



## Foundations of population-based SHM, Part III: Heterogeneous populations – Mapping and transfer



P. Gardner<sup>a</sup>, L.A. Bull<sup>a</sup>, J. Gosliga<sup>a</sup>, N. Dervilis<sup>a</sup>, K. Worden<sup>a</sup>

<sup>a</sup>Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK

### ARTICLE INFO

#### Article history:

Received 23 December 2019

Received in revised form 17 April 2020

Accepted 15 July 2020

Available online 14 August 2020

#### Keywords:

Population-based structural health monitoring

Transfer learning

Domain adaptation

### ABSTRACT

This is the third and final paper in a series laying foundations for a theory/methodology of *Population-Based Structural Health Monitoring* (PBSHM). PBSHM involves utilising knowledge from one set of structures in a population and applying it to a different set, such that predictions about the health states of each member in the population can be performed and improved. Central ideas behind PBSHM are those of *knowledge transfer* and *mapping*. In the context of PBSHM, knowledge transfer involves using information from a source domain structure, where labels are known for given feature sets, and mapping these onto the unlabelled feature space of a different, target domain structure. This mapping means a classifier trained on the transformed source domain data will generalise to the unlabelled target domain data; i.e. a classifier built on one structure will generalise to another, making Structural Health Monitoring (SHM) cost-effective and applicable to a wide range of challenging industrial scenarios. This process of mapping features and labels across source and target domains is defined here via *domain adaptation*, a subcategory of *transfer learning*. A mathematical underpinning for when domain adaptation is possible in a structural dynamics context is provided, with reference to topology within a graphical representation of structures. Subsequently, a novel procedure for performing domain adaptation on topologically different structures is outlined.

© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

This is the third and final paper in a series proposing foundations for a theory and methodology for *Population-Based Structural Health Monitoring* (PBSHM); it is preceded by [1,2]. PBSHM is the process of utilising information across a population of structures in order to perform and improve inferences that generalise for the complete population [3]. This approach to Structural Health Monitoring (SHM) clearly provides significant benefits; any knowledge, whether about the behaviour of features, or any damage-labelled data, obtained from any other members of the population, aids predictions across the whole population. For example, in the case of an aeroplane fleet, an assortment of damage-labelled data may be available for different members of the fleet, all under differing operational conditions. The concept of *Population-Based Structural Health Monitoring* (PBSHM) is to incorporate the feature and label data from each aeroplane to generate a machine learning-based approach that generalises across the complete fleet for all damage scenarios, especially when many members of the fleet have no labelled data associated with them.

Central concepts for performing PBSHM are those of *knowledge transfer* and *mapping*. These two processes are crucial for realising PBSHM for various reasons. Firstly, conventional approaches to data-driven SHM, whether supervised, unsuper-

E-mail address: [p.gardner@sheffield.ac.uk](mailto:p.gardner@sheffield.ac.uk) (P. Gardner)

vised or semi-supervised machine learning methods, assume that the training and test data are drawn from the same distribution. This assumption breaks down in PBSHM, as each member of the population will have differences in their data distributions, because of environmental variations, manufacturing and assembly differences, operational conditions etc. As a result of the differences, conventional machine learning approaches fail to generalise; for example, a classifier trained on data from one aeroplane in a fleet will generally fail to classify for a different aeroplane from the same population. Mapping data from one member of the population to another is therefore required, such that a general classifier can be generated. Another issue is that damage-labelled data, for all possible damage states of interest, are often not obtainable for each individual in the population. In fact, it is generally infeasible to obtain a complete label set for all possible damage states, even for one structure. However, it may be possible to obtain a relatively complete damage-label set by combining labels from all structures in a population, or from a historic database from other SHM campaigns. Leveraging this label knowledge from across a population and mapping it onto a consistent space, means that knowledge transfer is possible, aiding the generation of a general machine learning method for the complete population. By utilising these two processes, data-based SHM methods can be generated that are cost-effective and applicable across a wider variety of challenging industrial applications.

*Transfer learning* is one approach that aims to improve the performance of a learner by transferring knowledge between different domains. Within this branch of machine learning, various methods exist with differing assumptions about consistency between domains and what knowledge is being transferred. This paper outlines the mapping and knowledge transfer problems within PBSHM, with a particular focus on modal-based features. Specifically, the work presented here focuses on recent successes in the application of *domain adaptation* [4] – a sub-category of transfer learning – in PBSHM. Domain adaptation assumes labelled data are available in a *source* domain and that this can be used to aid classification in an unlabelled (or partially-labelled) *target* domain, by mapping the two domains onto a common *latent* space on which the data distributions coincide. However, conventional approaches to domain adaptation assume consistency between the feature and label space for each domain. This means that the features from one structure must be the same dimension as the other structure (i.e. the same number of natural frequencies), and that any damage labels for detection, localisation, classification or assessment can exist for both structures. These constraints can be limiting in the context of PBSHM. With reference to the topology of a graphical representation of a structure (discussed in Part II of this series [2]), this paper presents a mathematical underpinning for when domain adaptation is possible within the context of PBSHM.

The outline of this paper is as follows. Section 2 outlines the PBSHM contexts where transfer learning is applicable, providing key definitions and an overview of the mapping and knowledge transfer problems. Following the mapping problem descriptions, Section 3 demonstrates the contexts in which domain adaptation is applicable, referencing graphical representations of structures, through several applications on *n*-storey building structures. By exploring these applications, a novel process is developed that overcomes some of the restrictions of domain adaptation, namely in the context of label inconsistency. A further example is also provided in Section 3 demonstrating the applicability of utilising physics-based models in labelling real world data through transfer learning. Finally, conclusions are presented highlighting the effectiveness of a transfer learning approach to PBSHM.

## 2. Population-based structural health monitoring and transfer learning

PBSHM involves mapping data and labels from different structures within a population, such that a general learner can be inferred across the complete population. As a consequence, health monitoring can be performed, potentially online, for any member of the population. This section seeks to define the mapping and knowledge transfer scenarios within PBSHM and demonstrate the applicable forms of learning for each scenario, with an emphasis on transfer learning.

### 2.1. Population types

Key definitions are required in order to outline the mapping problems within PBSHM. A *population*, in the context of PBSHM, is a group of structures (the smallest being a group of two structures) that provides information required for performing health monitoring. This general definition of a population can be further divided into two categories: *homogenous* and *heterogeneous* populations [2]; these groupings relate to the level of dissimilarity within a population, where both population types benefit from knowledge transfer. Colloquially, a homogenous population is one in which any pair of structures in the population can be deemed nominally identical for a given context [1,2]. A heterogeneous population is thus a group of non-identical, and therefore different structures. Conceptually, differences will occur for a multitude of reasons, and structures will be deemed different due to various properties, based on specific contexts. This leads to a spectrum of different types of heterogeneous populations, which as similarities increase, will approach a homogenous population. One method of quantifying these differences, utilising Irreducible Element (IE) models and Attributed Graphs (AGs), has been discussed in Part II of this series of papers [2]. This approach highlights three main sources of difference within a structure, relating to the topology of the graph and types of attribute stored within the nodes and edges:

- *Geometry*: relates to the shape, size and scaling of a structure in a population, e.g. to a group of aluminium rectangular plates where each plate has a different length, width and thickness.

- *Material*: relates to different material classes, specific materials, and their properties for structures in a population, e.g. to a pair of the same size rectangular plates where one is made from aluminium and another from steel.
- *Topology*: relates to different topologies for graphical representations of structures in a populations, e.g. a pair of aluminium beams where one is a cantilever and the other is éncastre, where there is a difference between the two lumped-mass representations due to the difference in boundary conditions.

These classes of differences are summarised in Fig. 1, where each of these categories may overlap forming a different category of heterogeneous population. For example, consider a heterogeneous population of two beam-like structures, these may have geometric differences, e.g. Structure One is a beam with a tapered rectangular geometry and Structure Two is a beam of the same length with a uniform I-beam cross-section. In addition to these geometric differences, aspects of the materials may be different, e.g. Structure One is constructed from a unidirectional carbon-fibre layup and Structure Two is formed from the same carbon fibre but in a plain-weave layup. This population is therefore a heterogeneous population due to geometric and material differences, which are properties stored in the attributed graph, as defined in Part II [2], meaning the population would be located in the Venn diagram overlap between material and geometry.

More formal definitions may also be attached to these types of population, as discussed in paper I and II [1,2]. A homogeneous population can be defined as a group of structures that are *topologically homogeneous* where the geometric  $\Theta_g$  and material  $\Theta_m$  properties (collectively  $\Theta = \{\Theta_g, \Theta_m\}$ ) in the nodes and edges of the attributed graph are defined as being random draws from a concentrated base distribution i.e.  $p(\Theta)$ . It is noted that topologically homogeneous, as defined in paper II [2], relates to a group of structures that can be considered pairwise topologically equivalent, with respect their graphical representation, e.g. two structures that could be modelled as five degree-of-freedom lumped mass models. The probability mass in the base distribution  $p(\Theta)$  therefore defines the small differences between members in the population. A *strongly-homogeneous population* would have a unimodal base distribution with low dispersion for the geometric and material properties [1], where the strictest, perfect, form of homogeneous population is one in which the base distribution is a Dirac delta function, i.e. each member of the population is exactly the same. This latter scenario never occurs in reality, but is the assumption in applying conventional machine learning methods trained on one structure, to another. Using these population definitions and categories of difference helps determine the difficulty of the mapping problem for PBSHM.

It is also noted that differences observed in data may also occur outside the structural properties of individuals in a population, not captured in the attributes and topology<sup>1</sup> discussed above. These differences will relate to how data acquisition and any further processing to obtain features were performed. For example, differences in sensor placement will lead to differences in data distributions, even if those sensors are placed in the 'same' irreducible element of the structure. Properties of the data acquisition may also affect the obtained feature for each member of the population, e.g. the transmissibilities from two members of a homogeneous population may be different due to differences in the raw data sample rate. These sources of difference related to data acquisition and processing will need to be considered in defining the mapping problems in PBSHM; however, this paper simplifies the scenarios such that only structural differences in a population are considered.

## 2.2. Transfer learning

Transfer learning offers several techniques for dealing with scenarios where domains, tasks and distributions used within the training and testing of a learner are different [5]. Distinct from multi-task learning, where the objective is to learn multiple tasks across different domains [6], transfer learning leverages knowledge from *source* tasks in improving specific *target* tasks, i.e. the focus is on *target* tasks rather than all tasks (both source and target) equally [5]. This type of learning is applicable to PBSHM, even in the homogeneous population scenario, as variations in each structure within the population will lead to differences in the data distributions, meaning that a learner trained on one structure will not apply to another structure in the population. Before formally defining transfer learning, several other objects are required [5].

