparing mode shapes [204].

Since the objective is to facilitate transfer in scenarios where data in the target domain are sparse, it is proposed that similarity assessment can be performed only using mode shapes derived from the undamaged structure (normal-condition data). The general idea is that the mode shapes indicate areas of high strain for a given mode; thus, the locations where a given mode will be sensitive to damage. For example, the locations of the nodes will indicate the regions of a structure where damage will have the greatest influence on the modal features. As such, assuming that the nodal patterns of the mode shapes are similar between systems, it is hypothesised that vibration-based features (e.g. natural frequencies) will have a similar sensitivity to damage, thus meaning transfer should be possible. An illustrative example of the relationship between the influence of damage and the nodal pattern of a mode is presented in Appendix A.1.

It should be noted that the approach presented in this chapter only requires mode shapes from a limited time period where the structure is undamaged. As such, the high cost associated with obtaining a set of mode shapes – i.e. the requirement for a dense sensor network and the need for experts to perform modal analysis – could be reduced by performing a one-time analysis.

A popular tool for the comparison of mode shapes is the MAC [204]. The MAC is a normalised scalar product between each pair of modal vectors $\boldsymbol{\psi}_s^{(i)}$ and $\boldsymbol{\psi}_t^{(j)}$ from two modal matrices Ψ_s and Ψ_t, which in this chapter relate to the source and target domains respectively. The scalars are then arranged into a MAC matrix, assuming real-valued modal vectors; it is given by,

$$\text{MAC}(i,j) = \frac{|\boldsymbol{\psi}_s^{(i)T}\boldsymbol{\psi}_t^{(j)}|^2}{\boldsymbol{\psi}_s^{(i)T}\boldsymbol{\psi}_s^{(i)}\boldsymbol{\psi}_t^{(j)T}\boldsymbol{\psi}_t^{(j)}} \tag{5.1}$$

where $\text{MAC}(i,j) \in [0,1]$, with 0 indicating no correspondence and 1 is complete correspondence. If both modal matrices correspond to similar modes, the leading diagonal will be close to unity. Here, the MAC is compressed into a scalar; a measure can then be given by,

$$D_{\text{MAC}}(\Psi_s, \Psi_t) = \frac{1}{d}\sum_{i,j\in\mathcal{I}} \text{MAC}(i,j) \tag{5.2}$$

where $\mathcal{I} = (\mathbf{v}_s, \mathbf{v}_t) \mid \mathbf{v}_s, \mathbf{v}_t \in \mathbb{R}^d$ is the pairs of feature indices, where $\mathbf{v}_s, \mathbf{v}_t$ are vectors of integers representing the source and target indices corresponding to the features being compared respectively and d is the total number of features being compared in each domain, so the measure is normalised $D_{MAC} \in [0,1]$. This measure is called the MAC-discrepancy.

5.4 Motivating case study: evaluation of similarity measures

Measures that quantify the similarity between domains are central to any unsupervised transfer learning approach. Measures that do not require label information are of particular importance in unsupervised DA, as they typically form the basis of the cost function used to learn shared features. A pitfall of unsupervised DA is that it only leads to improved generalisation under the condition that the conditional distributions are sufficiently related between domains. In addition, ensuring that this condition is satisfied is challenging without using label information. This section presents results for attempting transfer, using unsupervised DA, within a numerical population. It is shown that the MAC-discrepancy is correlated with the accuracy resulting from transfer learning; thus, motivating its use in conjunction with DA to ensure features used for transfer have sufficiently-related conditional distributions; the methodology for feature selection using the MAC is presented in the proceeding section.

5.4.1 Numerical population: a classic SHM example

The population presented in this section considers a number of challenging transfer problems, where there are significant structural discrepancies. The population was generated via the lumped-mass approach to simulate 20 structures, with ten degrees of freedom (DoF). This chapter considers homogeneous transfer learning, where the label space is equivalent, i.e. $\mathcal{Y}_s = \mathcal{Y}_t$. Thus, connections between masses were kept consistent, and variation was introduced by randomly adding additional connections to the ground. An example of two structures is shown in Figure 5.2. Between one and three connections were added to the ten DoF chain of masses; these were added in random locations drawn independently for each structure. The central six masses were considered as candidate locations for extra connections.

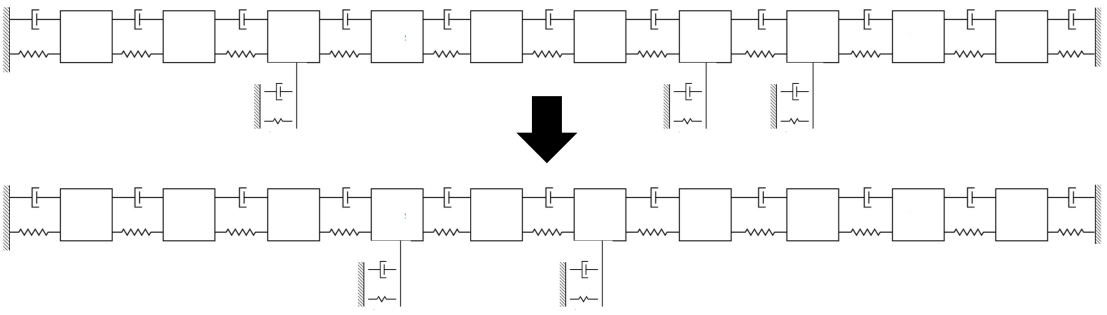


FIGURE 5.2: An example pair of numerical structures from the numerical case study.

Geometries for each structure were the same; geometries that the masses and stiffness were derived from, as well as details of the assumed material variation, are presented in Table 5.1. The approach for simulating data follows the methodology presented in Section 4.2.2. Briefly, each lumped mass was assumed to be a rectangular volume, parameterised by a length l_m, width w_m, thickness t_m, and density ρ, with the density sampled from a Gaussian distribution to represent manufacturing variation. Damage at a given degree of freedom was modelled as a 0.1m long open crack, located in the centre of each connection between masses; this was modelled as a reduction in stiffness following the model presented by Christides and Barr [183].

	Unit	Values
Beam geometry $\{l_b, w_b, t_b\}$	m	$\{5.6, 1.1, 6\}$
Mass geometry $\{l_m, w_m, t_m\}$	m	$\{5.6, 1.1, 6\}$
Crack geometry $\{l_{cr}, l_{loc}\}$	m	$\{0.1, 2.8\}$
Elastic modulus E	GPa	$\mathcal{N}(20, 1 \times 10^{-9})$
Density ρ	kg/m^3	$\mathcal{N}(2300, 20)$
Damping coefficient c	Ns/m	$\mathcal{G}(8, 0.8)$

TABLE 5.1: Properties for the numerical case study.

The damped natural frequencies ω_d and modes $\mathbf{\Phi}$ for 20 systems were calculated by solving the eigenvalue problem. Ten classes were generated, corresponding to no damage and damage to the central nine connections (springs). A total of 100 samples were generated for each damage state, giving 1000 samples in total, which were evenly divided into training and testing datasets. Each pair of structures that did not have identical or symmetrical ground connections was considered a transfer task, giving 360 transfer tasks. For each task, damage classification was attempted, considering the normal condition and nine damage locations, pertaining to damage to all nine springs between masses (not including any ground connections).

The natural frequencies were used as features $\mathbf{X} \in \mathbb{R}^{n \times 10}$, and the mode shapes of the normal conditions were utilised to calculate the MAC-discrepancy. For each task, the source structure was assumed to be labelled and only normal-condition data in the target were assumed to be labelled.

5.4.2 Transfer learning

To assess whether the MAC can be used to quantify joint-distribution similarity, first, this section establishes whether the MAC can indicate when unsupervised DA will result in low joint-distribution discrepancy. To this end, several measures are used to quantify the discrepancy between the source and target data after applying DA for each of the tasks in the numerical population.

The DA method selected here was NCA, as it presents a practical DA method for partial-DA scenarios where there are significant differences in the absolute values of features. As such, if the features have similar sensitivity to damage in a given location (i.e. the proportion that the frequency changes), a classifier trained on the source domain should generalise to the target after NCA. Following NCA, a k-nearest neighbours classifier (kNN), with one nearest neighbour, was trained on the source domain.

The MAC-discrepancy was evaluated for each transfer task to investigate whether it is capable of indicating when the source and target domains were related enough to perform unsupervised DA. In addition, two unsupervised measures, the MMD and the PAD, were used to measure the marginal-distribution discrepancy; for more details on these measures, the interested reader may refer to [59, 136]. These measures were chosen for their prevalence in DA and were implemented to verify whether a purely data-driven approach could effectively determine when unsupervised DA can be successfully applied. A fully supervised measure, the JMMD [104], was also used to show that joint-distribution discrepancy is the primary indicator of transfer robustness. It should be noted that this measure is not applicable in practice as it requires labels in the target domain. The objective of this comparison is to understand whether the MAC can provide additional information compared to using unsupervised data-based measures alone to motivate its application for feature selection, prior to the application of DA to facilitate more robust transfer.

Before applying the data-based measures, several parameters must be set. The PAD requires the specification of a suitably complex classifier to discriminate between domains. In this chapter, a support vector machine (SVM) with an RBF kernel was utilised. The SVM was trained using 70% of the test data, and the PAD was calculated using the error of the remaining 30% of the test data. The MMD projects data into an RKHS via a universal kernel; here, an RBF kernel was used, and the length scale was specified as the median of the pairwise distances between domains (the median heuristic), a common heuristic for unsupervised hyperparameter selection in DA [12, 96, 107, 136]. The MAC-discrepancy was calculated using the mode shapes obtained from a single observation of the undamaged structure. The accuracy of the test data was used as a classification measure.

5.4.3 Results

A comparison of the measures and accuracy is presented in Figure 5.3. Although within the population there are a number of structures with sufficient similarity to achieve accurate predictions, neither of the unsupervised measures (the PAD and MMD) were

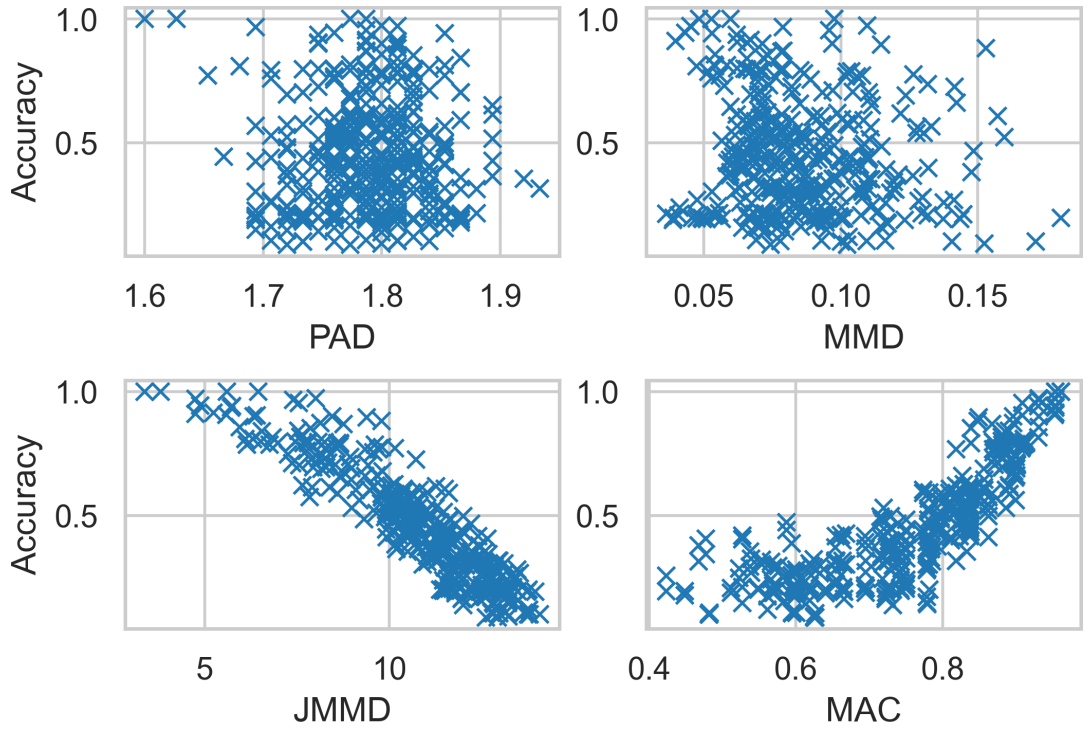


FIGURE 5.3: A comparison of the accuracy of a damage classifier after NCA for each pair of structures in the numerical population, with two unsupervised data-based measures, the MMD and PAD, a supervised data-based measure, the JMMD, and the MAC-discrepancy. The PAD, MMD, and JMMD are zero only when the source and target distributions are the same, and the MAC will be unity when all normal-condition mode shapes are identical in both domains.

indicative of the accuracy achieved via DA; this result is verified by their Pearson correlation coefficient with accuracy, given in Table 5.2. In addition, a number of tasks with low accuracy are associated with low values of the MMD, indicating that some cases follow similar marginal distributions but a number of clusters convey different contextual information (label switching); in some cases, this may have been caused by symmetry between the source and target structures. This result suggests that the population presented has several transfer tasks where conditional distribution similarity needs to be considered directly, and highlights the deficiency of the PAD and MMD in identifying related information in this setting. Moreover, this result illustrates a potential limitation with popular unsupervised transfer learning algorithms, such as TCA [202] and the DANN [110], as these methods rely on these measures for transfer. In many real cases, structures will have differences in connections and local stiffness; thus, the fact that these changes may lead to conditional-distribution shifts presents a major challenge to knowledge sharing with unsupervised transfer learning.

As expected, the supervised measure, the JMMD, is well correlated with accuracy; however, this measure cannot be applied in practice since it requires label information for

each damage state in the target domain. However, the MAC-discrepancy also shows a strong correlation with accuracy[2]. In addition, the MAC-discrepancy and JMMD have a Pearson correlation coefficient of -0.84, suggesting that the MAC of the healthy structure could be used as a proxy for the JMMD to measure joint-distribution discrepancy in scenarios where labelled data are sparse or unavailable. Furthermore, whereas the JMMD requires labelled data for all classes in the target domain, the MAC-discrepancy only requires access to data from the healthy structure (as does NCA – the DA method used here), meaning that it could be estimated before any damage has been observed in the target structure.

TABLE 5.2: The Pearson correlation coefficient between each of the measures and the Accuracy (Acc), as well as the MAC and JMMD.

	PAD-Acc	MMD-Acc	JMMD-Acc	MAC-Acc	MAC-JMMD
Correlation	0.04	-0.14	-0.91	0.82	-0.84

5.5 Physics-informed feature selection for transfer learning

Unsupervised transfer learning is reliable only in scenarios where there is strong correspondence between the source and target conditional distributions, motivating the selection of features that meet this criterion. While the full set of features may not satisfy this condition, a subset might [205]. In general, the task of feature selection involves finding a subset of features that can be used to effectively accomplish some downstream task (e.g. classification). In the case of unsupervised transfer learning, the goal of feature selection is to find a subset of features that satisfy the assumption that the features can be mapped into a shared feature space without using labelled target data. As discussed previously, supervised measures (e.g. the JMMD), are the gold-standard for indicating successful transfer. As it was shown in Section 5.4 that the MAC is correlated with the JMMD, the current section introduces a feature selection based on the MAC to facilitate transfer in scenarios where domain-adaptation assumptions hold only for a subset of features[3].

[2]The correlation for the JMMD is negative since the lower value of the JMMD indicates more similar domains.

[3]The typical use of data-based measures, such as the MMD, would be to find a projection to a feature spaces where these measures are directly minimised [96, 210]. However, the MAC cannot be used in this way, as it directly relates to the properties of a structure.

5.5.1 Physics-informed feature selection

Standard approaches to feature selection typically incorporate a selection criterion and a search strategy to select a set of non-redundant features that maximise discriminative information; thus, reducing issues related to high feature dimension [20]. This chapter introduces a transfer feature criterion (TFC) by incorporating the MAC-discrepancy into a feature-selection criterion to address the challenge of selecting a set of features that maximise the conditional distribution similarity between domains, such that conventional unsupervised DA methods can be applied reliably. In addition, balancing the trade-off between informative and domain-invariant features is a common challenge in transfer learning[48]; for this purpose, the source loss is included in the criterion. Thus, the aim is to select a set of features to maximise the following objective function,

$$\ell = -\frac{1}{n_s}\sum_{n=1}^{n_s} L(f_s(\mathbf{x}_{s,n}), y_{s,n}) + \lambda D_{MAC}(\Phi_s, \Phi_t) - \mu Q \tag{5.3}$$

where $\ell(\cdot)$ represents the loss for a source predictive function $f_s(\cdot)$, λ and μ are trade-off parameters, and Q represents a constraint to prevent the trivial solution of selecting the same feature multiple times; it is given by,

$$Q = \sum_{i=1}^{d}\sum_{i\neq j}[\mathbf{v}_{s,i} = \mathbf{v}_{s,j}] + [\mathbf{v}_{t,i} = \mathbf{v}_{t,j}] \tag{5.4}$$

where $[\cdot]$ represents the Iverson bracket, which takes the value 1 if the values are equal, otherwise, it is 0. To ensure that the most similar source and target features are in correspondence, the target features are selected as follows,

$$\mathbf{v}_t = \underset{j}{\operatorname{argmax}}\ \mathrm{MAC}(i, j) \tag{5.5}$$

A search strategy is required to find a set of source indices. Feature selection is an NP-hard optimisation problem, meaning an exhaustive search would be required to guarantee a globally optimal solution. In this chapter, only small feature sets are considered; thus, an exhaustive search is conducted. However, as the number of features increases, an exhaustive search will become computationally expensive, particularly for larger sets of features [211]. To extend the application of the TFC to high-dimensional feature spaces, heuristic search algorithms could be used [1].

An initial demonstration of this feature-selection approach between a plate generated using a 2D and 3D FE model is presented in Appendix A.2, showing that only corresponding modes can be used to share information between domains.

5.5.2 Case study: Numerical Population

In order to demonstrate the effectiveness of the TFC, the numerical population from Section 5.1 was evaluated by comparing its performance against naively applying DA to all the available features (natural frequencies). In addition, a multi-task approach is suggested to address issues with hyperparameter selection when target labels are unavailable.

5.5.2.1 Multi-task learning for hyperparameter selection

The lack of labels in unsupervised transfer learning means that traditional hyperparameter selection schemes, such as cross-validation, are challenging to apply. To ensure that the TFC and the benchmark algorithms are appropriately tuned for a given target task, a multi-task approach was taken, performing joint empirical risk minimisation over a number of labelled source domains [212]. A subset of structures was assumed to be labelled, representing a number of source structures. By considering each pair of source structures as a source/target pair, hyperparameter selection could be performed to find the best-performing model across all tasks. In this way, hyperparameters can be evaluated across P tasks by,

$$\theta = \arg\min_{\theta \in \Theta} \sum_{p=1}^{P} \sum_{i=1}^{n_{t,p}} \ell(f_s(\mathbf{x}_{t,i}^{p}), y_{t,i}^{p}) \tag{5.6}$$

where θ represents the vector of hyperparameters, while $\mathbf{x}_{t,i}^{p}$, $y_{t,i}^{p}$, and $n_{t,p}$ denote a feature vector, label, and the number of samples, respectively, for the target domain in task p. As discussed by Ando *et al.* [212], learning a set of hyperparameters on a single domain will likely overfit for the given task; however, utilising multiple related tasks allows for a general structure to be learnt, which can generalise to new tasks. In this chapter, a number of structures are used to learn a general model that performs well across the population. However, this method could also be used to select parameters for DA using a number of additional related tasks for each structure. Grid search was used for optimal parameter selection, but more efficient optimisation approaches could be used to reduce computation time.

5.5.2.2 Transfer learning

By applying the TFC, the objective is to find a set of features that satisfy the assumption that the conditional distributions are related, such that when it is used in conjunction

with conventional unsupervised DA methods, a feature space can be found where the distributions are sufficiently close to transfer a classifier trained using only source labelled data.

Following the results from the previous chapter, NCA was applied as an initial DA step. The performance of a classifier after only NCA was also compared to a classifier learnt on features resulting from NCA as well as TCA, and BDA. These methods were applied to show that more complex nonlinear DA mappings will not necessarily improve transfer without first selecting appropriate features. The use of TCA and BDA provides a comparison between methods that minimise the marginal-distribution divergence, and methods that also attempt to minimise the MMD between the class-conditional distributions $p_s(\mathbf{x}|y)$ and $p_t(\mathbf{x}|\hat{y})$. The benchmark algorithms were also applied to the TFC-selected features (also following NCA), to assess the potential benefit of further reducing distribution shift via more flexible mappings when more appropriate features have been selected for transfer.

Hyperparameter selection for all models was performed using the multi-task approach, with five structures being utilised. The associated tasks were considered "validation tasks" and were not included in the final results, reducing the total tasks for testing from 360 to 342.

Each model has a number of parameters that may influence the results; these were selected via the multi-task cross-validation scheme. For the TFC, this is the number of selected features and trade-off parameters for the MAC in the loss function (equation 5.3); these were selected as $m = 7$ and $\lambda = 0.1$. The kernel-based algorithms also require a regularisation term and the number of latent features to be specified; these were chosen as $m = 9$ and $\lambda = 0.1$ for TCA and $m = 5$ and $\lambda = 0.1$ for BDA. When applied after the TFC, TCA and BDA were chosen to reduce the feature dimension by one. Additionally, BDA has a trade-off parameter between the marginal and class-conditional distributions, but since the objective of implementing this algorithm was to investigate the use of pseudo-label-based class-conditional MMD, this was set to unity, so the marginal MMD was not used. Finally, as these methods utilise an RBF kernel, a length scale must be specified; to reduce the hyperparameter search space the median heuristic was used [136]. Classification performance was evaluated using accuracy scores, and the JMMD was used to quantify distribution discrepancy; each was calculated using the test data. It should be noted that the feature dimension may also influence the JMMD.

TABLE 5.3: Mean accuracy and JMMD for all tasks for the source and target test data on the numerical population of structures.

	No DA	NCA	TCA	BDA	TFC	TFC+TCA	TFC+BDA
Source test	1.00	1.00	1.00	0.99	1.00	1.00	0.99
Target test	0.10	0.45	0.45	0.45	0.58	0.56	0.61
JMMD	19.01	10.77	9.92	7.17	8.76	7.90	5.04

5.5.3 Results: unsupervised transfer learning

The mean accuracy on the target for all transfer tasks is presented in Table 5.3, including a comparison with performing no transfer learning (no DA). By only applying a linear transform estimated by NCA, it can be seen that classification accuracy improved significantly. Whilst the accuracy after applying NCA is still relatively low (45%), it is important to emphasise that this case study includes transfer tasks across a range of structures; thus, many transfer tasks are challenging as some pairs of structures have very different responses to damage. It can be seen that even with these challenging transfer tasks, NCA results in an improvement across the entire population with a low rate of negative transfer. Furthermore, when NCA and the TFC are applied together, further improvements to the average classification accuracy can be seen.

Here, both TCA and BDA do not provide a significant improvement compared to only applying the linear transform found via NCA, suggesting neither a more complex nonlinear mapping found via the MMD (TCA) nor the class-conditional MMD using pseudo-labels (BDA) can reliably provide further improvement to generalisation when some features have a different response to damage. This result is perhaps because these methods cannot find a shared space where the conditional distribution distance is low by only relying on unlabelled data to learn a mapping. On the other hand, the TFC (which was applied in conjunction with NCA), was able to significantly improve classification results on average, indicating that by selecting features with a similar response to damage via the MAC (similar conditional distributions), conventional approaches to DA (NCA) can more effectively facilitate transfer. Since conditional-distribution shift cannot typically be reduced using data-based methods without target labels, this result highlights a potential benefit of leveraging physics knowledge.

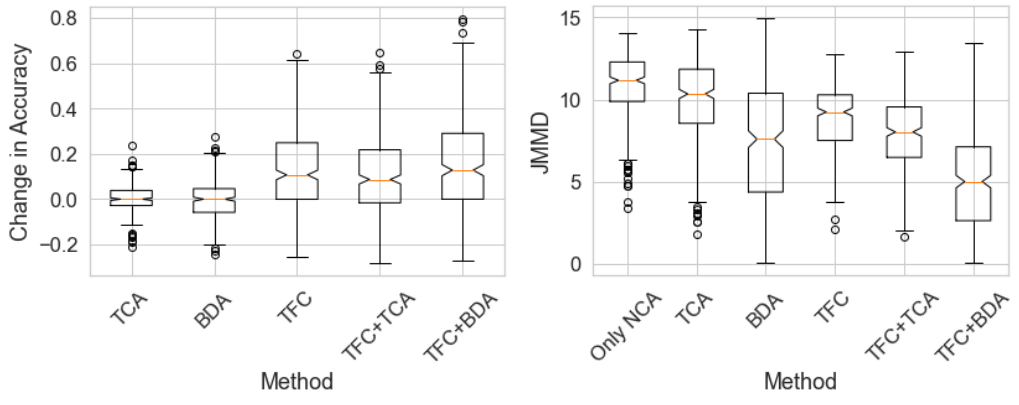
After finding a more similar set of features via the TFC, BDA was able to further improve the average accuracy of classification. This improvement may be because BDA relies on accurate estimation of pseudo-labels, which becomes more feasible after discarding less-related features.

Figure 5.4(a) presents a box-and-whisker diagram showing the range of change in accuracy of target classification compared to only applying NCA. It can be seen that for both TCA and BDA, there is a relatively high rate of negative transfer, meaning accuracy across the population is not improved. Conversely, the TFC leads to a low rate of negative transfer, improving generalisation for a much larger portion of the population, as well as providing larger maximum improvements to classification accuracy.

Another consideration here is that both TCA and BDA use (unlabelled) feature data for each damage class; this may not be a realistic assumption in engineering scenarios, where often only limited damage data are available in the target, i.e. common scenarios may present partial-DA problems. The TFC only relies on using the mode shapes from normal condition, as does the DA method used here (NCA); consequently, it could be applied to practical scenarios to facilitate real-time health-state prediction in structures with no previously-observed damage.

The JMMD after applying each method is presented in Figure 5.4(b) to show the discrepancy between the joint distributions after alignment. TCA and BDA both reduce distribution distance (Figure 5.4(b)). However, these methods do not improve classification results across the population, as classes may be misaligned if the conditional distributions are unrelated, but the marginal-distribution discrepancy will still be reduced. After applying the TFC, the JMMD still indicates a discrepancy between the data [57]. Applying the TFC and the DA algorithms concurrently further reduces discrepancy, particularly with BDA, which could lead to better generalisation.

To investigate the impact of the number of selected features and the sensitivity of the hyperparameter selection scheme, the mean accuracy was evaluated with varying numbers of selected features; Figure 5.5 presents these results. Overall, discarding dissimilar



(a) (b)

FIGURE 5.4: Box-and-whisker plots showing the change in accuracy compared to NCA (a) and the JMMD after each transfer method (b) for the numerical case study.

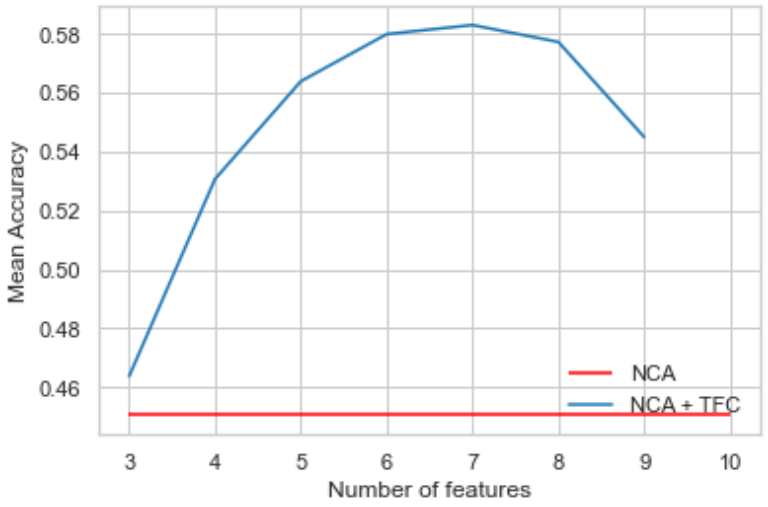


FIGURE 5.5: Mean accuracy for transfer when using the TFC to select a varying number of features on the numerical population. The accuracy of NCA using all the available features (ten natural frequencies) is shown in red and the accuracy for the TFC selecting a subset of features is shown in blue.

features is largely beneficial, improving performance across the population. It can also be seen that there is a balance between discarding dissimilar features and retaining informative features (discriminative information), where in this case study, an optimal average accuracy across the population is given by seven features. Using a small subset of five structures, the multi-task hyperparameter selection scheme was able to select this optimal number of features.

5.6 Experimental case study: heterogeneous helicopter blades

To further explore the application of using modal similarity to inform transfer, a case study consisting of two heterogeneous full-scale helicopter blades is presented. Specifically, the blades are from a Robinson R44 and a Gazelle helicopter. Both blades are similar in size and internal structure, suggesting that there is potential to share information. Importantly, there are several discrepancies, motivating the application of transfer learning; these differences are summarised in Table 5.4.

TABLE 5.4: Summary of the key differences between the Robinson R44 and Gazelle blades.

	Material	Mass (kg)	Length (m)	Width (m)	Lead edge thickness (mm)	Trail edge thickness (mm)
Metal blade	steel	26.95	4.88	0.255	32.70	4.30
Composite blade	carbon fibre	37.00	4.83	0.300	28.10	1.00

FIGURE 5.6: The experimental setup to perform modal testing on a metal (right) and composite (left) blade simultaneously.

The experimental set-up is shown in Figure 5.6. Modal testing was conducted on the blades in a free-free configuration. Data were collected via ten uniaxial 100 mV/g accelerometers, placed on the underside of each blade along the length. Since the mode shape vectors will correspond to modal displacement where accelerometers are placed, sensors were placed at positions corresponding to the same non-dimensionalised length and width to allow for direct comparison in the MAC calculation; the location of the sensors and shakers is given in Appendix A.3. Since the MAC assumes modal coordinates are in corresponding locations, ensuring sensor networks are proportional in both structure, as in this case, is required to use the MAC directly on the extracted mode shapes. Future work should focus on methods to perform comparisons in heterogeneous sensor networks.

The blades were excited using electrodynamic shakers attached in the flapwise direction, applying a continuous pink noise random excitation up to 800Hz, with a decay rate of 3dB/Octave and a sample rate of 1600Hz. To mitigate noise effects, ten frequency-domain averages were obtained. Testing was conducted on both blades simultaneously, assuring that data from both blades corresponded to the same environmental conditions.

Data were collected for five health states, including the normal condition and four pseudo-damage states, relating to adding small masses to specific locations along the length. The added masses were positioned at standardised lengths and widths of the blades, and the size of the mass was scaled to maintain a consistent ratio between the added mass and blade mass. As such, the locations of damage should correspond to similar points of a given mode shape and the extent of "damage" can be considered equivalent for both blades. A summary of the datasets is given in Table 5.5, where

TABLE 5.5: Summary of the blade datasets. The mass ratio between the metal and composite blade is 0.728.

Mass state	Repeats	Mass location (L*, W*)	Metal mass (g)	Comp mass (g)	Mass ratio
Normal condition	25	-	-	-	-
Damage 1	10	(0.627, 0.577)	76.6	105.8	0.724
Damage 2	10	(0.876, 0.577)	76.6	105.8	0.724
Damage 3	10	(0.627, 0.577)	250.0	350.0	0.714
Damage 4	10	(0.876, 0.577)	250.0	350.0	0.714

L* and W* refer to the non-dimensionalised length and width, respectively, which were measured from the root and leading edge. A more detailed outline of the test regime is presented in Appendix A.3. It should be noted that the relationship between damage sensitivity and the position of the nodes, as discussed in Section 5.4, should be inverted for added masses, with the masses having the largest influence on the modes when placed close to anti-nodes.

5.6.1 Experimental case study: transfer learning methodology

In this case study, the objective was to classify the normal and four damaged states of the blades, with the aim of transferring the acquired knowledge from one blade to another. Two tasks were considered, wherein each blade was considered as both the source and the target domain. These tasks will be referred to as M→C when considering the metal blade as the source and composite blade the target, and C→M for the opposite case.

An example of a frequency response function (FRF) of each blade for the normal condition, taken from the sensor closest to the tip (sensor 10, see Appendix A.3) is presented in Figure 5.7. Initially, it can be seen that there are significant differences between the responses of the blades; the peaks are shifted and their amplitudes vary. To visualise the datasets, NCA was performed on the raw FRF data to correct for differences in mean and scale of the FRF amplitude.

Following adaptation via NCA, PCA was learnt on a single domain and was applied to both datasets; the first two principal components are presented for PCA learnt using data from the metal blade in Figure 5.8(a) and on the composite blade in Figure 5.8(b). It can be seen in both cases that the average distance between the normal condition and damage classes is generally larger in the source, suggesting that PCA trained on one set of FRF features does not effectively capture the variance in the other. This discrepancy may largely be influenced by features not being in correspondence i.e. a frequency with a large contribution from a given mode in one blade is often in direct

comparison with a frequency that has little contribution from any modes in the other blade, meaning they will have different sensitivities to damage. Thus, the frequencies with high damage-sensitivity (discriminative features) differ when the FRF frequencies are not processed to account for the difference in natural frequencies.

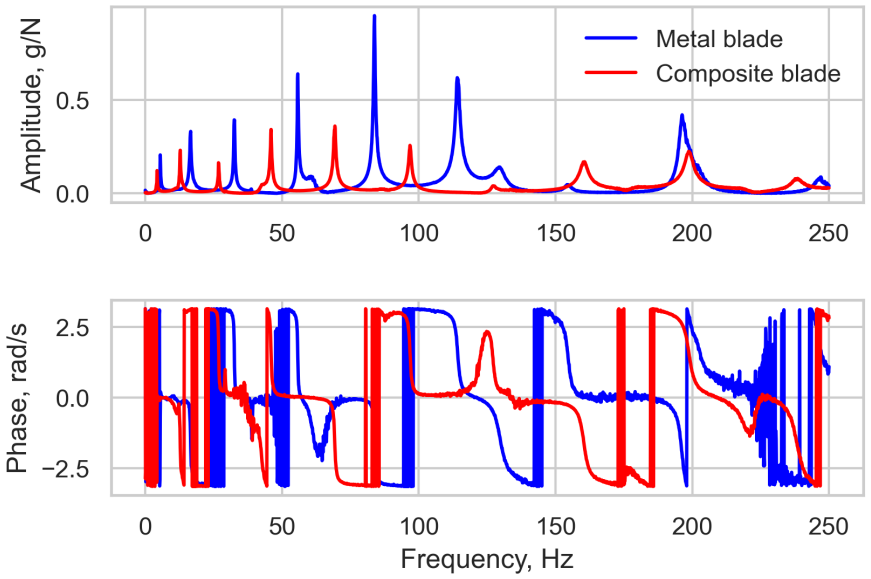


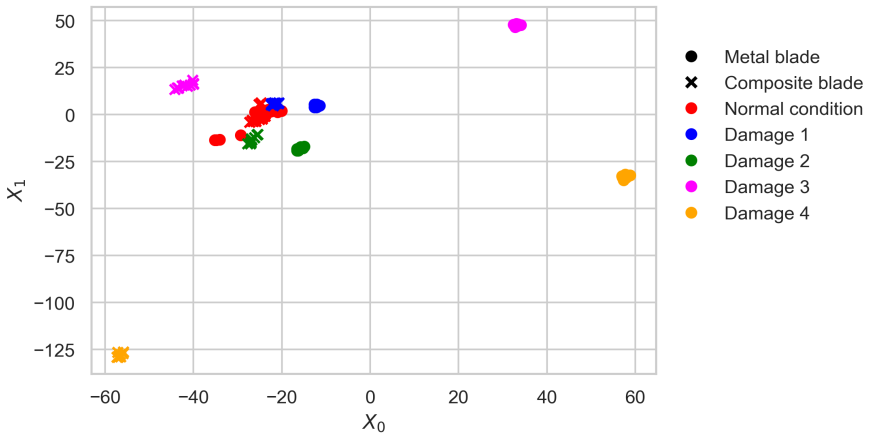
FIGURE 5.7: An example of the FRFs for the metal blade (blue) and composite blade (red) with no added masses.

The FRF amplitude data from the sensor closest to the tip were utilised as features to expedite the modal analysis process. A subset of the FRF was determined by selecting a window of 20 features centred around the natural frequencies; thus, in using these features, peaks of the FRF are compared across domains, which means features correspond to a similar physical phenomenon. As such, modal analysis was only conducted three times on the normal-condition data, in order to identify the regions of the FRF that should be put in correspondence for effective transfer, which was assumed to predominantly correspond to the respective mode; an example of this feature space is shown in Figure 5.9 and the natural frequencies are given in Appendix A.3, along with natural frequencies for the damage states, which were not used in this analysis.

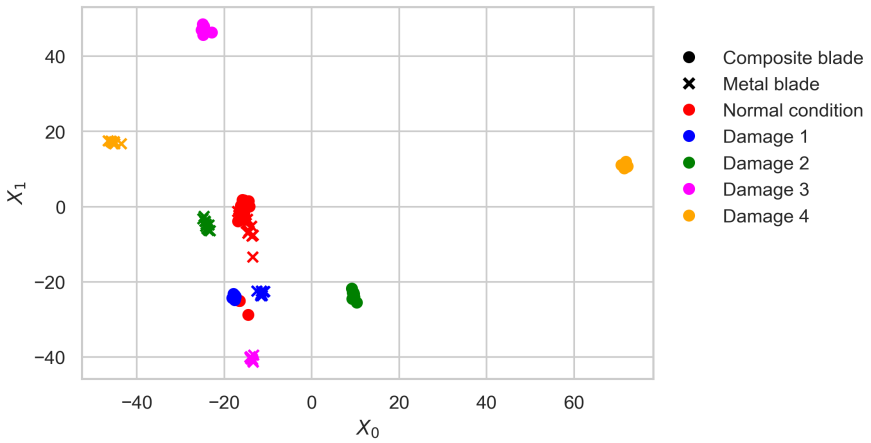
In addition, the mode shapes were used to calculate the MAC between the blades, which were then used to inform feature selection; the MAC matrix is given in Figure 5.10. Nine modes were identified in this range in the composite blade. However, since one of the benchmarks in the following section aims to test the efficacy of the TFC by including benchmark results for algorithms that use all potential features, the ninth mode identified in the composite blade was removed to maintain a homogeneous feature space between domains. Note that the first eight modes were already in correspondence,

although this may not always be the case, and in these scenarios, using the MAC to bring modes into correspondence may be even more critical.

The transfer-learning methodology closely adhered to the one outlined in Section 6, with the inclusion of PCA, trained on the source domain, as a benchmark for a typical dimensionality-reduction technique; feature dimension was selected to maintain 90% of the variance. The primary deviation in the transfer-learning strategy was in the hyper-parameter selection and testing scheme. As the population lacked multiple tasks that could facilitate multi-task hyperparameter selection, the regularisation hyperparameters were selected as the values found in the numerical case ($\lambda = 0.1$ for all cases). The number of features for each method was determined via leave-one-out (LOO) validation using the source data, under the assumption that if the modes are sufficiently discriminative



(a)



(b)

FIGURE 5.8: PCA of all FRF features up to 250Hz, learning the principal components on the metal blade data (a) and composite blade data (b). The source and target data are represented by (○) and (×) respectively. Normal condition is shown by red markers, while the pseudo-damage states induced with the 76.6g mass are indicated by blue and green markers, and the pseudo-damage induced with the 250g masses are shown in magenta and yellow.

in the source domain, they should also be in the target domain. This approach selected two modes for all methods. Since the data were limited, the test results were determined using LOO validation, excluding the JMMD values, which were found on all the data since this measure requires a sample of data. It should be noted that the limited sample size of this case study may impact the reliability of the JMMD. The transfer learning methods are benchmarked against a kNN trained without transfer learning, where "no DA" refers to a kNN learnt using the untransformed features, as presented in Figure 5.9.

5.6.2 Experimental case study: results

The initial features are visualised using PCA in Figure 5.11. It can be seen that the damage classes (shown by their respective colour) are not necessarily close between domains, indicating differences in the data distributions, meaning it is unlikely a classifier would generalise well. This result may be because this initial feature space includes information from a less-similar mode (Mode 6) and a number of modes that are likely not damage-sensitive (e.g. the lower modes). It can also be seen that four normal-condition data are shifted in both domains. These data relate to the last four measurements, which were taken after the majority of tests (refer to Appendix A.3 for more details). As such, these differences may be caused by changes in the boundary conditions, from interacting with the blades during testing or variations in temperature. The higher scatter in the normal-condition data means that it is pertinent that features are both discriminative and have high cross-domain similarity.

Table 5.6 shows the test accuracy on the target obtained from LOO validation and the JMMD, which was obtained using all the source and target data. It can be observed that, in comparison to naively trying to apply a classifier trained using the source labels (no DA), applying DA via NCA led to an improvement in classification. While all test

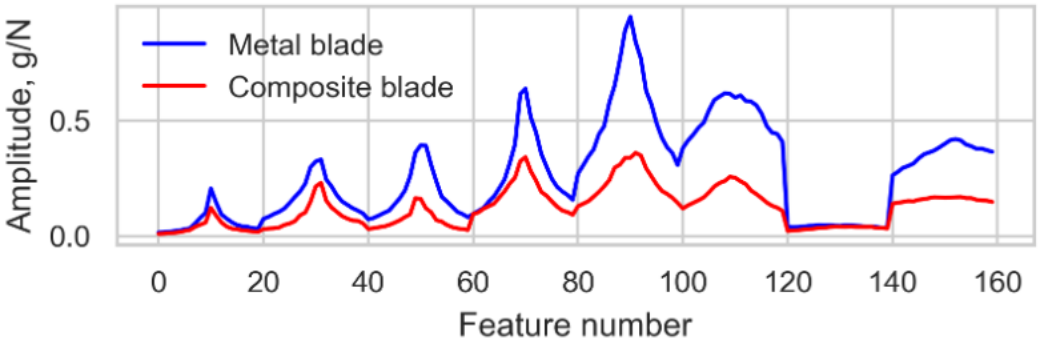


FIGURE 5.9: Example of the feature space after selecting a window of 20 frequencies centred around the natural frequencies for the metal blade (blue) and composite blade (red).

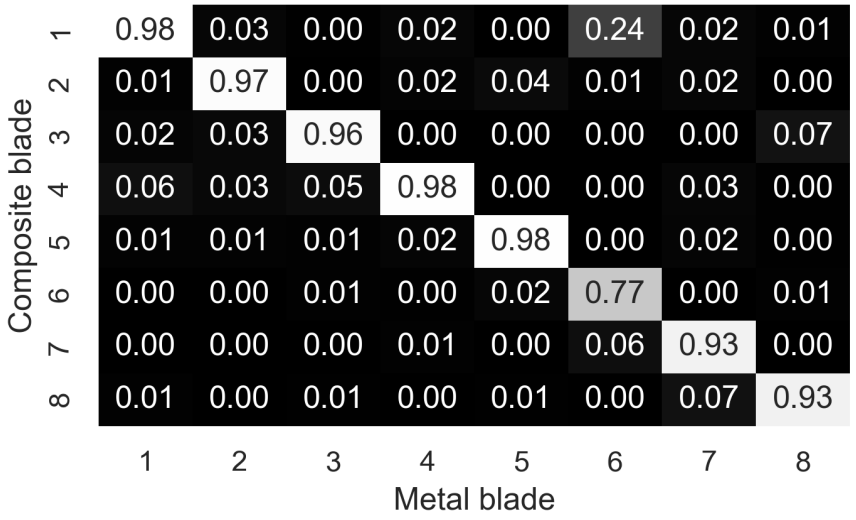


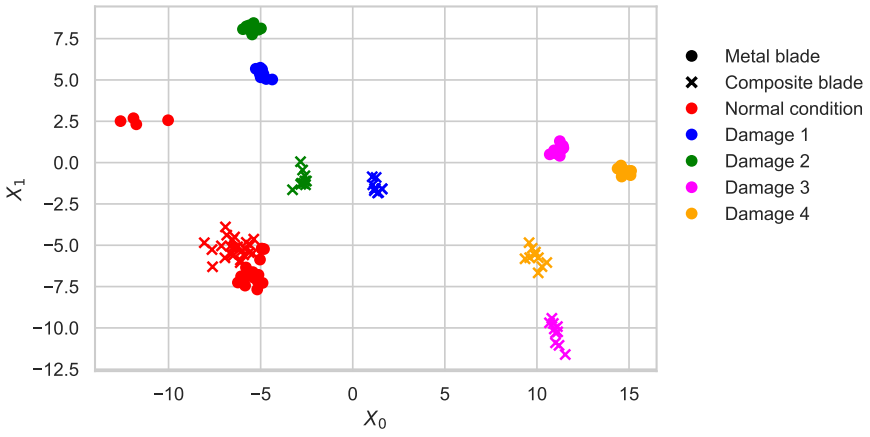
FIGURE 5.10: MAC matrix between the modes of the metal blade and composite blade normal condition.

TABLE 5.6: Mean accuracy for the source and target test data and the JMMD between all data for transfer between the metal and composite blade datasets.

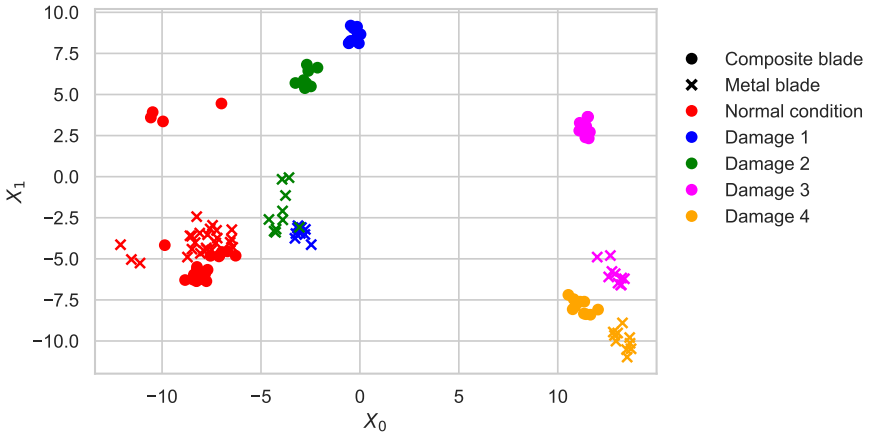
	No DA	NCA	PCA	TCA	BDA	TFC	TFC+ TCA	TFC+ BDA
M→C: Source test accuracy	1.00	0.98	0.98	1.00	1.00	1.00	0.98	1.00
M→C: Target test accuracy	0.15	0.71	0.69	0.54	0.75	0.85	0.88	1.00
M→C: JMMD	9.66	5.50	5.90	5.04	0.73	3.58	0.74	0.27
C→M: Source test accuracy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C→M: Target test accuracy	0.38	0.58	0.58	0.77	0.85	0.89	0.85	1.00
C→M JMMD	9.66	5.42	4.24	3.76	1.73	3.85	1.59	0.26

samples were correctly classified in the source domains for each task, the accuracy in the target domain for no DA and NCA was significantly compromised. This result suggests that transferring a classifier from the source domain generalises significantly worse than supervised learning using the target data, which is indicative of domain shift leading to large generalisation errors. Furthermore, the results achieved after also applying PCA with NCA (PCA in Table 5.6), suggest that this issue cannot be alleviated by reducing the feature dimension alone. In addition, further unsupervised DA after applying NCA (TCA and BDA), did further improve classification in the target in some cases, although TCA led to negative transfer in M→C.

Using the TFC in combination with NCA (TFC in Table 5.6), led to a significant improvement in generalisation compared to using NCA alone. Moreover, when additional DA was applied generalisation was further improved (TFC+TCA and TFC+BDA in Table 5.6). Furthermore, using BDA (TFC+BDA in Table 5.6), perfect classification could be achieved in both cases using a classifier that was trained only using source labels, and resulted in features with a low JMMD. This result shows that comparing



(a)



(b)

FIGURE 5.11: PCA visualisation of all of the features selected in a window centred around the natural frequencies. (a) is for M→C. The source data are shown by (○) markers and the target data are shown by (×) markers, respectively. Both the training and testing data are plotted for a one iteration of LOO validation.

features that correspond to similar damage-sensitive modes can increase the similarity of their conditional distributions, facilitating transfer via unsupervised DA.

For C→M, TCA led to negative transfer, whereas BDA improves generalisation in all instances, perhaps suggesting that using pseudo-labels to estimate the JMMD is a more robust objective for learning a mapping compared to using the MMD in cases where initial classification accuracy is high. This result may be because using the JMMD-based objective in BDA iteratively reduces the MMD between specific classes, meaning that it can further reduce small discrepancies between classes, as suggested by the JMMD in Table 5.6.

The subset of features selected via the TFC and aligned using NCA are visualised in Figure 5.12, using PCA. In this feature space, mass states are in close correspondence

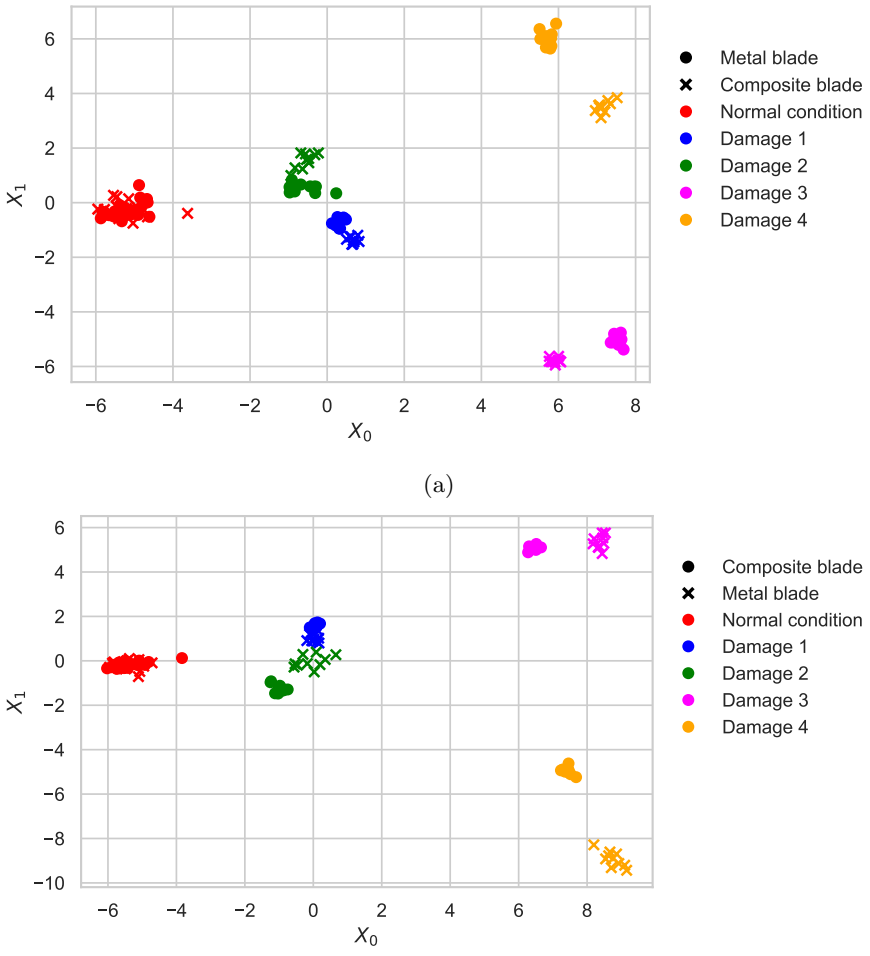


FIGURE 5.12: PCA visualisation of the TFC-selected frequencies, corresponding to the fourth and fifth modes, for M→C (a) and C→M (b). The source data are shown by (○) markers and the target data are shown by (× markers), respectively. Both the training and testing data are plotted for a one iteration of LOO validation.

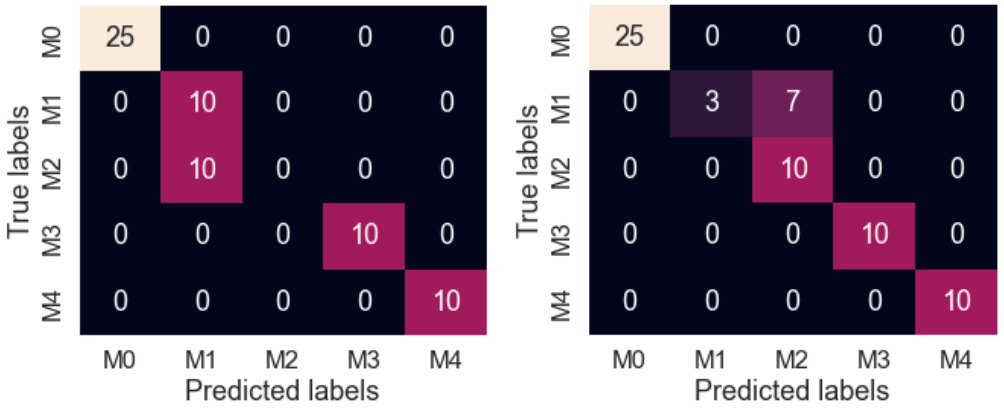
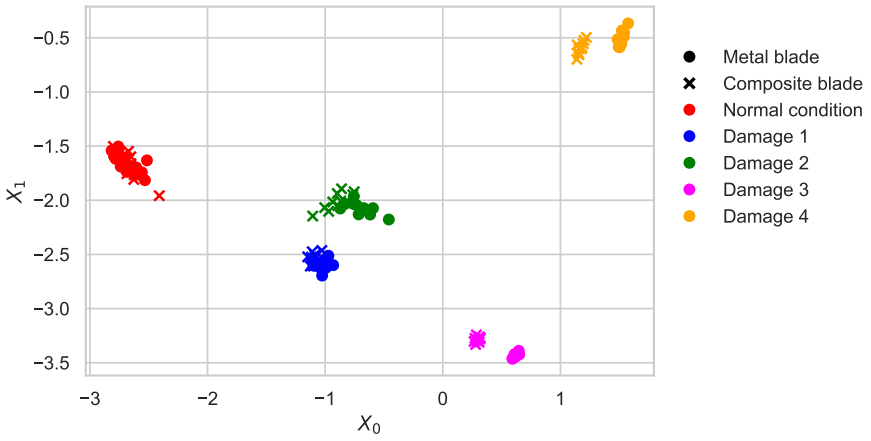
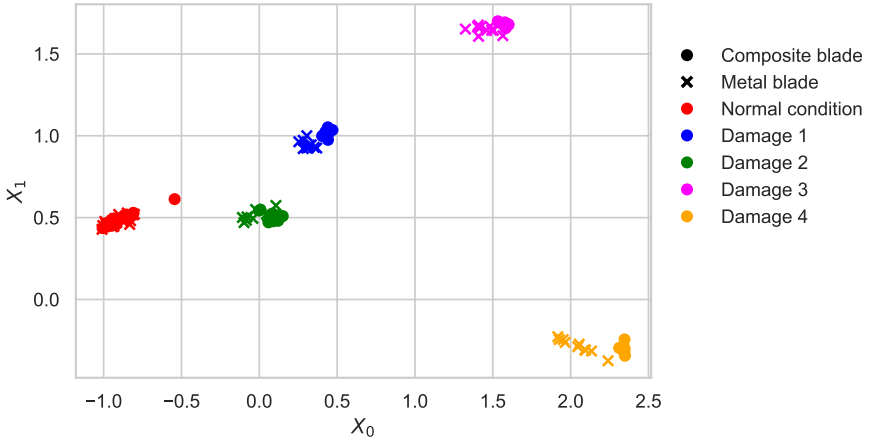


FIGURE 5.13: Confusion matrices for the test data on the TFC-selected features using a kNN for M→C (a) and C→M (b).



(a)



(b)

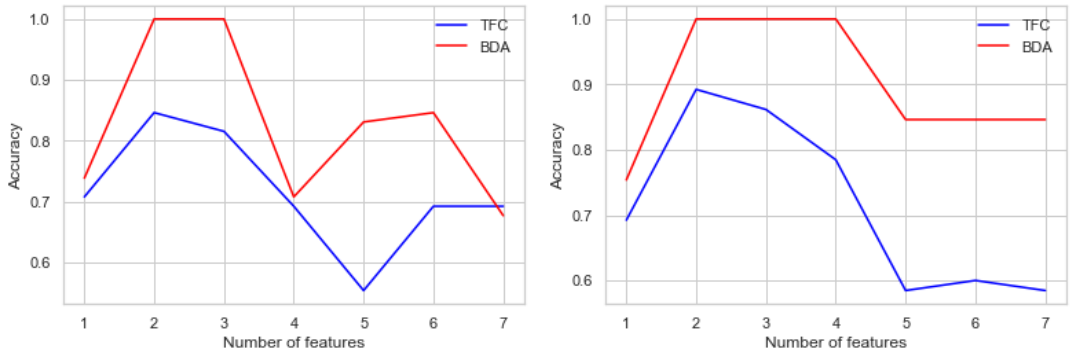
FIGURE 5.14: Features found via BDA applied to the TFC-selected frequencies, corresponding to the fourth and fifth modes, for M→C (a) and C→M (b). The source data are shown by (○) markers and the target data are shown by (×) markers, respectively. Both the training and testing data are plotted for a one iteration of LOO validation.

with their respective state between domains and the features are discriminative. However, the minor-damage classes (Damage 1 and Damage 2) are close; thus, a small shift in the target led to a drop in classification performance, shown in the confusion matrices given in Figure 5.13. This result motivates additional DA to further reduce these discrepancies.

Applying additional DA, specifically BDA, successfully reduced this shift in both tasks, resulting in perfect classification (shown in Table 5.6). This methodology resulted in a two-dimensional feature space, down from high-dimensional raw FRF features. An example of the BDA features for training and testing data for both transfer tasks is presented in Figure 5.14. It should be noted that a potential limitation of BDA is that it assumes that the label space is homogeneous, which may not always be the case in realistic scenarios; this issue requires further research into partial-DA algorithms [158].

The test accuracy for varying the number of selected features using the TFC (shown in blue), as well as the result of using both the TFC and BDA (shown in red), is presented in Figure 5.15. In this case, cross-validation with the source data selected a suitable number of features. It was found that increasing the number of selected features beyond two resulted in worse classification. Although there exist other similar modes, such as Modes One to Three, these modes exhibit less sensitivity to damage; hence, their inclusion leads to a negative impact on performance.

The mode shapes are visualised in Figure 5.16. The TFC-selected frequencies corresponding to the fourth and fifth modes, and the nodal patterns suggest a clear similarity in these modes. In addition, masses were located at anti-nodes in Mode Four, and Damage 1 and Damage 3 were located at an anti-node in Mode Five, suggesting that these modes would be sensitive to damage in this location. These results verify that the TFC was able to effectively select modes that would be expected to be both sensitive to the pseudo-damage investigated in this study and similar between domains.



(a) (b)

FIGURE 5.15: Mean accuracy on the target test data for selecting a varying number of features, with (a) representing the transfer from metal to composite (M→C) and (b) representing the transfer from composite to metal (C→M).

5.7 Discussion and conclusions

Unsupervised transfer learning in a PBSHM framework has the potential to reduce costs and facilitate more in-depth diagnostics for SHM systems. However, these methods require data from different structures to have similar underlying distributions, such that a shared feature space can be found using limited labelled target data. To ensure that the distributions are similar, structures, and their corresponding features, must have a similar response to the damage states of interest. As such, this chapter presents an approach to answer "what to transfer?" when considering frequency-based features by proposing a feature-selection criterion leveraging the MAC.

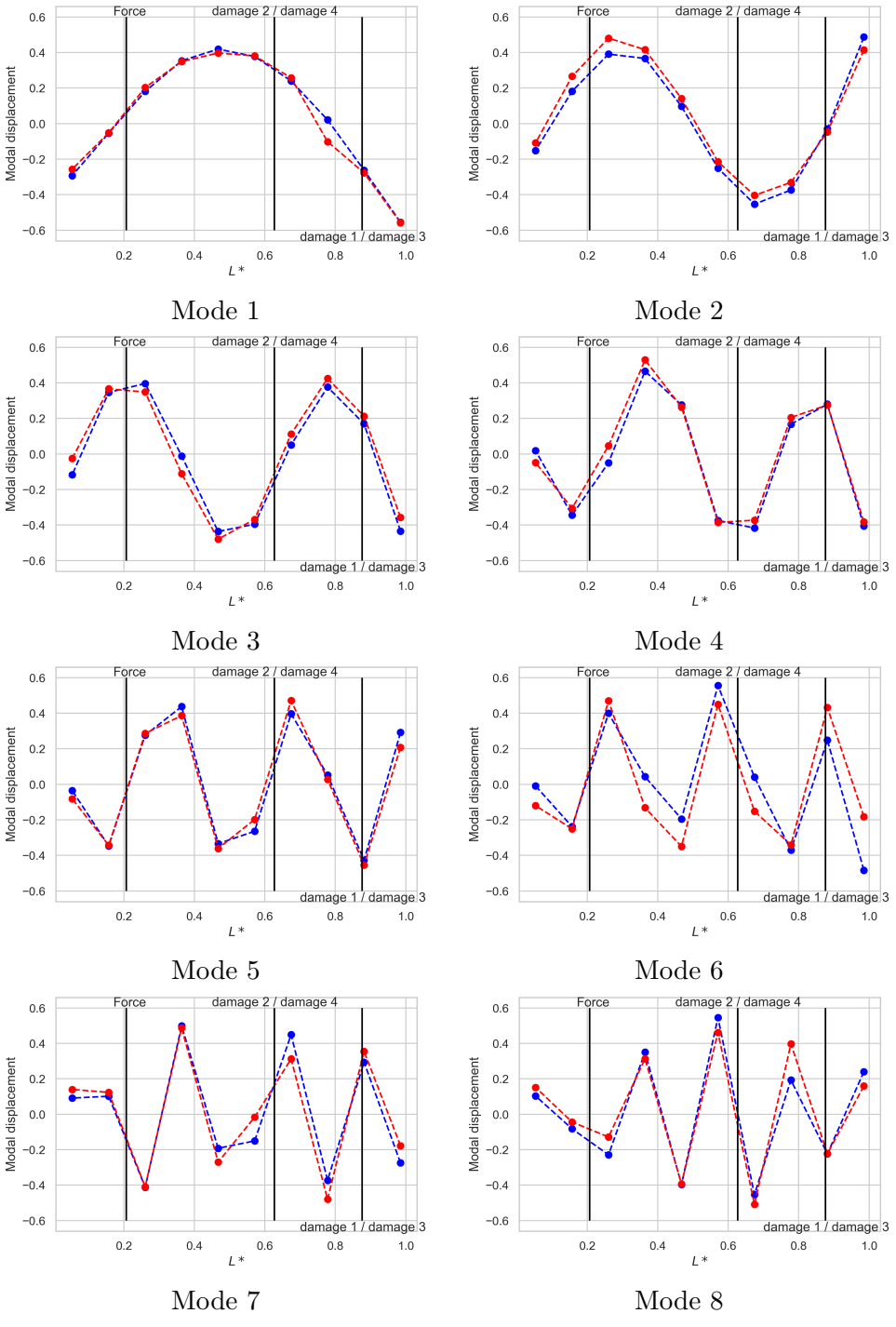


FIGURE 5.16: The first eight identified mode shapes for the metal blade (blue) and composite blade (red). The sensor locations are shown by •, and the forcing location, as well as the mass locations, are indicated by the vertical lines.

An initial motivating case study demonstrated that unsupervised data-based measures (the PAD and the MMD) are not necessarily indicative of the similarity of damage response (conditional-distribution similarity), between two sets of natural frequencies. However, the MAC between the mode shapes corresponding to the undamaged structure was shown to be strongly related to a supervised measure that directly measures the joint-distribution similarity – the JMMD. Thus, a feature-selection criterion incorporating the MAC was proposed to identify sets of features that meet the conditional-distribution similarity assumption required by unsupervised transfer learning methods. By selecting a subset of features, the assumption that all features are strongly related is relaxed, allowing for this assumption to hold for only a subset of features [205]. Furthermore, this measure only requires data from the undamaged structure, making it applicable to a wide range of PBSHM scenarios.

The proposed feature-selection criterion was validated via two case studies. The first involved a heterogeneous numerical population with 342 transfer tasks (transfer of a damage classifier). Using the MAC to select a subset of related features (natural frequencies) led to significantly improved generalisation across the population. In contrast, two popular nonlinear DA algorithms (TCA and BDA) did not improve upon a linear transform learned via NCA, on average across the population. This study showed that, for a diverse population, selecting features with similar undamaged mode shapes can greatly enhance generalisation between source and target domains using NCA, with minimal risk of negative transfer. Moreover, after applying the TFC, BDA improved generalisation further, suggesting that it can be used to select suitable features for use with conventional DA methods.

The second case study investigated transferring a damage classifier to predict different damage locations and extents in a heterogeneous pair of helicopter blades. This case study highlights aspects regarding the importance of selecting related features. FRF frequencies were used as features. As an initial processing step, frequencies in a window around the natural frequencies were put in correspondence across the source and target domains. Using the proposed feature criterion in conjunction with unsupervised DA (NCA and BDA), a set of features corresponding to similar, and damage-sensitive modes could be selected, and a low-dimensional feature space could be inferred, resulting in perfect classification in the target domain using a classifier trained using only source data for both transfer tasks. Thus, this case study demonstrates a methodology, combining prior knowledge about the physical behaviour of frequency-based features, with unsupervised DA methods, as capable of extracting a shared low-dimensional feature space from high-dimensional frequency data.

The proposed method is a step towards demonstrating how engineering knowledge can be used to inform what features can be transferred. Nevertheless, there are several potential issues with using the MAC, and future work is required to extend the findings in this chapter. First, the requirements for the sensor networks should be investigated, considering both optimal sensor locations for individual structures and identifying corresponding sensor locations between structures. Previous work has focused on selecting informative subsets of sensors [206], which could be extended to considering sensors across multiple structures. Another interesting approach would be to interpolate between sensors, either using FE or statistical models. In addition, this approach should be evaluated in structures under realistic operating conditions to evaluate robustness to noisy identification of the modes, the influence of environmental and operating conditions, and nonlinearity.

This chapter highlights the importance of selecting features with similar mode shapes for effective transfer. While lower modes are more likely to be similar across structures, they tend to be less sensitive to damage. In contrast, the most discriminative vibration-based features for damage localisation often correspond to local modes, which are typically harder to consistently identify across different structures. Further research should investigate which damage identification tasks can feasibly transfer using vibration-based features. In addition, investigating additional damage-sensitive features and corresponding similarity measures is likely also an essential extension to the work presented in this chapter; the findings of this chapter suggest that future similarity measures would benefit from including prior physics knowledge.

In this chapter, it was shown that labels relating to added masses could be transferred between a steel and a composite blade. While it is reasonable to expect these structures would both have a similar relationship to an added mass, it is also clear that some damage types would exhibit different behaviours; for example, crack propagation would not be similar in these two different materials. In this regard, further work is required to identify the conditions certain damage types can be used for transfer between heterogeneous structures.

A related limitation is that labels that correspond to specific extents of damage (e.g. a crack of a certain length) will not necessarily correspond to an equivalent change in response, as discussed in the experimental case study. Thus, methods to label damage such that extent corresponds across different domains should be developed, ideally such that damage extent would relate to similar maintenance decisions. In this chapter, prior knowledge of the mass of the blades was used to find corresponding labels for different pseudo-damages. A similar approach should be identified for a range of damages; for

example, equating crack-length labels between structures with varying cross-sectional areas, by using knowledge of the geometry.

This chapter demonstrated a transfer strategy with initial results indicating robustness to negative transfer. However, the availability of similar features is predicated on the need for structures to have a related response to damage. Thus, prior to the application of any transfer-learning strategy, domain similarity should be used to guide operators on "when to transfer?". The following chapter discusses the use of similarity measures for predicting the outcomes of transfer and introduces another application of the MAC-discrepancy.

Chapter 6

Predicting the outcomes of transfer using a physics-informed measure

A major limitation for the practical implementation of transfer learning is that improved predictive performance is not guaranteed, and in the worst-case scenario, it may result in negative transfer. In addition, given the lack of labelled target data, assessment of prediction quality for transfer learning-based SHM models is challenging. Thus, this chapter presents a regression framework for predicting the classification rates that would result from performing transfer learning, given a specific source/target domain pair. Since joint-distribution similarity is the key indicator of whether a classifier can be generalised across domains, this regression framework leverages a joint-distribution similarity measure that could be feasibly used without obtaining data relating to various health states in the target domain; in this chapter, the MAC-discrepancy is used following the promising results presented in the previous chapter.

6.1 Introduction

Previous research has demonstrated the application of transfer learning in various health monitoring applications [12, 15, 41, 135, 142, 153, 155], and the preceding chapters have presented a methodology to address limitations related to "what to transfer?" and "how to transfer?" when target data are sparse. Another critical stage in a transfer-learning pipeline is evaluating which source domains are suitable, such that the likelihood of negative transfer is mitigated [57]. Previous applications of transfer learning have largely addressed this stage by assuming that a suitable source domain can be selected

using domain knowledge [48, 57, 180]. However, in many scenarios discerning when a population of complex structures will have a similar damage response is not trivial. This issue motivates the development of methods able to quantify similarity to assist in deciding "when to transfer?" in PBSHM.

In this chapter, the decision of "when to transfer?" is discussed with regard to the probability of negative transfer. However, evaluating when negative transfer is likely to occur prior to observing damage (and acquiring labels), is challenging since the target risk cannot be directly evaluated [48, 57]. To address this issue, this chapter proposes predicting the outcomes of transfer using similarity measures.

The core idea of this chapter is that if a similarity measure can be evaluated prior to the acquisition of target labels, it could be used to predict whether transfer is expected to reach a certain performance criterion for a given source/target pair. An example of a performance criterion may be an acceptable probability of negative transfer. Following the results of the previous chapter, the MAC-discrepancy is chosen as a similarity measure in this chapter. However, additional similarity measures may also prove informative and provide additional information, such as AG-based similarity measures, and should be considered in future work.

To the author's knowledge, while many works exist investigating similarity measures in relation to transfer learning [6, 44, 46], the only paper using similarity measures to guide the selection of source/target domains in SHM is [135]. This paper used unsupervised data-based measures to select a similar source/target pair. However, as it uses unsupervised data-based measures, this approach may not be robust in all scenarios, as discussed in the previous chapter. In addition, the approach proposed in this chapter also gives predictions for transfer outcomes, providing additional information to inform "when to transfer?". This capability offers a significant advantage over previous approaches that rely on data-based distribution similarity measures, such as the method in [135], which selects the domain with the highest similarity score, since it can also estimate more interpretable measures, such as accuracy.

6.2 Predicting the outcomes of transfer

Negative transfer has potentially critical consequences on the decision-making process. For example, if there is misclassification between a damage class and an undamaged class, the transfer learning-based model may lead to increased costs by prompting unnecessary inspections, or cause critical interventions to be missed, potentially resulting

in more severe damage or, in the worst-case-scenario, structural failure. Therefore, before applying these methods, it is crucial to assess whether the chosen transfer strategy is suitable for a given source/target domain pair.