**Definition 1.** A domain  $\mathcal{D} = \{\mathcal{X}, p(X)\}$  is an object that consists of a feature space  $\mathcal{X}$  and a marginal probability distribution  $p(X)$  over the feature data  $X = \{\mathbf{x}_i\}_{i=1}^N \in \mathcal{X}$ , a finite sample from  $\mathcal{X}$ .

**Definition 2.** A task  $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$  is an object that consists of a label space  $\mathcal{Y}$  and a predictive function  $f(\cdot)$  ( $p(Y|X)$  in a probabilistic setting) which can be inferred from training data  $\{\mathbf{x}_i, y_i\}_{i=1}^N$  where  $\mathbf{x}_i \in \mathcal{X}$  and  $y_i \in \mathcal{Y}$ .

It is noted for clarity that both  $\mathcal{X}$  and  $\mathcal{Y}$  are *spaces* and not actual observations, which instead are denoted by the finite samples sets  $X$  and  $Y$  (for the feature and label respectively). In the case of a single source and target domain, there will be a source domain dataset  $D_s = \{\mathbf{x}_{s,i}, y_{s,i}\}_{i=1}^{N_s}$  where  $\mathbf{x}_{s,i} \in \mathcal{X}_s$  and  $y_{s,i} \in \mathcal{Y}_s$ , and a target domain dataset, formed generally as  $D_t = \{\mathbf{x}_{t,i}, y_{t,i}\}_{i=1}^{N_t}$  where  $\mathbf{x}_{t,i} \in \mathcal{X}_t$  and  $y_{t,i} \in \mathcal{Y}_t$  [5]. In practice the target domain dataset will often have no label information attached to *any* feature observation (hence the need for transfer) i.e.  $D_t = \{\mathbf{x}_{t,i}\}_i$ ; although it can still be said that each feature

<sup>1</sup> It is noted that topology includes boundary condition and loading definitions.

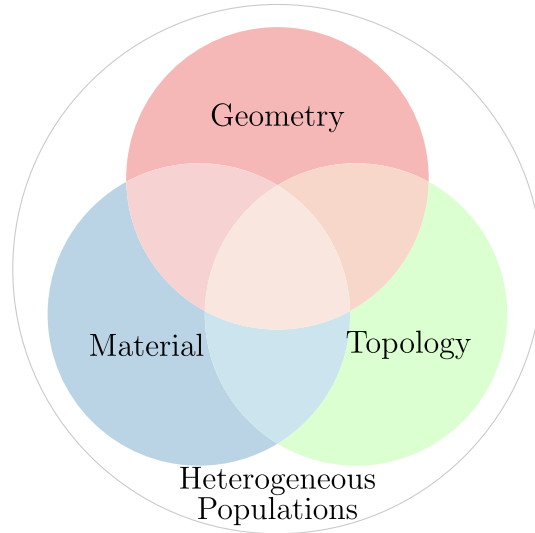


Fig. 1. Categories of heterogeneous population within population-based SHM.

observation has a 'true', but unknown, label associated with it in the label space  $\mathcal{Y}_t$  (where the goal of transfer learning is to identify the 'true' label). Lastly, it is generally assumed that there will be more observations in the source domain than target domain, i.e.  $0 \leq N_t \ll N_s$ . Given these objects, one can formally introduce transfer learning.

**Definition 3.** *Transfer learning* states that given a source domain  $\mathcal{D}_s$  and task  $\mathcal{T}_s$ , and a target domain  $\mathcal{D}_t$  and task  $\mathcal{T}_t$ , it is the process of improving the target predictive function  $f_t(\cdot)$  in  $\mathcal{T}_t$  by using knowledge from  $\mathcal{D}_s$  and  $\mathcal{T}_s$ , whilst assuming  $\mathcal{D}_s \neq \mathcal{D}_t$  and/or  $\mathcal{T}_s \neq \mathcal{T}_t$ .

Transfer learning methods then differ in their assumptions about whether  $\mathcal{X}$ ,  $p(X)$ ,  $\mathcal{Y}$  or  $p(y|X)$  are consistent across the source and target. Another distinction is also made within the transfer learning literature about the dimensions  $d$  of the feature spaces [7,8].

**Definition 4.** *Homogeneous transfer learning* assumes that the feature and label spaces represent the same attributes,  $\mathcal{X}_s = \mathcal{X}_t$  and  $\mathcal{Y}_s = \mathcal{Y}_t$ , and therefore that the dimensions of the feature space are equal,  $d_s = d_t$ .

**Definition 5.** *Heterogeneous transfer learning* assumes that the feature spaces are non-equivalent,  $\mathcal{X}_s \neq \mathcal{X}_t$ , and often that the source and target domains share no common features, meaning  $d_s \neq d_t$ . In addition, heterogeneous transfer learning can also assume that  $\mathcal{Y}_s \neq \mathcal{Y}_t$ .

To contextualise these categories, PBSHM-based examples are provided. Homogenous transfer learning is the case where both feature and label spaces are equal. An example where these techniques are applicable is a localisation problem between two aircraft. In this scenario the labels for both aircraft are undamaged ('0'), or damage in the left ('1') or right wing ('2') i.e.  $\mathcal{Y}_s = \mathcal{Y}_t \in \{0, 1, 2\}$ . The feature spaces for both aircraft are transmissibilities of 1024 spectral lines across the same frequency range, i.e.  $\mathcal{X}_s = \mathcal{X}_t$  as  $d_s = d_t = 1024$ . Heterogeneous transfer learning is applicable when at least the features are not the same. An example where both the features and labels are not equal is a localisation problem between a two-span cable stay bridge (source) and a five-span truss bridge (target), where the goal is to detect which span is damaged. Here there would be three and six possible labels for the source and target structures respectively, i.e.  $\mathcal{Y}_s \neq \mathcal{Y}_t$ . In addition, the feature spaces for these bridges might be different, e.g. the source features are the first ten deck natural frequencies i.e.  $d_s = 10$  and the target feature data may be 512 point acceleration time series i.e.  $d_t = 512$ . In this context it can still be expected that there is information that can be transfer between these structure (potentially where the mapping passes through some latent space).

Clearly heterogeneous transfer learning is the more challenging, but more general category of approach. This is because the techniques require a mapping to account for dimensionality transformations, a more complex mapping of the joint distributions, as well as the potential to handle inconsistent label spaces.

In order to clarify terminology, it is noted that the categories of homogeneous and heterogeneous *transfer learning* do not directly relate to homogeneous and heterogeneous *populations*, i.e. *homogeneous transfer learning* is not the only way of performing knowledge transfer for a *homogeneous population* and *heterogeneous transfer learning* is not the only way of performing knowledge transfer for a *heterogeneous population*. This is clarified further in Section 2.3, in which different transfer

learning approaches are linked to a variety of population types, and in Section 3 with the application of homogeneous transfer learning to both homogeneous and heterogeneous populations.

### 2.2.1. Transfer learning technologies

There are a wide variety of transfer learning technologies within the literature. This section briefly describes a variety of these approaches (although the aim here is not to be exhaustive) such that the scope of approaches to knowledge transfer in PBSHM is made clear; the interested reader is referred to the following review papers for more details [5,7,8]. In particular, this section will discuss *parameter*, *instance* and *feature-based* approaches to transfer learning.

Typically, within the realm of deep learning and convolutional neural networks, transfer learning techniques have been developed that are parameter-based or fine-tuning approaches. These approaches seek to learn some parameter set, typically a set of weights at a particular set of layers in a neural network, for some source domain, believing that these can be fixed when learning a target task. When considering a convolutional neural network the idea is to use some pre-trained convolution neural network, where the belief is that some of the convolutional layers (and their weights) have captured some feature extraction process that will generalise to the target domain. Examples of this approach can be found within the SHM literature, all focussing on image classification [9–11].

Another approach to transfer learning has been to form the problem from an instance-based view. Here the aim is to determine the degree of similarity between source domains and the target domain, where the measure of similarity helps adjust weights on the source domains (even on a point or cluster basis) and hence on the level of information transferred [12].

The last form of transfer learning, and the one utilised in this paper, is a feature-based approach. These methods seek to find some feature representation (typically through some mapping) where the differences between the source and target domains are reduced. Several of these feature-based approaches have been utilised within an SHM context [4]. A large number of approaches exist within the literature that approach transfer learning in this manner [13–18], all with varying assumptions about the amount of information available in learning the mapping to the feature representation, and what the differences are between the source and target.

### 2.2.2. Other approaches for knowledge transfer

Outside of transfer learning, there exists alternative methods that also seek to transfer knowledge between different systems and datasets. Specially this section will consider the concepts of *ontologies* and *knowledge graphs*.

*Ontologies* aim to be representations of entities (e.g. information and knowledge) describing all their interdependences and relationships. As such, ontologies are useful in outlining all knowledge about a specific domain [19], e.g. what a company knows about SHM and its application to their structures. The benefits of creating an ontology are not only that they are helpful in sharing or explaining concepts, but that if a new project or domain is envisaged, previous ontologies can be reused or transferred to these purposes, helping identify better or more effective processes. One can envisage how an ontology might also form part of a theory on PBSHM, providing several benefits such as helping to transfer what techniques and methods may be appropriate from one system to another etc. Ontologies have been explored in several disciplines such as computer science, semantics and natural language processing, and have also been considered in an SHM context [20–22]. The reason ontologies are not explored in detail in this series (although this is an area to be explored in the future), is that ontologies do not form a method themselves (although transfer learning could be part of an ontology) for transferring label information directly to feature data, which is the key aim of mapping and transfer in this current paper.

*Knowledge graphs*, in a similar manner, also aim to be objects that define entities and their interrelationships. These have primarily been developed for aiding search engines, by allowing semantic searches. However, in recent years, knowledge graphs have been integrated as training data for machine learners [23,24]. Obviously, parallels to Part II [2] can be drawn, where the structures being represented as graphs can more formally be used in training machine learners, rather than using them to isolate the SHM problem, and indicate where transfer learning is applicable. However, even if graphs were formally used in training a machine learner for PBSHM, some form of transfer learning would still be required, as the problem remains of moving information from a source graph to a target graph. For this reason, this paper focuses mainly on transfer learning as the key tool in performing knowledge transfer.

## 2.3. Mapping scenarios

Before defining mapping scenarios within a *population-based* view of SHM, it is first important to define the categories of problem established in conventional SHM (where an SHM problem is denoted  $\mathcal{SP}$ ) as this will help in defining both *what* knowledge is being transferred and *when* transfer is possible. There are several types of challenge within SHM, categorised by Rytter into a hierarchy [25], where generally difficulty increases with each level in the hierarchy.