To assess the potential outcomes of transfer in the absence of labelled target test datasets, this chapter proposes predicting the outcomes of a given transfer-learning strategy $\mathcal{T}$ (encompassing feature extraction, the transfer learning algorithm $\mathcal{A}$ and the method used to select hyperparameters), using a similarity measure $\mathcal{S}$ that is correlated with a measure of the quality of prediction, quantified via a quality measure $\mathcal{Q}$. As discussed in the previous chapter, obtaining a suitable similarity measure that can be computed prior to testing a model learnt in itself may be challenging. To briefly summarise, it is important that these measures can be applied without representative labelled target datasets – particularly for unsupervised transfer learning – and it should capture structural properties that indicate when the damage response between two structures will deviate, such that it is strongly correlated with quality measures, such as accuracy.

Assuming that such a similarity measure is available, past examples of transfer can be used to learn a distribution over transfer outcomes, given the similarity between a new source/target pair. Specifically, for a given transfer strategy, a predictive function could be learnt to map the similarity measure to a quality measure as follows,

$$p(\mathcal{Q}|\mathcal{S},\mathcal{T}) = f(\mathcal{S}) \tag{6.1}$$

where $f(\cdot)$ denotes a probabilistic regression function, which is learnt using previous examples of transfer using a given transfer strategy.

While it may appear that obtaining training datasets consisting of many examples of transfer would in itself be infeasible, as generally target domains have limited labelled data, it is proposed that these tasks could be learnt using "pseudo-target domains". Specifically, pairs of available source domains could be selected, where the labels in one domain are held out for testing, with this domain considered as a target domain. In this way, results for transfer within a population of interest could be obtained. In addition, since a transfer task can be constructed from any pair of source domains, the number of transfer tasks grows exponentially with the number of source domains. For N_s source domains, the number of transfer tasks N_{TL} is given by $N_{TL} = (N_s - 1)^2$, removing cases where the source and target domains correspond to the same structure. Nevertheless, the current methodology does require diverse datasets from multiple structures, which may require available data from previous monitoring campaigns, which would be currently challenging to obtain. However, it is becoming more common to embed sensors on structures, meaning suitable datasets may become available in the future.

Modal displacements are used as a structural representation in this chapter, which facilitates similarity quantification via the MAC as discussed in the previous chapter. This measure was chosen because it was shown to address limitations of data-based measures, providing a measure correlated to the outcomes of transfer without labelled target data. However, any informative measure could be applied in this framework; for example, an AG could be used to represent structures, allowing measures of graph-similarity, such as the Jaccard index [6], to be used to quantify the similarity. The integration of the proposed approach with other measures is left to future work.

There is also a range of quality measures $\mathcal{Q}$, that could be leveraged to assess the performance of a transfer learning on the target domain. Common examples would include accuracy, F1 scores, rates of true positive/negative rates, and false positive/negative rates; in this chapter, accuracy is used.

Regardless of the chosen quality measure, most quality measures will be bounded as they represent rates of classification. Thus, the proceeding sections present a method using generalised regression models with a beta likelihood to ensure predictions remain within the bounded range. This generalised model leverages Gaussian processes (GPs), which were chosen as they provide a non-parametric framework for efficiently modelling complex functions and quantifying uncertainty in predictions through a Bayesian formulation. The following sections provide the necessary background on standard GPs and the beta-likelihood GP implemented in this chapter.

6.2.1 Gaussian process regression

Before outlining the beta-likelihood GP, this section provides a brief introduction to the standard GP. For a more detailed explanation, the interested reader is referred to [20]. The GP was chosen because it provides a flexible, non-parametric Bayesian modelling approach capable of capturing uncertainty in predictions [20]. A standard GP typically assumes a Gaussian likelihood, assuming that the regression problem follows the form $y = f(\mathbf{x}) + \epsilon$, where $f(\cdot)$ maps $X \to Y$, and ϵ is independent additive Gaussian white noise, i.e. $\epsilon \sim \mathcal{N}(0, \sigma^2)$. The GP can be seen as placing a prior directly over $f(\mathbf{x})$; it is defined as,

$$f \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \tag{6.2}$$

where $m(\mathbf{x})$ represents a mean function, and $k(\mathbf{x}, \mathbf{x}')$ is the kernel function that encodes the covariance structure. It can be seen that the GP is fully defined by a mean function $m(\mathbf{x})$ and a covariance function $k(\mathbf{x}, \mathbf{x}')$[1]

[1]It is common to assume that the mean function is zero, as it simply offsets predictions.

A key consequence of the GP formulation is that kernels enable the specification of a particular family of functions. Thus, rather than placing a prior over model weights, as is standard in Bayesian machine learning, the GP can be viewed as directly defining prior distributions over possible functions using the kernel function, conditioning only on the training data. Kernels allow for the expression of various prior beliefs about the function's form – such as periodicity, the number of times a function can be differentiated, or a functional form (i.e. a certain polynomial order) – or the kernel function can be specified to encode limited prior knowledge by defining flexible models capable of representing any smooth function [213]. Another key advantage of the GP is that the posterior-predictive function can be derived in closed form, making it a computationally-efficient method for small datasets.

The standard GP has two limitations in the context of predicting quality measures – it assumes the output is Gaussian distributed and that noise is constant. The following sections introduce a generalisation of the standard GP to instead assume that the output is beta distributed.

6.3 Beta-likelihood GP

The formulation of the beta-likelihood GP implemented in this chapter was first presented in [214], and has been demonstrated in SHM applications for modelling wind turbine power curves [215]. Furthermore, this model is able to capture heteroscedastic noise, meaning that instead of assuming that noise is independent and constant, it is able to estimate a changing degree of uncertainty in the input space. This model may lead to more accurate uncertainty quantification, as it is hypothesised that predictions near the bounds of the similarity measure are likely to be more certain than those in intermediate ranges.

The Beta likelihood is a suitable likelihood function for predicting rates [216], as it is a continuous probability distribution bounded in the interval [0,1][2]. The generalised model, $y \sim \text{Beta}(a, b)$, is defined by two parameters, shape a, and rate b, with $a, b \in R^+$, and it can be defined by assigning independent GP priors on each parameter as follows,

$$y \sim \text{Beta}(a = e^{f(\mathbf{x}_i)}, b = e^{g(\mathbf{x}_i)}) \tag{6.3}$$

where $f(\mathbf{x}_i) = \mathcal{GP}(0, k_f(\mathbf{x}_i, \mathbf{x}_i))$ and $g(\mathbf{x}_i) = \mathcal{GP}(0, k_g(\mathbf{x}_i, \mathbf{x}_i))$, which are defined by independent kernel functions $k_f(\cdot)$ and $k_g(\cdot)$, respectively. The outputs from each latent

[2] Any bounded random variable can be represented in this range by using min-max normalisation.

GP are exponentiated to ensure positivity of the shape and rate parameters. This model form allows for the output to be modelled as a beta distribution which varies in shape dependent on $\mathbf{x}_i$; thus, it can also capture variations in the beta distribution across the input space, allowing for heteroscedastic noise quantification.

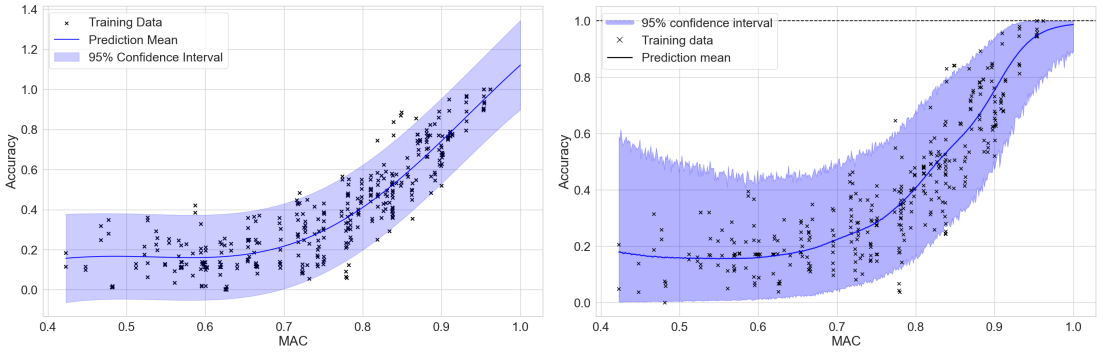
A major drawback of this model formulation is that there no longer exists a closed-form solution for the posterior or posterior-predictive distributions. Fortunately, inference can be performed via variational inference, which is a popular method for approximating intractable posteriors by minimising the KL-divergence between a set of candidate distributions and the true posterior (for more details, the interested reader may refer to [20]). In addition, variational inference can be performed using automatic differentiation in several GP packages – this chapter utilises GPflow [217]. The posterior predictive is obtained by sampling from the posterior. For more details on the beta-likelihood GP, the interested reader may refer to [214].

6.4 Case study: predicting transfer outcomes in a heterogeneous numerical population

The numerical case study presented in the previous chapter (Section 5.4.1), was used to demonstrate the proposed regression framework. To briefly review this case study, it consisted of twenty heterogeneous structures, which were varied by randomly adding extra connections to ground. By considering each pair of structures within this population of twenty numerical structures, 360 transfer tasks can be considered.

In this chapter, the same transfer procedure as outlined in the previous chapter was applied (Section 5.4.2). Briefly, the transfer strategy $\mathcal{T}$ used NCA to estimate a mapping to project the target data into the source feature space, followed by training a conventional supervised classifier (a kNN). This classifier, trained solely on source data, was then used to classify instances into one of ten classes and accuracy was used as a quality measure $\mathcal{Q}$. This process was repeated for each unique transfer task to generate a training dataset for the beta-likelihood GP; all transfer outcomes were used to train the beta-likelihood GP. In addition, a conventional GP was also trained using the same data to highlight the importance of constraining the predictions to a physically meaningful range.

The training data, mean predictions and 95% confidence intervals for the conventional GP and beta-likelihood GP are presented in Figure 6.1(a) and Figure 6.1(b), respectively. While the mean predictions within the range of observed data appear reasonable, it can be seen that the conventional GP led to results without physical meaning when data



(a) (b)

FIGURE 6.1: Predictions of accuracy given the MAC-discrepancy for a conventional GP (a), and the beta-likelihood GP. The training data are shown by (×), the mean predictions by solid blue lines and the 95% confidence intervals are given by the shaded blue regions.

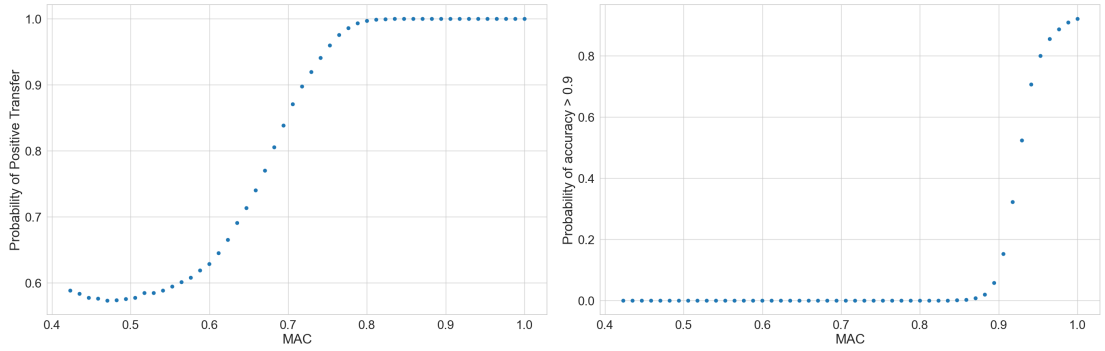
approach the boundaries. For example, at the lower ranges, the confidence interval suggests accuracy may become negative, and at higher values that it may exceed unity. In addition, when extrapolating outside of the range of the observed data, the mean prediction suggests that accuracy will be above unity. On the other hand, the beta-likelihood GP limited predictions to values with physical meaning (Figure 6.1(b)).

Prediction of the distribution of transfer outcomes given an informative-similarity measure $\mathcal{S}$ has the potential to facilitate informed reasoning about the feasibility of transfer given a new target structure. The results in Figure 6.1(b) demonstrate that the beta-likelihood GP presents a suitable model for predicting quality measures for two main reasons. First, as mentioned, the beta distribution is a suitable candidate for predicting rates as it is bounded between [0,1]; in Figure 6.1(b), this behaviour is reflected by the mean prediction tending to one as the MAC-divergence reaches unity, and the confidence intervals at the limits of the observed data reaching zero and unity.

Secondly, as the model uses latent GPs to directly predict parameters of the beta distribution over the quality measure, this model can accurately capture changing levels of uncertainty, allowing for decision making to be more reflective of the complete range of potential outcomes. For example, in Figure 6.1(b) it can be seen that that as the MAC-discrepancy approaches unity, the confidence intervals contract, indicating increased levels of confidences in the outcomes of transfer. In addition, it can be seen that the probability distribution is skewed towards higher accuracy values, as shown by the mean prediction.

In this case, the confidence intervals appear particularly wide at the lower range of observed data. This result may be because only a few transfer tasks had such low MAC-discrepancy values in this population, meaning only a few data were available in

this region of the input space. In practice, obtaining many examples of transfer across the entire range of a similarity measure may be challenging. However, it may only be necessary to train a model capable of generalising to similarity values that produce desired results (i.e. above a minimum threshold), if it is assumed that transfer outcomes will have a monotonic relationship with the similarity measure. For example, if a model were trained on a more limited range and it was determined that transfer was not feasible for MAC-discrepancy values below 0.6, asset managers could be confident that values below this threshold would also not lead to suitable predictive models.



(a) (b)

FIGURE 6.2: Examples for predicting the probability of achieving positive transfer (a) and an accuracy greater than 0.9 (b), via the beta-likelihood GP.

Once a suitable regression model has been learnt, principled analysis could be performed to answer the fundamental question "when to transfer?". In the context of PBSHM, one way to answer this question would be to only transfer when the probability of negative transfer is low. Figure 6.2(a) shows the expected probability of positive transfer found using the beta-likelihood GP. Given that the prediction task contains ten possible labels, a weak-random model would result in a 10% chance of correct classification, so to evaluate the likelihood of positive transfer, the beta GP could be used to estimate the probability that transfer would lead to an accuracy greater than 0.1. In this population, it can be seen that MAC-discrepancy values above approximately 0.77 have an expected probability of positive transfer of unity.

Another way of deciding "when to transfer?" may be to predict when transfer would likely result in a required criterion for the quality of predictions. To illustrate this idea, the probability of transfer yielding an accuracy above 0.9 is presented in Figure 6.2(b). For the given transfer strategy and prediction task, it can be seen that a high probability of achieving an accuracy above 0.9 requires high levels of similarity, although an accuracy of 0.9 may be challenging to achieve in general in a ten-class classification problem. In practice, a more full decision analysis may consider the cost associated with correct classification or misclassification of data; however, this is left to future work.

6.5 Discussion and conclusions

Deciding "when to transfer?" is a fundamental question that must be answered before allocating resources to train and deploy predictive functions using transfer learning. Answering this question is particularly challenging without sufficient labelled target data to validate the quality of predictions produced by transfer learning-based models. As a proxy for directly evaluating the performance of each model learnt via transfer learning, this chapter proposes leveraging a regression model to predict the range of outcomes given an informative-similarity measure. To generate datasets to train such models, it is proposed that source/target pairs could be generated from a set of candidate source domains, where labels for the "target" domain would be hidden during training, but could be used for obtaining a measure of quality for classification. In addition, the results in this chapter suggest that using a beta-likelihood, such as in the beta-likelihood GP, would be advantageous for this task, as it ensures that rate predictions are physically meaningful and that uncertainty is realistically quantified.

This regression framework was demonstrated using the numerical case study first presented in Section 5.4.1. The MAC-discrepancy was used to predict the accuracy of a classifier following NCA for a given source/target pair. These initial results showed the beta-likelihood GP could effectively capture this relationship, while constraining predictions to physically-meaningful values. In addition, it was found that as the MAC-discrepancy increased, the beta-likelihood GP predictions tended towards unity, and prediction uncertainty decreased, suggesting high certainty in positive transfer. It was also demonstrated how such models could be used to estimate the probability of negative transfer. These results demonstrate how informative-similarity measures, such as the MAC-discrepancy, could be used to guide operators on "when to transfer?".

While this chapter presents promising initial results for this methodology, there are several key limitations of this study and many interesting potential directions for future work. The main objective of this framework is to guide operators in relation to "when to transfer?". One limitation of the approach presented in this paper is that classification rates were predicted only as an overall average across all classes using accuracy. In practice, the impact of misclassification depends on whether the resulting misinformation would lead to a suboptimal maintenance decision, and the associated consequences. To address these limitations, this approach was incorporated into a decision framework in Hughes *et al* [218]. In this paper, a constrained regression model using a Dirichlet likelihood was used to predict true-positive, false-positive and false-negative rates using the MAC-discrepancy, and the decision "when to transfer?" was framed in relation to the value of information gained via transfer learning. Furthermore, in [172, 218], the authors further validated this decision-based framework with an experimental population

of eight experimental structures representing aeroplanes using a similarity measure based on structural properties. These studies suggest that the expected value of information could be a valuable means to justify "when to transfer?" to decision makers, motivating further research to both develop and validate a framework for estimating the value to be gained via transfer learning.

The proposed approach could also be extended by generating results for multiple transfer strategies, offering a practical alternative to conventional model selection in unsupervised transfer-learning scenarios. In this way, optimal feature sets and transfer-learning algorithms could be selected for a given similarity level. Such an approach could be seen as related to the multi-task hyperparameter selection scheme used in the previous chapter, where instead of finding parameters that provide the best classification across the entire set of source datasets, using this framework, transfer strategies and corresponding hyperparameters could be selected given the degree of similarity of the transfer task.

Another important direction of future research involves the development of more informative similarity measures. In this chapter, the MAC-discrepancy was demonstrated as a potential candidate for scenarios where there are no target labels. However, only using this measure may result in high uncertainty predictions of transfer outcomes in some cases, and it may be misleading in others, as it will not be indicative of all structural differences relevant to damage prediction. For example, in the previous chapter, the MAC indicated that a metal and composite helicopter blade were highly similar, allowing for a damage classifier to be transferred. However, if the transfer task involved predicting crack propagation, the difference in materials would be a more critical consideration. An interesting direction for future research could be in considering additional structural information to reduce variance in predictions; for example, the MAC could be used with knowledge-based measures, such as those based on AGs [171].

The beta-likelihood GP was shown to have beneficial properties for the prediction of quality measures; however, it is also important to highlight several limitations. First, this model requires the estimation of two latent GPs, meaning there are more hyperparameters compared to a standard GP. In general, GPs are known to be sensitive to hyperparameters [213], meaning increasing the number of hyperparameters may require larger datasets to prevent overfitting. Second, inference must be performed by approximate inference methods. Here, variational inference was used, which results in a computationally-efficient inference scheme, but it only approximates the true posterior and can be susceptible to convergence to local minima. In addition, predictions are also more computationally expensive, as calculating the predictive posterior requires numerical integration via sampling. In the application presented in this chapter, computational efficiency may not be a limiting factor, as the predictions only need to be made once per

pair of structures in a population. However, it would be highly beneficial to further develop methods that can robustly estimate constrained regression functions with limited data.

While the framework presented in this chapter presents a potential solution to the issue where labelled data relating to the target structure are not available for validation, it also introduces the question – "would it be feasible obtain sufficient data from a representative population, such that a model will generalise to unseen target structures?" This question could be considered in three parts: "is it feasible to obtain a diverse source dataset from a single structure representing each health state of interest?", "is it possible to ensure the training population is representative of new target structures?", and "could data from many structures be obtained?".

The first of these issues has been discussed in Chapter 4 and is further discussed in Chapter 8. To briefly reiterate, it is not likely that a single real structure would experience every health state of interest, and methods such as multi-source transfer learning or transfer from numerical models may be required. If future work finds these methods necessary and feasible, these new transfer learning methods could still be used in the presented framework; however, if the source domains are numerical structures, additional validation of the method would be required to ensure that the relationship between numerical structures effectively captures the relationship from numerical to real structures.

The second issue is that these models must be trained on representative training populations. Thus, future work should focus on identifying the conditions under which the training population can be considered representative of target structures. To ensure the model has not over-fit, a typical approach would be to use a test dataset relating to a set of "pseudo-target" structures. However, this solution would still require an understanding of the variation within the population and a method to ensure the target structure is well represented by the test dataset. This issue further motivates the further development of principle similarity measures, with the aim of discovering what variations in structures affect transfer for given transfer tasks. It may also be the case that investigating all forms of structural variation is not feasible, and the proposed approach could require some grouping of structures based on engineering judgement. Using the example of the metal and composite helicopter blades from the previous chapter, a solution using engineering judgement may involve restricting a population to only include structures with similar materials.

The issue of obtaining data from many structures for training may also limit the application of the proposed framework. In some cases, a single operator is responsible for many similar structures, as in wind farms or national bridge infrastructure, while in other cases, datasets may only be available across multiple organisations. This issue

highlights the importance of methods to facilitate data sharing, which could include methods to promote privacy of certain aspects of the data or to seamlessly share large datasets, and might also need to represent a cultural shift in data sharing between organisations for some industries. In addition, while sensing of structures is becoming more common, there are limited historic monitoring campaigns that could be used, which perhaps suggests these methods could not be applied using real source datasets before more SHM datasets are collected.

Chapter 7

Active transfer learning for SHM with an application to bridge monitoring

An outstanding issue for transfer learning in SHM is that previously investigated methods assume that no labelled data are available in the target domain. Consequently, they do not address how such technologies can be incorporated into an online framework – updating as labels are acquired throughout the monitoring campaign. This chapter proposes a Bayesian model for DA in PBSHM that enables continual improvement of unsupervised mappings using a limited quantity of labelled target data. Furthermore, the model is integrated into an active learning strategy designed to guide inspections to select the most informative observations to label; therefore, further reducing the quantity of labelled data required to train robust target classifiers.

7.1 Introduction

One of the core contributions of this thesis has been the development of a transfer strategy to select features and perform DA where only data from the undamaged target structure are available. This strategy addresses limitations with previous approaches by addressing the partial-DA problem; thus, it could facilitate the use of labelled source data to learn target classifiers from near the start of a new monitoring campaign. However, the lack of information used to learn mappings by NCA may require high structural similarity to mitigate the likelihood of negative transfer [57]. SHM data are typically acquired online, meaning that as the monitoring campaign progresses, available target data will gradually become more abundant, providing additional data to update DA mappings

[56]. Thus, this chapter investigates how online data can be used to incrementally improve a DA mapping.

In practice, SHM data would be sequentially observed and potentially labelled via inspections [1]. Consequently, DA algorithms that address two specific research challenges must be developed for application to online SHM data. Firstly, at the onset of the monitoring campaign, the target structure may only have data related to normal operation, while the source dataset(s) should encompass a wider range of the health states of interest, a situation requiring methods robust to imbalanced data scenarios [48]. Second, these methods should be capable of adapting, as contextual information is acquired during the monitoring campaign [2]; this may improve generalisation and reduce the likelihood of negative transfer as the monitoring campaign progresses [57], when maintenance decisions become more critical.

Given the cost of acquiring labels, it would be beneficial to schedule inspections to coincide with the most informative data. *Active learning* has been demonstrated to significantly reduce the label requirements in SHM by using a predictive model to infer which unlabelled data would provide the largest improvement if they were labelled [2, 191, 219–221]. In the transfer-learning literature, guided sampling strategies have been proposed to leverage source data to improve the initial model [167–169] and have been demonstrated to mitigate the class imbalance issue in DA [170]; these methods will be referred to as *active transfer learning* methods. However, to the author's knowledge, these methods have not been investigated in SHM.

This chapter proposes the first online transfer-learning strategy for PBSHM by incorporating a novel Bayesian DA method into an active-learning framework. The core distinction between previous DA methods and the active transfer learning method presented in this paper is demonstrated in Figure 7.1. Figure 7.1 shows the conventional unsupervised DA setting, where target data are unlabelled, but representative of all classes in the source domain, and a single mapping is learnt to classify target data, which represents the approach taken by previous DA methods applied to SHM [12, 41, 135, 140, 143, 144, 146, 147]. In comparison, active transfer learning assumes that initial mappings must be estimated with limited initial target data, and mappings will be updated with labelled data as it is obtained during the monitoring campaign via inspections. To facilitate active transfer learning in sparse target data scenarios, a Bayesian transfer learning model is proposed that addresses limitations with previous methods proposed in the transfer learning literature [131–133], by allowing for a probabilistic mapping, regularised using engineering knowledge, to be inferred in conjunction with a flexible classifier trained using source data. The proposed approach is validated using an experimental dataset consisting of three laboratory-scale bridges with

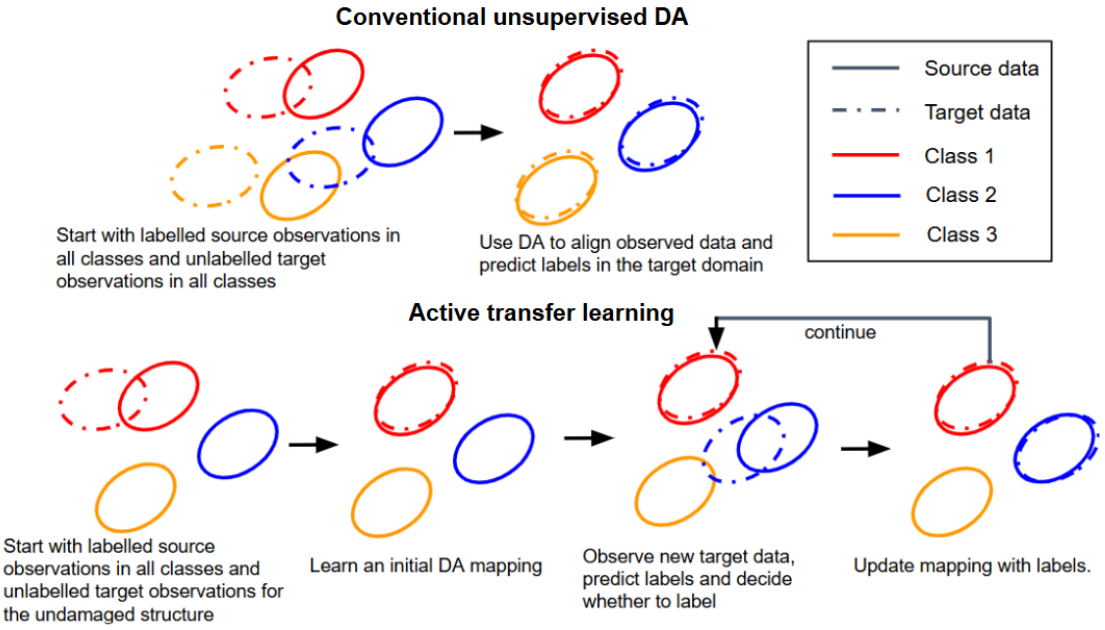


FIGURE 7.1: A demonstration of the assumptions made by conventional unsupervised DA (top), where a labelled source dataset and an unlabelled target dataset representative of all classes are used, and of the active transfer learning approach proposed in this chapter (bottom), which assumes that initial target data are limited and not representative of all classes in the source domain, and that target data will continue to be observed and occasionally labelled, allowing for mappings to be updated.

varied support locations; these structures were subjected to a range of damage states and environmental conditions using an environmental chamber.

This chapter is structured as follows. Section 7.2 outlines the necessary background. Section 7.3 introduces the proposed DA methodology, and Section 7.4 presents the experimental datasets and demonstrates the transfer of a damage classifier using the proposed method. Finally, conclusions are presented in Section 7.5 and potential future work is highlighted.

7.2 Towards an online framework for transfer learning in PBSHM

Transfer learning is dependent on the assumption that structures will have a sufficiently "similar" response to damage, such that ϕ can be learnt without labels; this assumption is particularly critical for unsupervised transfer learning [57]. If this assumption is not satisfied, unsupervised transfer learning can lead to negative transfer. Furthermore, unsupervised transfer learning may be particularly challenging in SHM, as in many cases only data from the undamaged, and perhaps a few damage states, will be available

in the target domain, as illustrated at the start of the active transfer learning example in Figure 7.1.

Many of the previous applications of transfer learning to SHM and condition monitoring have focused on unsupervised DA to transfer source labels in the absence of any target labels [12, 41, 135, 140, 142, 143, 145, 147, 153–156]. However, these methods do not address the two research challenges required for online transfer learning. First, they assume there are unlabelled observations for each of the damage states of interest in the target domain. Moreover, if damage in the target structure is detected, a few labels could be collected; however, these previous applications do not incorporate any labels.

While in SHM it often unfeasible to obtain comprehensive labelled datasets because of budget constraints and/or safety/accessibility issues, it may still be feasible to obtain labels for a few health states throughout the operation of the target structure via periodic, or guided, inspections. In such cases, supervised transfer learning could be used to increase the available information to learn shared regularities between domains [57]. Generally, the likelihood of negative transfer is lower when labelled data are available and it may have the potential to enable transfer between less similar structures [57]. In addition, when labelled data are available, some issues related to class imbalance can be mitigated by aligning data by class. One approach to supervised DA involves using a shared classifier to perform DA [131, 133]. Using a discriminative classifier for DA may further reduce class imbalance issues, as it only estimates class boundaries rather than the underlying distributions.

As discussed in Section 3.2.8, a few examples exist where target labels have been used for transfer learning in SHM. However, most of these applications require labelled data from all classes of interest in the target domain. In practice, damage will be observed and (potentially) labelled sequentially throughout the monitoring campaign. Thus, to transfer a classifier trained using data from multiple damage states in the source domain, transfer must be performed using no target labels, or a limited set of target labels which only represent a subset of all classes in the source, i.e $\mathcal{Y}_t \subseteq \mathcal{Y}_s$. As far as the authors are aware, Gardner *et al.* [53] is the only example of supervised domain adaptation for PBSHM. In [53], kernelised Bayesian transfer learning (KBTL) was applied to learn a shared classifier and a shared feature space across multiple structures with different feature dimensions. It was shown that this approach could classify damage states where there were no labels in that specific domain; however, the case studies assume most classes included labels.

To summarise: this chapter aims to propose a practical online framework for transfer learning in PBSHM. To achieve this objective, a transfer learning model is proposed to address the following two core limitations of previous methods:

1. At the start of a monitoring campaign, a transfer learning method must be capable of learning a mapping that allows for a classifier to generalise to the target domain using limited data corresponding to the undamaged target structure. Given that this mapping must be learnt with limited data, it is also proposed that this issue motivates considering uncertainty arising from transfer.

2. As labels are acquired throughout the operation of a structure, a transfer learning method should be able to update, to leverage this additional information. Furthermore, it should be able to use this information to improve the prediction of classes which have only been observed (and labelled) in the source domain.

7.2.1 Selecting informative labels: probabilistic active learning

As labels are acquired throughout the target structure's monitoring campaign, it may be possible to improve generalisation and mitigate the likelihood of negative transfer of a transfer learner. However, budget restrictions will limit the number of observations in the target that can be labelled. Thus, it would be beneficial for inspections to coincide with the most informative samples to label. One approach to guide inspections is to use an initial model to classify (online) streams of data, and use the predictions to inform which samples should be labelled; generally, this is the main objective of *stream-based active learning* [222].

Active learning typically aims to develop approaches for two main settings: stream-based and pool-based [222]. In stream-based active learning, data are acquired sequentially, and the active learner must determine whether to label, or *query*, the current observation; generally, if the observation is not labelled in this instance, it cannot be labelled retrospectively. Alternatively, pool-based methods aim to label the more informative data from a previously obtained unlabelled dataset. In SHM, it is typically not possible to obtain labels of previously obtained data; only the current condition can be investigated. Thus, stream-based methods are the focus of this chapter.

The specification of the sampling strategy is crucial, as it determines which data are most likely to be selected for labelling. One of the most widely used approaches is uncertainty sampling [222]. For example, maximum-entropy sampling (MES) selects data with the highest entropy, prioritising queries for observations where the current model yields the most uncertain or "confused" label probabilities [222]. Commonly, uncertainty is measured using the Shannon entropy [223] of the posterior-predictive-distribution,

$$H(\hat{y}_i) = -\sum_{c=1}^{C} p(\hat{y}_i = c|\mathbf{x}_i, \mathcal{D}_l)\log\, p(\hat{y}_i = c|\mathbf{x}_i, \mathcal{D}_l) \tag{7.1}$$

Entropy-based sampling typically results in labelling samples that lie close to the boundaries of the classifier, which, intuitively, should be the most informative data for defining classification boundaries between previously observed classes. A weakness of this approach is that for most classifiers, observations at the extremities of the model will not be queried, meaning it may not query data corresponding to novel classes. When using generative models, another approach to uncertainty-based sampling is to sample observations with low-likelihood values [222]. These queries would appear more "novel" to the model, rather than confused; thus, this query strategy is well-suited for novelty detection. To combine the benefits of either approach, these measures can be incorporated into a joint strategy to obtain more varied labelled datasets and reduce sampling bias [2]. Other approaches aim to label samples that are expected to improve the model as quickly as possible. These methods often select samples that would lead to the largest reduction in entropy of the posterior distribution of a Bayesian model [222].

Labelling data based on a criterion has been shown capable of reducing overall labelling efforts [222]; however, training datasets will not be representative of the underlying distributions - a phenomenon known as *sampling bias*. For example, data may be over-represented near boundaries using MES. This issue may lead to worse performance than random sampling, particularly as larger datasets are obtained [221, 224]. In the worst cases, poor initial models can cause suboptimal model convergence, where data relating to the optimal model will never be sampled under the selection criterion [225]. This issue is generally dependent on the labelling criterion used; development of criteria that mitigate sampling bias is a major research focus in active learning [3, 222].

A few previous studies have demonstrated that active learning can reduce the number of labels required to train conventional machine-learning models for SHM. For example, generative mixture models have been used with a mixture of entropy- and likelihood-based [2, 14] and decision-theoretic sampling strategies [191]. In addition, to further reduce label requirements and mitigate the effects of sampling bias, the combination of semi-supervised and active learning has been investigated [14, 221], as well as the use of efficient discriminative classifiers (the relevance vector machine) [221]. Uncertainty sampling has also been used with neural networks to classify images of defects [220] and in [219], a Bayesian convolutional neural network was used for tool monitoring. Finally, [226] proposed a probabilistic framework for active sampling for a damage-progression model. However, active learning has only recently been considered in the context of PBSHM for multi-task learning (using hierarchical modelling) for regression [227], and has not been investigated in the context of classification or transfer learning.

7.2.2 Active transfer learning

Active transfer learning can address the drawbacks of considering either transfer learning or active learning separately. As previously discussed, from a transfer learning perspective, incorporating labels can reduce the likelihood of negative transfer and improve generalisation where unsupervised transfer learning alone achieves insufficient classification performance [57]. Using more informative labelled datasets selected via active sampling could achieve these improvements with fewer labels [169]; thus, facilitating the application of supervised transfer learning with smaller labelling budgets.

From the perspective of active learning, leveraging source data has several advantages. First, using transfer to initialise the active learner may result in a stronger initial model, meaning that it can select more meaningful samples from the start of the process [167]. However, it should be noted that there is also a risk that important target samples will not be labelled if initial transfer is poor; this will be discussed more in the following sections. In addition, uncertainty-based methods will likely require fewer samples from classes with abundant source data before they can be classified with low uncertainty – leading to fewer samples overall [169]. Furthermore, while conventional active learning allows for classifiers to be learnt without a fully-labelled dataset *a-priori*, observations can only be labelled as classes that have been previously observed. However, using an appropriate transfer learning strategy, classification of classes that have only been observed in the source domain could also be attempted.

7.3 Classifier-based Bayesian domain adaptation

This section presents a novel Bayesian model for DA, the DA-RVM (domain adaptation relevance vector machine), which aims to perform transfer by leveraging a prior DA mapping and limited labelled target data. The model has two core components – a classifier that is learnt using both source and (limited) target data, and a linear mapping that aims to project the target data into the source feature space. To achieve a high likelihood of classification in both domains using a single classifier, domain divergence must be low [59]; thus, the latent mapping learns to minimise the domain shift between domains. A graphical model depicting the proposed model is shown in Figure 7.2.

It is also proposed that the prior of this mapping can be defined via unsupervised DA, which assumes that suitable generalisation of a source classifier can be achieved by only accounting for marginal-distribution shift. A consequence of the Bayesian formulation is, as labelled data become more abundant in the target, the posterior mapping becomes less influenced by the prior mapping, meaning it relies less on the strict assumptions made

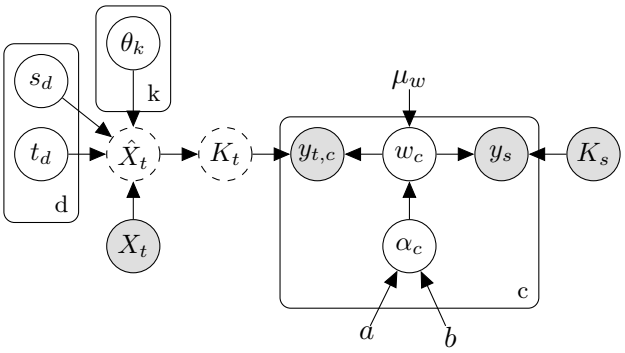


FIGURE 7.2: Graphical model representation of the proposed DA-RVM. Nodes correspond to variables: shaded nodes denote observed variables, solid outlines indicate random variables, and dotted outlines represent deterministic nodes. Arrows without a connected parent node indicate prior distributions. Plates represent replicates over dimensions for the mapping variables and classes for classifier weights.

by unsupervised DA. Furthermore, by performing adaptation via a joint discriminative classifier, this approach does not rely on assumptions about the underlying generative process of the data or require measures between the data distributions.

In this chapter, the main objective is to learn a general mapping using limited target labels that minimises domain shift between the domains, allowing a classifier to generalise to classes which have only been observed in the source domain. Thus, the mapping is restricted to a linear transformation, which is decomposed into a scale $\mathbf{s} \in \mathcal{R}^d$, translation $\mathbf{t} \in \mathcal{R}^d$, and rotation $\boldsymbol{\theta} \in \mathcal{R}^k$, where $k = \frac{1}{2}d(d-1)$. Compared to the nonlinear mappings found by many popular DA methods, a less flexible mapping was selected as it would likely require less data to prevent overfitting[1]. In addition, the mapping projects target data into the source feature space, maintaining the interpretability of the original feature space, i.e. in structural terms, increases in natural frequency values can still be interpreted as a stiffness increase. In addition, decomposing the mapping in this way promotes interpretability of the mapping itself, allowing for engineering judgement to be used to define prior mapping uncertainty and verify whether posterior mappings are reasonable.

In the context of MES, the ability to define prior uncertainty directly on the mapping parameters also provides some control over the sampling process. For example, if there is high uncertainty about the quality of initial transfer, prior uncertainty can be assumed to be high, meaning that at the start of the active-learning process, predictions in the target domain will be less certain, leading to more queries via the active-sampling procedure,

[1]It should be noted that the proposed mapping does imply a strong prior assumption about the form of the shift between domains; in many scenarios this assumption may be too strict and it could be relaxed by kernelising the data prior to finding the mapping or including additional transformation terms.

even if the source classifier is able to predict most source data with high confidence. The modelling assumptions for the mapping are given by,

$$\mathbf{s} \sim \mathcal{TN}(\boldsymbol{\mu}_s, \boldsymbol{\sigma}_s, a_s, b_s) \tag{7.2}$$

$$\mathbf{t} \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\sigma}_t) \tag{7.3}$$

$$\boldsymbol{\theta} \sim \mathcal{TN}(\boldsymbol{\mu}_\theta, \boldsymbol{\sigma}_\theta, a_\theta, b_\theta) \tag{7.4}$$

where $\mathcal{TN}$ is the truncated normal distribution, with parameters $\boldsymbol{\mu}_s$, $\boldsymbol{\sigma}_s$, a_s, b_s, which are the mean, standard deviation, lower- and upper-bound for the prior scale; $\boldsymbol{\mu}_\theta$, $\boldsymbol{\sigma}_\theta$, a_θ, b_θ are the same parameters for the rotation angles and $\mathcal{N}$ represents a normal distribution with $\boldsymbol{\mu}_t$, $\boldsymbol{\sigma}_t$ denoting the mean and standard deviation for the translation parameters. In this chapter, natural frequencies are considered as features, so the scale is restricted to be positive ($a_s = 0$ and $b_s = \infty$) and rotation is limited to $[-\frac{\pi}{4}, \frac{\pi}{4}]$ to reflect the assumption that their relationship with changing stiffness does not reverse across domains. The rotation angles, scale and translation are assembled into matrix form, denoted by $\boldsymbol{\Theta}$, $\mathbf{S}$, and $\mathbf{T}$ respectively; the target is therefore, aligned to the source by,

$$\hat{\mathbf{X}}_t = \mathbf{X}_t \cdot \boldsymbol{\Theta}^T \cdot \mathbf{S} + \mathbf{T} \tag{7.5}$$

where $\hat{\mathbf{X}}_t$ denotes the transformed target features. The classifier used in this model is a relevance vector machine (RVM), a sparse vector learner first proposed by Tipping *et al.* [228], and later extended to a multi-class setting in [229]. In the RVM, data are projected into a reproducing kernel Hilbert space (RKHS) via a kernel embedding, and sparse weights are learnt over the samples. Thus, test data are classified given their similarity (via the kernel function) to the samples corresponding to non-zero weights; these samples are referred to as *relevance vectors*. In this model, the kernel matrix is found by,

$$\mathbf{K} = \left[k(\mathbf{x}_i, \mathbf{x}_{s,j})\right]_{i \in n, j \in n_s} \tag{7.6}$$

where $k(\cdot)$ represents a kernel function and n is the total number of labelled samples $n = n_s + n_{t,l}$. Note that the target data are projected into the kernel space after they are mapped to the source feature space via equation (7.5). Thus, the mapping remains linear, while the classifier can be specified as a flexible nonlinear classifier with a suitable kernel function. The probabilistic modelling assumptions for the classifier are given by,

$$\boldsymbol{\alpha_c} \sim \Gamma(\boldsymbol{a}, \boldsymbol{b}) \tag{7.7}$$

$$\mathbf{w}_c \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\alpha_c}^{-1}) \tag{7.8}$$

where $\mathbf{w}_c$ are the weights for class $c \in C$, and prior precision values are specified by $\boldsymbol{\alpha_c}$, which are drawn from a Gamma distribution $\Gamma(\cdot)$, with shape and rate parameters $\mathbf{a}$ and $\mathbf{b}$. By specifying a and b, so the gamma distribution results in an uninformative prior on the precision, this prior promotes large precision values, effectively restricting the values for irrelevant weights to near zero. Here, relevance vectors are restricted to the source samples, which forces the mapping to align the data such that these relevance vectors are representative of both the source and target domains, implying that divergence between the domains must be low. This choice was to prevent the potential solution of finding domain-specific relevance vectors, which may lead to an arbitrary mapping. Multi-class classification is achieved via the softmax link function, given by,

$$P(y = c|\mathbf{k}_i) = \frac{e^{\gamma_c}}{\sum_{j=1}^{C} e^{\gamma_j}} \quad \text{where} \quad \gamma_c = \mathbf{k}_i \mathbf{w_c}^T \tag{7.9}$$

where $\mathbf{k}_i$ represents a kernelised sample. For the target data, classification is achieved via the same procedure, but features are first transformed via equation (7.5) before being projected into the RKHS. Since both source and target data are used to learn a classifier, this model is also related to multi-task learning [132].

While there are a variety of suitable classifiers that could be used in this framework, the RVM was chosen for three core reasons. First, the RVM is a flexible nonparametric classifier which has been shown to learn efficiently with sparse datasets [228, 229]. The RVM also produces tight decision boundaries, and prediction probabilities converge to a

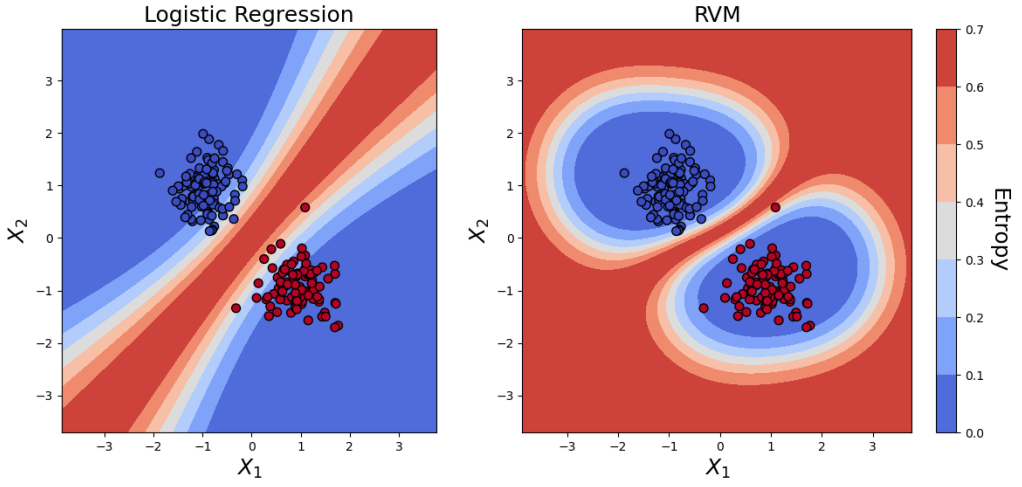


FIGURE 7.3: Toy example showing the entropy in label predictions produced by a Bayesian logistic regression model (left) and an RVM (right).

uniform distribution as samples get further from the relevance vectors (when a Gaussian kernel is used). To demonstrate the behaviour of the RVM, a toy example showing the entropy for a Bayesian logistic regression model and an RVM is presented in Figure 7.3. From the perspective of transfer, high classification likelihood would be achieved when data are mostly distributed in low entropy regions, on the correct side of the classification boundaries; thus, more restrictive boundaries restrict the possible mappings significantly – potentially leading to better alignment when only a few classes are labelled in the target domain. In the example in Figure 7.3, for the RVM, it can be seen that the region of low-entropy is significantly more restricted compared to logistic regression. There is also an additional benefit when considering this model for MES. As previously discussed, MES typically leads to sampling near the boundaries, but often will not sample novel data at the extremities of the model. However, the RVM only has high confidence in data near the relevance vectors, as can be seen in Figure 7.3; thus, it will also assign high entropy to observations at the extremities of the model, combining the benefits of both MES and low-likelihood sampling [221].

The model was implemented in a general-purpose probabilistic programming language – NumPyro [230]. The parameters of the model are inferred via MCMC using the no-U-turn (NUTS) implementation of Hamiltonian Monte Carlo [231]. The parameters for the RVM were initialised using only the source data with the RVM_2 expectation-maximisation algorithm outlined in [229]. Weights with values below 10^{-5} were pruned from the initial model to reduce the computational complexity of learning this model via sampling.

7.3.1 Related transfer learning methods

Performing DA via a joint classifier has been investigated in a few previous studies [131–133]. In [133], a shared feature space was found via a joint binary support vector machine (SVM). This approach differs from the proposed method since it cannot learn a shared space common to multiple classes, so it cannot be used to predict classes that have previously not been observed in the target domain. Hoffman *et al.* proposed the first approach to learn a shared feature space common to multiple binary SVM classifiers, to predict classes in the target domain which have not been previously labelled [131]. However, this method requires both the mapping and classifier to be learnt in an RKHS, and does not use a probabilistic model. The most similar approach is KBTL [132], which finds a projection into a latent space shared between multiple domains in a Bayesian framework. The main differences to the proposed approach are that KBTL learns a nonlinear mapping (via a kernel mapping), and does not maintain the interpretability of

the original feature space. This mapping is powerful in scenarios where feature dimensions differ between domains, or where dimensionality reduction is required. However, the flexibility of the mapping may lead to overfitting to observed target classes, it does not result in an interpretable feature space, and defining an informative prior mapping may be challenging.

7.3.2 Inferring a prior mapping with distribution alignment

In practice, to learn a discriminative classifier, it is required that the underlying conditional distributions of the training and testing data are the same, i.e. $p_s(y|\mathbf{x}) = p_t(y|\mathbf{x})$. However, the lack of labels and limited samples of data means that learning a mapping that directly aligns the conditional distributions is often not possible. As previously discussed, unsupervised DA generally assumes that the underlying conditional distributions can be aligned by minimising a distribution-distance metric between a sample of data – often these approaches aim to minimise a marginal-distribution distance, assuming that labels are unavailable or sparse [48]. Such mappings will only result in invariant conditional distributions if both the domains are sufficiently related, and there is a suitable DA method to find a mapping with the available data [57]. In addition, as discussed in the previous chapter, testing the outcomes of transfer is challenging, as in many cases labelled data will not be sufficient in the target domain to perform conventional validation, such as cross-validation. As such, current approaches to DA may need to be applied to testing data prior to direct validation.

Determining when these assumptions apply, without traditional model validation, is a critical challenge for the practical application of DA. Without validation, assessing the reliability of predictions becomes even more challenging, highlighting the importance of research into validation and prediction of transfer outcomes for PBSHM [218]. This chapter proposes that these mappings can be treated as a "prior mapping" in the DA-RVM, where prior uncertainty is defined to ensure that the reliability of predictions is reflected in the label probabilities. In this way, assumptions made when estimating the initial DA mapping can be considered as prior assumptions, where the posterior mapping parameters are updated with data directly relating to the quantity of interest – the likelihood of classification in the target domain.

In this chapter, the mapping is defined by scale, translation and rotation parameters. Thus, an appropriate set of DA techniques would be *statistic alignment* [48]. Since engineering datasets are prone to class imbalance, *normal condition alignment* (NCA) was used to mitigate issues related to class imbalance. NCA was used to define the prior expected translation and scale, and prior rotation was assumed to be zero.

7.3.3 Active sampling scheme

While incorporating labels into a DA framework may be beneficial, it is pertinent that the number of samples are minimised to reduce the associated cost of the monitoring system. To this end, an active-sampling strategy is proposed to ensure that the most informative data are labelled. This chapter utilises an MES strategy first proposed in [14]. Sampling is performed in a stream-based setting following the procedure outlined in Figure 7.4.

To decide when to query a sample, first entropy is obtained for test data using equation (7.1). To constrain the mapping prior to the acquisition of damage labels, additional classes that are not critical for maintenance decisions could be considered; in this chapter, the probabilities of all classes related to the undamaged structure are added to prevent unnecessary labelling at this boundary (see Section 7.4.2 for more details). A more general approach would be to use different weightings for each class to reflect their importance to decision making [232] – this approach should be a focus of future work. The information efficiency [223] is then used to normalise entropy between zero and one,

$$\eta(\mathbf{x}_i) = \frac{H(\hat{y}_i)}{\log(C)} \tag{7.10}$$

The information efficiency $\eta(\mathbf{x}_i)$, reflects the confidence in the label prediction compared to a uniformly distributed label prediction. Following [14], $\eta(\mathbf{x}_i)$ can be treated as a pseudo-probability that observation i should be labelled. An observation is then labelled if a random draw q from a uniform distribution $q \sim \mathcal{U}(0,1)$ is less than $\eta(\mathbf{x}_i)$. Since the probability of sampling any observation will never be zero, this sampling scheme provides some protection against sampling bias.

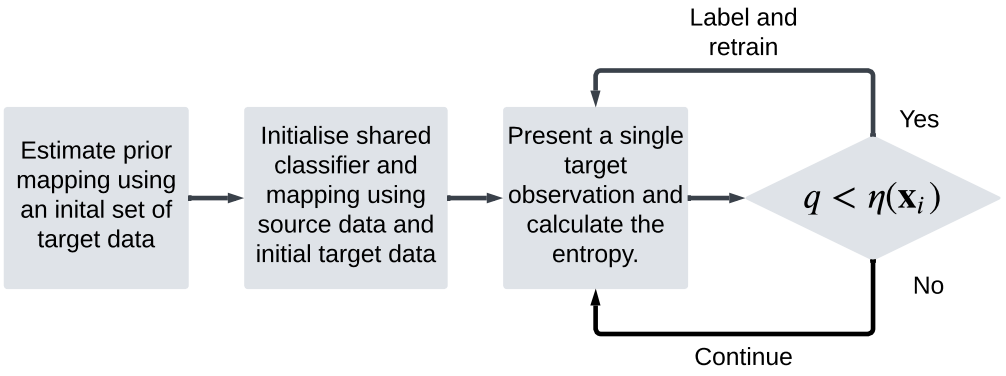


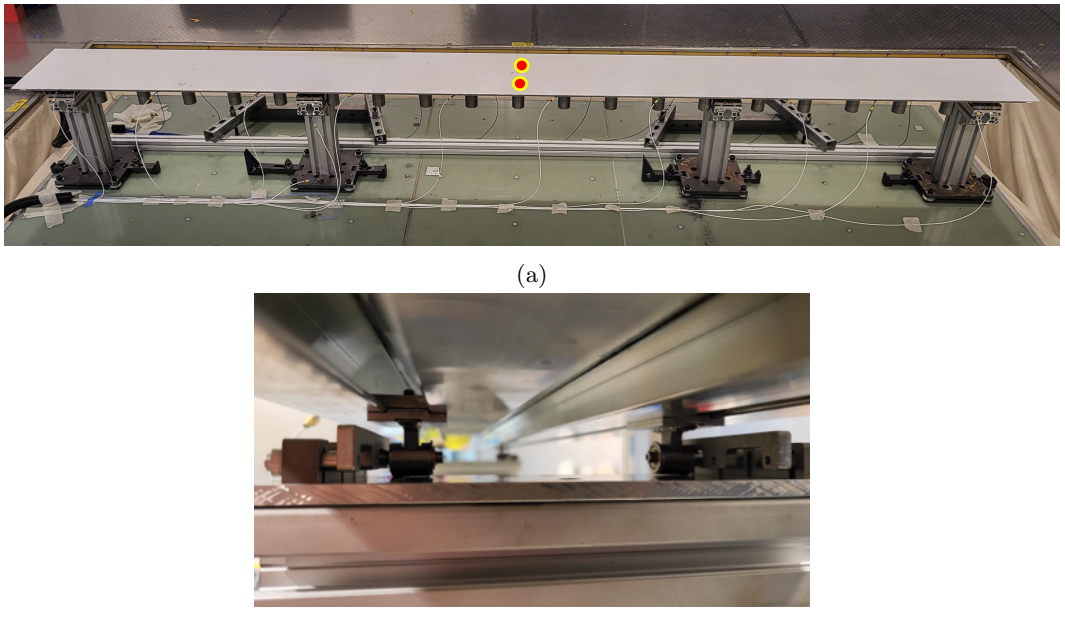
FIGURE 7.4: Flow chart to illustrate the active learning process with DA.

7.4 Transfer between laboratory-scale bridges

This section presents an experimental dataset collected to investigate the active transfer-learning approach for damage classification using a population of laboratory-scale beam-and-slab bridges. Specifically, data for three bridges with varying span lengths were obtained across changing temperatures and the same four pseudo-damage states.

The inspection and maintenance of populations of bridges presents a major challenge, and there are significant safety concerns as bridges are operated towards the end of their design life. In addition, the scale and cost of these structures will often limit available SHM data to streaming data obtained throughout the operation of the structure. While it is uncommon for two bridges to have a nominally-identical design, there exist many examples of large heterogeneous populations with slight variations in geometry (i.e. with different lengths and support locations), managed by a single asset manager. For example, the main highways agency in the UK, National Highways, was responsible for managing 9,392 bridges in 2020 [18]. This motivates the application of the proposed active transfer-learning framework to bridge-monitoring applications.

7.4.1 Experimental dataset



(a)

(b)

FIGURE 7.5: The experimental set-up to perform modal testing for one configuration (B1), showing the full bridge (a) and the connection between the deck and supports via roller bearings (b).

A population of three bridges, each with three spans, was constructed using a bespoke modular bridge kit that facilitates changing of the deck length, number of supports and

support locations; thus, allowing for controlled variation between structures. Figure 7.5(a) presents an example of one of the configurations used in these experiments. The kit consists of a set of four supports and a deck, supported by two I-beams, which are connected via a pair of roller bearings at each support, shown in Figure 7.5(b). The bearings at one end are locked, such that they behave as pin joints. The location of the supports was varied between bridges to produce a heterogeneous population. The locations of the supports for each bridge are presented in Table 7.1. The bridges are referred to as B1, B2, and B3; this abbreviation will be used for the remainder of the chapter.

The bridges were attached to a six-axis shaker table via bolts at the base of each support, within an environment chamber, as can be seen in Figure 7.5(a). The bandwidth of excitation from the shaker table is approximately 90Hz; therefore, a set of masses was uniformly distributed along the underside of the deck, shown in Figure 7.5(a), to reduce the natural frequencies and aid modal identification. Modal testing was conducted by applying a continuous white-noise random excitation via the shaker table. Data were collected via twenty uniaxial 100 mV/g accelerometers, organised in two rows of ten on each edge of the underside of the deck, and the response was measured at a sample rate of 256Hz.

To investigate challenges presented by changing environmental conditions, the first two bridges, B1 and B2, were subjected to a range of temperature effects; B3 was only tested at ambient temperatures. Specifically, the response of the bridges was measured across two temperature cycles: from 15°C down to -15°C for B1, and from 15°C to -5°C for B2. A thermocouple was attached to the deck surface to monitor its temperature. To emulate a bi-linear stiffness relationship, which can be observed in concrete bridges [185], a fabric sheet was attached to the surface of the deck and saturated with water for the second temperature cycle, such that when frozen, its stiffness would sharply increase. Data were also acquired at ambient temperatures (between 23°C and 31°C), as well as for four pseudo-damage states which correspond to two masses (damage extents), 21.6g

TABLE 7.1: Summary of the configuration for each experimental bridge structure. Support locations indicate the position of the bearings connecting the deck and the supports.

	Deck length (m)	Support 1 location (m)	Support 2 location (m)	Support 3 location (m)	Support 4 location (m)
B1	3.00	0.14	0.725	2.28	2.86
B2	3.00	0.14	0.82	2.19	2.86
B3	3.00	0.14	0.86	2.15	2.86

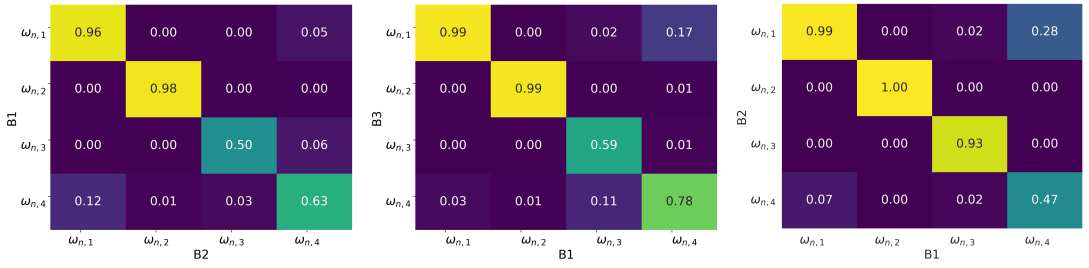
TABLE 7.2: Number of samples available after SSI per class for each bridge.

	Ambient	Freezing	21.6g off-centre	21.6g centre	64.4g off-centre	64.4g centre
B1	179	138	10	9	10	10
B2	129	54	5	5	7	10
B3	36	0	5	5	10	10

and 64.4g masses, placed in the centre of the central span in two locations, indicated by the red circles in Figure 7.5.

To obtain natural frequencies for use as features, output-only modal analysis (OMA) was performed using covariance stochastic-subspace identification (SSI); natural frequencies were extracted, based on a reference set, selected via an automated pole-selection algorithm, using the software presented in [233]. Several samples were unidentified, and a few visually obvious outliers were removed from the normal condition; however, more robust modal analysis and principled approaches for removing outlying data from the training datasets should be a focus of further work. Table 7.2 shows the number of samples per class for each dataset following modal analysis. As is often characteristic of SHM datasets, the data are imbalanced, with larger quantities of undamaged data.

The full experimental dataset is openly available (https://doi.org/10.15131/shef.data.27732792.v1). For more details, the interested reader may refer to [234].



(a) (b) (c)

FIGURE 7.6: Visualisation of the MAC scores between each pair of structures used as a source/target pair.

7.4.2 Transfer tasks and methodology

Four transfer tasks were investigated in this chapter, which are split into two case studies. The first case study investigates transfer between structures under changing temperatures using the B1 and B2 datasets, considering each structure as a source and target, resulting in two transfer tasks; these tasks will be referred to as B1→B2 and B2→B1.

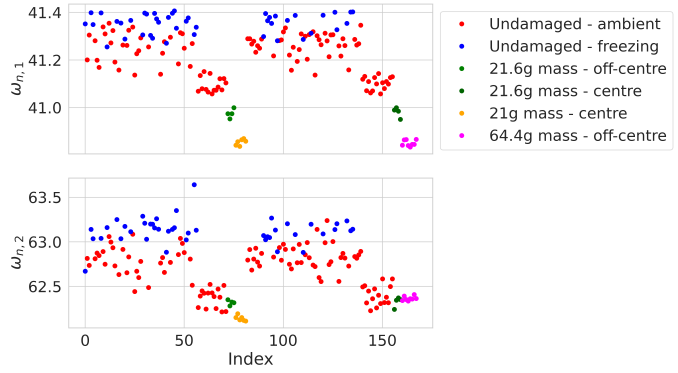
The second case study investigates transfer from datasets with comprehensive temperature data and a target with limited data, considering B1 and B2 as source domains, and using B3 as a target, resulting in another two transfer tasks, referred to as B1→B3 and B2→B3.

The results from modal analysis were used to select features for transfer via the MAC [204], following the results of the previous chapters. The MAC matrices for the first four identified natural frequencies are presented in Figure 7.6. In this chapter, the first two natural frequencies were selected as features, as they have high MAC scores for each pair of structures; however, the third mode also has a high MAC value between B2 and B3, and may be utilised for transfer in future work. It should also be noted that this comparison was facilitated by the homogeneous sensor networks on each bridge; to perform this analysis using heterogeneous sensor networks, future work is required, as discussed in Section 5.6.1 and Section 5.7.

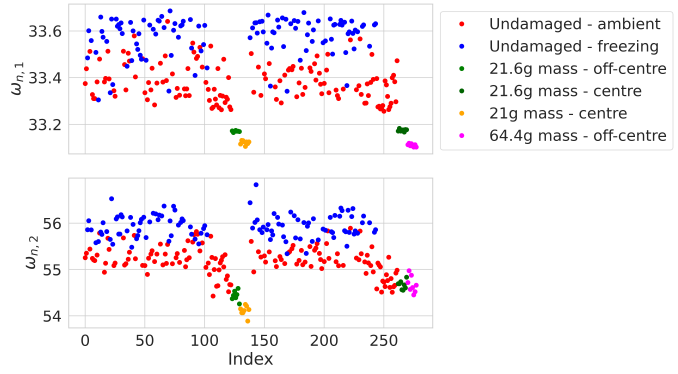
In each transfer task, the objective was to transfer a damage classifier capable of predicting the normal condition, and the four mass-states. The location of the 21.6g masses was not discriminative using the identified natural frequencies; thus, these two locations were considered as a single class, resulting in three damage classes. Furthermore, to constrain the mapping in the initial model (before damage is observed), the healthy data were split by ambient ($T > 0^{\circ}C$) and freezing temperatures ($T < 0^{\circ}C$). Therefore, the classifier was trained to discriminate between five classes - ambient and freezing normal-condition data, pseudo-damage caused by adding a 21.6g mass (Damage 1) and pseudo-damage resulting from a 64.4g mass placed off-centre (Damage 2) and at the centre (Damage 3) of the central span.

To emulate an active-sampling process for SHM, with the structure's state gradually transitioning from undamaged to damaged states, the target data were presented to the model as follows. First, data were split into training and testing datasets at a ratio of 80:20 using stratified sampling to ensure that the proportion of damage and undamaged data was consistent; the dataset was randomly shuffled and 100 training/testing datasets were generated in this way to test for differences in initial data used to learn the NCA mapping and the effect of presenting streaming data in different orders. The damage states were organised into two damage scenarios, where in a single location, damage is initialised with minor damage (the 21.6g mass) and progresses to more severe damage (the 64.4g mass). Data were then ordered so as to present undamaged data collected at changing temperatures, followed by undamaged data collected under ambient conditions, and subsequently by a damage scenario. Thus, each target domain includes two cycles of normal-condition data, followed by a damage scenario. Figure 7.7 presents an example of a single repeat of the training data (the first two natural frequencies), for each target

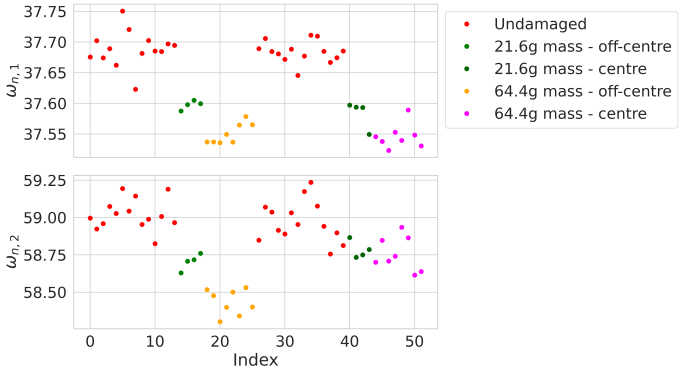
domain considered. It can be seen that the expected range of values for both the first and second natural frequencies does not overlap between domains, motivating the application of mapping-based DA for transfer.



(a)



(b)



(c)

FIGURE 7.7: Example of the ordered training data used for the active sampling process when B1, B2, and B3 are considered as target domains, presented in panel (a), panel (b) and panel (c) respectively.

At the start of the active-sampling process, the model was initialised using the source training dataset and only a subset of target training data, representing data corresponding to the undamaged structure. In each case, the initial data used for NCA were

selected to correspond to similar temperatures to account for class imbalance, following the results in Chapter 4. Specifically, for B1→B2 and B2→B1, 70 initial data were used, and the ambient undamaged class was used to estimate the NCA mapping, as B1 contains data corresponding to lower temperatures. For B1→B3 and B2→B3, only 14 initial data were used as there were fewer normal-condition data. In addition, data in B3 were only collected at room temperature; therefore, NCA was learn using data above 23°C in both domains.

The remaining data were presented sequentially, being labelled using the probabilistic sampling strategy discussed in Section 3.3; uncertainty between the ambient and freezing classes was not considered to prevent unnecessary sampling of normal-condition data. Since there is significant class imbalance, the macro F1-score was used to assess classification performance on the entire test data, which included data from each class (for details see [12]).

The Gaussian kernel was utilised in the RVM as it is flexible and well-studied [20, 229]. Following [221, 229], the bandwidth was defined as $\frac{1}{d}$. To demonstrate the benefits of incorporating transfer into the active-learning procedure, results were obtained for an active learner, using the RVM_2 algorithm from [229], trained solely with target data. Since the target-only RVM requires multiple target classes to be initialised, three random samples from the first damage scenario were selected to initialise the classifier. In addition, the RVM_2 algorithm was used with labelled target data to provide a comparison to supervised learning, and using no DA and source labelled data to demonstrate the requirement for transfer learning.

The specification of variances on the prior mapping parameters reflects the confidence in the NCA mapping, given the initial quantity of data. Thus, variances for all mapping parameters were set as $\sigma_t = \sigma_s = \sigma_\theta = 0.1$ in the first case study, and the variance for translation and scale were increased to $\sigma_t = \sigma_s = 1$ for the second case study, since only very few data were used to learn the NCA mapping, leading to large discrepancies in mean and scale between domains, as discussed in Section 4.5.

7.4.3 Case study: active transfer learning under changing temperatures

Figures 7.8(a) and 7.8(b) show the F1 scores across the test set after each unlabelled observation was presented, with solid lines representing the mean F1 scores from 100 repeats and the shaded region indicating the 10th to 90th percentiles. It can be seen in both cases that naively applying a source-only classifier led to poor generalisation in the target domain, indicated by the orange line. This motivates the application

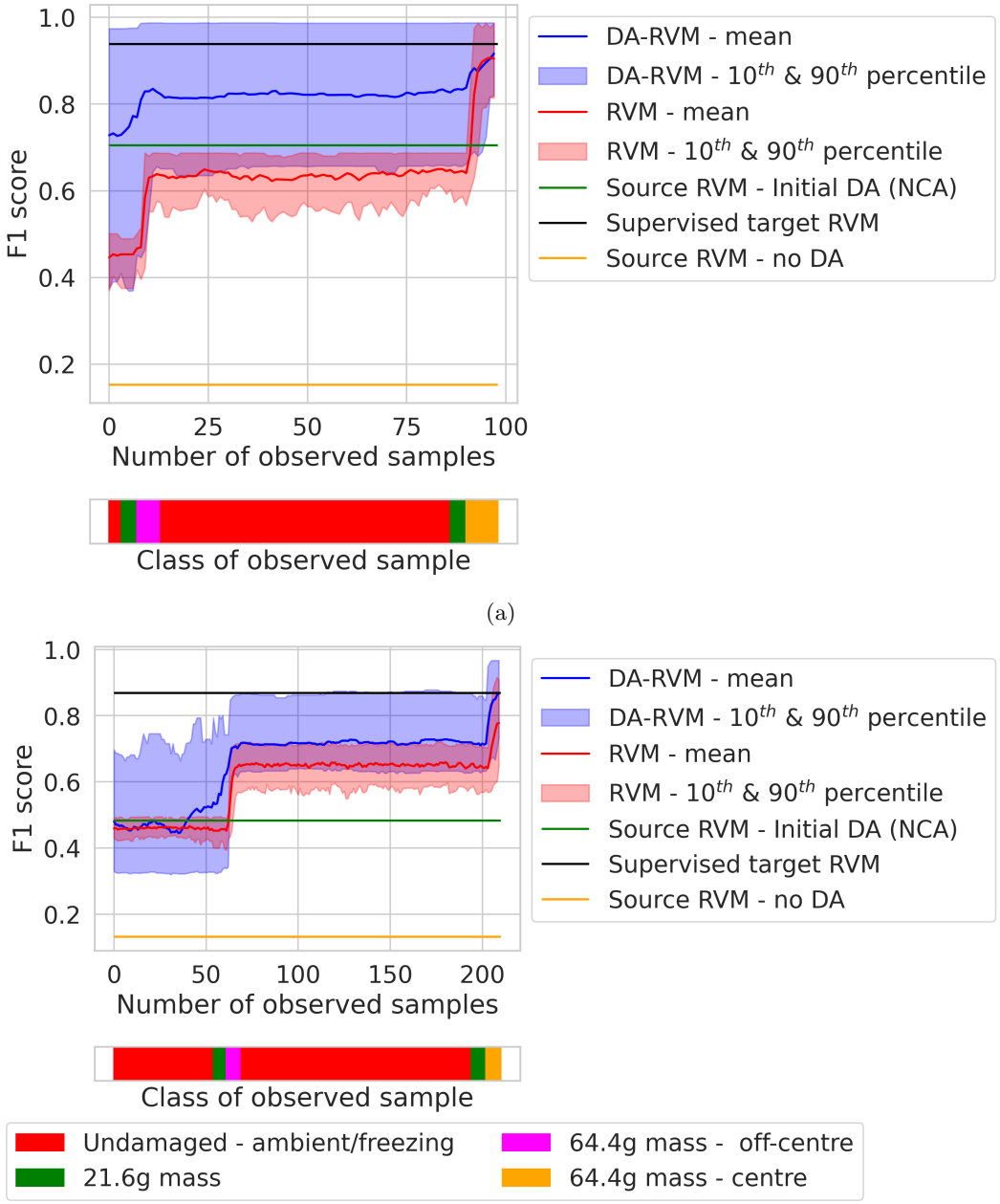


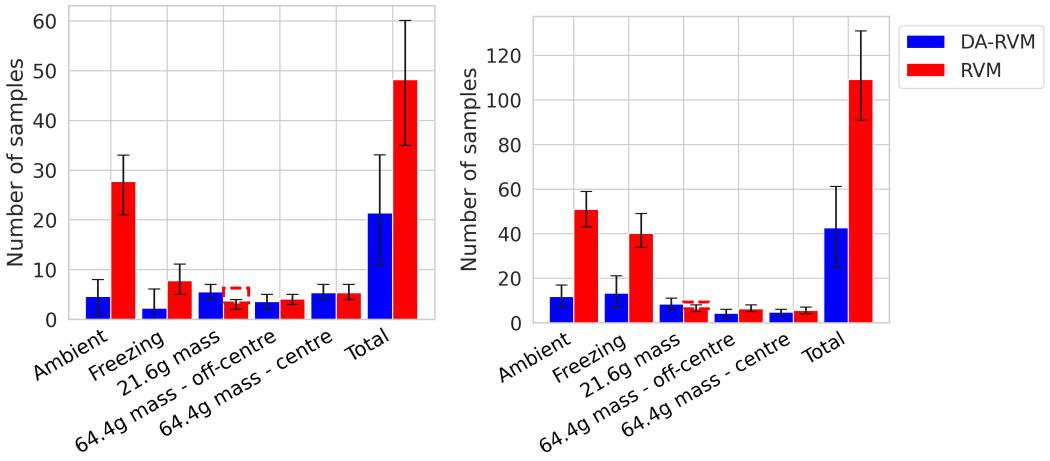
FIGURE 7.8: The test F1 score vs the number of unlabelled samples presented to the active learners for B1→B3, shown in (a), and B2→B3 presented in (b). Mean F1 scores are shown by solid lines, and the region between the 10th and 90th percentiles is shown by the shaded region, with the DA-RVM given in blue and the target-only RVM in red.

of transfer learning; it can be seen that applying NCA improves generalisation of the source classifier to the target domain (the mean F1 score is indicated by the green line). However, NCA still exhibits significantly worse performance compared to a fully-supervised classifier learnt using target data, indicated by the black line in Figure 7.8, motivating the incorporation of additional information to further improve the initial NCA mapping.