- *Detection*: identify the presence of damage.
- *Localisation*: identify the position of damage.
- *Classification*: identify the type of damage.
- *Assessment*: identify the extent of damage.
- *Prognosis*: identify the safety and residual life of the structure.

This hierarchy provides a framework for defining the mapping problems within a population-based SHM context, as each level describes the required label agreement. By grounding the mapping scenarios within PBSHM to the hierarchy of SHM problems, one shows that the degree of similarity required is problem dependent. This approach will be useful in leveraging as much prior engineering knowledge as possible in performing transfer learning for PBSHM (although it is noted that other divisions of SHM problems may also provide different insight into transfer learning in PBSHM). Table 1 states the level of feature and label<sup>2</sup> consistency for the categories of population outlined in Section 2.1 from a physical viewpoint. The table considers modal features and that the localisation label resolution is at the edge/node level of the graphical representation of the structure [2]. Each column in the heterogeneous population category assumes that all other columns are equivalent, and that more complex heterogeneous populations are formed by combining columns, as indicated by Fig. 1.

Two types of heterogeneous population in Table 1, material and topology, are stated as having scenarios where label consistency may or may not occur. Both of these categories have a hierarchy of additional attributes (discussed in Part II [2]) that will define whether these labels are consistent. For example, if populations are heterogeneous in terms of their material class, the classification and assessment labels will be inconsistent, e.g. a metallic structure will not experience delamination, whereas there are delamination labels associated with a composite. Furthermore, even when the material class and specific materials are the same, the grade or batch properties will cause variation that could cause assessment label inconsistency. In a similar manner, classification and assessment label consistency can change within the topology category. If the topologies between a population are different, but all the joints within the graph are equivalent, then the classification and assessment labels will be consistent. In contrast, if the joints change between structures, e.g. one structure is assembled via bolted joints and the other structure welded, then the classification and assessment label spaces will be inconsistent.

In scenarios where the label spaces are not consistent, some form of label space matching is required. This is particularly important in an unlabelled (or partially-labelled) target space where information on cluster relationships between the source and target domain are not defined in the training data and the label space mismatch means that the problem is ill-defined; for example, in a SHM localisation problem, where some locations do not actually exist on both structures.

Even when the damage label space between members of the population is inconsistent globally, a subset of these labels may be consistent, as shown in Fig. 2. This demonstrates the power of graph and attribute matching between members of a population (see Part II [2]), as locating the common attributes/subsystems  $S_c$  between members of a population helps identify the level of information that can be exchanged through a mapping  $\phi$  (where this mapping may be through a latent space [5]). Within an IE and AG representation of structures, these common subsystems  $S_c$  become common subgraphs in the AGs, which can be determined by graph matching, as demonstrated in Part II [2]. There will be SHM scenarios in which only part of a population needs to be similar for mapping to be possible. For example, if the SHM problem  $\mathcal{SP}$  is performing damage localisation on a five and three degree-of-freedom structure, then *locally* features and localisation labels can be considered consistent. This would occur if, for example, the first three natural frequencies are considered, and that damage labels are only used for the first three degrees-of-freedom for each structure. It is also noted that global feature consistency can be obtained, even for topologically-heterogeneous populations, by selecting an appropriate feature. For example, if in the previous example the features were transmissibilities, with the same number of frequencies bins over the same frequency range, then the feature spaces are consistent. However, this leads to questions over whether the key damage sensitive information is contained within the feature for both structures, and may lead to a phenomenon known as *negative transfer*, discussed in Section 2.4.

It is clear that the path to performing SHM on a population level requires a different workflow to that of conventional SHM, where health monitoring is performed on an individual structure. Fig. 3 presents the differences in these workflows, where *graph matching* and *transfer learning* become analogous stages to *feature extraction*, i.e. a damage sensitive feature for a population will be one which captures structural similarities in the population and is sensitive to the same type of damage (at least up to the SHM problem  $\mathcal{SP}$  being performed), and which can be transferred appropriately between each member of the population. This poses the question as to whether an appropriate feature for a conventional SHM problem will apply in the population case, as scenarios will appear in which a feature will become inappropriate for a population context.

From Table 1 it is clear that the start of any population-based SHM campaign is considering what SHM scenario is required and then obtaining the level of consistency and similarities between members of the population. Once this has been obtained, there are then four mapping problems that occur: consistency in feature and label spaces, consistency in the feature space and inconsistency in the label space, inconsistency in the feature space and consistency in the label space, and inconsistency in feature and label spaces. Table 2 defines the transfer learning approach appropriate for each assumption. Label-space matching refers to the process of identifying and pairing equivalent label classes in the transfer learning mapping, where some classes will be left unmatched between the source and target domains. This process must be performed correctly to prevent negative transfer of class labels, and is challenging, as often the task domain is unlabelled (or minimally labelled with undamaged labels only) requiring some form of semi-supervised learning [27]. It is important to state that this risk of negative transfer increases when label inconsistency is assumed, for both homogenous and heterogeneous transfer learning, and so as far as possible, a PBSHM scenario should be posed as one that is consistent in the label space.

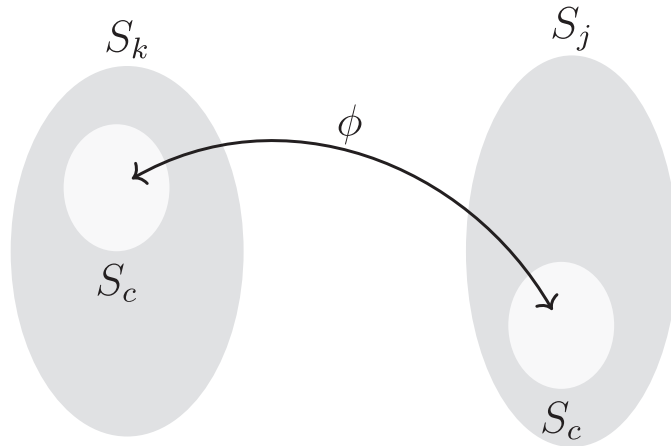
At this point it is useful to compare transfer learning to other forms of learning that are also appropriate for a population-based approach to SHM. As mentioned in the first paper of this series [1], a *form* is one approach to PBSHM for a homoge-

<sup>2</sup> Prognosis labels are not considered here, as generally a physics-based model of damage progression is required in order to perform prognosis [26].

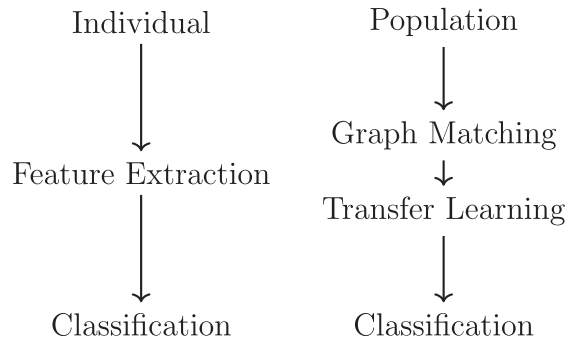
**Table 1**

Feature and label consistencies for types of population when considering modal features and localisation resolution at the graphical representation level.

	Homogeneous	Heterogeneous		
		Geometry	Material	Topology
Feature Consistency	Yes	Yes	Yes	No
Detection Label Consistency	Yes	Yes	Yes	Yes
Localisation Label Consistency	Yes	Yes	Yes	No
Classification Label Consistency	Yes	Yes	Yes/No	Yes/No
Assessment Label Consistency	Yes	Yes	Yes/No	Yes/No



**Fig. 2.** The potential mapping  $\phi$  between common attributes/subsystems  $S_c$  for a population of two structures  $S_k$  and  $S_j$ .



**Fig. 3.** Comparison of workflows for conventional (left) and population-based (right) SHM.

**Table 2**

The type of transfer learning approach for each feature and label consistency scenario.

	Label Consistency	Label Inconsistency
Feature Consistency	Homogeneous Transfer Learning	Homogeneous Transfer Learning with Label Space Matching
Feature Inconsistency	Heterogeneous Transfer Learning	Heterogeneous Transfer Learning with Label Space Matching

neous population, and will assume consistent label spaces (with respect to the normal conditions) and can assume either consistent or inconsistent feature spaces. This technique seeks to create a general representation of a population within the data-domain. A form can be inferred from a group of structures and used to transfer this knowledge to remaining members of a population. When used in this manner, it can be seen as a type of transfer learning. However, a form can also be inferred across a complete population and in this scenario aims to infer an improved general learner for the complete population, in which it is a type of multi-task learning. The other difference for the form approximation in Part I [11] is that the

model is usually learnt in the data-domain, and therefore the variance of a general learner will often be inflated, unlike transfer learning or multi-task learning that often project to a latent space, with the aim of removing uncertainty not related to the latent process. As stated previously, multi-task learning differs from transfer learning in that the aim is to infer a general learner over all domains rather than transfer from one group to another. Multi-task learning can be both consistent and inconsistent in the feature space but will be consistent in the label space. In a PBSHM context, multi-task learning may be useful in removing the effects of confounding influences [28] and improving a learner for a structure with minimal labels within a population.

#### 2.4. Negative transfer

A major concern when performing transfer learning, is if information is incorrectly mapped from one domain to another and reduces the performance of a learner when compared to learning from the target domain alone [5]. This phenomenon is known as *negative transfer*, and often occurs when source domains are *very* dissimilar to the target domain [29]. The fundamental idea behind transfer learning is that the source domains contain useful and relevant information about a target domain. This can be hard to determine before transfer, and becomes especially problematic when label inconsistency is assumed, the target domain is unlabelled, or that an example for each label class is not available. A binary classification scenario where negative transfer occurs is when source domain classes are mapped on to the opposite classes on the target domain, due to the target domain being unlabelled, and the source classes being dissimilar to the target domain.

Negative transfer highlights the question of when to transfer, and motivates the reasoning behind developing a measure of similarity between structures [2], as this provides an informative method for determining if a dataset will harm the learning process. Within the machine learning literature, several approaches have been proposed that seek to address negative transfer [30–32]. Graph-based methods that define the relationships between source domains by stating them graphically using a transferability metric have been developed [30]. This provides similar motivation for the Irreducible Element and Attributed Graph approach [2]; however, in an SHM context these graphical representations can be formed from a physics-based viewpoint, aiding the strength of knowledge about similarity. Other approaches in avoiding negative transfer have sought to weight each source domain based on its relevance to the target domain; this is known as instance weighting [12]. Local cluster-based weighting has also been proposed, meaning that for each class on a source domain, an individual weighting is provided, stating that informativeness may not be globally shared in a particular source domain [31]. As a result, it is important to consider and account for the possibility of negative transfer when identifying what structures, and their corresponding datasets, to use in transfer learning, as well as developing algorithms to reduce or remove the possibility of negative transfer within PBSHM.