In both cases, updating the NCA mapping with the DA-RVM using labelled target data significantly improved the F1 scores. While, before any damage was observed, the DA-RVM produces similar results to NCA (the expected initial mapping), after a few observations from the first damage scenario (green and magenta regions on the colour bar), the mean F1 scores improve significantly. In addition, the DA-RVM consistently produces better classification than the target-only RVM prior to observing all classes, indicating that leveraging both an informative initial mapping and a few labels, the DA method is able to learn a classifier that can extrapolate to yet to be observed classes in the target. Moreover, the 10th percentile does not generally produce lower F1 scores than the mean result of the target-only RVM, demonstrating robustness to negative transfer. Finally, after all data are presented to the DA-RVM, it achieves a mean F1 score close to the fully-supervised RVM, matching the target-only RVM in B1→B2 (Figure 7.8(a)) and exceeding it in B2→B1 (Figure 7.8(b)).

Although the DA-RVM achieves similar performance to the fully supervised target model at the end of the active-sampling procedure, it is able to achieve this result using fewer labelled observations, as shown in Figure 7.9[2]. The DA-RVM used 10.2% and 12.0% of samples for B1→B2 and B2→B1, respectively, compared to 23.4% and 31.5% for the target-only RVM. While both methods reduce the number of labels required compared to a fully-supervised RVM, the DA-RVM results in fewer queries and a smaller reduction

[2]The dashed line indicates the samples used to initialise the target-only RVM; these were not selected during the sampling process.



(a) (b)

FIGURE 7.9: The number of observations queried via active sampling using the DA-RVM (blue) and target-only RVM (red), for B1→B2, shown in (a), and B2→B1 presented in (b). The black lines indicate the range of samples, showing the 10th and 90th percentiles, while the red dashed lines above the 21.6g mass bar represent additional samples used to initialise the target-only RVM.

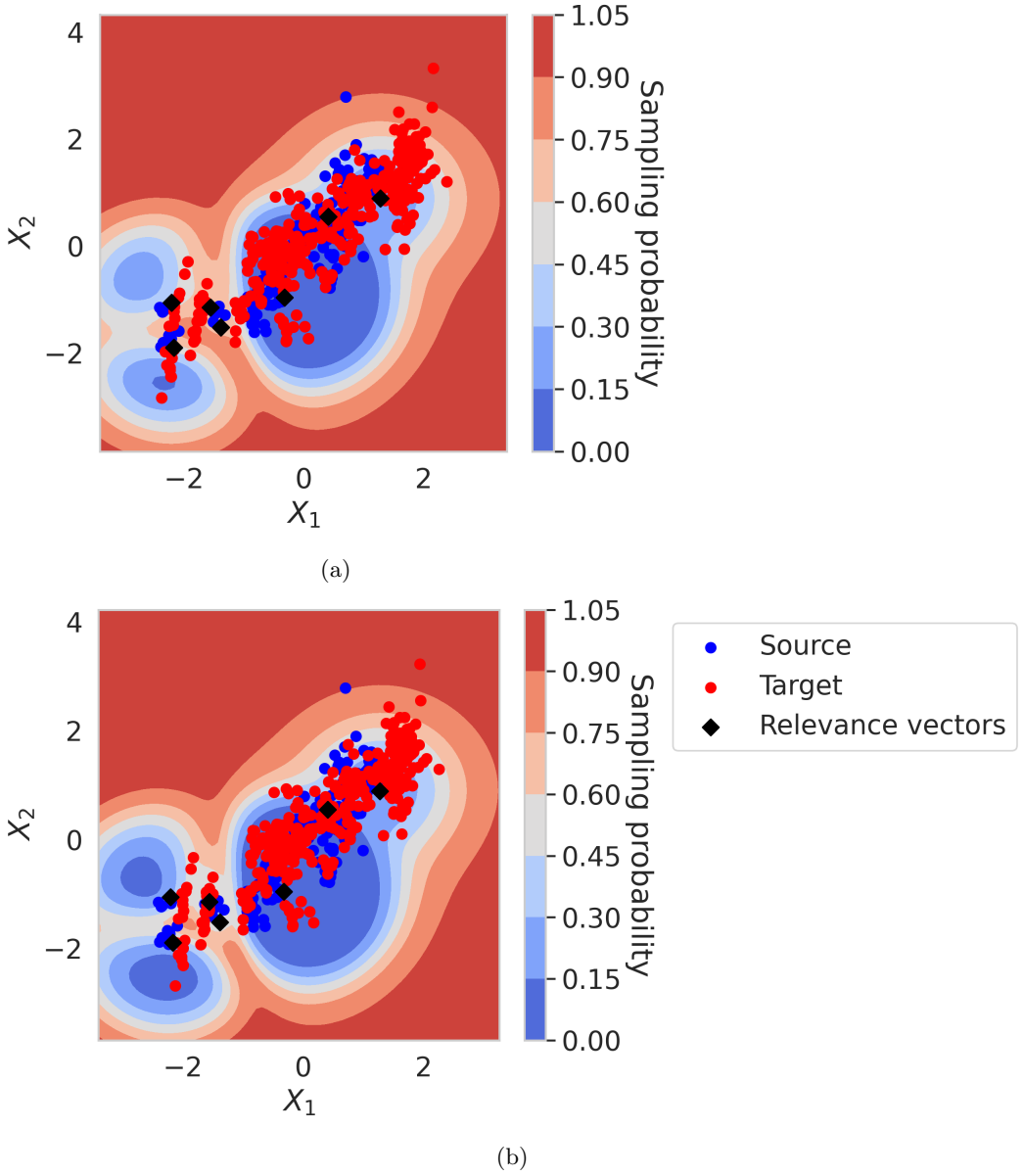


FIGURE 7.10: An example of the sampling probability for one test repeat (training and testing data), with the target data and sampling probabilities mapped to the source domain via the expected posterior mapping after the DA-RVM was presented with all data. For B1→B2, shown in (a), and B2→B1 presented in (b).

in classification performance. In addition, the DA-RVM resulted in far fewer normal-condition data being labelled by increasing confidence in predictions of the undamaged state by leveraging source data, which in practice would reduce unnecessary inspections.

Compared to conventional active learning, leveraging source data not only presents the opportunity to classify data in yet to be observed classes, but also facilitates more efficient querying behaviour. This improvement can be explained by inspecting the sampling probability of the DA-RVM; an example of the sampling probability in the

initial classifier for the B2→B1 task is given in Figure 7.10(a), with the target data and entropy mapped to the source feature space via the expected posterior mapping. It can be seen that there are already regions where sampling probability is low when the DA-RVM is initialised, particularly for samples from the largest cluster, which represent the undamaged data. As such, it appears that the source data allow for the initial model to have better-defined boundaries, guiding the labelling process to prioritise damage classes, where classes are less separable and there are fewer data in the source domain. It can also be seen that by using an RVM as a classifier, the model would effectively sample novel data, as the DA-RVM effectively produces low-entropy regions near observed data, whereas the extremities of the model have a sampling probability near unity. The sampling probability for the DA-RVM after the active sampling process is shown in Figure 7.10(b), where it can be seen that obtaining labels in the target has led to a further reduction in sampling probability in some regions. This reduction in sampling probability is particularly evident for the damage classes – the three smaller clusters.

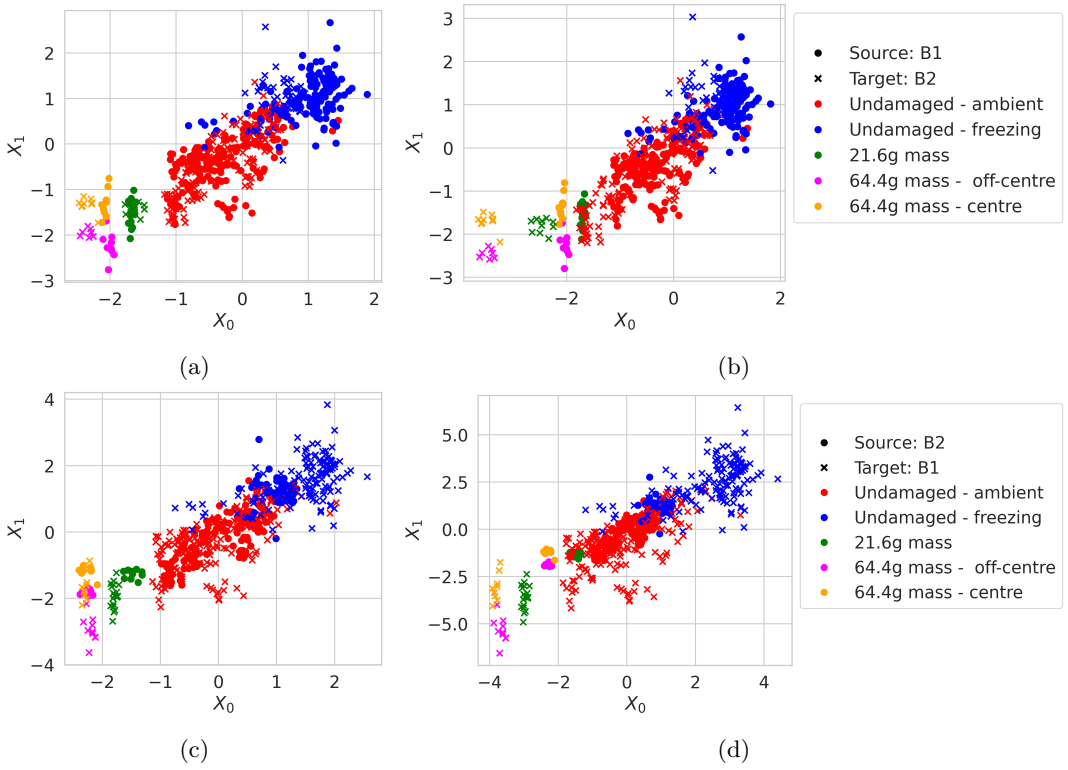


FIGURE 7.11: An example of the data (training and testing data), after the NCA mappings, which resulted in the highest and lowest JMMD values. The NCA mappings for B1→B2, shown in (a) and (c), and B2→B1 are presented in (b) and (d) for the lowest and highest JMMD values, respectively.

A core advantage of the DA-RVM over conventional DA is its ability to incrementally correct poor initial alignment using labels, which is particularly useful when sampling bias prevents accurate estimation of distribution divergence or when domain similarity is insufficient for unsupervised DA. This advantage is demonstrated in this case study,

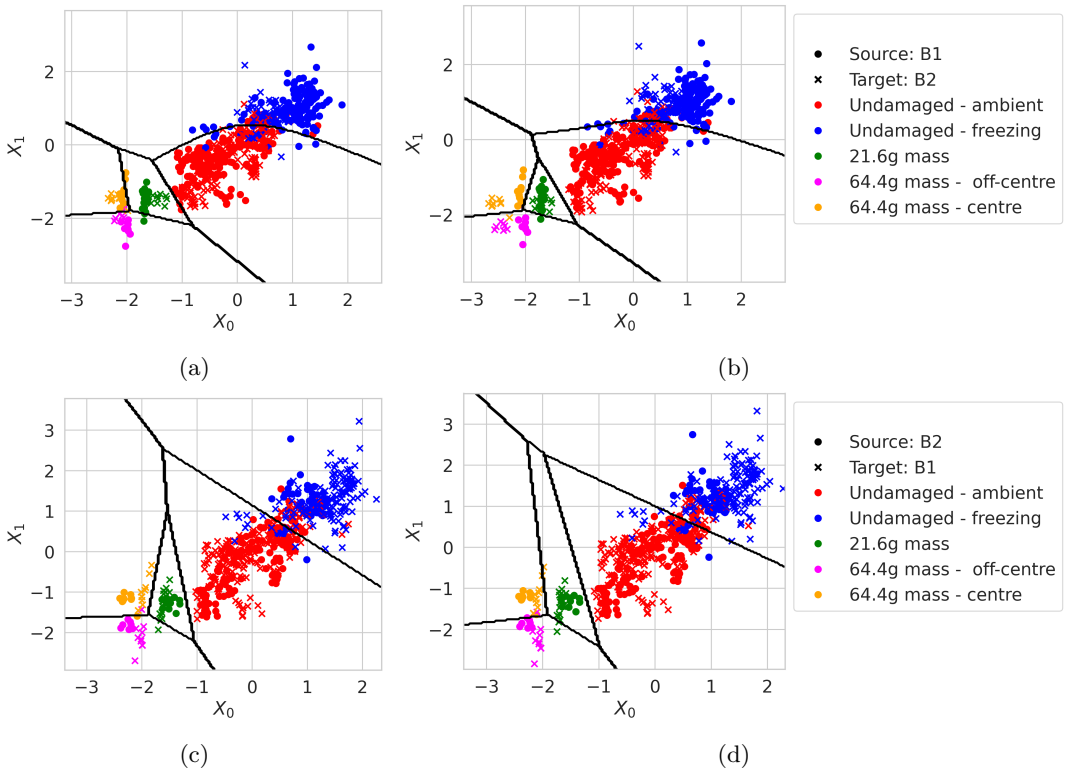


FIGURE 7.12: An example of the data (training and testing data), after the final DA-RVM mappings for the same test repeats, which resulted in the highest and lowest JMMD values after the NCA mapping. The DA-RVM mappings for B1→B2, shown in (a) and (c), and B2→B1, presented in (b) and (d), for the lowest and highest JMMD values, respectively.

since the small initial random sample across various temperatures used to learn the NCA mapping resulted in varying alignment quality. To demonstrate this impact, the JMMD was used to identify the "best" and "worst" NCA mappings, corresponding to the lowest and highest joint distribution distances, respectively. Features from the "best" mapping for B2→B1 (Figure 7.11(a)) show the target classes closely aligned with the corresponding classes in the source domain. In contrast, the "worst" mapping (Figure 7.11(b)) shows differences in scale, with source and target damage data occupying distinct regions in the feature space. These differences in initial alignment quality potentially contribute to the higher variability observed in the DA-RVM results compared to the target-only RVM (Figure 7.8).

The features found following the active-sampling process for these same "best" and "worst" repeats are presented in Figure 7.12. It can be seen that alignment could be improved in both cases, resulting in shared classifiers that can predict samples from both domains, indicated by the classification boundaries shown in black. This result supports the idea that labels from a few damage classes can be used to improve poor initial mappings, mitigating the likelihood of negative transfer later in the monitoring campaign. In addition, in the final feature space, the freezing temperature data for

both bridges can be observed as an increase in the values of these features, following an increase in stiffness, and each damage (mass) class produces a reduction in the value of these features, showing the features derived via this DA process maintain their physical interpretability, while also facilitating shared visualisation of both datasets.

7.4.4 Case study: active transfer to a target domain with limited data

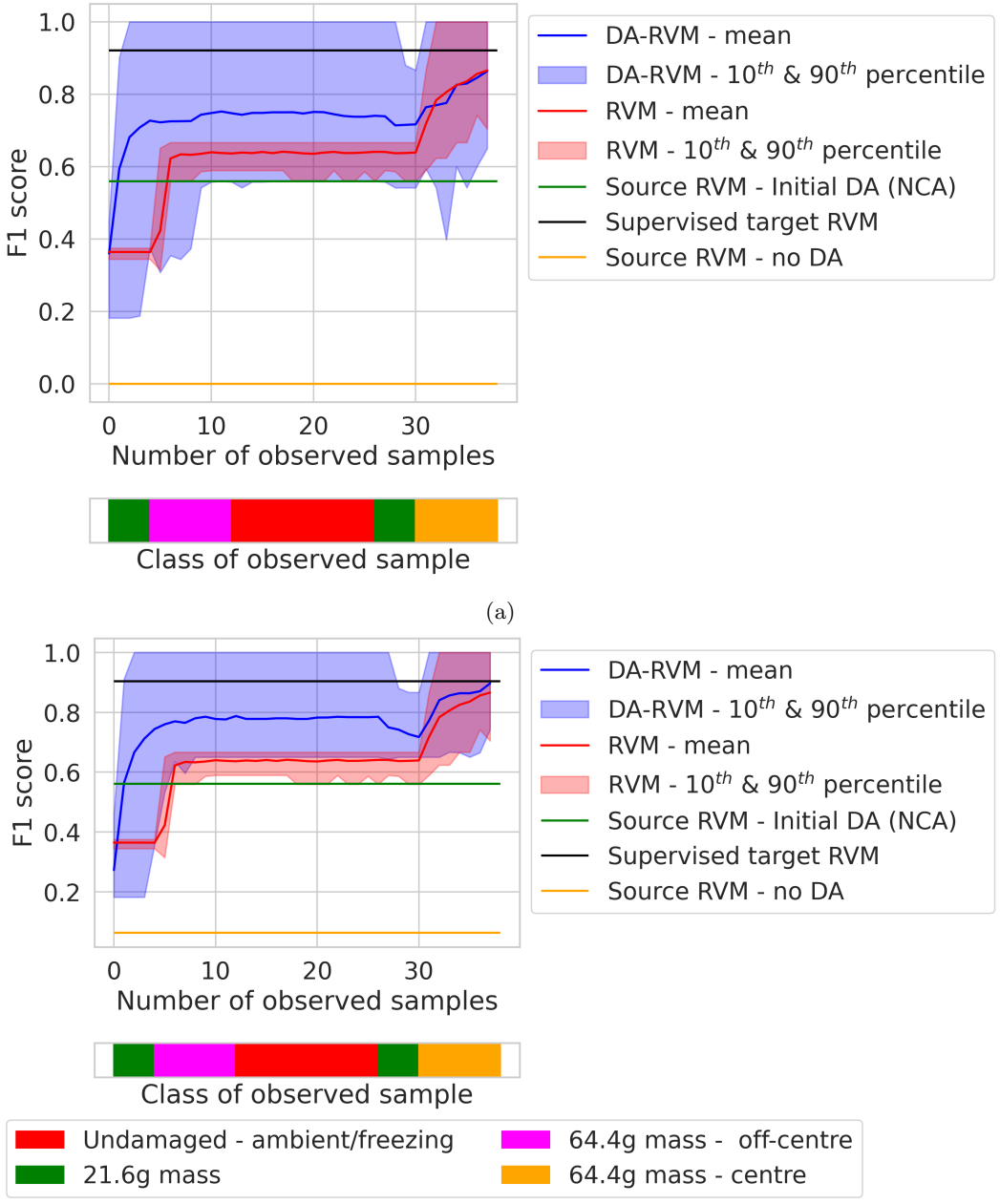
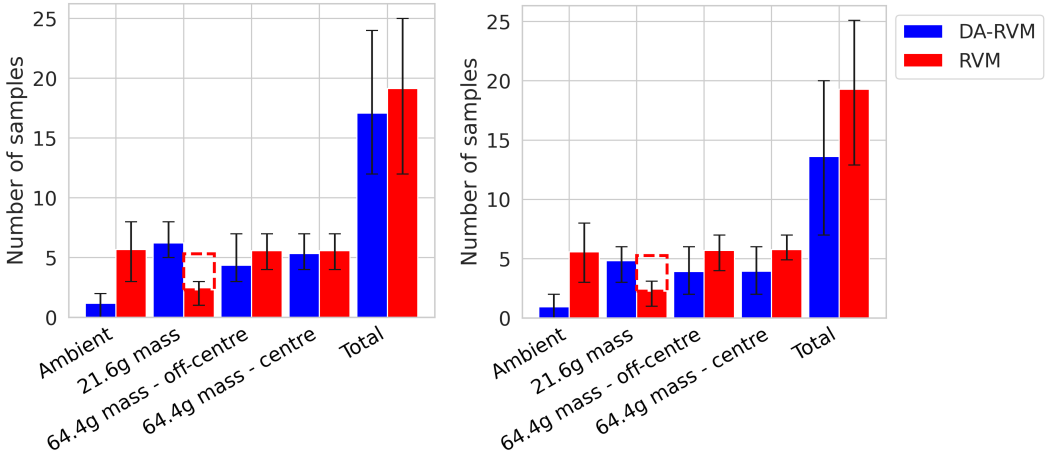


FIGURE 7.13: The test F1 scores for the DA-RVM against the number of labelled samples selected via active sampling (shown in blue) and random sample (shown in red); B1→B2 is shown in (a), and B2→B1 is given in (b).

The second case study presents a scenario where the target data to estimate initial DA mappings are extremely sparse and not representative of the same environmental effects as the source dataset. It can be seen in this case study, the initial classification rate of the DA-RVM is worse than NCA in both cases, shown in Figure 7.13. There are two potential reasons for these results. First, in this case study, only one class is available at the start of the sampling process, meaning there is limited information to constrain the posterior mapping parameters. Second, the prior variance on the translation and scale parameters was increased to $\sigma_t = \sigma_s = 1$. Thus, since the initial information available to reduce the posterior variance is limited, this choice of priors seems to have increased the uncertainty of prediction for all the damage classes to near uniform; this can been seen by the initial sampling probabilities in Figure 7.15(a), where it can be seen that the regions of the feature space away from the normal condition (the large cluster at the origin), have a sampling probability of near unity (corresponding to a uniform labelling probability). This behaviour would often be desirable, and it highlights the importance of considering both classification and mapping uncertainty, as this increased prediction uncertainty could limit the impact of incorrect classifications on decision-making in scenarios where trust in initial transfer is low. Note that in B2→B3, the target-only RVM achieves a slightly higher initial F1 score, likely because of being initialised with three data points from the 21.6g mass state, enabling classification of this class prior to observation of these data during the active-sampling process; however, in practice, such data would be unavailable at this stage.



(a) (b)

FIGURE 7.14: The number of observations queried via active sampling using the DA-RVM (blue) and target-only RVM (red) is shown for B1→B3 in (a) and for B2→B3 in (b). The black lines indicate the range of samples, showing the 10th and 90th percentiles, while the red dashed lines above the 21.6g mass bar represent additional samples used to initialise the target-only RVM.

Similarly to the previous case study, it can be seen that observing small quantities of labelled target data allowed for significant improvements in the F1 score, as shown in

Figure 7.13. Following only a few observations of the 21.6g mass-state, the rise in mean F1 score for classification of all classes is particularly pronounced in this case study, and the target-only RVM only achieves similar F1 scores after observing data from all classes. This result provides further evidence that by using only data from a minor damage extent in one location, the mapping parameters can be updated to allow for classification of classes where labelled data are only available in the source domain. Furthermore, after only observing a few observations from the 21.6g mass-state, the 90th percentile reaches an F1 score of unity, showing that in some cases, only a few data from a minor damage

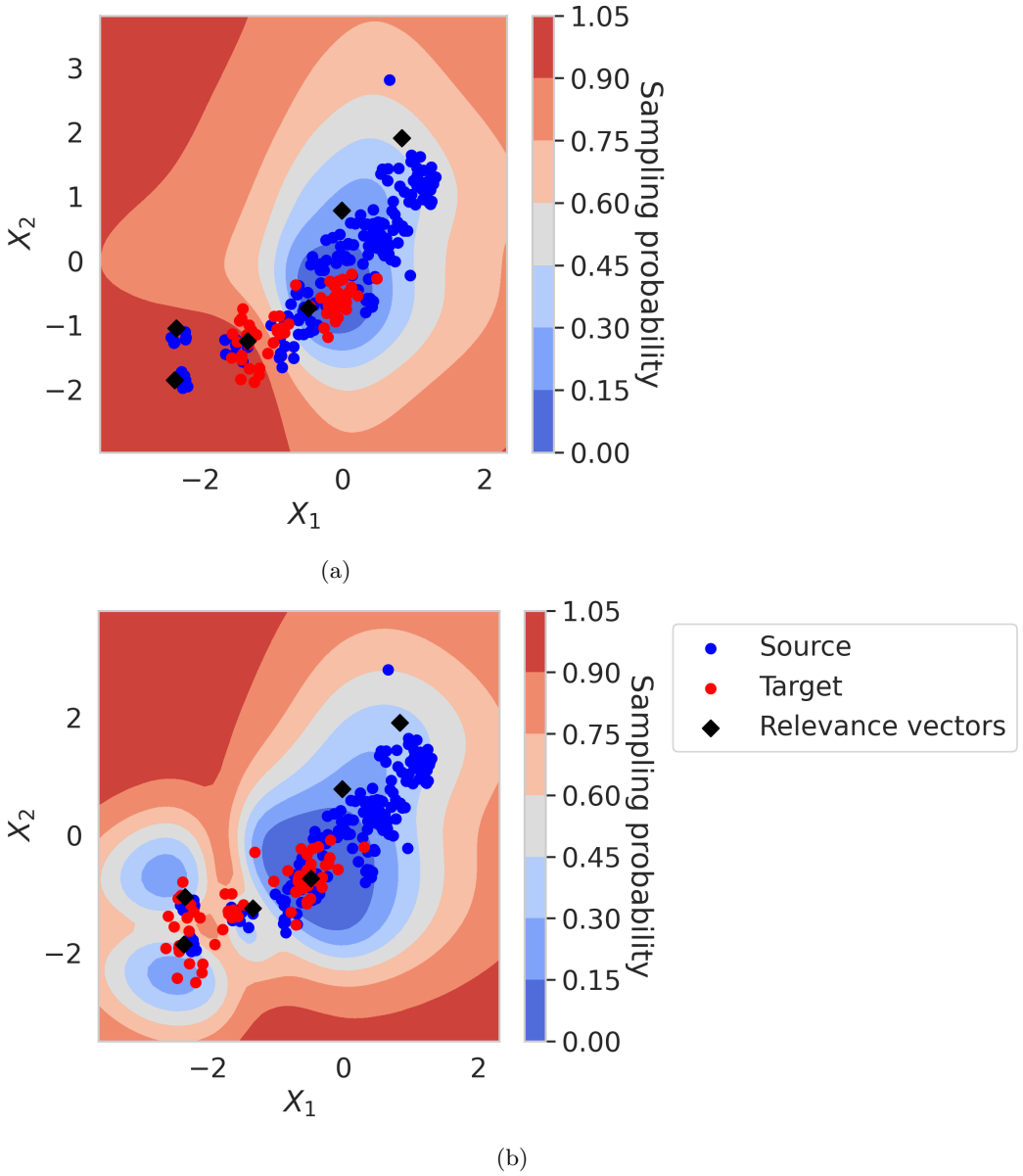


FIGURE 7.15: An example of the sampling probability for a single test repeat (training and testing data), with the target data and sampling probabilities mapped to the source domain via the expected posterior mapping after the DA-RVM was presented with all data; for B1→B3, shown in (a), and B2→B3 presented in (b).

state allow effective transfer of the labelled data in the source domain. In addition, at the end of the sampling process, both the DA-RVM and the target-only RVM achieved similar mean F1 scores, which are approaching the result of a fully-supervised RVM, as shown in Figure 7.13, while using far fewer labelled data, as shown in Figure 7.14.

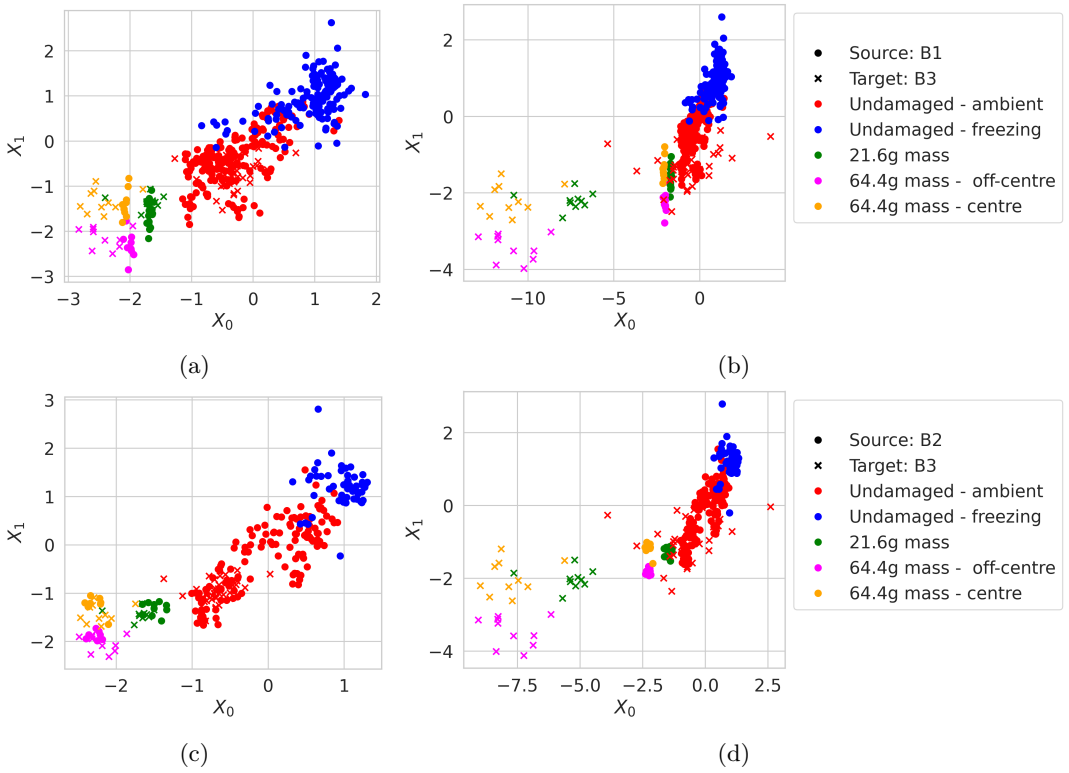


FIGURE 7.16: An example of the data (training and testing data), after the NCA mappings which resulted in the highest and lowest JMMD values. The NCA mappings for B1→B2, shown in (a) and (c), and B2→B1 is presented in (b) and (d) for the lowest and highest JMMD values respectively.

Examining the "best" (Figure 7.16(a)) and "worst" (Figure 7.16(b)) NCA mappings, selected using the JMMD as in the previous section, it can be seen that there are significant discrepancies between the initial mappings. While visually the "best" mapping (Figure 7.16(a)) appears to align the target such that classes in the source are close to the corresponding target class, the "worst" mapping seems to have a large difference in scale – visually, this discrepancy in scale appears to be larger than the "worst" example from the previous case study (shown in Figure 7.11(b)). This result is perhaps caused by the small sample of normal-condition data used to estimate the target mean and standard deviation being insufficient to produce unbiased estimates of the statistics. However, following observation of all data, the DA-RVM was able to correct the poor initial mapping, as shown by the expected DA-RVM posterior mapping found after observing all data shown in Figure 7.17.

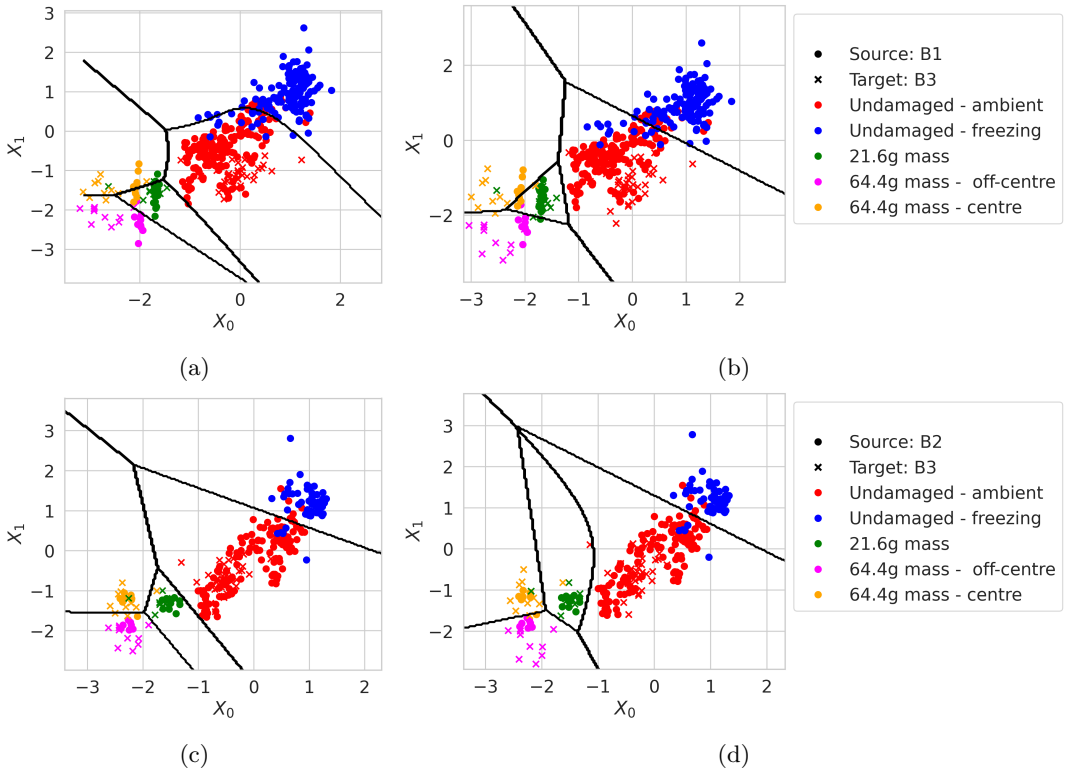


FIGURE 7.17: An example of the data (training and testing data), after the final DA-RVM mappings for the same test repeats which resulted in the highest and lowest JMMD values after the NCA mapping. The DA-RVM mappings for B1→B2, shown in (a) and (c), and B2→B1, presented in (b) and (d), for the lowest and highest JMMD values respectively.

It is worth noting here that the specification of the prior variance has an important effect on the final mapping. Given the small number of labelled data, if the mapping variance is assumed to be small and large-scale differences are present (as seen in Figure 7.17(a) and Figure 7.17(b)), the DA-RVM may struggle to learn the large-scale values needed to correct this misalignment, as such values are unlikely under the prior. Appendix B.1 presents the same examples where the variance of the mapping parameters was chosen to be $\sigma_t = \sigma_s = 0.1$, to demonstrate this issue, showing that the worst mappings are unchanged, even with labels. This highlights the important balance when defining the prior variance: it must be high enough to avoid over-constraining the mapping parameters, yet low enough to prevent the mapping from overfitting to the limited target data.

As with the previous case, the variation in results for the DA-RVM is higher than the target-only RVM. Furthermore, while the DA-RVM introduces the potential of increasing the test F1 score beyond the maximum possible F1 score for a target-only model; in contrast to all other case studies, in B1→B3 the 10^{th} percentile of the DA-RVM also drops below the target-only RVM at some stages of the active-sampling process, as

shown in Figure 7.13(a). The few test repeats producing worse F1 scores with the DA-RVM suggest that the negative effect of poor initial mappings can persist even after the inclusion of labels. This result is likely caused by poor initial NCA mappings, and shows that even though leveraging labels may reduce the likelihood of negative transfer, it may still be a critical issue, and the selection of a suitable source structure, features, and data preprocessing are crucial considerations.

7.4.5 A comparison between random and active sampling

A final consideration is that the active sampling procedure should select a more informative label set compared to a random sample; to demonstrate the effectiveness of the active sampling procedure used in this chapter a comparison with random sampling is presented in this section.

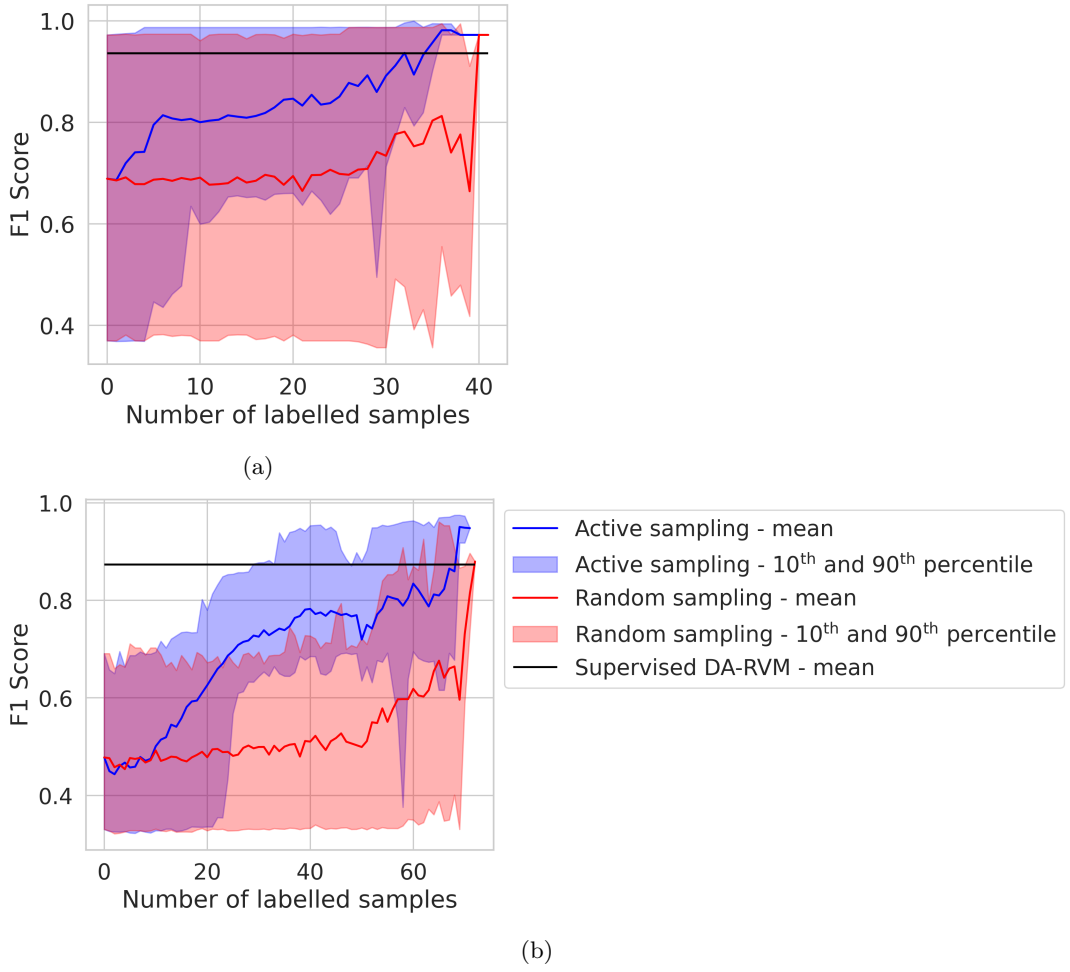


FIGURE 7.18: The test F1 scores for the DA-RVM against the number of labelled samples selected via active sampling (shown in blue) and random sample (shown in red); B1→B2 is shown in (a), and B2→B1 is given in (b).

To verify the effectiveness of the active-sampling strategy in the first case study, a comparison with random sampling is presented in Figures 7.18(a) and 7.18(b), for B1→B2 and B2→B1, respectively. Random sampling results were generated by selecting samples at random from the entire target training dataset, selecting the number chosen by the active-sampling strategy for the given test repeat. In both cases, uncertainty sampling caused a sharper rise in F1 score and higher performance in the final model. There is a significant increase in the F1 score with random sampling and a reduction in the inter-percentile range at the higher end of the labelled sample count; however, this occurs because only a few test repeats led to this many queries, and these test repeats correspond to those with high F1 scores.

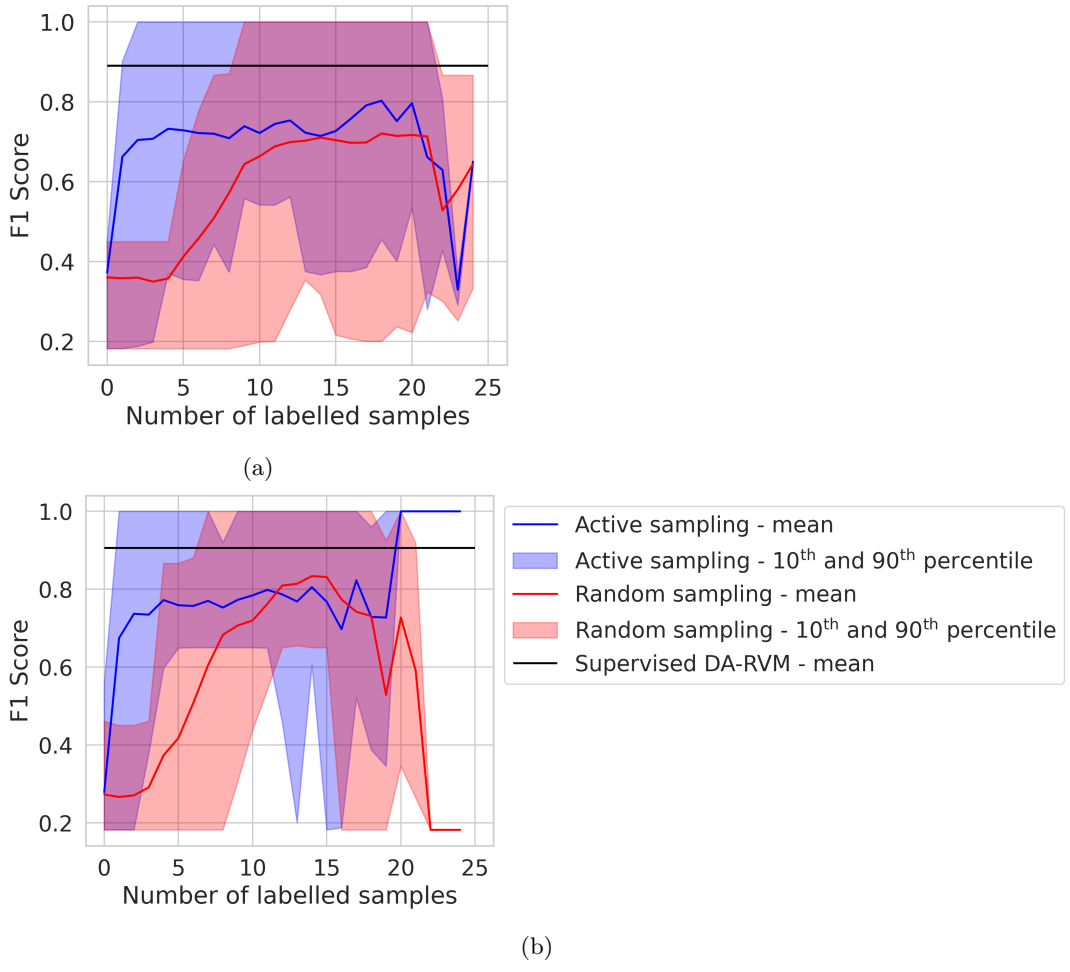


FIGURE 7.19: The test F1 scores for the DA-RVM against the number of labelled samples selected via active sampling (shown in blue) and random sample (shown in red); B1→B3 is shown in (a), and B2→B3 is given in (b).

The active-sampling strategy is also benchmarked against random sampling for the DA-RVM in the second case study, as shown in Figure 7.19. As with the previous case, active sampling results in improvements in the F1 scores with far fewer samples. In

addition, the sudden changes in mean F1 scores seen as the number of labelled samples increases were caused by the small number of repeats that queried above 20 samples.

7.5 Discussion and conclusions

A critical limitation of conventional data-driven approaches to SHM is that supervised machine-learning methods require a fully labelled dataset with examples representing each health state of interest; this is often costly and/or unfeasible. One technology for reducing data requirements in a target domain is transfer learning. However, in scenarios where data are sparse, issues such as class imbalance make many unsupervised transfer learning methods prone to negative transfer in certain scenarios.

In this thesis, methods have been proposed to transfer, where only normal-condition data are available. However, SHM data are typically acquired online, and data to learn mappings will increase incrementally. Thus, this chapter developed a method for updating a prior DA mapping, specified using unsupervised DA, and limited labelled data. It was shown that NCA mappings could be treated as a prior, allowing for a classifier to reflect the prior uncertainty in these mappings and providing a method to incrementally update these mappings. Furthermore, this model was incorporated into an active-learning framework to further reduce the label requirements of this model by guiding the labelling process.

Four transfer tasks were used to demonstrate the proposed framework by transferring a damage classifier between laboratory-scale bridge structures subject to various temperatures and pseudo-damage states. In all cases, leveraging labelled source data enabled the DA-RVM to classify health states that had not been observed in the target, even when the target dataset only contained data corresponding to a subset of the classes in the source dataset – showing robustness to class imbalance. On the other hand, conventional active-learning approaches can only classify data from previously observed health states. Specifically, an initial DA mapping, estimated using limited data from the undamaged target structure, was shown to be able to improve target classification even before any labels were obtained. By labelling a few samples, the DA mapping could be updated, improving the likelihood of correctly classifying target classes for which labelled data were only available in the source dataset. The ability to classify health states prior to their observation in the target domain has significant implications in SHM, as predictions about health states critical to decision making could be achieved before these health states are observed in the target structure and without repeating labelling efforts. Furthermore, the active transfer-learning approach resulted in fewer

overall queries compared to conventional active learning, which in practice would result in a reduction in inspections; hence, lower costs for the SHM systems.

There are several interesting potential directions for future work. One of the main limitations of the current approach is that the number of observations labelled is not directly related to a labelling budget. In practice, an operator would have a limited budget for inspections; thus, the current approach may exhaust the budget prior to observing all data. Ensuring the labelling budget is not depleted early in the sampling process is a common challenge in stream-based active learning [222]. A potential solution would be to ensure that labels correspond to health states useful for decision making, i.e. using a decision-based sampling procedure as in [191, 221]. While this approach can still request more labels than a budget allows, it could result in fewer queries overall, as typically data corresponding to minor damage states are labelled less, since they are not critical for decision making. In addition, it would provide a more interpretable output based on the expected value of information (EVOI), allowing decision makers to make informed decisions about when inspections are appropriate.

This chapter assumes that labelled data are available for all classes in the source domain. In practice, however, obtaining such comprehensive source datasets could be challenging, and a more feasible approach might be to assume that data are distributed across multiple source domains. Previous DA methods are often prone to negative transfer when aligning datasets with only a subset of shared classes, whereas the presented framework can effectively align data using a limited number of shared classes. This capability supports the extension to a multi-source scenario, which could allow for the aggregation of class information from multiple source monitoring campaigns by aligning the target domain to each source domain using a shared subset of classes. Even with data from multiple structures, novel classes may still arise in the target domain. Given that the proposed approach assigns high sampling probability to unexplored regions in the feature space, it should facilitate querying and inclusion of new target classes, though this requires validation in future studies.

In this chapter, the data corresponding to the undamaged structure were split into ambient and freezing temperature classes to constrain mappings learnt with limited target data, encouraging mappings to also ensure temperature effects are consistent across domains. Learning a mapping using temperature labels is an initial example of leveraging EoVs to increase the available contextual knowledge to learn mappings. This approach could also be extended to include more granular information about temperatures or other environmental effects; for example, the DA-RVM could be extended to a multi-task approach, which could include a shared damage classifier and regression model to predict temperature, using a single mapping for both tasks. This approach also

motivates investigation into methods for data normalisation in the context of transfer; specifically, it would be interesting to investigate whether the influence of EoVs should be removed prior to transfer, as is in common in SHM [28], or whether measurements relating to EoVs should be used to provide additional information for transfer, with perhaps methods being developed to jointly remove their influence after transfer.

In addition, this chapter presents an example of transferring information between distinct bridges with multiple spans. A related approach would be two consider individual spans as domains. An initial case study with the modular bridge dataset is presented in [198], demonstrating a single class classifier categorising added mass data being transferred between the short and long spans of the bridges presented in Table 7.1. The idea of transferring between sub-structures could be a rich area for future research, as it would be applicable to a wide range of structure with multiple components. In addition, it may also allow the availability of data to be increased without operators requiring information from multiple different structures, which could lead to challenges with data sharing when these assets are operated by different organisations.

Active learning in the context of transfer learning presents several interesting considerations. In some cases, source domains may represent structures still in operation meaning that sampling strategies could consider querying data from multiple structures. An additional extension could involve a multi-task approach with a latent DA mapping, where multiple structures stream data with the objective of enhancing performance across all structures.

Finally, this chapter applies strict assumptions about the mapping form to minimise the risk of overfitting more complex mappings when limited labelled target data are available. A drawback of this approach is if these mapping assumptions are too restrictive, it may not be possible to learn a suitable mapping to facilitate label sharing. Further work could investigate ways to extend the proposed approach to accommodate more complex mappings while minimising the risk of overfitting; for example, perhaps a number of transformation operations could be selected from a set of candidates using sparsity inducing priors.

Chapter 8

Conclusions and future work

SHM systems have the potential to enhance both the safety and efficiency of the operation of structures. However, a critical challenge to the practical application of data-driven SHM models is the availability and/or cost of data, particularly informative labelled data. The cost and sparsity of data are core motivations of PBSHM, which seeks to extend the value of data from a single structure by leveraging it across a population of structures.

Data acquired from distinct systems invalidates the assumption made in conventional machine learning – that training data are representative of testing data. Consequently, PBSHM introduces several key research challenges. These issues may be categorised into three areas. First, to ensure that transfer between a given source/target pair is viable, methods must be developed to quantify the similarity between domains ("when to transfer?"). Second, feature extraction for PBHSM must develop methods to extract relevant damage-sensitive and transferable features ("what to transfer?"). Finally, transfer learning algorithms that can be applied in sparse-data situations need to be implemented and adapted for specific PBSHM applications ("how to transfer?").

The core objective of this thesis was to develop transfer strategies for training damage classifiers with a minimal label budget for a target structure. As such, methods relating to each of these key areas of research for transfer learning have been investigated. A summary of the core contributions of this thesis is given as follows:

1. Statistic alignment (SA) methods were developed to allow for the generalisation of multi-class classifiers when only data from the undamaged structure are available in the target dataset. Thus, these methods address limitations with previous DA methods applied to SHM that require target datasets to be representative of all damage states[12, 41, 135, 140, 143, 144, 146, 147].

2. The MAC between the source and target mode shapes was formulated into a discrepancy measure for transfer learning – the MAC-discrepancy. It was shown to be strongly correlated with transfer outcomes when transferring damage localisation labels; meanwhile, unsupervised distribution-divergence measures were found to not always indicate when transfer is possible. Thus, this measure presents a potential solution for justifying transfer learning, where previous methods from the transfer learning literature cannot [59, 60, 136].

3. A transfer feature selection criterion (TFC) was developed, implementing the MAC and source loss to select vibration-based features that are more likely to satisfy the similarity conditions between source/target features assumed by unsupervised DA methods. The TFC presents the first principled approach to answer "what to transfer?" in PBSHM, whereas previous approaches either do not discuss feature selection in relation to transferable features [12, 135, 140, 144, 146, 147], or use engineering judgement [41, 143].

4. A framework for predicting classification rates resulting from transfer was proposed, using a similarity measure between data and/or structures as an indicator for the success of transfer. The MAC was demonstrated as one potential similarity measure for this application, and a beta-likelihood GP was demonstrated to constrain rate predictions to physically-meaningful values. This framework addresses limitations with previous data-based similarities, which are challenging to interpret, and presents the first approach to provide task-relevant predictions to support decisions relating to "when to transfer?" in SHM, where previous approaches have assumed source and target domains can be selected using engineering judgement [12, 41, 140, 143, 144, 146, 147] or using unsupervised measures [135].

5. SA methods presented earlier in the thesis were extended to be incorporated into an online framework, allowing for the mappings to be updated with labelled data, obtained sequentially throughout the monitoring process. In addition, an active-learning procedure was implemented. Results showed the proposed transfer learning method can result in better classification accuracy, while using fewer labels compared to conventional (target-only) active learning. The proposed approach is the first to allow for mappings to be improved online with labels, where previous approaches can only utilise unlabelled target data and estimate static mappings [12, 41, 135, 140, 144, 146, 147]. Furthermore, the proposed model (the DA-RVM), addresses limitations with previous methods proposed in the transfer learning literature [131–133], by allowing for a probabilistic mapping that can be effectively regularised to be used in conjunction with a flexible classifier trained using source data.

6. The thesis also presents two experimental datasets collected to validate the developed methods. As far as the author is aware, these case studies allowed for the first examples of successfully transferring a multi-class damage classifier between heterogeneous populations of blades, as well as bridges, using experimental data and corresponding damage states.

8.1 Summary

Statistic alignment for transfer with sparse target data

During a typical monitoring campaign, it is likely that only a few damage states of interest will be observed, and these will mostly occur towards the end of a structure's design life. Thus, target data for learning transfer-learning methods will often be restricted to data relating to the healthy target structure. To transfer multi-class source damage classifiers between domains in these scenarios, Chapter 4 investigates using statistic alignment methods. Issues relating to class imbalance were addressed by selecting data acquired directly after inspections, under the assumption that data during these periods were generated by the undamaged structures.

These SA methods were initially investigated using a population of numerical multi-storey structures and datasets collected from two real bridges – the Z24 Bridge and KW51 Bridge. Initial results found that in sparse data scenarios with a large shift in the mean of the features, the proposed methods, NCA and NCORAL, were able to facilitate generalisation of a source classifier to target data. Meanwhile, previously used DA algorithms were shown to provide limited improvements under this mean shift, particularly in the partial-DA scenario. Furthermore, when mean shift was high, previously applied DA methods required the use of NCA/NCORAL as a preprocessing step to provide consistent improvements, even under conventional DA label-space assumptions. The results presented in this chapter have since been built-upon, suggesting that NCA is capable of improving transfer in various additional applications, including transfer between FE models and real monitoring data [197], lab-scale bridge [196], full-scale bridges [198], masts [195], aircraft wings [135], and lab representations of aircraft [199].

Unsupervised DA has the potential to leverage rich source datasets to deploy SHM systems in scenarios with sparse target data, enabling the classification of common failure modes using source knowledge alone. This lack of target information motivated the estimation of linear mappings in Chapter 4. A major implication of relying on these linear mappings is that they may only be suitable for application to similar structures. Further research should investigate when these linear mappings are appropriate, and

how to apply more complex mappings when required; these issues are discussed further, using BDA in Chapter 5 and labels to update NCA parameter estimates in Chapter 7. A related challenge involves developing ways of representing structural information and quantifying similarity to guide when similarity between sets of source and target features is "sufficient"; this issue was discussed in Chapter 6.

Physics-informed transfer learning via feature selection

Considering the requirement for domains, and their corresponding features, to have related joint distributions, the preceding two chapters focused on identifying methods for determining similarities between source and target features using limited observations. One of the main objectives of Chapter 5 was to develop a data-based similarity measure to address "what to transfer?" and "when to transfer?". A potential treatment of these questions is discussed in Chapters 5 and 6, respectively.

Initially, a numerical population was investigated with the objective of transferring a damage classifier. The results of transfer between each pair of structures were used to evaluate several popular unsupervised and supervised data-based metrics – the MMD, the PAD, and the JMMD. Additionally, a MAC-based measure was proposed for similarity quantification when only limited data from the target normal condition are available. Among the investigated data-based metrics, the JMMD was found to be the only metric strongly correlated with the accuracy of classifiers transferred using NCA. This limitation presents a critical challenge for similarity quantification in PBSHM, as the JMMD cannot be estimated without labelled target data. Fortunately, for damage localisation, the MAC-based measure was also found to be a strong indicator of transfer outcomes. This finding motivates the application of the MAC to PBSHM, as it may provide insights comparable to the JMMD while requiring only measurements from the undamaged target structure.

To guide the selection of both discriminative and transferable features, the MAC-discrepancy was incorporated into a transfer feature selection criterion (TFC). Using the same numerical population, consisting of structures with varying similarity, both NCA and TFC were shown to consistently improve the generalisation of a source classifier. In contrast, finding a low-dimensional feature space via TCA or BDA without first selecting transferable features was shown to be challenging and led to a higher rate of negative transfer. Additionally, an experimental case study was presented, including data obtained via vibration testing of a composite and a metal helicopter blade. A damage classifier using FRF amplitudes as features was transferred using the TFC and NCA, facilitating improved generalisation of a source structure using only normal-condition data. Furthermore, following the application of TFC and NCA, additional transfer via BDA was

demonstrated to be beneficial. This approach showcased a transfer-learning strategy capable of extracting transferable features and performing dimensionality reduction from high-dimensional FRF data.

This chapter demonstrated that the MAC could be a useful tool for guiding the selection of transferable features in vibration-based SHM. However, significant research is still needed to improve similarity assessment between pairs of source and target features, and further work should investigate the use of different types of damage-sensitive features. Additionally, a prerequisite for selecting transferable features is that the source and target structures have similar responses to damage. Thus, the following chapter discusses the use of similarity measures to determine when structures are sufficiently related for transfer, suggesting that the MAC is a candidate for one type of similarity quantification.

Predicting the outcomes of transfer using a physics-informed metric

Fundamentally, the decision to transfer depends on whether the expected improvement from transfer learning exceeds the costs associated with implementing it. However, the absence of labelled test data makes directly evaluating the performance of transfer-learning models challenging. This challenge motivates developing a proxy for direct validation, which is the objective of the regression framework introduced in Chapter 7. Specifically, it was proposed that similarity measures could be used to predict the expected performance of a given transfer-learning strategy and source/target pair.

A training dataset was obtained by generating accuracy values for multiple examples of transfer between a population of numerical structures. To demonstrate the regression framework, the MAC-discrepancy was used as a feature that is indicative of accuracy. A beta-likelihood GP was used to ensure physically meaningful prediction and uncertainty of classification rates, allowing for a beta distribution over classification rates to be generated. In the numerical case study, it was found that the expected prediction increased monotonically, and uncertainty decreased as (MAC) similarity increased. Furthermore, it was discussed how these predictions could be used to estimate the probability of negative transfer or achieving a certain performance criterion.

The main motivation of the work in this chapter was to provide information to guide "when to transfer?". A pragmatic way to view this decision would be to attempt to quantify the reduction in costs resulting from improved maintenance decisions, compared to the cost of developing and deploying a transfer learning-based SHM system. To extend the work presented in this thesis, Hughes *et al.* quantify the consequences of true positives, false positives, and false negatives in terms of cost, allowing the expected value of transferred information to be quantified [218]. Further work should extend this

analysis to a wider range of applications and investigate the use of additional similarity measures to encode knowledge about the structural similarity. Methods should also be investigated to reduce the dependence on previous examples of transfer between real structures from a specific population. For example, the use of FE models or related populations could be investigated.

Active transfer learning for SHM

The preceding chapters mainly focused on the unsupervised transfer-learning setting, motivated by the cost of each labelled datum. Chapter 7 extends the methods for transfer developed in earlier chapters to incorporate label information, which was assumed to be obtained sequentially online. To this end, a novel Bayesian DA model was proposed – the DA-RVM – to learn with sparse target data, while considering the uncertainty resulting from both a probabilistic classifier and DA mapping. This model was incorporated into an active-learning strategy to demonstrate how transfer-learning-based models could be used to guide the labelling process and produce more informative labelled target datasets.

The DA-RVM offers two main advantages for active learning in PBSHM. First, it enables the specification of linear mappings, allowing priors to be defined using SA. Meanwhile, it still leverages a flexible nonparametric classifier, the RVM, which is well-suited for learning with limited labelled data. Second, the DA-RVM quantifies uncertainty on the mapping parameters. Consequently, classification uncertainty will be influenced not only by the source training data, as in many previous deterministic DA approaches, but also by uncertainty on the posterior mapping parameters. Modelling this mapping uncertainty has significant implications for uncertainty sampling, as it helps mitigate the effects of overconfident predictions.

To demonstrate the application of the DA-RVM, three experimental datasets were collected via vibration testing of three distinct lab-scale bridges. Using these datasets, the DA-RVM was shown to facilitate the classification of unseen damage classes in the target, initially transferring using only normal-condition data, while improving extrapolation to these unseen classes by updating the mapping and classifier as labelled target data were observed sequentially. In comparison, the same active-sampling scheme using a target-only RVM required more samples to achieve the same level of classification performance. It was also shown that by defining uncertainty on prior mapping parameters, uncertainty in prior DA mappings can be reflected in classification probabilities.

The active transfer learning scheme presented in Chapter 7 aims to provide a framework for DA in SHM, where an uncertain prior mapping is updated as additional information is acquired from a target structure online. The DA-RVM was demonstrated to have several

advantages for sparse data PBSHM scenarios; however, there are several limitations which should be a focus of future work. The mapping form in the DA-RVM presented in Chapter 7 enforces the strict assumption that a constrained linear mapping is sufficient to align the domains. In practice, it may be challenging to know the ideal mapping form; as such, different mapping forms and methods for constraining mappings should be investigated. Several potentially interesting approaches include choosing mapping forms using FE models, from data via sparsity-inducing priors or regularisation, or by encoding physics knowledge.

Another potential drawback of the DA-RVM model for use online is that it cannot be directly updated, and requires sampling procedures (MCMC) to infer a new model with new labelled data. This drawback is a consequence of using a linear mapping prior to kernelisation and the use of the softmax function to facilitate a multi-class formulation. This formulation is beneficial for the presented application, and the computational cost is not limiting in this low-dimensional problem, although methods that can be efficiently updated should be investigated for scenarios where computational budgets are limited.

There are also several research issues related to active sampling in PBSHM. For example, the approach presented in this paper is not directly related to an inspection budget, meaning it may greedily select data at the start of the monitoring campaign and exhaust the budget before observing more informative data. In practice, the active-sampling scheme should be related to a budget, to ensure that it is well allocated across the lifecycle of a structure. Furthermore, this budget may be allocated for a population of structures in many cases; for example, a windfarm operator would typically have a shared budget for maintenance of all wind turbines. A related challenge involves determining a "threshold" to decide when a label should be obtained. Chapter 7 probabilistically decides when to label data based on the information efficiency; however, a more realistic approach may modify the probability of sampling such that it relates to the label budget, and perhaps also enforces a minimum/maximum periodicity for inspections.

8.2 Limitations and future work

This thesis presents some of the first studies investigating "when to transfer?", "what (features) to transfer?" and "how to transfer?" when target data are sparse and presented sequentially online in PBSHM. Nevertheless, there exist several limitations that must be addressed before transfer learning can be applied in these sparse-data SHM scenarios. As a result, there is a large scope for future work to extend the findings in this thesis; this section aims to discuss a few key areas.

8.2.1 Investigating the effects of structural variation

In the author's opinion, one of the most pressing challenges for the practical implementation of transfer learning is the development of principled similarity measures. To this end, research is needed to identify the structural information that influences transferability and to determine how it can be represented to enable the computation of similarity measures between domains.

This issue first motivates further validation of transfer learning across large populations of structures, with the aim of investigating how specific types of structural variation – i.e. material type/properties, geometry and boundary conditions – influence damage-sensitive features in the context of transfer. Extensive studies are required to determine "how similar" structures should be to produce reliable transfer. As part of this research, data from multiple populations will need to be acquired experimentally, or via FE modelling. For example, the modular bridge kit presented in Chapter 7 could be used to investigate the effects of changing the geometry, boundary conditions, number/location of supports, and materials for multiple bridges.

Future work should also focus on incorporating important structural information into the development of novel similarity metrics to encode information related to these variations; these measures could leverage AG representations as in [172]. Similarly, other representations of physics knowledge, such as high-fidelity FE models, could be used to compare structures and inform what features are likely to result in robust generalisation with transfer learning.

A related issue involves discerning how these similarity measures could be interpreted, as it is challenging to interpret a similarity measure in terms of it's effect on a decision making process. One solution proposed in this thesis suggests that classification rates could be predicted, given that previous examples of transfer are available to train a regression model. In addition, Hughes *et al.* extended this framework to enhance interpretability by quantifying the expected value of information (EVOI) when applying transfer learning [218]. These frameworks should be further extended in large populations with additional similarity measures.

8.2.2 Identifying damage-sensitive features and equivalent labels for transfer learning

When considering transfer across structures with differences in geometry and materials, several considerations arise relating to "what to transfer?". One issue relates

to selecting corresponding labels. For example, two structures may have different geometries, meaning that direct comparison of damage location is not possible; however, there may be equivalences that can be leveraged to allow for label transfer, such as non-dimensionalised damage locations. This would be an example of using physical insight to equate heterogeneous damage labels, converting absolute damage labels to non-dimensionalised damage locations as demonstrated in Chapter 5 using helicopter blades. In practice, these label shifts may arise for several reasons; additional examples include transfer between cracks in different-width beams or different failure modes for different materials. These label shifts motivate the development of methods to identify equivalent label spaces in heterogeneous populations.

Another major consideration is the balance between transferable and discriminative features. This issue is highlighted in Chapter 5, as lower frequency modes with higher MAC values were found to correspond to more transferable features. However, it may be the case that higher frequency modes carry important information relating to damage, suggesting in some cases transfer-learning based SHM models may be less sensitive to damage than conventional data-based models. This issue motivates the investigation of a pool of SHM features that can be used with transfer learning, along with developing associated criteria to select transferable features; these features could include strain, transmissibilities, acoustic emissions, images etc. Furthermore, to further increase the sensitivity of SHM models, it may be beneficial to investigate feature fusion [1] or multi-view learning methods [52] to leverage information from several sensor types.

A related issue is that sensor networks may differ in their placement and resolution, potentially introducing additional differences in the corresponding features. This variation may cause particular issues in signals that are sensitive to sensor location, such as spectral lines, or features that reflect spatial information, such as mode shapes. In the context of this thesis, natural frequencies could be robust features to sensor location, assuming sensor placement allows the identification of suitable natural frequencies, whereas mode shape comparison would not be directly feasible as it relates to modal displacement at a given sensor location. Furthermore, in many applications, it will not be feasible to ensure sensor networks are homogeneous. Thus, methods should be developed to transfer between heterogeneous sensor networks; for example, methods could be developed to interpolate between sensors to obtain so-called "virtual sensors" [235], either using data-based or numerical models.

8.2.3 Transfer learning with sparse, incomplete datasets

In many engineering scenarios source and target data will be sparse, providing limited information to leverage with transfer learning. This limitation motivates the use of transfer learning with strong inductive biases. Two interesting directions of research would be to develop transfer learning methods that incorporate methods from multi-source transfer learning [48], semi-supervised learning [34] and/or physics-informed learning [39, 236].

One opportunity to improve learning in sparse data scenarios would be to incorporate physics to regularise or constrain transfer learning models. Chapter 7 presents an initial example of constraining mappings to maintain an assumed relationship between temperature and stiffness using a shared classification boundary. Future work could consider additional contextual information relating to EoVs, such as temperature, loading and wind speeds/directions to learn DA mappings. These data could be leveraged in a similar method to the DA-RVM, using a single mapping to jointly maximise the performance across multiple classification and regression models. Another interesting direction would be to use physics to identify appropriate functions for mappings. These mappings may derived from simplified numerical models of the source and target structures. Similarly, predictive functions could be constrained/regularised with physics, and population data could be used to more robustly learn model parameters. For example, a model with a predictive function derived from an SDoF system [237], or an extension to MDoF systems, could naturally be extended into a multi-task framework to further regularise population-level parameters, such as natural frequencies. Using physics-based models in this context may also make the task of identifying related relationships easier as parameters relate to physical phenomena; thus, presenting a potential method to reduce the likelihood of negative transfer.

This thesis has mainly assumed that the source dataset will be diverse and representative of multiple health states of interest. In practice, obtaining such diverse source datasets may be challenging, as a single structure will likely not naturally experience this wide range of damage states. For some applications, the source dataset may be obtained by choosing to directly damage one structure. While in conventional SHM, this approach may be too expensive, if the obtained source dataset could be used to improve the maintenance and operation of large populations/fleets, the value of this data would be significantly increased. However, this solution would likely be expensive and/or infeasible for large-scale infrastructure.

Another approach would be to leverage an FE model to generate a source dataset; a few studies have demonstrated the possibility of using numerical source datasets [144, 198]. However, generating realistic damage scenarios is often challenging and can demand high

computational resources to solve high-fidelity models. Furthermore, realistic variation caused by EoVs and sensor acquisition noise may be challenging to simulate, potentially causing source models to misrepresent these effects.

An alternative approach would be to extend the approaches developed in this thesis to a multi-source approach. Leveraging multiple source domains could allow for multi-class classifiers, where damage-state data are sparse in each source domain, but more complete across multiple source domains. Using multiple source datasets would facilitate the application of transfer learning in populations where individual structures are unlikely to have damage-state data, but a few structures may experience damage. This scenario presents an interesting challenge for DA, as multiple domains would need to be mapped to a shared space using datasets with limited and varied labelled damage-state data.