### 3. Domain adaptation

Domain adaptation is one form of transfer learning that seeks to transfer feature spaces between source and target domains, assuming that their marginal distributions over the finite sample set are not equal  $p(X_s) \neq p(X_t)$  (and often that the joint distributions are different  $p(Y_s, X_s) \neq p(Y_t, X_t)$ ). These techniques are primarily designed for homogeneous transfer learning, where the feature space and label space are consistent [5,13,15,17]; however, heterogeneous transfer learning forms of domain adaptation do exist [14,16,7] for handling inconsistent feature spaces through a projection matrix.

This section outlines domain adaptation and its assumptions, before demonstrating its applicability to four case studies. In order to keep this section concise one algorithm, Joint Domain Adaptation (JDA) [15] (and a novel modification of the method), are used to illustrate the effectiveness of domain adaptation to different population types, as it contains the general assumptions required for this discussion, and its learnt mapping is straightforward to visualise. It is noted that other algorithms may show variations in performance, but the general challenges outlined in this section will still apply.

The first case study in this section performs PBSHM on a homogeneous population, with the second to fourth case studies involving heterogeneous populations of differing degrees; geometric and material differences, topological differences, and finally a numerical to experimental scenario. These case studies illustrate what aspects of PBSHM are currently achievable, highlighting the required areas of further research in making PBSHM applicable across the complete range of problems outlined in Section 2.3.

Domain adaptation is formally defined (for a single source and target domain) as:

**Definition 6.** *Domain adaptation* applies when a given inference is required for a target domain  $\mathcal{D}_t$  and task  $\mathcal{T}_t$ , and is the process of improving the target predictive function  $f_t(\cdot)$  in  $\mathcal{T}_t$  given a source domain  $\mathcal{D}_s$  and task  $\mathcal{T}_s$ , whilst assuming  $\mathcal{X}_s = \mathcal{X}_t$  and  $\mathcal{Y}_s = \mathcal{Y}_t$  but that  $p(X_s) \neq p(X_t)$ , and one can further assume  $p(Y_s|X_s) \neq p(Y_t|X_t)$ .

Various algorithms have been developed for this scenario [13,15,17] and are a type of feature-based approach to transfer learning, as discussed in Section 2.2.1. In addition, several of these approaches have been applied in a PBSHM context [4].

#### 3.1. Joint domain adaptation

In the applications below Joint Domain Adaptation (JDA) is implemented in order to demonstrate the performance of domain adaptation to different population types in PBSHM. The application of this algorithm is used to visually demonstrate



the ideas outlined in Section 2. In keeping with the focus of the paper, only a brief outline of JDA is provided; the interested reader is referred to [15,4] for more mathematical details. JDA is a homogeneous transfer learning technique introduced by Long et al. [15] that assumes the joint distributions between the source and target are different  $p(Y_s, X_s) \neq p(Y_t, X_t)$ , and finds the optimal latent mapping  $\phi$  in which the distance between the marginal distributions  $p(\phi(X_s)) \approx p(\phi(X_t))$  and the class conditional distributions  $p(\phi(X_s)|Y_s = c) \approx p(\phi(X_t)|Y_t = c)$  for each class  $c \in \{1, \dots, C\}$  in  $\mathcal{Y}$  are minimised – the goal is that the source and target datasets map on top of each other. This objective is assumed to approximate minimising joint distributions from the source and target domains, where the class conditionals are matched, as the conditionals are often challenging and computationally expensive to compute.

The cost function used to minimise the distance between the source and target joint distributions is formed from the (squared) Maximum Mean Discrepancy (MMD) distance [33]. This distance is the difference between the two empirical means of the feature data through a nonlinear mapping induced by a kernel  $K$  into a Reproducing Kernel Hilbert Space (RKHS), i.e.  $K = k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$  (where a variety of kernels can be implemented, e.g. a linear kernel  $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$ ). The MMD distance is defined as,

$$\text{Dist}(p(\phi(X_s)), p(\phi(X_t))) = \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} \phi(\mathbf{x}_{s,i}) - \sum_{i=1}^{N_t} \phi(\mathbf{x}_{t,i}) \right\|_{\mathcal{H}}^2 = \text{tr}(KM) \quad (1)$$

where  $K = k(X, X) \in \mathbb{R}^{(N_s+N_t) \times (N_s+N_t)}$  given  $X = X_s \cup X_t \in \mathbb{R}^{(N_s+N_t) \times d}$  and  $M$ , the MMD matrix, defines the empirical mean,

$$M(i, j) = \begin{cases} \frac{1}{N_s^2}, & \mathbf{x}_i, \mathbf{x}_j \in X_s \\ \frac{1}{N_t^2}, & \mathbf{x}_i, \mathbf{x}_j \in X_t \\ -\frac{1}{N_s N_t}, & \text{otherwise.} \end{cases} \quad (2)$$

By leveraging the low-rank empirical kernel embedding  $\tilde{K} = KWW^T K$  [34] the (squared) MMD distance for both the marginal and class-conditional distributions (i.e. the summation of the two distances) can be formed in terms of a set of weights  $W \in \mathbb{R}^{(N_s+N_t) \times k}$ ,

$$\text{Dist}(p(\phi(X_s)), p(\phi(X_t))) + \text{Dist}(p(\phi(X_s)|Y_s), p(\phi(X_t)|Y_t)) \approx \text{tr}(W^T K M_c K W) \quad (3)$$

where  $M_c$  is the MMD matrix for the complete distance, defining the empirical mean embedding; this summation can be minimised to find the optimal latent mapping. It is noted that JDA assumes a *completely unlabelled* target domain and utilises a simplistic form of semi-supervised learning, using a classifier trained on the source domain projection to predict pseudo-labels  $\hat{Y}_t$  for the projected target domain. The MMD matrix is defined as,

$$M_c(i, j) = \begin{cases} \frac{1}{N_s^{(c)} N_s^{(c)}}, & \mathbf{x}_i, \mathbf{x}_j \in \mathcal{D}_s^{(c)} \\ \frac{1}{N_t^{(c)} N_t^{(c)}}, & \mathbf{x}_i, \mathbf{x}_j \in \mathcal{D}_t^{(c)} \\ \frac{-1}{N_s^{(c)} N_t^{(c)}}, & \begin{cases} \mathbf{x}_i \in \mathcal{D}_s^{(c)} \mathbf{x}_j \in \mathcal{D}_t^{(c)} \\ \mathbf{x}_j \in \mathcal{D}_s^{(c)} \mathbf{x}_i \in \mathcal{D}_t^{(c)} \end{cases} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where  $\mathcal{D}_s^{(c)} = \{\mathbf{x}_i : \mathbf{x}_i \in \mathcal{D}_s \wedge y(\mathbf{x}_i) = c\}$  are the instances that belong in class  $c$  given the true source label  $y(\mathbf{x}_i)$  of  $\mathbf{x}_i$  and  $N_s^{(c)} = |\mathcal{D}_s^{(c)}|$ ; and  $\mathcal{D}_t^{(c)} = \{\mathbf{x}_i : \mathbf{x}_i \in \mathcal{D}_t \wedge \hat{y}(\mathbf{x}_i) = c\}$  are the instances that belong in class  $c$  given the pseudo-target label  $\hat{y}(\mathbf{x}_i)$  of  $\mathbf{x}_i$  and  $N_t^{(c)} = |\mathcal{D}_t^{(c)}|$  (where  $\wedge$  is the logical AND symbol). The optimisation problem is then formed (subject to the regularisation constraint, where  $\mu$  governs the level of regularisation, and kernel Principle Component Analysis (PCA) removes the trivial solution),

$$\min_{W^T K H K W = \mathbb{I}} = \sum_{c=0}^C \text{tr}(W^T K M_c K W) + \mu \text{tr}(W^T W) \quad (5)$$

where  $H = \mathbb{I} - 1/(N_s + N_t)\mathbf{1}$  is the centring matrix,  $\mathbb{I}$  is an identity matrix and  $\mathbf{1}$  a matrix of ones. Using a Lagrangian approach the optimisation problem can be formed as an eigenvalue problem where the optimal weights  $W$  are obtained from the eigenvectors corresponding to the  $k$  smallest eigenvalues from,

$$\left( K \sum_{c=0}^C M_c K + \mu \mathbb{I} \right) W = K H K W \phi. \quad (6)$$

Due to the pseudo-labelling of the target features, [15] recommends running several iterations of the optimisation to find the optimal  $W$ . A  $k$ -dimensional transformed feature space can be calculated by  $Z = KW \in \mathbb{R}^{(N_s+N_t) \times k}$ , and a classifier trained on the transformed source data can be applied to the transformed target data.

### 3.2. Homogeneous populations

A homogenous population is one in which the label spaces will be consistent, as stated in Table 1, and depending on the feature utilised, may also have consistent feature spaces. This makes homogenous populations an ideal scenario for applying JDA, and demonstrating the benefits of transfer learning. This section presents a homogenous population of two five-storey shear structures, as depicted in Fig. 4. For both structures, the stiffness elements  $\{k_i\}_{i=1}^5$  are constructed from the summation of the tip stiffness of four rectangular cantilever beams in bending i.e.  $4k_b$ , presented in Fig. 4b, where damage is applied to one beam of the four beams at a given degree-of-freedom, in the form of an open crack using the reduction in stiffness model from [35]. The geometric and material properties are nominally the same and are shown in Table 3.