This thesis has focused on scenarios where no, or sparse, labelled target data are available, motivating the application of DA. These methods have been shown to facilitate transfer between heterogeneous populations. However, using such limited information, directly leveraging label information from the source via DA may require the structures to be strongly related. However, this does not necessarily mean that other methods for transfer would not be beneficial in scenarios where structures are not sufficiently related to use unsupervised DA. As such, an interesting extension to the work in this study would be to investigate transferring via various approaches to determine what features and relationships can be transferred for given structural similarities and data availability.

8.2.4 Opportunities to incorporate transfer into decision frameworks

A final consideration related to how transfer learning models can be incorporated into decision-making processes. Using model predictions to inform decisions is currently an ongoing topic in SHM [232, 238], and introducing the opportunities and risks associated with transfer learning would require several additional considerations.

One potential avenue for research would be to develop a framework that provides an expected value of information (EVOI) for PBSHM systems, as discussed in [239]. A framework for using EVOI in SHM has been developed in [232], which proposes modelling failure modes as Bayesian networks representing fault trees and assigning associated costs or utilities to these failure events. This framework requires specification of a damage classifier, a degradation model, a set of actions, and associated costs for actions/taking no action. Extending this framework into a population-based framework presents the opportunity to augment the classifier, and potentially the degradation

model using data across a population. As demonstrated in this thesis, transfer learning could facilitate classifiers that can predict a wider range of health states of interest, which presents the opportunity to address an outstanding challenge discussed in [232], as decision-frameworks often assume knowledge of all relevant health-states a-priori. However, negative transfer could have detrimental impacts on decision processes, meaning it would be important to incorporate the probability of negative transfer, as discussed in [218].

The degradation model should provide insight into the progression of damage, given a certain damage state has been identified via the classifier. These degradation models present an important challenge for these decision frameworks, as modelling damage degradation is challenging, and if these models are inaccurate, decisions will be poor. Transfer learning/multi-task learning may also increase the feasibility to learn these models with a data-based approach; for example, in [240], a meta-learning approach was use to improve the prediction of crack progression. However, even when considering population data, many scenarios obtaining data relating progressive damage will be challenging; therefore, numerical models may need to be incorporated for modelling degradation. In addition, transfer-based damage progression models would also require an associated probability of negative transfer to safely include them in a decision framework.

When considering a operation of population, a natural extension for EVOI decision frameworks would be to consider how resources can be allocated to optimise the profitability/safety of an entire population [239, 241]. Extending these frameworks to consider costs and utilities across entire populations may align more closely with common decision processes that are constrained to maintenance budgets defined for an entire population, which is often the case for wind farms or national infrastructure.

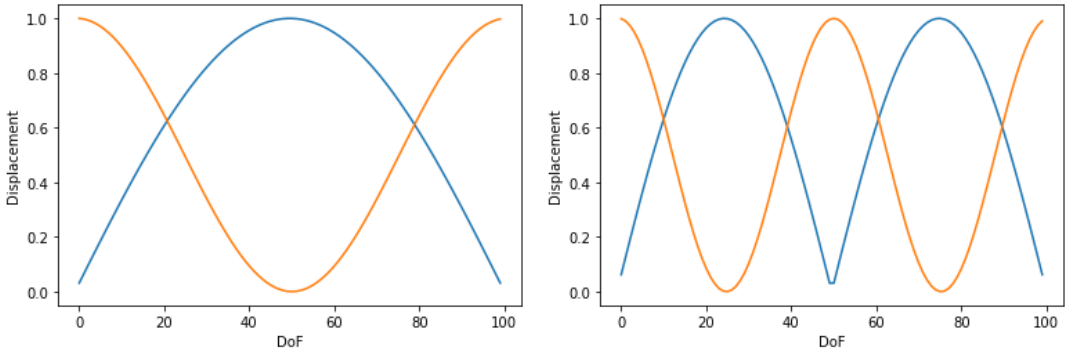
Appendix A

Chapter 5 - additional material

A.1 Motivating example: relationship between mode shapes and damage

To illustrate this relationship, a 100 degree-of-freedom (DoF) chain of masses connected with springs and dampers was considered. The mode shapes and natural frequencies were found via the eigenvalue problem for the undamaged structure, as well as the natural frequencies for a given stiffness reduction (damage) at each DoF. The absolute value of the mode shapes were compared to the difference between damaged and undamaged natural frequency ($\omega_{d,normal} - \omega_{d,damage}$), each normalised such that $x \in [0, 1]$. The comparison of the first two modes (blue) and the scaled discrepancy of the associated natural frequencies (orange) are shown in Figure A.1. It can be seen that the deviation of the natural frequencies is inversely related to the modal displacement, suggesting that they could provide an indication of which damage locations will correspond between structures[1]. This example shows the scaled discrepancy; thus, it suggests that if the modal displacement for a given location corresponds between structures, the only differences will be caused by the scale of the features and differences in damage extent; the former problem can be addressed via unsupervised TL [194].

[1] Only the natural frequencies are considered in this study, but more generally the mode shapes indicate areas of sensitivity to damage for each mode, so may also be indicative of similarity of other vibration-based features.



(a) (b)

FIGURE A.1: A comparison between the (absolute) mode shape of an undamaged 100DoF chain of masses (blue) and the relative discrepancy in the associated natural frequency caused by a stiffness reduction at each DoF (orange) for the first (a) and second mode (b); both are normalised between 0 and 1.

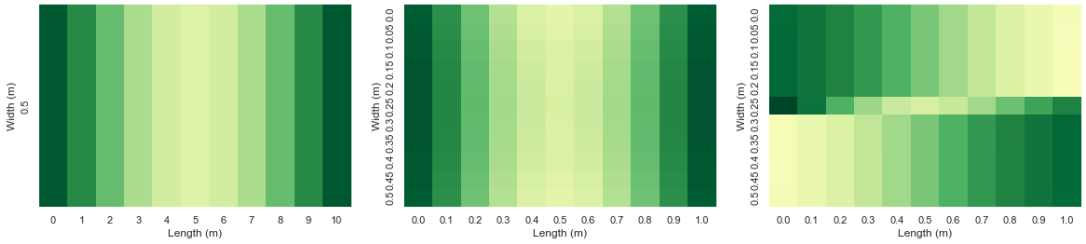
A.2 Supplementary case study: finite-element beams

To demonstrate the idea of using mode shapes to select similar damage-sensitive features, an example consisting of two beams generated via finite-element analysis (FEA) is presented. The task is to transfer between a simple one-dimensional beam, simulated using Euler-Timoshenko elements, to a beam simulated with three-dimensional solid elements with the same geometry; the node plots are shown in Figures A.3a and A.3b respectively. Both beams were simulated with a length of 1m, width of 0.5m and thickness of 0.1m.

This case presents a number of interesting considerations for transfer using the mode shapes. Firstly, variation between the two beams will be present because the three-dimensional beam can simulate varying displacement across the thickness and width of the beam, whereas the one-dimensional beam cannot; thus, only pure bending modes will be in direct correspondence. The first bending mode (in the z-direction) of the source and target is given in Figures A.2a and Figure A.2b respectively, which are in correspondence, as well as the first torsional mode in the target (in the z-direction); this mode cannot be simulated by the source. In addition, the three-dimensional beam contains a much larger number of elements (sensing locations), so direct comparison of the mode shapes is challenging; this could be a prominent issue when transferring between two real structures with different sensor arrays. Here, the line of nodes along the centre line is considered, which should be able to identify pure bending modes.

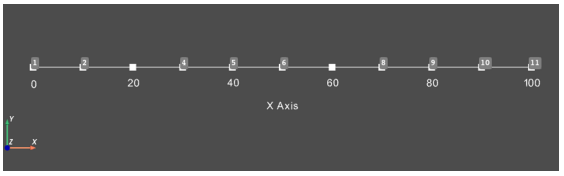
In addition, a mismatch in elements presents a problem where the label spaces are heterogeneous, as damage can not be identified across the width of the one-dimensional

beam; however, it is hypothesised that such a model could still be used to locate damage along the length of the beams by using bending modes. As such, the transfer task considered is a damage-localisation task, where damage was simulated as a 5% reduction of Young's Modulus in the elements along the width at $0.25L$, $0.5L$ and, $0.75L$, where L is the length of the beam. Fifty samples for the undamaged beams and each damage class were simulated, with variation being introduced by adding Gaussian noise to the Young's Modulus; the beams were assumed to have the material properties of steel, with a Young's modulus of 210GPa and density of 7800kg/m^3.

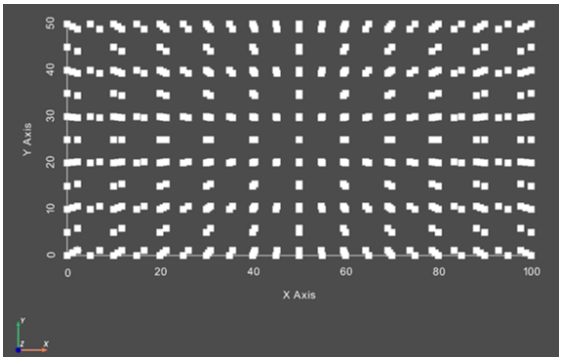


(a) (b) (c)

FIGURE A.2: The first bending mode in the 1D (a) and 3D FE beam (b), as well as a torsional mode in the 3D beam (c) in the z-axis.



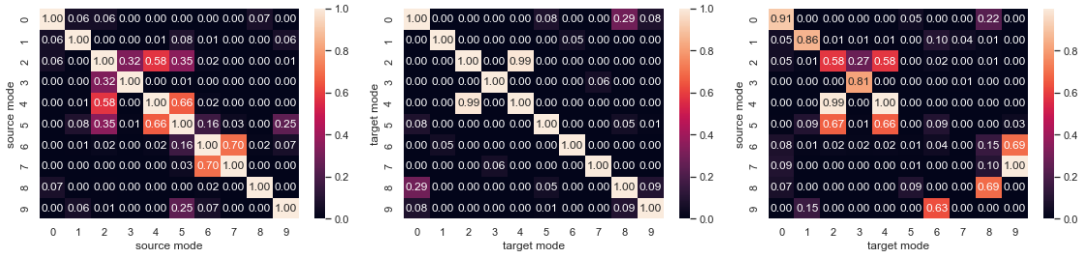
(a)



(b)

FIGURE A.3: Node plots for the 2D and 3D FE models of a beam.

The mode shapes in the x,y and z coordinates were concatenated and the MAC matrices for the source (1D beam), target (3D beam), as well as, between the source and target are presented in Figures A.4(a), A.4(b), and A.4c respectively. Initially, it can be seen that two modes in the target are identified as almost identical (Mode 2 and Mode 4), which is because target Mode 2 is predominantly a torsional Mode in z, whereas Mode



(a) (b) (c)

FIGURE A.4: The MAC matrix between the first ten modes identified from the 1D (a) and 3D beams (b), using a subset of nodes along the centre line for the 3D beam, and between the 1D and 3D beams (c).

4 is the first bending Mode in y, but as only the centre line nodes are considered, the contribution of this torsional mode is negligible in the MAC, as displacement in the centre is small. Here, potential misidentification of the mode shapes because of insufficient sensor resolution illustrates a major challenge for selecting domain-invariant features, as utilising source Mode 4 and target Mode 2 would introduce a dissimilar feature into the set, potentially causing negative transfer, as shown by the kernel density plot (KDE; the interested reader may refer to [20]), of the natural frequencies in Figure A.5.

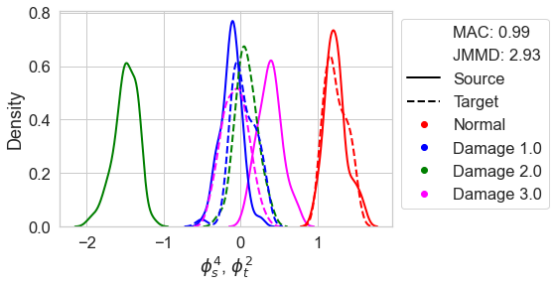


FIGURE A.5: A KDE plot comparing the natural frequencies of the second source mode with the fourth target mode, demonstrating the problem of comparing mode shapes with limited sensor networks. Density is normalised for each class independently.

The TFC was applied to extract the related features; here five features were selected, as there are five independent pairs of modes with MAC values above 0.8. In addition, NCA was applied to reduce distribution shift. The selected features are visualised using KDE plots, as shown in Figure A.6. By selecting corresponding bending modes, the associated natural frequencies follow similar joint distributions, shown by the near-zero JMMD values. The MAC discrepancy, accuracy for damage localisation and JMMD, for all the features, TFC-selected features and the features not selected by the TFC are presented in Table 2. While the original set of features could achieve perfect classification for this task, the JMMD value suggests that the data were generated from two distinct distributions, whereas the TFC features have a JMMD near zero, indicating the opposite. Furthermore, the remaining features have a significantly higher JMMD,

suggesting that the TFC was able to extract the shared information.

TABLE A.1: The MAC discrepancy (d_{MAC}), accuracy and JMMD values for the first ten natural frequencies, GA-selected frequencies and the unselected frequencies for the 1D and 3D FE beams.

	d_{MAC}	Accuracy	JMMD
All features	0.49	1.00	1.52
TFC features	0.92	1.00	0.15
Unselected features	0.07	0.63	2.94

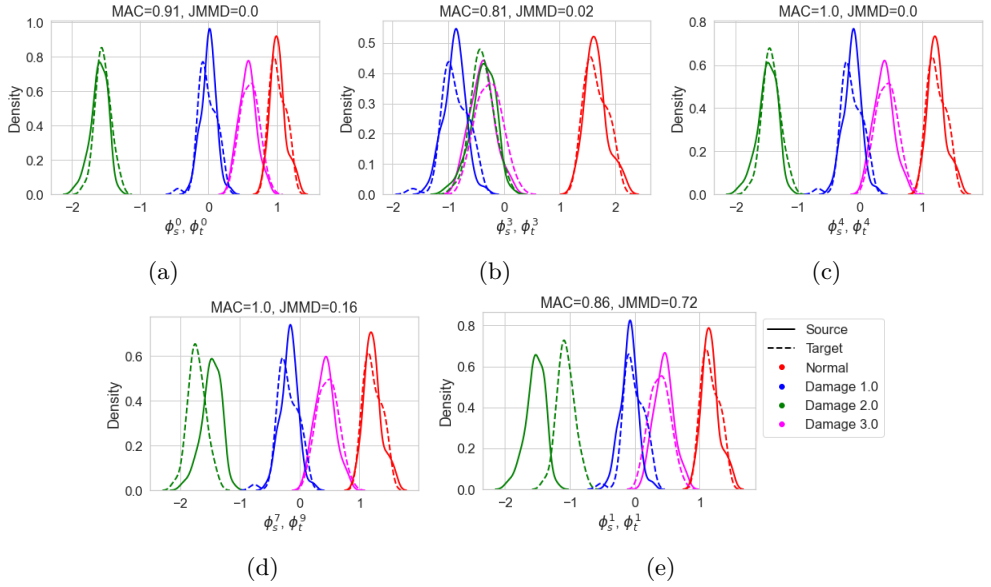


FIGURE A.6: KDE plots of the TFC-selected features for a beam generated as a 1D chain of beam elements (source) and a beam generated with a 3D geometry (target). Density is normalised for each class independently.

A.3 Experimental case study: heterogeneous blades - additional details

TABLE A.2: Table summarising sensor locations for the experiments on the blades, where L* is the non-dimensionalised length and W* is the non-dimensionalised width. The final two sensors in the composite blade change in location as the blade gets thinner close to the root.

Sensor no.	1	2	3	4	5	6	7	8	9	10	Force
L*	0.053	0.157	0.260	0.364	0.467	0.571	0.674	0.778	0.881	0.985	0.315
W* metal	0.666	0.666	0.666	0.666	0.666	0.666	0.666	0.666	0.666	0.666	0.275
W* comp	0.666	0.666	0.666	0.666	0.666	0.666	0.666	0.666	0.5	0.133	0.283

TABLE A.3: Table presenting the natural frequencies identified for the metal and composite blades, for the normal condition and four mass states.

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8
Metal Blade: Normal	5.46	16.52	32.56	55.76	83.76	114.24	154.61	197.07
Comp Blade: Normal	4.28	12.70	26.83	45.95	69.26	96.84	127.13	160.37
Ratio: Normal	1.28	1.30	1.21	1.21	1.21	1.18	1.22	1.22
Metal Blade: Damage 1	5.44	16.55	32.64	55.63	83.51	113.82	154.14	196.04
Comp Blade: Damage 1	4.28	12.87	26.79	45.86	68.87	96.4	126.64	159.89
Ratio: Damage 1	1.27	1.29	1.22	1.21	1.21	1.18	1.22	1.23
Metal Blade: Damage 2	5.44	16.74	32.65	55.47	83.68	113.77	153.90	196.01
Comp Blade: Damage 2	4.24	12.85	26.81	45.75	69.09	96.52	126.36	160.04
Ratio: Damage 2	1.28	1.30	1.22	1.21	1.21	1.18	1.22	1.22
Metal Blade: Damage 3	5.46	16.52	32.49	55.44	82.97	113.55	154.03	195.77
Comp Blade: Damage 3	4.28	12.72	26.73	45.64	68.28	98.09	126.6	160.01
Ratio: Damage 3	1.28	1.3	1.22	1.21	1.22	1.16	1.22	1.22
Metal Blade: Damage 4	5.47	16.31	32.53	54.92	83.54	112.74	152.58	196.06
Comp Blade: Damage 4	4.24	12.67	26.81	45.30	68.89	97.15	125.74	160.29
Ratio: Damage 4	1.29	1.29	1.21	1.21	1.21	1.16	1.21	1.22

TABLE A.4: Table providing the sequence of experiments on the metal and composite blades.

Test no.	Mass state	Repeats
1	Normal	15
2	Damage 3	10
3	Normal	3
4	Damage 4	10
5	Normal	3
6	Damage 1	10
7	Normal	1
8	Damage 2	10
9	Normal	3

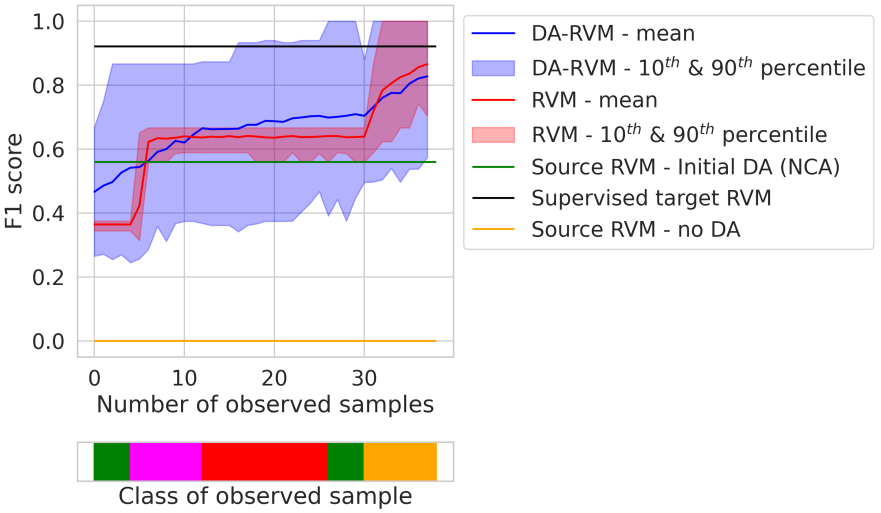
Appendix B

Chapter 7 - additional material

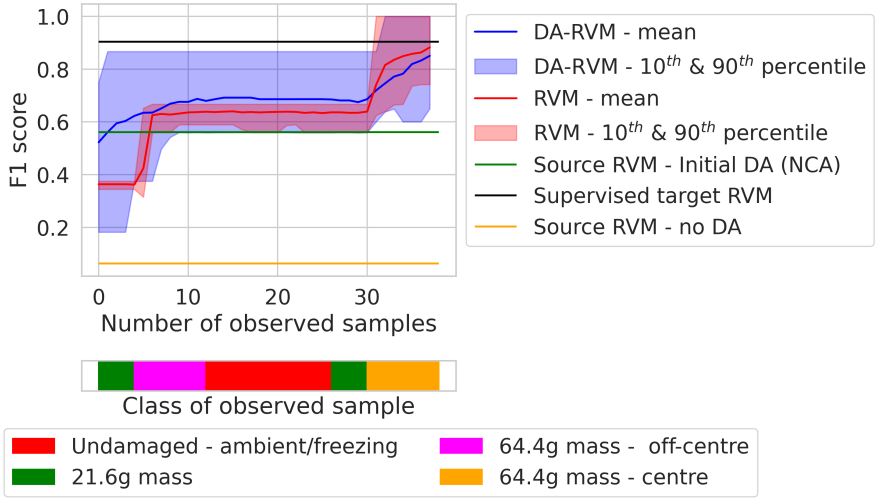
B.1 Supplementary case study: Low prior mapping variance for active transfer to a target without changing temperatures

This section provides the results for the second case study when prior variance on the scale and translation is defined to be more restrictive: $\sigma_t = \sigma_s = 0.1$. The F1 scores throughout the active-learning process are presented in Figure B.1. It can be seen in this case that, although the initial F1 scores are higher when defining lower prior variance, the mean F1 score was not improved to the same extent after observing the first damage scenario, and the final F1 score is slighly lower than the target-only RVM. This result is likely caused by the low prior variance preventing the mapping from learning large enough scale and translation values to update poor initial mappings found in some test repeats.

The features found via the expected posterior mapping after observing all data for the same "best" and "worst" test repeats as Figure 7.16, are presented in Figure B.2. It can be seen that while the "best" mappings resulted in close alignment in the final feature space, shown in Figure B.2(a) and Figure B.2(b), the "worst" test repeat for B2→B1, shown in Figure B.2(d) remains largely unchanged from the original NCA mapping (Figure 7.16(d)). This result suggests that the prior mapping was too restrictive. It can also be seen that there is only a boundary between the ambient and freezing data in Figure B.2(d); this is because high mapping variance means data in most regions of the feature space were assigned a uniform label probability.



(a)



(b)

FIGURE B.1: Mean test F1 score vs the number of samples previously presented to the active learners for B1→B2, shown in (a), and B2→B1 presented in (b).

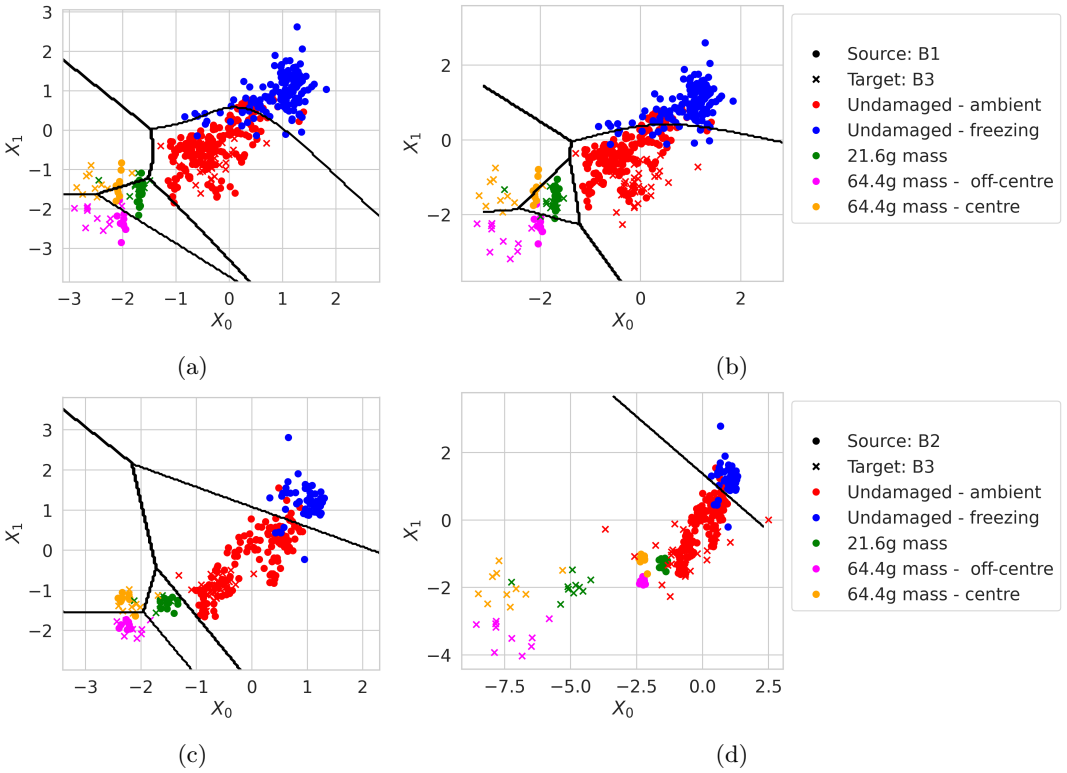


FIGURE B.2: Visualisation of the data (training and testing data), mapped via the expected posterior mapping after being presented will all data, for B1→B2, shown in (a), and B2→B1 presented in (b).

Bibliography

[1] C. Farrar and K. Worden, *Structural Health Monitoring: A Machine Learning Perspective*. Chichester: Wiley, 2013.

[2] L. Bull, K. Worden, T. Rogers, C. Wickramarachchi, E. Cross, T. McLeay, W. Leahy, and N. Dervilis, "A probabilistic framework for online structural health monitoring: active learning from machining data streams," in *Journal of Physics: Conference Series*, vol. 1264, no. 1. IOP Publishing, 2019, Art. 012028.

[3] T. Rogers, K. Worden, R. Fuentes, N. Dervilis, U. Tygesen, and E. Cross, "A bayesian non-parametric clustering approach for semi-supervised structural health monitoring," *Mechanical Systems and Signal Processing*, vol. 119, pp. 100–119, 2019.

[4] A. Malekloo, E. Ozer, M. AlHamaydeh, and M. Girolami, "Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights," *Structural Health Monitoring*, vol. 21, no. 4, pp. 1906–1955, 2022.

[5] L. Bull, P. Gardner, J. Gosliga, T. Rogers, N. Dervilis, E. Cross, E. Papatheou, A. Maguire, C. Campos, and K. Worden, "Foundations of population-based SHM, Part I: Homogeneous populations and forms," *Mechanical Systems and Signal Processing*, vol. 148, Art. 107141, 2021.

[6] J. Gosliga, P. Gardner, L. A. Bull, N. Dervilis, and K. Worden, "Foundations of Population-based SHM, Part II: Heterogeneous populations – Graphs, networks, and communities," *Mechanical Systems and Signal Processing*, vol. 148, Art. 107144, 2021.

[7] P. Gardner, L. Bull, J. Gosliga, N. Dervilis, and K. Worden, "Foundations of population-based SHM, Part III: Heterogeneous populations – Mapping and transfer," *Mechanical Systems and Signal Processing*, vol. 149, Art. 107142, feb 2021.

[8] A. Rytter, "Vibrational based inspection of civil engineering structures," *Dept. of Building Technology and Structural Engineering, Aalborg University*, 1993.

[9] C. Farrar, N. Dervilis, and K. Worden, "The past, present and future of structural health monitoring: An overview of three ages," *Strain*, vol. 61, no. 1, Art. e12495, 2025.

[10] J. E. Mottershead and M. Friswell, "Model updating in structural dynamics: a survey," *Journal of sound and vibration*, vol. 167, no. 2, pp. 347–375, 1993.

[11] E. J. Cross, S. J. Gibson, M. R. Jones, D. J. Pitchforth, S. Zhang, and T. J. Rogers, "Physics-informed machine learning for structural health monitoring," *Structural Health Monitoring Based on Data Science Techniques*, pp. 347–367, 2022.

[12] P. Gardner, X. Liu, and K. Worden, "On the application of domain adaptation in structural health monitoring," *Mechanical Systems and Signal Processing*, vol. 138, Art. 106550, 2020.

[13] A. Pagani, M. Enea, and E. Carrera, "Component-wise damage detection by neural networks and refined fes training," *Journal of Sound and Vibration*, vol. 509, Art. 116255, 2021.

[14] L. A. Bull, N. Dervilis, K. Worden, E. J. Cross, and T. J. Rogers, "A sampling-based approach for information-theoretic inspection management," *Proceedings of the Royal Society A*, vol. 478, no. 2262, Art. 20210790, 2022.

[15] P. Gardner, T. J. Rogers, C. Lord, and R. J. Barthorpe, "Learning model discrepancy: A Gaussian process and sampling-based approach," *Mechanical Systems and Signal Processing*, vol. 152, Art. 107381, 2021. [Online]. Available: https://doi.org/10.1016/j.ymssp.2020.107381

[16] L. A. Bull, P. Gardner, J. Gosliga, T. J. Rogers, N. Dervilis, E. J. Cross, E. Papatheou, A. E. Maguire, C. Campos, and K. Worden, "Foundations of population-based SHM, Part I: Homogeneous populations and forms," *Mechanical Systems and Signal Processing*, vol. 148, Art. 107141, feb 2021.

[17] D. for Business and Trade, "Offshore wind: Investing in the uk," https://www.great.gov.uk/international/investment/sectors/offshore-wind/, 2024, accessed: 2024-11-27.

[18] Highways England, "Freedom of information request on bridges," Response to FOI Request Ref: 101497, Nov. 2020, Highways England Official Correspondence Team, Bedford, UK.

[19] Railway Accident Investigation Unit (RAIU), "Malahide viaduct collapse on the dublin to belfast line, on the 21st august 2009," Railway Accident Investigation

Unit, Dublin, Ireland, Tech. Rep. 2010-R004, August 2010, accessed: 2024-11-29. [Online]. Available: https://www.raiu.ie/

[20] K. P. Murphy, *Machine learning: a Probabilistic Perspective*. MIT press, 2012.

[21] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[22] D. Barber, *Bayesian Reasoning and Machine Learning*. Cambridge, UK: Cambridge University Press, 2012.

[23] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer science & business media, 2013.

[24] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York :Springer, 2006.

[25] N. Dervilis, M. Choi, S. Taylor, R. Barthorpe, G. Park, C. Farrar, and K. Worden, "On damage diagnosis for a wind turbine blade using pattern recognition," *Journal of Sound and Vibration*, vol. 333, no. 6, pp. 1833–1850, 2014.

[26] N. Dervilis, K. Worden, and E. Cross, "On robust regression analysis as a means of exploring environmental and operational conditions for SHM data," *Journal of Sound and Vibration*, vol. 347, pp. 279–296, 2015.

[27] L. Bull, K. Worden, R. Fuentes, G. Manson, E. Cross, and N. Dervilis, "Outlier ensembles: A robust method for damage detection and unsupervised feature extraction from high-dimensional data," *Journal of Sound and Vibration*, vol. 453, pp. 126–150, 2019.

[28] E. J. Cross, K. Worden, and Q. Chen, "Cointegration: a novel approach for the removal of environmental trends in structural health monitoring data," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 467, no. 2133, pp. 2712–2732, 2011.

[29] M. R. Jones, T. J. Rogers, K. Worden, and E. J. Cross, "A bayesian methodology for localising acoustic emission sources in complex structures," *Mechanical Systems and Signal Processing*, vol. 163, Art. 108143, 2022.

[30] G. Manson, K. Worden, and D. Allman, "Experimental validation of a structural health monitoring methodology: Part iii. damage location on an aircraft wing," *Journal of Sound and Vibration*, vol. 259, no. 2, pp. 365–385, 2003.

[31] N. Mechbal, J. S. Uribe, and M. Rébillat, "A probabilistic multi-class classifier for structural health monitoring," *Mechanical Systems and Signal Processing*, vol. 60, pp. 106–123, 2015.

[32] J. Chen, S. Yuan, and H. Wang, "On-line updating gaussian process measurement model for crack prognosis using the particle filter," *Mechanical Systems and Signal Processing*, vol. 140, Art. 106646, 2020.

[33] C. Sbarufatti, M. Corbetta, A. Manes, and M. Giglio, "Sequential Monte-Carlo sampling based on a committee of artificial neural networks for posterior state estimation and residual lifetime prediction," *International Journal of Fatigue*, vol. 83, pp. 10–23, 2016.

[34] F. Schwenker and E. Trentin, "Pattern classification and clustering: A review of partially supervised learning approaches," *Pattern Recognition Letters*, vol. 37, pp. 4–14, 2014.

[35] L. Bornn, C. R. Farrar, G. Park, and K. Farinholt, "Structural health monitoring with autoregressive support vector machines," *Journal of Sound and Vibration*, 2009.

[36] D. Sen, A. Aghazadeh, A. Mousavi, S. Nagarajaiah, R. Baraniuk, and A. Dabak, "Data-driven semi-supervised and supervised learning algorithms for health monitoring of pipes," *Mechanical Systems and Signal Processing*, vol. 131, pp. 524–537, 2019.

[37] Y. Huang, L. Gong, S. Wang, and L. Li, "A fuzzy based semi-supervised method for fault diagnosis and performance evaluation," in *2014 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*. IEEE, 2014, pp. 1647–1651.

[38] M. R. Jones, T. J. Rogers, and E. J. Cross, "Constraining Gaussian processes for physics-informed acoustic emission mapping," *Mechanical Systems and Signal Processing*, vol. 188, Art. 109984, 2023.

[39] M. Haywood-Alexander, W. Liu, K. Bacsa, Z. Lai, and E. Chatzi, "Discussing the spectrum of physics-enhanced machine learning: a survey on structural mechanics applications," *Data-Centric Engineering*, vol. 5, Art. e30, 2024.

[40] G. Mariniello, T. Pastore, C. Menna, P. Festa, and D. Asprone, "Structural damage detection and localization using decision tree ensemble and vibration data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 9, pp. 1129–1149, 2021.

[41] L. Bull, P. Gardner, N. Dervilis, E. Papatheou, M. Haywood-Alexander, R. Mills, and K. Worden, "On the transfer of damage detectors between structures: An experimental case study," *Journal of Sound and Vibration*, vol. 501, Art. 116072, 2021.

[42] L. A. Bull, D. Di Francesco, M. Dhada, O. Steinert, T. Lindgren, A. K. Parlikad, A. B. Duncan, and M. Girolami, "Hierarchical Bayesian modeling for knowledge transfer across engineering fleets via multitask learning," *Computer-Aided Civil and Infrastructure Engineering*, vol. 38, no. 7, pp. 821–848, 2023.

[43] T. Dardeno, L. Bull, R. Mills, N. Dervilis, and K. Worden, "Hierarchical Bayesian modelling of a family of FRFs," in *Structural Health Monitoring 2023*. Destech Publications, Inc., 2023, pp. 2897–2905.

[44] D. Brennan, T. Rogers, E. Cross, and K. Worden, "On quantifying the similarity of structures via a graph neural network for population-based structural health monitoring," in *Conference: ISMA2022*, 2022, Art. 9.

[45] C. T. Wickramarachchi, J. Gosliga, A. Bunce, D. S. Brennan, D. Hester, E. J. Cross, and K. Worden, "Similarity assessment of structures for population-based structural health monitoring via graph kernels," *Structural Health Monitoring*, Art. 14759217241265626, 2024.

[46] D. S. Brennan, C. Kent, D. Hester, K. Worden, and C. O'Higgins, "Incorporating specialist engineering knowledge into IE models to facilitate enhanced SHM-based transfer learning within PBSHM," *e-Journal of Nondestructive Testing*, 2024.

[47] D. S. Brennan, T. J. Rogers, E. J. Cross, and K. Worden, "Calculating structure similarity via a graph neural network in population-based structural health monitoring: Part II," in *Society for Experimental Mechanics Annual Conference and Exposition*. Springer, 2023, pp. 151–158.

[48] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.

[49] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, pp. 1–40, 2016.

[50] Y. Zhang and Q. Yang, "An overview of multi-task learning," *National Science Review*, vol. 5, no. 1, pp. 30–43, 2018.