The two structures are a homogeneous population as they are topologically<sup>3</sup> and structurally<sup>4</sup> equivalent [2] in their lumped-mass representation and the material and geometry parameters can be defined by a unimodal distribution with low dispersion. The SHM problem  $\mathcal{SP}$  considered here is a localisation problem, where a crack that is 10% of the beam width is applied at a distance 15% of the beam length from the base of the beam. This damage scenario is simulated for each of the degrees-of-freedom meaning that there are six labelled scenarios: undamaged labelled '1', damage at  $k_1$  labelled '2', damage at  $k_2$  labelled '3', damage at  $k_3$  labelled '4', damage at  $k_4$  labelled '5', and damage at  $k_5$  labelled '6', i.e.  $\mathcal{Y}_s = \mathcal{Y}_t \in \{1, 2, 3, 4, 5, 6\}$ . The features considered in this case study, depicted in Fig. 5, are the five damped natural frequencies ( $\mathcal{X}_s = \mathcal{X}_t$ ) meaning  $X_s \in \mathbb{R}^{N_s \times 5}$  and  $X_t \in \mathbb{R}^{N_t \times 5}$ , where  $N_s = 1800$  and  $N_t = 1200$  and each class has an equal weighting of data points. In order to test the inferred mapping, a separate test dataset from the target domain was obtained where  $X_{test} \in \mathbb{R}^{N_{test} \times 5}$  and  $N_{test} = 1500$ , again where each class has an equal number of data points. It can be seen from Fig. 5 that there is a shift in the damped natural frequencies from the source to the target domain (in the south-west direction) highlighting the need for transfer learning. It is noted that Fig. 5 shows two-dimensional visualisations of the five-dimensional feature space where each quadrant is a comparison of two features, i.e. the top right is a comparison of the fifth and first damped natural frequencies – this format for displaying feature spaces is utilised throughout this paper. JDA was implemented with ten iterations, a linear kernel, and a  $k$ -Nearest Neighbour ( $k$ -NN) classifier (five neighbours), such that the emphasis is on the domain adaptation mapping and not the classifier. Cross-validation (using five-folds) was implemented in order to identify the regularisation parameter,  $\mu = 1 \times 10^{-3}$ , and the number of transfer components,  $k = 4$ . The inferred mapping is presented in Fig. 6, where it can be seen that the source and target datasets have been successfully mapped onto each other, where the histograms of the marginal distributions are shown on the diagonal. The classification results are given in Table 4, where it can be seen that compared to a  $k$ NN without JDA, domain adaptation provides a substantial improvement in classification performance.

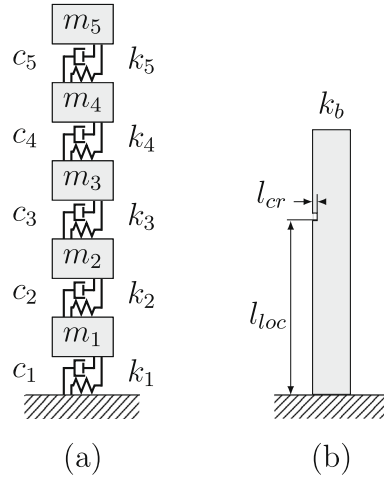
### 3.3. Heterogeneous populations: geometric and material differences and consistent label spaces

The first heterogeneous population case study considers a population with geometry and material differences where the label space between the source and target domain are consistent. The two structures in the population are structurally equivalent and are both five-storey shear structures (Fig. 4). The material differences exist as the first structure is steel and the other aluminium; the geometric dissimilarities occur due to differences in the dimensions of the structure, displayed in Table 5. The SHM problem  $\mathcal{SP}$  is the same as in the homogeneous case study i.e. a six-class localisation problem, i.e.  $\mathcal{Y}_s = \mathcal{Y}_t \in \{1, 2, 3, 4, 5, 6\}$ , where damage is introduced as open cracks located at each degree-of-freedom. The features are the five damped natural frequencies ( $\mathcal{X}_s = \mathcal{X}_t$ ) such that  $X_s \in \mathbb{R}^{N_s \times 5}$  and  $X_t \in \mathbb{R}^{N_t \times 5}$ ,  $X_{test} \in \mathbb{R}^{N_{test} \times 5}$ ;  $N_s = 1800$ ,  $N_t = 1200$  and  $N_{test} = 1500$  all with equal weighting of data points in each class. The source and target domain features are presented in Figs. 7 and 8, the magnitudes of which are very different highlighting the need for transfer learning.

In a similar manner to the homogeneous case study, JDA was implemented with ten iterations, a linear kernel and a  $k$ NN classifier (five neighbours). Cross-validation (using five-folds) identified the regularisation parameter,  $\mu = 1 \times 10^{-3}$ , and number of transfer components,  $k = 4$ . The identified transfer components from the JDA mapping are depicted in Fig. 9, where the mapping has grouped the source and target domain clusters correctly together; although the source and target domains are less close than in the homogeneous population example. Once more, domain adaptation in the form of JDA, shows vast improvements in classification performance when compared to training a classifier on the source domain and applying it to the target, which has a classification performance equivalent to random guessing, as demonstrated in Table 6. This case study demonstrates that for a selection of heterogeneous problems, homogeneous transfer learning, in the form of domain adaptation, is still applicable. As a result, PBShM scenarios that can be formed as feature and label consistent, are achievable with current transfer learning technologies.

<sup>3</sup> Topologically-equivalent structures are those with the same graphical representation. In terms of IE and AGs, two graphs are topologically-equivalent if the underlying unattributed graphs are identical.

<sup>4</sup> Structurally-equivalent structures are those with the same graphical representation with the same locations of ground nodes. In terms of IE and AGs, two graphs are structurally equivalent if they are topologically equivalent with ground nodes in the AGs occurring in the same place.



**Fig. 4.** Schematic of the five degree-of-freedom shear structure. Panel (a): Schematic of the full system. Panel (b): Schematic of the cantilever beam component where  $\{k_i\}_{i=1}^5 = 4k_b$  i.e. the stiffness coefficients in (a) are generated from four times the tip bending stiffness in (b).

### 3.4. Heterogeneous populations: topological differences and inconsistent label spaces

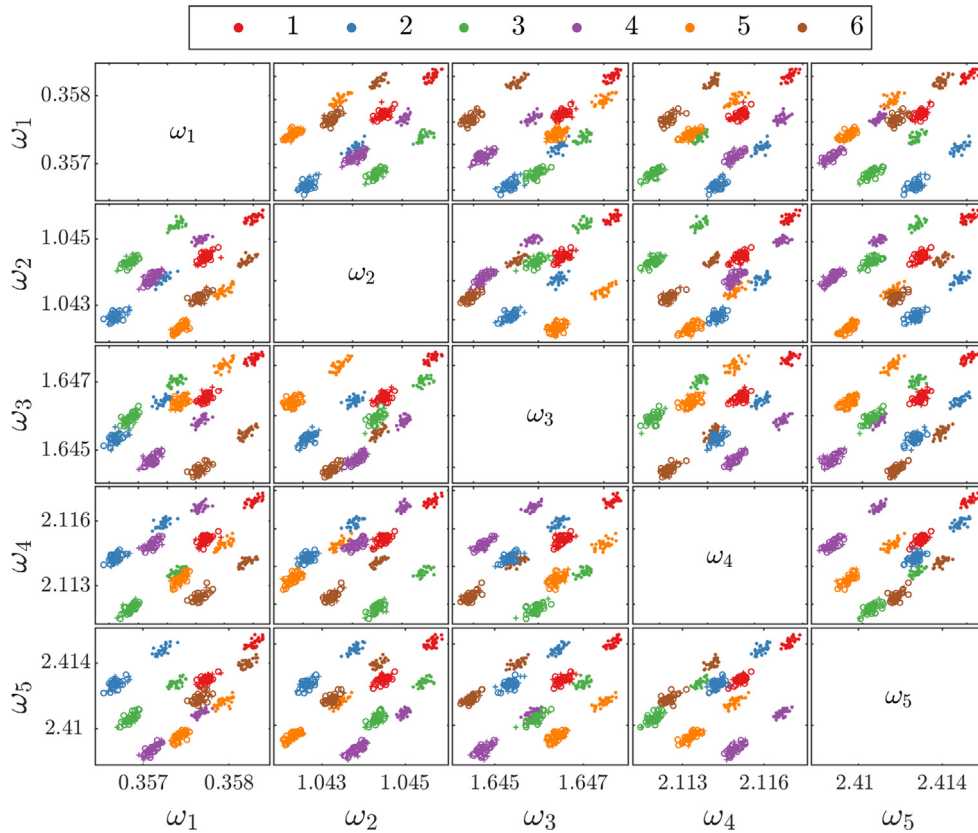
A significant challenge in PBSHM is performing damage identification on populations with inconsistent label spaces, i.e.  $\mathcal{Y}_s \neq \mathcal{Y}_t$ . However, it may be the case that a subset of the label space from the source and target structures could be considered consistent; for example if considering a damage localisation problem for a population of building structures with a different number of storeys, e.g. a three-storey and a five-storey structure, then a set of the localisation labels are physically equivalent, e.g. the labels about damage located in first three storeys from both structures would exist in both the source and target domains. In these scenarios the challenge is determining what subset of the label space is consistent, which can be particularly difficult as data from the target structure are often unlabelled or at best partially labelled – typically with labels only being available for the undamaged condition. Again, this issue highlights the need for identifying similarities between structures, particularly in terms of matching label spaces. This consideration provides additional motivation for the development of an IE and AG approach in [2], where structural similarities (and therefore matching label spaces) can be found for a given SHM problem  $\mathcal{SP}$ . This section explores and develops an approach for performing domain adaptation in these scenarios.

It is helpful to formalise the requirements on transfer learning in these scenarios with mathematical notations such that the problem is made specific. Each structure will have a label space  $\mathcal{Y}$  composed from the possible damage states being considered by the SHM problem,  $\mathcal{SP}$ . In these scenarios it is assumed that there is a common subset of labels  $\mathcal{Y}^{sub}$  that occur in both the source  $\mathcal{Y}_s$  and target  $\mathcal{Y}_t$  label spaces, i.e.  $\mathcal{Y}^{sub} \subset \mathcal{Y}_s$  and  $\mathcal{Y}^{sub} \subset \mathcal{Y}_t$ . A special case occurs when the common subset of the label space is equal to the source label space, i.e.  $\mathcal{Y}_s \subset \mathcal{Y}_t$ , called the  $(L + N)$ -problem; as  $N$  extra labels are present in the target domain that are not found in the source domain. In this scenario all the information from the source domain should be transferred to the target domain, aligning the source labels to their equivalent target labels with the remaining  $N$  labels being left ‘unpaired’ with the source domain. The simplest form of this problem is the scenario where the target label space is equivalent to the source label space with the addition of an extra class, i.e.  $\mathcal{Y}_t \in \{\mathcal{Y}_s, Y_{+1}\}$ ; this is referred to as an  $(L + 1)$ -problem. These scenarios demonstrate that, for some local set, the label space between the source and target structures may be considered consistent and transfer learning, although challenging, is achievable.

In terms of PBSHM, heterogeneous populations, where individual structures are not topologically (or structurally) equivalent, lead to inconsistency in the label space for a damage localisation problem. In keeping with the shear structure illustrations, an  $(L + N)$ -problem could be formed by considering a source structure with three-storeys where the label space is  $\mathcal{Y}_s \in \{1, 2, 3, 4\}$  (i.e. ‘1’ is undamaged and ‘2’ to ‘4’ represent damage at each of the three storeys) and a target structure with

**Table 3**  
Properties of the source and target structures for the homogenous population case study.

Property	Unit	Source	Target
Beam geometry, $\{l_b, w_b, t_b\}$	mm	{5000.0, 350.0, 350.0}	{4999.0, 351.2, 349.7}
Mass geometry, $\{l_m, w_m, t_m\}$	mm	{12000.0, 12000.0, 350.0}	{12001.0, 11998.8, 351.6}
Elastic modulus, $E$	GPa	$\mathcal{N}(210.00, 1 \times 10^{-9})$	$\mathcal{N}(210.89, 1 \times 10^{-9})$
Density, $\rho$	kg/m <sup>3</sup>	$\mathcal{N}(8000.0, 50)$	$\mathcal{N}(8019.4, 10)$
Damping coefficient, $c$	Ns/m	$\mathcal{G}(8.0000, 0.8)$	$\mathcal{G}(7.9981, 0.8)$



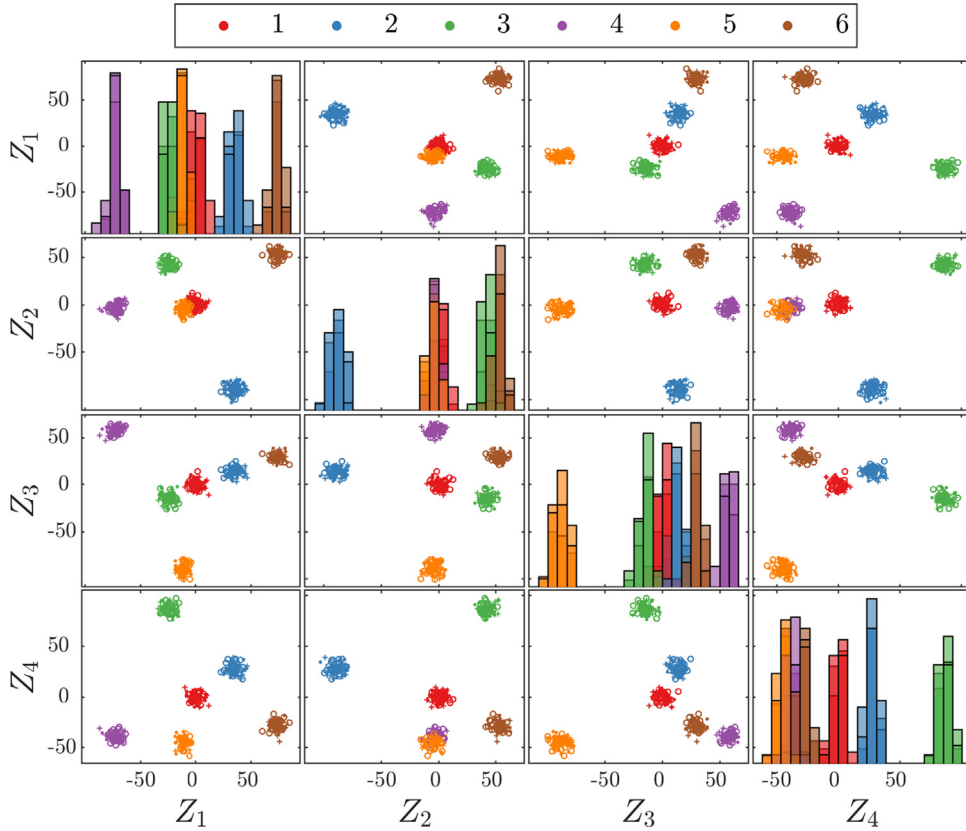
**Fig. 5.** Subset of source and target domain features (every tenth data point) for the homogeneous population case study (where damped natural frequencies are in Hz). The source domain data are denoted by ( $\cdot$ ), the target-domain training and testing data are denoted by (+) and (o) respectively.

$N$ -storeys where the label space is  $\mathcal{Y}_t \in \{1, 2, 3, 4, \dots, N+1\}$  (where labels '5' to ' $N+1$ ' refer to damage at storeys not physically present in the source domain). An  $(L+1)$ -problem could be a scenario where the source structure is a three-storey structure and the target is a four-storey structure, meaning  $\mathcal{Y}_s \in \{1, 2, 3, 4\}$  and  $\mathcal{Y}_t \in \{1, 2, 3, 4, 5\}$ . As stated the mapping problem becomes more complex and prone to negative transfer as the source and target labels spaces are considered increasingly inconsistent. For this reason the examples presented in this case study are all  $(L+1)$ -problems where a novel procedure is proposed for performing domain adaptation for this specific scenario.

It is worth noting that the IE and AG representation of structures in [2] provides a natural way of expressing the consistency of the label spaces. This is because a label set is attached to each node and edge of the attributed graph. As a consequence, for a given SHM problem  $SP$ , a graph matching algorithm can be used to identify the level of label space agreement, where an ideal scenario for a heterogeneous population is that they produce problems that are  $(L+N)$  in nature.

The novel procedure proposed in Algorithm 1 is designed for an  $(L+1)$ -problem. In terms of the source and target datasets, the approach assumes a fully labelled source domain where feature examples are present for each class label, and that the target domain is unlabelled but separable enough to perform unsupervised clustering such that unlabelled clusters can be identified (where the true label for each cluster remains unknown before transfer). The approach subsequently assumes that there are geometrical similarities between the feature data relating to the  $L$ -label space in both the source and target datasets. This means that the mapping between the corresponding  $L$  labels in the source and target domain should be 'closer' than the complete source dataset to data from a different set of  $L$  labels in the target domain – this can be thought of as a naive form of manifold assumption, i.e. it is expected that the manifold of the source and target clusters is the 'same'. The algorithm identifies  $L+1$  clusters in the target domain (through an unsupervised clustering approach), it then iterates through each of the  $L$  combinations of clusters (in a leave-one-out sense), inferring the JDA mapping for each unique set of clusters and calculating the MMD distance of the mapping – using  $\text{tr}(W^T K M_c K W)$  – in the transformed space<sup>5</sup>. The combination of  $L$  target clusters that produce the smallest MMD distance are then assumed to be the matching label space to the source domain. At this stage, the complete source and target datasets are projected through the inferred mapping and an unsu-

<sup>5</sup> It is noted that the number of transfer components should be the same for each JDA mapping, such that the MMD distance is calculated for domains of the same dimension.



**Fig. 6.** Subset of transfer components for the source and target domains (every tenth data point) in the homogeneous population case study. The source domain transfer components are denoted by (.), the target-domain training and testing transfer components are denoted by (+) and (o) respectively.

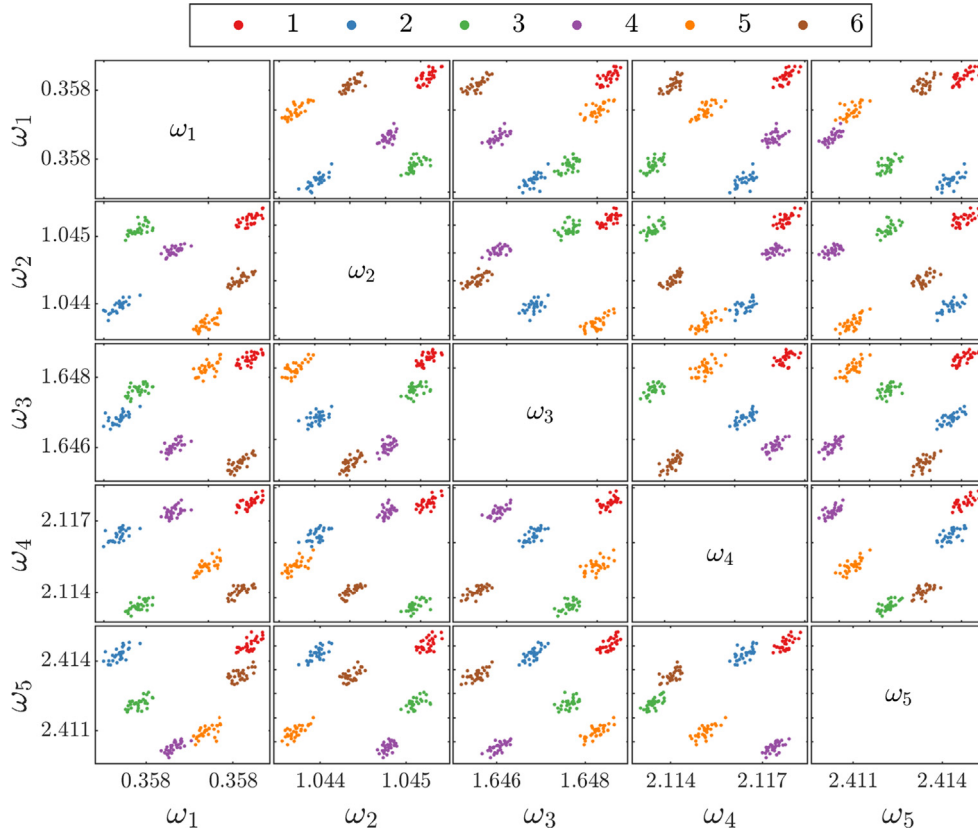
**Table 4**  
Classification results for the homogenous population case study trained on the labelled source domain and applied to the unlabelled target domain.

Method		kNN	JDA
Mapping Training	Accuracy	83.8%	100.0%
Mapping Testing	Accuracy	83.9%	100.0%

**Table 5**  
Properties of the source and target structures for the heterogeneous population case study with geometric and material differences.

Property	Unit	Source	Target
Beam geometry, $\{l_b, w_b, t_b\}$	mm	{5000, 350, 350}	{4300, 500, 500}
Mass geometry, $\{l_m, w_m, t_m\}$	mm	{12000, 12000, 350}	{10000, 10000, 500}
Elastic modulus, $E$	GPa	$\mathcal{N}(210, 1 \times 10^{-9})$	$\mathcal{N}(71, 75 \times 10^{-7})$
Density, $\rho$	kg/m <sup>3</sup>	$\mathcal{N}(8000, 10)$	$\mathcal{N}(2700, 2.5)$
Damping coefficient, $c$	Ns/m	$\mathcal{G}(50, 0.1)$	$\mathcal{G}(8, 0.8)$

pervised clustering algorithm is used to cluster the source and target datasets in the transformed space; where the source data should align with the correct  $L$ -target domain clusters, where the extra target domain cluster remains unlabelled. At this stage the unsupervised clustering labels can be matched to the  $L$ -source labels, where  $L$  identified unsupervised clusters are given the source label with which they are in most correspondence.



**Fig. 7.** Subset of source domain features (every tenth data point) for the heterogeneous population case study with geometric and material differences (where damped natural frequencies are in Hz).