[51] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and S. Y. Philip, "Generalizing to unseen domains: A survey on domain generalization," *IEEE transactions on knowledge and data engineering*, vol. 35, no. 8, pp. 8052–8072, 2022.

[52] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.

[53] P. Gardner, L. Bull, N. Dervilis, and K. Worden, "On the application of kernelised Bayesian transfer learning to population-based structural health monitoring," *Mechanical Systems and Signal Processing*, vol. 167, Art. 108519, 2022.

[54] M. F. Silva, A. Santos, R. Santos, E. Figueiredo, and J. C. Costa, "Damage-sensitive feature extraction with stacked autoencoders for unsupervised damage detection," *Structural Control and Health Monitoring*, vol. 28, no. 5, Art. e2714, 2021.

[55] S. Bee, E. Papatheou, M. Haywood-Alexander, R. Mills, L. Bull, K. Worden, and N. Dervilis, "Better together: Using multi-task learning to improve feature selection within structural datasets," *arXiv preprint arXiv:2303.04486*, 2023.

[56] S. J. Pan, "Transfer learning," *Learning*, vol. 21, pp. 1–2, 2020.

[57] Z. Wang, Z. Dai, B. Póczos, and J. Carbonell, "Characterizing and avoiding negative transfer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 293–11 302.

[58] L. Torrey and J. Shavlik, "Transfer learning," *Data Classification: Algorithms and Applications*, pp. 537–570, 2014.

[59] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representations for domain adaptation," *Advances in Neural Information Processing Systems*, pp. 137–144, 2007.

[60] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine Learning*, vol. 79, no. 1-2, pp. 151–175, 2010.

[61] A. Gretton, B. Sriperumbudur, D. Sejdinovic, H. Strathmann, S. Balakrishnan, M. Pontil, and K. Fukumizu, "Optimal kernel choice for large-scale two-sample tests," *Advances in Neural Information Processing Systems*, vol. 2, pp. 1205–1213, 2012.

[62] M. Sugiyama, M. Krauledat, and K.-R. Müller, "Covariate shift adaptation by importance weighted cross validation." *Journal of Machine Learning Research*, vol. 8, no. 5, 2007.

[63] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two-sample-problem," *Advances in Neural Information Processing Systems*, pp. 513–520, 2007.

[64] M. Sugiyama and M. Kawanabe, *Machine learning in non-stationary environments: Introduction to covariate shift adaptation.* MIT press, 2012.

[65] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *Advances in Neural Information Processing Systems*, vol. 4, no. January, pp. 3320–3328, 2014.

[66] A. Antoniou, H. Edwards, and A. Storkey, "How to train your MAML," in *Seventh International Conference on Learning Representations*, 2019.

[67] H. Daumé, "Frustratingly easy domain adaptation," *ACL 2007 - Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pp. 256–263, 2007.

[68] A. Argyriou, T. Evgeniou, and M. Pontil, "Multi-task feature learning," *Advances in Neural Information Processing Systems*, vol. 19, 2006.

[69] A. Gelman, "Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper)," *Bayesian Anal.*, 2006.

[70] T. Tommasi, F. Orabona, and B. Caputo, "Safety in numbers: Learning categories from few examples with multi model knowledge transfer," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 3081–3088.

[71] G. C. Cawley, "Leave-one-out cross-validation based model selection criteria for weighted LS-SVMs," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*. IEEE, 2006, pp. 1661–1668.

[72] T. Tommasi and B. Caputo, "The more you know, the less you learn: from knowledge transfer to one-shot learning of object categories," in *Proceedings of the British Machine Vision Conference*, 2009, pp. 80–1.

[73] T. Evgeniou, C. A. Micchelli, M. Pontil, and J. Shawe-Taylor, "Learning multiple tasks with kernel methods." *Journal of Machine Learning Research*, vol. 6, no. 4, 2005.

[74] L. Duan, D. Xu, and I. W.-H. Tsang, "Domain adaptation from multiple sources: A domain-dependent regularization approach," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 3, pp. 504–518, 2012.

[75] L. Duan, I. W. Tsang, D. Xu, and T.-S. Chua, "Domain adaptation from multiple sources via auxiliary classifiers," in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp. 289–296.

[76] B. Chu, V. Madhavan, O. Beijbom, J. Hoffman, and T. Darrell, "Best practices for fine-tuning visual classifiers to new domains," in *Computer Vision–ECCV 2016*

Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14. Springer, 2016, pp. 435–442.

[77] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2019.

[78] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational Intelligence and Neuroscience*, vol. 2018, no. 1, Art. 7068349, 2018.

[79] Y. Guo, H. Shi, A. Kumar, K. Grauman, T. Rosing, and R. Feris, "Spottune: transfer learning through adaptive fine-tuning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4805–4814.

[80] L. Yuan, D. Chen, Y.-L. Chen, N. Codella, X. Dai, J. Gao, H. Hu, X. Huang, B. Li, C. Li *et al.*, "Florence: A new foundation model for computer vision," *arXiv preprint arXiv:2111.11432*, 2021.

[81] T. Wolf, "Transformers: State-of-the-art natural language processing," *arXiv preprint arXiv:1910.03771*, 2020.

[82] B. Min, H. Ross, E. Sulem, A. P. B. Veyseh, T. H. Nguyen, O. Sainz, E. Agirre, I. Heintz, and D. Roth, "Recent advances in natural language processing via large pre-trained language models: A survey," *ACM Computing Surveys*, vol. 56, no. 2, pp. 1–40, 2023.

[83] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong *et al.*, "A survey of large language models," *arXiv preprint arXiv:2303.18223*, vol. 1, no. 2, 2023.

[84] Y. Liang, H. Wen, Y. Nie, Y. Jiang, M. Jin, D. Song, S. Pan, and Q. Wen, "Foundation models for time series analysis: A tutorial and survey," in *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, 2024, pp. 6555–6565.

[85] M. Awais, M. Naseer, S. Khan, R. M. Anwer, H. Cholakkal, M. Shah, M.-H. Yang, and F. S. Khan, "Foundation models defining a new era in vision: a survey and outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[86] J. Gao, W. Fan, J. Jiang, and J. Han, "Knowledge transfer via multiple model local structure mapping," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008, pp. 283–291.

[87] F. Zhuang, P. Luo, S. J. Pan, H. Xiong, and Q. He, "Ensemble of anchor adapters for transfer learning," *International Conference on Information and Knowledge Management, Proceedings*, vol. 24-28-Octo, pp. 2335–2340, 2016.

[88] M. Sugiyama, T. Suzuki, S. Nakajima, H. Kashima, P. Von Bünau, and M. Kawanabe, "Direct importance estimation for covariate shift adaptation," *Annals of the Institute of Statistical Mathematics*, vol. 60, no. 4, pp. 699–746, 2008.

[89] J. Huang, A. Gretton, K. Borgwardt, B. Schölkopf, and A. Smola, "Correcting sample selection bias by unlabeled data," *Advances in Neural Information Processing Systems*, vol. 19, 2006.

[90] A. Gretton, A. Smola, J. Huang, M. Schmittfull, K. Borgwardt, and B. Schölkopf, "Covariate shift by kernel mean matching," *Dataset Shift in Machine Learning*, vol. 3, no. 4, Art. 5, 2009.

[91] Q. Sun, R. Chattopadhyay, S. Panchanathan, and J. Ye, "A two-stage weighting framework for multi-source domain adaptation," pp. 1–9.

[92] X. Liao, Y. Xue, and L. Carin, "Logistic regression with an auxiliary data source," in *Proceedings of the 22nd International Conference on Machine learning*, 2005, pp. 505–512.

[93] W. Dai, Q. Yang, G. R. Xue, and Y. Yu, "Boosting for transfer learning," *ACM International Conference Proceeding Series*, vol. 227, pp. 193–200, 2007.

[94] Y. Yao and G. Doretto, "Boosting for transfer learning with multiple sources," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1855–1862, 2010.

[95] J. Jiang and C. Zhai, "Instance weighting for domain adaptation in NLP." ACL, 2007.

[96] S. Jialin Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, 2011.

[97] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2200–2207, 2013.

[98] B. Chen, W. Lam, I. Tsang, and T. L. Wong, "Location and scatter matching for dataset shift in text mining," *Proceedings - IEEE International Conference on Data Mining, ICDM*, no. December, pp. 773–778, 2010.

[99] S. Si, D. Tao, and K. P. Chan, "Evolutionary cross-domain discriminative Hessian Eigenmaps," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 1075–1086, 2010.

[100] C. Wang and S. Mahadevan, "Heterogeneous domain adaptation using manifold alignment," *IJCAI International Joint Conference on Artificial Intelligence*, pp. 1541–1546, 2011.

[101] H. Yan, Y. Ding, P. Li, Q. Wang, Y. Xu, and W. Zuo, "Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2272–2281, 2017.

[102] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, "Balanced distribution adaptation for transfer learning," *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 1129–1134, 2017.

[103] L. Duan, D. Xu, and I. W. Tsang, "Learning with augmented features for heterogeneous domain adaptation," *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*, vol. 1, pp. 711–718, 2012.

[104] M. Long, J. Wang, G. Ding, S. J. Pan, and P. S. Yu, "Adaptation regularization: A general framework for transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 5, pp. 1076–1089, 2014.

[105] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, 2018.

[106] G. Wilson and D. J. Cook, "A survey of unsupervised deep domain adaptation," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 11, no. 5, pp. 1–46, 2020.

[107] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," *32nd International Conference on Machine Learning, ICML 2015*, vol. 1, pp. 97–105, 2015.

[108] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," *34th International Conference on Machine Learning, ICML 2017*, vol. 5, pp. 3470–3479, 2017.

[109] N. Xiao and L. Zhang, "Dynamic weighted learning for unsupervised domain adaptation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 15 242–15 251.

[110] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," *International Conference on Machine Learning*, pp. 1180–1189, 2015.

[111] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, pp. 7167–7176.

[112] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," *Advances in Neural Information Processing Systems*, vol. 2018-Decem, no. NeurIPS, pp. 1640–1650, 2018.

[113] H. Zhao, S. Zhang, G. Wu, J. P. Costeira, J. M. F. Moura, and G. J. Gordon, "Adversarial multiple source domain adaptation," *Advances in Neural Information Processing Systems*, vol. 2018-Decem, no. NeurIPS, pp. 8559–8570, 2018.

[114] M. Harel and S. Mannor, "Learning from multiple outlooks," *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, pp. 401–408, 2011.

[115] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*. Springer, 2016, pp. 443–450.

[116] H. He and D. Wu, "Transfer Learning for Brain-Computer Interfaces: A Euclidean Space Data Alignment Approach," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 399–410, 2020.

[117] Y. Li, N. Wang, J. Shi, X. Hou, and J. Liu, "Adaptive Batch Normalization for practical domain adaptation," *Pattern Recognition*, vol. 80, pp. 109–117, aug 2018.

[118] W.-G. Chang, T. You, S. Seo, S. Kwak, and B. Han, "Domain-specific batch normalization for unsupervised domain adaptation," pp. 7354–7362, 2019.

[119] S. Seo, Y. Suh, D. Kim, G. Kim, J. Han, and B. Han, "Learning to optimize domain specific normalization for domain generalization," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16*. Springer, 2020, pp. 68–83.

[120] Z. X. Conti, R. Choudhary, and L. Magri, "A physics-based domain adaptation framework for modeling and forecasting building energy systems," *Data-Centric Engineering*, vol. 4, Art. e10, 2023.

[121] N. Makondo, M. Hiratsuka, B. Rosman, and O. Hasegawa, "A non-linear manifold alignment approach to robot learning from demonstrations," *Journal of Robotics and Mechatronics*, vol. 30, no. 2, pp. 265–281, 2018.

[122] P. L. C. Rodrigues, C. Jutten, and M. Congedo, "Riemannian procrustes analysis: transfer learning for brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 8, pp. 2390–2401, 2018.

[123] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," *2011 International Conference on Computer Vision*, pp. 999–1006, 2011.

[124] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2066–2073, 2012.

[125] H. Wang, B. Li, and H. Zhao, "Understanding gradual domain adaptation: Improved analysis, optimal path and beyond," in *International Conference on Machine Learning*. PMLR, 2022, pp. 22 784–22 801.

[126] S. Uguroglu and J. Carbonell, "Feature selection for transfer learning," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2011, pp. 430–442.

[127] C. Persello and L. Bruzzone, "Kernel-based domain-invariant feature selection in hyperspectral images for transfer learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2615–2626, 2015.

[128] L.-l. Chen, A. Zhang, and X.-g. Lou, "Cross-subject driver status detection from physiological signals based on hybrid feature selection and transfer learning," *Expert Systems with Applications*, vol. 137, pp. 266–280, 2019.

[129] B. H. Nguyen, B. Xue, and P. Andreae, "A particle swarm optimization based feature selection approach to transfer learning in classification," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2018, pp. 37–44.

[130] R. K. Sanodiya, M. Tiwari, J. Mathew, S. Saha, and S. Saha, "A particle swarm optimization-based feature selection for unsupervised transfer learning," *Soft Computing*, vol. 24, pp. 18 713–18 731, 2020.

[131] J. Hoffman, E. Rodner, J. Donahue, T. Darrell, and K. Saenko, "Efficient learning of domain-invariant image representations," *arXiv preprint arXiv:1301.3224*, 2013.

[132] M. Gönen and A. Margolin, "Kernelized Bayesian transfer learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 28, no. 1, 2014.

[133] L. Duan, D. Xu, and I. Tsang, "Learning with augmented features for heterogeneous domain adaptation," *arXiv preprint arXiv:1206.4660*, 2012.

[134] W. Li, L. Duan, D. Xu, and I. W. Tsang, "Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 36, no. 6, pp. 1134–1148, 2013.

[135] P. Gardner, L. 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," *Mechanical Systems and Signal Processing*, vol. 172, Art. 108918, 2022.

[136] A. Gretton, "A Kernel Two-Sample Test," *The Journal of Machine Learning Research*, vol. 13, pp. 723–773, 2012.

[137] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 3, no. January, pp. 2672–2680, 2014.

[138] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[139] S. M. Brealy, A. J. Hughes, T. A. Dardeno, L. A. Bull, R. S. Mills, N. Dervilis, and K. Worden, "Multitask learning for improved scour detection: A dynamic wave tank study," *arXiv preprint arXiv:2408.16527*, 2024.

[140] E. Figueiredo, M. Omori Yano, S. Da Silva, I. Moldovan, and M. Adrian Bud, "Transfer learning to enhance the damage detection performance in bridges when using numerical models," *Journal of Bridge Engineering*, vol. 28, no. 1, Art. 04022134, 2023.

[141] G. Marasco, I. Moldovan, E. Figueiredo, and B. Chiaia, "Unsupervised transfer learning for structural health monitoring of urban pedestrian bridges," *Journal of Civil Structural Health Monitoring*, pp. 1–17, 2024.

[142] G. Michau and O. Fink, "Domain adaptation for one-class classification: monitoring the health of critical systems under limited information," *arXiv preprint arXiv:1907.09204*, 2019.

[143] P. Gardner, L. A. Bull, N. Dervilis, and K. Worden, "Domain-adapted Gaussian mixture models for population-based structural health monitoring," *Journal of Civil Structural Health Monitoring*, vol. 12, no. 6, pp. 1343–1353, 2022.

[144] M. Omori Yano, E. Figueiredo, S. da Silva, A. Cury, and I. Moldovan, "Transfer learning for structural health monitoring in bridges that underwent retrofitting," *Buildings*, vol. 13, no. 9, Art. 2323, 2023.

[145] P. Gardner, L. Bull, N. Dervilis, and K. Worden, "Overcoming the problem of repair in structural health monitoring: Metric-informed transfer learning," *Journal of Sound and Vibration*, Art. 116245, 2021.

[146] S. da Silva, M. Omori Yano, R. d. O. Teloli, G. Chevallier, and T. G. Ritto, "Domain adaptation of population-based of bolted joint structures for loss detection of tightening torque," *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, vol. 10, no. 1, 2024.

[147] S. Xu and H. Y. Noh, "Phymdan: Physics-informed knowledge transfer between buildings for seismic damage diagnosis through adversarial learning," *Mechanical Systems and Signal Processing*, vol. 151, Art. 107374, 2021.

[148] S. Zhang, Q. Zhang, J. Gu, L. Su, K. Li, and M. Pecht, "Visual inspection of steel surface defects based on domain adaptation and adaptive convolutional neural network," *Mechanical Systems and Signal Processing*, vol. 153, Art. 107541, 2021.

[149] Y. Narazaki, W. Pang, G. Wang, and W. Chai, "Unsupervised domain adaptation approach for vision-based semantic understanding of bridge inspection scenes without manual annotations," *Journal of Bridge Engineering*, vol. 29, no. 2, Art. 04023118, 2024.

[150] Z.-D. Li, W.-Y. He, W.-X. Ren, Y.-L. Li, Y.-F. Li, and H.-C. Cheng, "Damage detection of bridges subjected to moving load based on domain-adversarial neural network considering measurement and model error," *Engineering Structures*, vol. 293, Art. 116601, 2023.

[151] J. Liu, S. Xu, M. Bergés, and H. Y. Noh, "Hiermud: Hierarchical multi-task unsupervised domain adaptation between bridges for drive-by damage diagnosis," *Structural Health Monitoring*, vol. 22, no. 3, pp. 1941–1968, 2023.

[152] D. Peng, C. Liu, W. Desmet, and K. Gryllias, "Deep unsupervised transfer learning for health status prediction of a fleet of wind turbines with unbalanced data," in *Proceedings of the Annual Conference of the PHM Society 2021*, 2021.

[153] X. Wang and F. Liu, "Triplet loss guided adversarial domain adaptation for bearing fault diagnosis," *Sensors*, vol. 20, no. 1, Art. 320, 2020.

[154] X. Li, W. Zhang, Q. Ding, and J. Q. Sun, "Multi-Layer domain adaptation method for rolling bearing fault diagnosis," *Signal Processing*, vol. 157, pp. 180–197, 2019.

[155] Y. Li, Y. Song, L. Jia, S. Gao, Q. Li, and M. Qiu, "Intelligent fault diagnosis by fusing domain adversarial training and maximum mean discrepancy via ensemble learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2833–2841, 2020.

[156] J. Jiao, M. Zhao, J. Lin, and K. Liang, "Residual joint adaptation adversarial network for intelligent transfer fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 145, Art. 106962, 2020.

[157] Y. Ding, M. Jia, and Y. Cao, "Remaining useful life estimation under multiple operating conditions via deep subdomain adaptation," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[158] Z. Cao, M. Long, J. Wang, and M. Jordan, "Partial transfer learning with selective adversarial networks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2724–2732, 2018.

[159] J. Jiao, M. Zhao, J. Lin, and C. Ding, "Classifier inconsistency-based domain adaptation network for partial transfer intelligent diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 5965–5974, 2019.

[160] Y. Deng, D. Huang, S. Du, G. Li, C. Zhao, and J. Lv, "A double-layer attention based adversarial network for partial transfer learning in machinery fault diagnosis," *Computers in Industry*, vol. 127, Art. 103399, 2021.

[161] X. Wang and Y. Xia, "Knowledge transfer for structural damage detection through re-weighted adversarial domain adaptation," *Mechanical Systems and Signal Processing*, vol. 172, Art. 108991, 2022.

[162] M. Zhou and Z. Lai, "Structural damage classification under varying environmental conditions and unknown classes via open set domain adaptation," *Mechanical Systems and Signal Processing*, vol. 218, Art. 111561, 2024.

[163] Y. Gao and K. M. Mosalam, "Deep Transfer Learning for Image-Based Structural Damage Recognition," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 9, pp. 748–768, sep 2018.

[164] S. Dorafshan, R. J. Thomas, and M. Maguire, "Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete," *Construction and Building Materials*, vol. 186, pp. 1031–1045, 2018.

[165] N. Bao, T. Zhang, R. Huang, S. Biswal, J. Su, and Y. Wang, "A deep transfer learning network for structural condition identification with limited real-world

training data," *Structural Control and Health Monitoring*, vol. 2023, no. 1, Art. 8899806, 2023.

[166] C. Han, Z. Wang, Y. Fu, S. Dyke, and A. Shahriar, "Transfer-ae: A novel autoencoder-based impact detection model for structural digital twin," *Applied Soft Computing*, Art. 112174, 2024.

[167] P. Rai, A. Saha, H. Daumé III, and S. Venkatasubramanian, "Domain adaptation meets active learning," in *Proceedings of the NAACL HLT 2010 Workshop on Active Learning for Natural Language Processing*, 2010, pp. 27–32.

[168] A. Saha, P. Rai, H. Daumé, S. Venkatasubramanian, and S. L. DuVall, "Active supervised domain adaptation," in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2011, Athens, Greece, September 5-9, 2011, Proceedings, Part III 22*. Springer, 2011, pp. 97–112.

[169] B. Xie, L. Yuan, S. Li, C. H. Liu, X. Cheng, and G. Wang, "Active learning for domain adaptation: An energy-based approach," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 8, 2022, pp. 8708–8716.

[170] X. Ma, J. Gao, and C. Xu, "Active universal domain adaptation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 8968–8977.

[171] D. S. Brennan, C. Kent, D. Hester, K. Worden, and C. O'Higgins, "Expanding irreducible element models: beams and composites," *e-Journal of Nondestructive Testing*, 2024.

[172] 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," *Structural Health Monitoring*, pp. 1597–1606, 2023.

[173] K. Rombach, G. Michau, and O. Fink, "Controlled generation of unseen faults for partial and open-partial domain adaptation," *Reliability Engineering & System Safety*, vol. 230, Art. 108857, 2023.

[174] Z. Cao, L. Ma, M. Long, and J. Wang, "Partial adversarial domain adaptation," *In Proceedings of the European Conference on Computer Vision (ECCV)*, vol. 11212, pp. 135–150, 2018.

[175] Z. Cao, K. You, M. Long, J. Wang, and Q. Yang, "Learning to transfer examples for partial domain adaptation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2985–2994, 2019.

[176] S. Li, C. H. Liu, Q. Lin, Q. Wen, L. Su, G. Huang, and Z. Ding, "Deep residual correction network for partial domain adaptation," *arXiv*, vol. XX, no. XX, 2020.

[177] W. Zhang and D. Wu, "Manifold embedded knowledge transfer for brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 5, pp. 1117–1127, 2020.

[178] J. Maeck and G. De Roeck, "Description of Z24 benchmark," *Mechanical Systems and Signal Processing*, vol. 17, no. 1, pp. 127–131, 2003.

[179] K. Maes and G. Lombaert, "Monitoring railway bridge KW51 before, during, and after retrofitting," *Journal of Bridge Engineering*, vol. 26, no. 3, Art. 04721001, 2021.

[180] P. Spitzer, D. Martin, L. Eichberger, and N. Kühl, "Towards a problem-oriented domain adaptation framework for machine learning," *arXiv preprint arXiv:2501.04528*, 2025.

[181] N. Dervilis, I. Antoniadou, R. J. Barthorpe, E. J. Cross, and K. Worden, "Robust methods for outlier detection and regression for SHM applications," *International Journal of Sustainable Materials and Structural Systems*, vol. 2, no. 1/2, Art. 3, 2015.

[182] E. Cross, K. Koo, J. Brownjohn, and K. Worden, "Long-term monitoring and data analysis of the Tamar bridge," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 16–34, 2013.

[183] S. Christides and A. D. S. Barr, "One-dimensional theory of cracked Bernoulli-Euler beams," *International Journal of Mechanical Sciences*, vol. 26, no. 11-12, pp. 639–648, 1984.

[184] J. Poole, P. Gardner, N. Dervilis, L. Bull, and K. Worden, "On normalisation for domain adaptation in population-based structural health monitoring," *In Proceedings of the 13th International Workshop on Structural Health Monitoring, 2021.*

[185] J. Maeck, B. Peeters, and G. De Roeck, "Damage identification on the Z24 bridge using vibration monitoring," *Smart Materials and Structures*, vol. 10, no. 3, Art. 512, 2001.

[186] B. Peeters and G. De Roeck, "One-year monitoring of the Z24-bridge: environmental effects versus damage events," *Earthquake Engineering & Structural Dynamics*, vol. 30, no. 2, pp. 149–171, 2001.

[187] G. D. Roeck, "The state-of-the-art of damage detection by vibration monitoring: the SIMCES experience," *Journal of Structural Control*, vol. 10, no. 2, pp. 127–134, 2003.

[188] A. Teughels and G. De Roeck, "Structural damage identification of the highway bridge Z24 by FE model updating," *Journal of Sound and Vibration*, vol. 278, no. 3, pp. 589–610, 2004.

[189] E. Reynders and G. De Roeck, "A local flexibility method for vibration-based damage localization and quantification," *Journal of sound and vibration*, vol. 329, no. 12, pp. 2367–2383, 2010.

[190] R. Langone, E. Reynders, S. Mehrkanoon, and J. A. Suykens, "Automated structural health monitoring based on adaptive kernel spectral clustering," *Mechanical Systems and Signal Processing*, vol. 90, pp. 64–78, 2017.

[191] A. J. Hughes, L. A. Bull, P. Gardner, R. J. Barthorpe, N. Dervilis, and K. Worden, "On risk-based active learning for structural health monitoring," *Mechanical Systems and Signal Processing*, vol. 167, Art. 108569, 2022.

[192] K. Maes, L. Van Meerbeeck, E. Reynders, and G. Lombaert, "Validation of vibration-based structural health monitoring on retrofitted railway bridge KW51," *Mechanical Systems and Signal Processing*, vol. 165, Art. 108380, 2022.

[193] C. Shui, Q. Chen, J. Wen, F. Zhou, C. Gagné, and B. Wang, "Beyond $\{H\}$-divergence: Domain adaptation theory with Jensen-Shannon divergence," *arXiv preprint arXiv:2007.15567*, 2020.

[194] J. Poole, P. Gardner, N. Dervilis, L. Bull, and K. Worden, "On statistic alignment for domain adaptation in structural health monitoring," *Structural Health Monitoring*, pp. 1581–1600.

[195] C. T. Wickramarachchi, P. Gardner, J. Poole, C. Hübler, C. Jonscher, and R. Rolfes, "Damage localisation using disparate damage states via domain adaptation," *Data-Centric Engineering*, vol. 5, Art. e3, 2024.

[196] V. Giglioni, J. Poole, R. Mills, I. Venanzi, F. Ubertini, and K. Worden, "Transfer learning in bridge monitoring: Laboratory study on domain adaptation for population-based SHM of multispan continuous girder bridges," *Mechanical Systems and Signal Processing*, vol. 224, Art. 112151, 2025.

[197] V. Giglioni, J. Poole, I. Venanzi, F. Ubertini, and K. Worden, "A domain adaptation approach to damage classification with an application to bridge monitoring," *Mechanical Systems and Signal Processing*, vol. 209, Art. 111135, 2024.

[198] V. Giglioni, J. Poole, R. Mills, I. Venanzi, F. Ubertini, and K. Worden, "On the application of domain adaptation for knowledge transfer and damage detection

across bridge spans: an experimental case study," *Procedia Structural Integrity*, vol. 62, pp. 887–894, 2024.

[199] G. Delo, A. J. Hughes, C. Surace, and K. Worden, "On the influence of attributes for assessing similarity and sharing knowledge in heterogeneous populations of structures," *Available at SSRN 5006357*.

[200] C. T. Wickramarachchi, "Statistical alignment in transfer learning to address the repair problem: An experimental case study," 2023.

[201] T. A. Dardeno, L. A. Bull, N. Dervilis, and K. Worden, "When does a bridge become an aeroplane?" *arXiv preprint arXiv:2411.18406*, 2024.

[202] S. J. Pan, X. Ni, J. T. Sun, Q. Yang, and Z. Chen, "Cross-domain sentiment classification via spectral feature alignment," *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, pp. 751–760, 2010.

[203] W. Zhang, L. Deng, and D. Wu, "Overcoming negative transfer: A survey," *arXiv*, pp. 1–15, 2020.

[204] R. J. Allemang, "The modal assurance criterion–twenty years of use and abuse," *Sound and Vibration*, vol. 37, no. 8, pp. 14–23, 2003.

[205] M. Rojas-Carulla, B. Schölkopf, R. Turner, and J. Peters, "Invariant models for causal transfer learning," *Journal of Machine Learning Research*, vol. 19, no. 36, pp. 1–34, 2018.

[206] K. Worden and J. Dulieu-Barton, "An overview of intelligent fault detection in systems and structures," *Structural Health Monitoring*, vol. 3, no. 1, pp. 85–98, 2004.

[207] C. R. Farrar, P. J. Cornwell, S. W. Doebling, and M. B. Prime, "Structural health monitoring studies of the alamosa canyon and i-40 bridges," Los Alamos National Lab.(LANL), Los Alamos, NM (United States), Tech. Rep., 2000.

[208] J. Kim, H. Jeon, and C. Lee, "Applications of the modal assurance criteria for detecting and locating structural faults," in *Proceedings of the International Modal Analysis Conference*. SEM Society for Experimental Mechanics Inc, 1992, pp. 536–536.

[209] C. R. Farrar and D. Jauregui, *Damage Detection Algorithms Applied to Experimental and Numerical Modal Data from the I-40 Bridge*. Los Alamos National Laboratory, 1996.

[210] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[211] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence.* MIT press, 1992.

[212] R. K. Ando, T. Zhang, and P. Bartlett, "A framework for learning predictive structures from multiple tasks and unlabeled data." *Journal of Machine Learning Research*, vol. 6, no. 11, 2005.

[213] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning.* MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[214] A. D. Saul, J. Hensman, A. Vehtari, and N. D. Lawrence, "Chained Gaussian processes," in *Artificial intelligence and statistics.* PMLR, 2016, pp. 1431–1440.

[215] J. H. Mclean, M. R. Jones, B. J. O'Connell, E. Maguire, and T. J. Rogers, "Physically meaningful uncertainty quantification in probabilistic wind turbine power curve models as a damage-sensitive feature," *Structural Health Monitoring*, vol. 22, no. 6, pp. 3623–3636, 2023.

[216] S. Ferrari and F. Cribari-Neto, "Beta regression for modelling rates and proportions," *Journal of applied statistics*, vol. 31, no. 7, pp. 799–815, 2004.

[217] A. G. d. G. Matthews, M. van der Wilk, T. Nickson, K. Fujii, A. Boukouvalas, P. León-Villagrá, Z. Ghahramani, and J. Hensman, "GPflow: A Gaussian process library using TensorFlow," *Journal of Machine Learning Research*, vol. 18, no. 40, pp. 1–6, apr 2017. [Online]. Available: http://jmlr.org/papers/v18/16-537.html

[218] A. J. Hughes, G. Delo, J. Poole, N. Dervilis, and K. Worden, "Quantifying the value of positive transfer: An experimental case study," *arXiv preprint arXiv:2407.14342*, 2024.

[219] C. Feng, M.-Y. Liu, C.-C. Kao, and T.-Y. Lee, "Deep active learning for civil infrastructure defect detection and classification," in *Computing in Civil Engineering 2017*, 2017, pp. 298–306.

[220] D. Chakraborty, N. Kovvali, A. Papandreou-Suppappola, and A. Chattopadhyay, "An adaptive learning damage estimation method for structural health monitoring," *Journal of Intelligent Material Systems and Structures*, vol. 26, pp. 125–143, 2015.

[221] A. J. Hughes, L. A. Bull, P. Gardner, N. Dervilis, and K. Worden, "On robust risk-based active-learning algorithms for enhanced decision support," *Mechanical Systems and Signal Processing*, vol. 181, Art. 109502, 2022.

[222] B. Settles, "Active learning literature survey," 2009.

[223] D. J. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge university press, 2003.

[224] S. Dasgupta and D. Hsu, "Hierarchical sampling for active learning," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 208–215.

[225] L. Bull, K. Worden, G. Manson, and N. Dervilis, "Active learning for semi-supervised structural health monitoring," *Journal of Sound and Vibration*, vol. 437, pp. 373–388, 2018.

[226] G. Martinez Arellano and S. Ratchev, "Towards an active learning approach to tool condition monitoring with Bayesian deep learning," in *ECMS 2019: 33rd International ECMS Conference on Modelling and Simulation*, 2019.

[227] D. R. Clarkson, L. A. Bull, C. T. Wickramarachchi, E. J. Cross, T. J. Rogers, K. Worden, N. Dervilis, and A. J. Hughes, "Active learning for regression in engineering populations: A risk-informed approach," *arXiv preprint arXiv:2409.04328*, 2024.

[228] M. Tipping, "The relevance vector machine," *Advances in Neural Information Processing Dystems*, vol. 12, 1999.

[229] T. Damoulas, Y. Ying, M. A. Girolami, and C. Campbell, "Inferring sparse kernel combinations and relevance vectors: an application to subcellular localization of proteins," in *2008 Seventh International Conference on Machine Learning and Applications*. IEEE, 2008, pp. 577–582.

[230] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. A. Szerlip, P. Horsfall, and N. D. Goodman, "Pyro: deep universal probabilistic programming," *J. Mach. Learn. Res.*, vol. 20, pp. 28:1–28:6, 2019. [Online]. Available: http://jmlr.org/papers/v20/18-403.html

[231] M. D. Hoffman, A. Gelman *et al.*, "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo," *J. Mach. Learn. Res.*, vol. 15, pp. 1593–1623, 2014.

[232] A. J. Hughes, R. J. Barthorpe, N. Dervilis, C. R. Farrar, and K. Worden, "A probabilistic risk-based decision framework for structural health monitoring," *Mechanical Systems and Signal Processing*, vol. 150, Art. 107339, 2021.

[233] E. García-Macías and F. Ubertini, "MOVA/MOSS: Two integrated software solutions for comprehensive structural health monitoring of structures," *Mechanical Systems and Signal Processing*, vol. 143, Art. 106830, 2020.

[234] J. Poole, V. Giglioni, K. Worden, and R. Mills, "Transfer learning for bridge monitoring: model testing of four lab-scale multi-span continuous girder bridges under changing temperatures and damage conditions," 11 2024. [Online]. Available: https://doi.org/10.15131/shef.data.27732792.v1

[235] S. Vettori, E. Di Lorenzo, B. Peeters, and E. Chatzi, "A virtual sensing approach to operational modal analysis of wind turbine blades," in *Proceedings of ISMA2020 international conference on noise and vibration engineering, Leuven, Belgium*, 2020, pp. 3579–3590.

[236] E. J. Cross, T. J. Rogers, D. J. Pitchforth, S. J. Gibson, S. Zhang, and M. R. Jones, "A spectrum of physics-informed gaussian processes for regression in engineering," *Data-Centric Engineering*, vol. 5, Art. e8, 2024.

[237] E. J. Cross and T. J. Rogers, "Physics-derived covariance functions for machine learning in structural dynamics," *IFAC-PapersOnLine*, vol. 54, no. 7, pp. 168–173, 2021.

[238] A. Kamariotis, E. Chatzi, and D. Straub, "A framework for quantifying the value of vibration-based structural health monitoring," *Mechanical Systems and Signal Processing*, vol. 184, Art. 109708, 2023.

[239] A. J. Hughes, P. Gardner, and K. Worden, "Towards risk-informed pbshm: Populations as hierarchical systems," in *Society for Experimental Mechanics Annual Conference and Exposition*. Springer, 2023, pp. 117–127.

[240] G. Tsialiamanis, C. Sbarufatti, N. Dervilis, and K. Worden, "On a meta-learning population-based approach to damage prognosis," *Mechanical Systems and Signal Processing*, vol. 209, Art. 111119, 2024.

[241] G. Tsialiamanis, K. Worden, N. Dervilis, and A. J. Hughes, "Towards an active-learning approach to resource allocation for population-based damage prognosis," *arXiv preprint arXiv:2409.18572*, 2024.