---

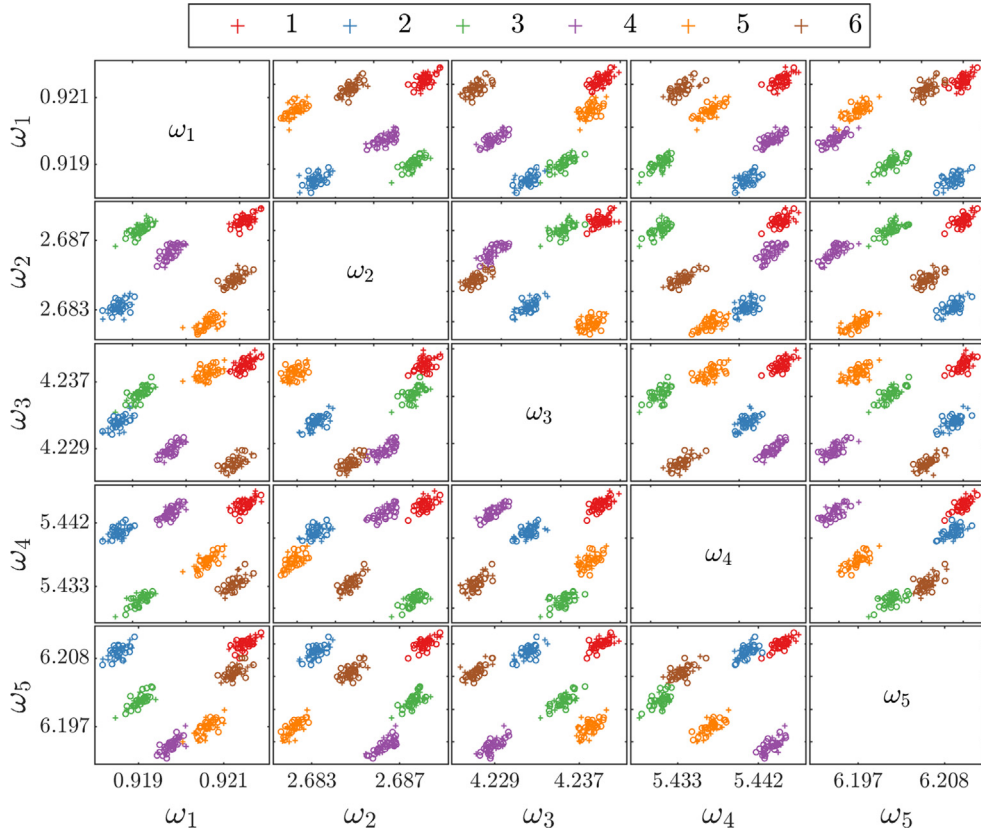
#### Algorithm 1 MMD approach to the $(L + 1)$ -problem

---

$Y_t^{pred} = \text{CLUSTER}(X_t)$       ▷ Find  $L + 1$  clusters in  $X_t$   
**for**  $i = 1 : L + 1$       ▷ For the number of clusters in the target domain  
 $X_t^{sub} = X_t(Y_t^{pred} \neq i)$       ▷ Select target data for  $L$  clusters  
 $K^{sub} = \text{KERNEL}(X_s, X_t^{sub})$       ▷ Embed the source and subset of target data in a kernel  
 $\{W, M_c\} = \text{JDA}(K^{sub}, Y_s, k, \mu)$       ▷ Find weights and MMD matrix from JDA mapping  
 $K = \text{KERNEL}(X_s, X_t)$       ▷ Embed the source and complete target datasets in a kernel  
 $Z[i] = KW$       ▷ Calculate the transfer components  
 $mmd[i] = \text{tr}(W^T K^{sub} M_c K^{sub} W)$       ▷ Calculate the MMD distance for the  $i$ th mapping  
**end for**  
 $n = \min(mmd)$       ▷ Find the minimum MMD distance  
 $Z^{opt} = Z[n]$       ▷ Find the 'optimal' mapping  
 $Y_t^{pred} = \text{CLUSTER}(Z^{opt}, Y_s)$       ▷ Find  $L + 1$  clusters from transformed data

---

The effectiveness of the proposed approach is demonstrated on several examples of the  $(L + 1)$ -problem. These examples help illustrate the limitations of the assumption that the feature data for the  $L$  clusters between the source and target datasets are geometrically similar. This demonstrates when the approach is applicable and motivates further research required to solve SHM problems for general heterogeneous populations. All of the following examples consider populations of an  $n$ - and an  $m$ -storey structure, where the SHM problem  $\mathcal{SP}$  is the localisation of open cracks (10% of the beam width at a distance 15% of the beam length) at each storey. Unless otherwise stated, the properties of the structures are shown in Table 7 where the differences between each member of the population are solely the number of storeys. It is noted that the clustering algo-



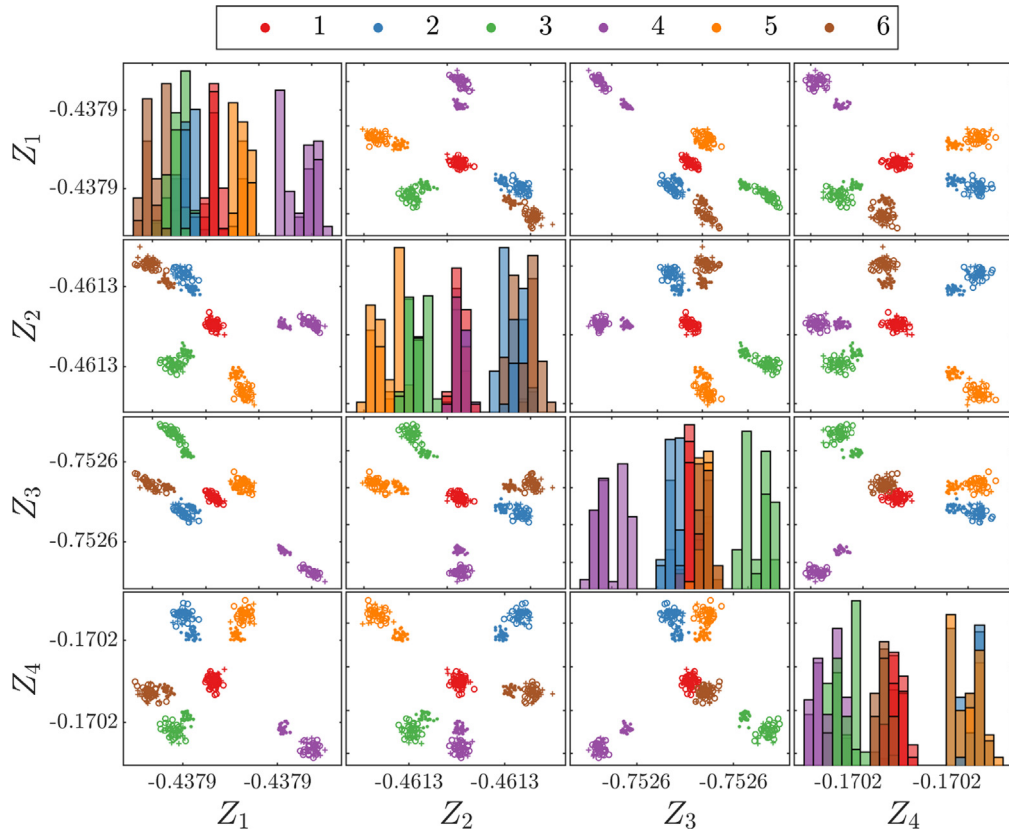
**Fig. 8.** Subset of target domain features (every tenth data point) for the heterogeneous population case study with geometric and material differences (where damped natural frequencies are in Hz). The target-domain training and testing data are denoted by (+) and (o) respectively.

rithm implemented here is a Gaussian Mixture Model (GMM) [36], although other methods could also be used. JDA is implemented with ten iterations, where pseudo-labelling is performed using a *k*NN, where other parameters are stated depending on the application.

**3.4.1. Three-storey to three-storey: missing cluster**

The first demonstration of Algorithm 1 is for the scenario where  $\mathcal{Y}_s = \mathcal{Y}_t$ , but feature data for a single class label is not present in the source domain. This situation represents a form of the  $(L + 1)$ -problem where the assumption of geometric similarity of the feature data for the source and target domain holds, as the two labels spaces are equal. In this example the source and target structures have different geometries and material properties (and are therefore a heterogeneous population), stated in Table 8, but are topologically-equivalent; both being three-storey shear structures. In order to generate an  $(L + 1)$ -problem, the cluster associated with damage at the second storey is considered not available for the source domain. The label sets are  $Y_s \in \{1, 2, 4\}$  and  $Y_t \in \{1, 2, 3, 4\}$ , and it can therefore be considered that the label sets are inconsistent (where ‘1’ is the undamaged class and labels ‘2’ to ‘4’ relate to damage at each degree-of-freedom i.e. ‘2’ denotes damage between the ground and first storey etc.). The first two damped natural frequencies of the self-normalised source and target domains are presented in Fig. 10a highlighting the differences in the label spaces. It is noted that  $X_s \in \mathbb{R}^{N_s \times 3}$ ,  $X_t \in \mathbb{R}^{N_t \times 3}$ ,  $X_{test} \in \mathbb{R}^{N_{test} \times 3}$  (i.e. the first three damped natural frequencies are used) where  $N_s = 900$ ,  $N_t = 800$  and  $N_{test} = 1000$  (where each class has an equal weighting of data points).

Algorithm 1 aims to select the combination of target domain clusters that produce the smallest MMD distance – assuming that this set of target domain clusters will produce positive transfer. An unsupervised GMM (where the number of clusters was four) was used to cluster the target domain such that unique combinations of clusters could be implemented in inferring the JDA mappings (using a linear kernel, where  $k = 2$  and  $\mu = 0.5$ ). Fig. 10b states the MMD distances when each cluster has been removed (in a leave-one-out sense), where it can be seen that the MMD is smallest when cluster ‘3’ is removed from the target domain – showing that the algorithm’s assumptions hold for this problem. To illustrate the JDA mappings, two examples are presented in Fig. 11, where the JDA mappings that produce the largest and smallest MMD distances are shown; i.e. when cluster ‘2’ is removed and cluster ‘3’ respectively. It is clear that Fig. 11a illustrates negative transfer has occurred, which is expected as a different combination of labels are present in the source and target domains.



**Fig. 9.** Subset of transfer components for the source and target domains (every tenth data point) in the heterogeneous population case study with geometric and material differences. The source domain transfer components are denoted by ( $\cdot$ ), the target-domain training and testing transfer components are denoted by ( $+$ ) and ( $\circ$ ) respectively.

**Table 6**  
Classification results for the heterogeneous population case study with geometric and material differences trained on the labelled source domain and applied to the unlabelled target domain.

Method		kNN	JDA
Mapping Training	Accuracy	16.7%	100.0%
Mapping Testing	Accuracy	16.7%	100.0%

**Table 7**

Properties of the source and target structures for the heterogeneous population case study with topological differences.

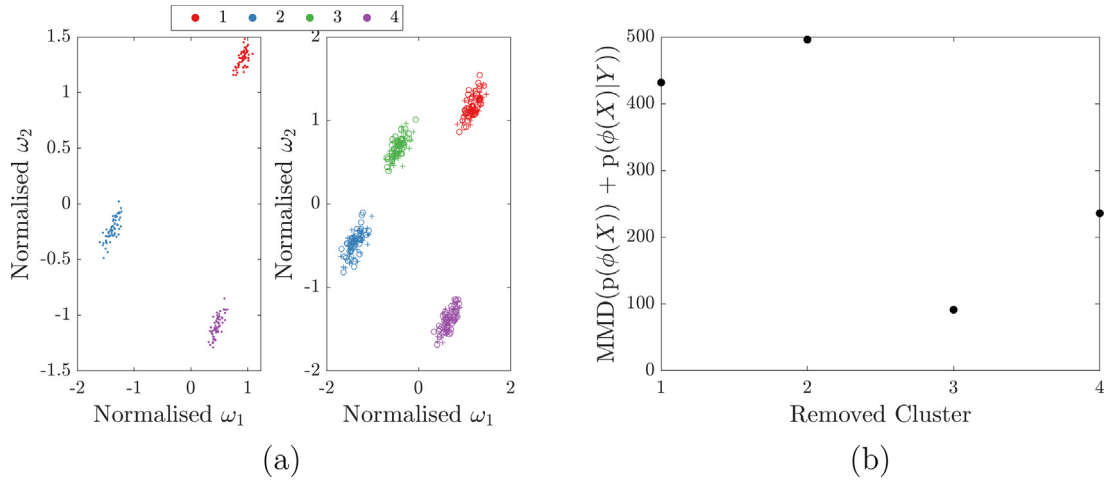
Property	Unit	Source	Target
Beam geometry, $\{l_b, w_b, t_b\}$	mm	{5000, 500, 500}	{5000, 500, 500}
Mass geometry, $\{l_m, w_m, t_m\}$	mm	{12000, 12000, 500}	{12000, 12000, 500}
Elastic modulus, $E$	GPa	$\mathcal{N}(210, 1 \times 10^{-9})$	$\mathcal{N}(210, 1 \times 10^{-9})$
Density, $\rho$	kg/m <sup>3</sup>	$\mathcal{N}(8000, 10)$	$\mathcal{N}(8000, 10)$
Damping coefficient, $c$	Ns/m	$\mathcal{G}(50, 0.1)$	$\mathcal{G}(50, 0.1)$

In contrast, Fig. 11b shows positive transfer, and that the source and target data are placed closer than in Fig. 11a – therefore validating the assumptions of Algorithm 1. It is interesting to note that in Fig. 11b the cluster not utilised in the mapping, i.e. class ‘3’, has been placed near the origin. This placing is helpful in producing separable clusters in the transformed domain, where the unsupervised GMM (with four components) trained on the source and target transformed space has correctly identified four unique clusters, as shown in Fig. 12. The identified unlabelled clusters are given the labels of the source

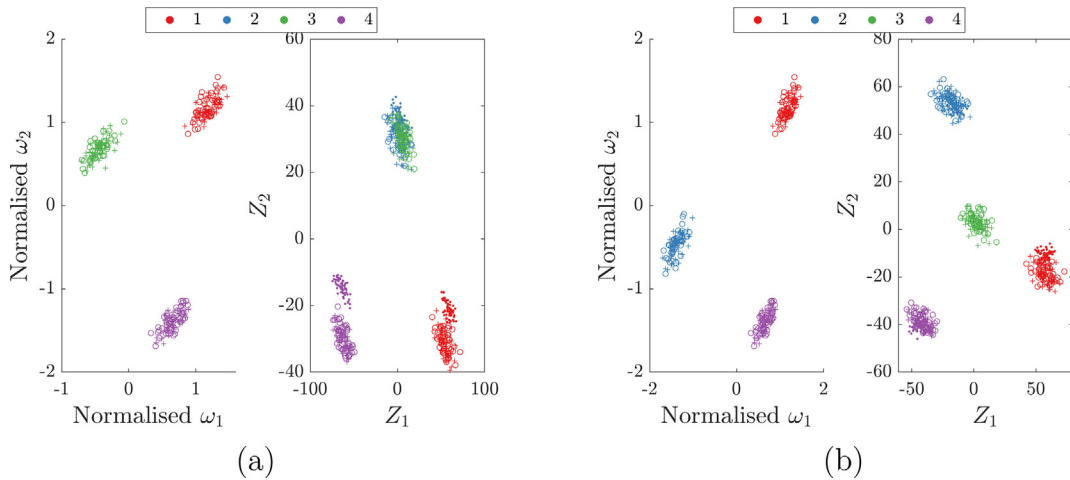


**Table 8**  
Properties of the source and target structures for the three-storey to three-storey case study.

Property	Unit	Source	Target
Beam geometry, $\{l_b, w_b, t_b\}$	mm	{5000, 500, 500}	{6000, 400, 400}
Mass geometry, $\{l_m, w_m, t_m\}$	mm	{12000, 12000, 500}	{11000, 11000, 400}
Elastic modulus, $E$	GPa	$\mathcal{N}(200, 1 \times 10^{-9})$	$\mathcal{N}(210, 1 \times 10^{-9})$
Density, $\rho$	kg/m <sup>3</sup>	$\mathcal{N}(8000, 10)$	$\mathcal{N}(7800, 10)$
Damping coefficient, $c$	Ns/m	$\mathcal{G}(50, 0.1)$	$\mathcal{G}(25, 0.2)$



**Fig. 10.** Panel (a): Subset of normalised first and second natural frequencies for source (left) and target (right) domain (every fifth data point) for the three-storey to three-storey case study. Panel (b): MMD distances from JDA mappings where one cluster has been removed from the target domain.

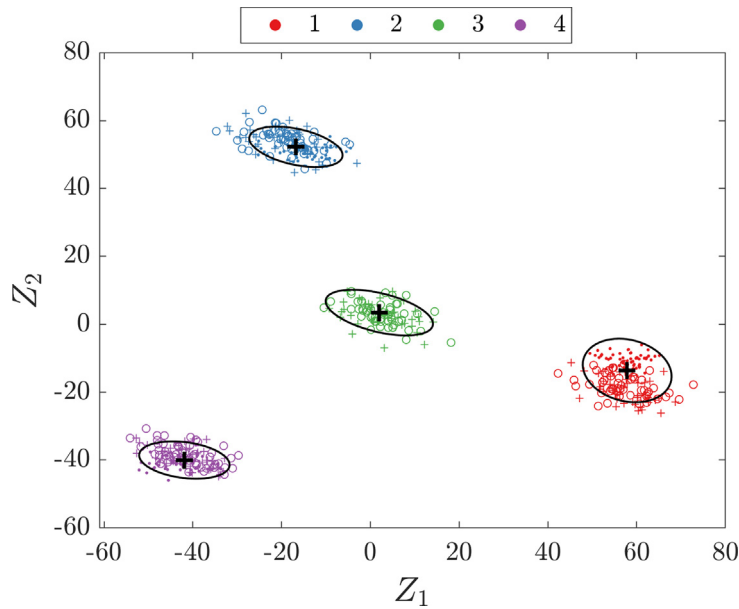


**Fig. 11.** Examples of JDA mapping for the three-storey to three-storey case study. Panel (a): Subset of normalised first and second natural frequencies for the target domain (every fifth data point) where cluster '2' is removed (left) and JDA mapping (right). Panel (b): Subset of normalised first and second natural frequencies for the target domain (every fifth data point) where cluster '3' is removed (left) and JDA mapping (right).

domain data within the cluster, resulting in 100% classification accuracy of the known  $L$ -label sets in the training-target and test-target datasets, and the target domain data relating to class '3' is completely captured by its own Gaussian cluster.

3.4.2. Three-storey to four-storey: labelling from the fixed end

In order to demonstrate the applicability of Algorithm 1 to an  $(L + 1)$ -problem where  $\mathcal{Y}_s \neq \mathcal{Y}_t$  a second example is presented. In this example the SHM problem is performing damage localisation between a three- (source) and a four- (target) storey shear structure – a heterogeneous population that is not topologically-equivalent, where its geometric and material



**Fig. 12.** Unsupervised GMM trained on the ‘optimal’ JDA transfer components for the three-storey to three-storey case study.

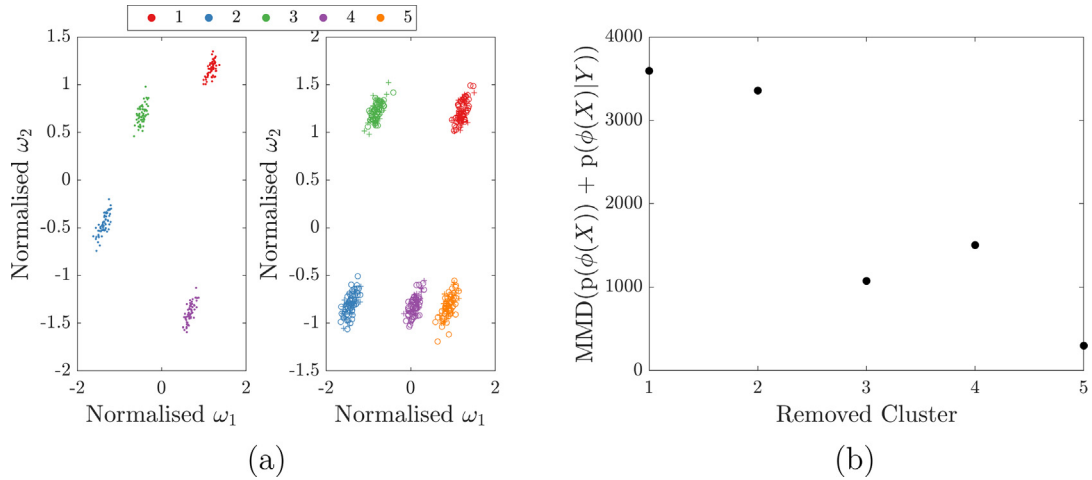
properties are shown in Table 7. The two label spaces are therefore  $\mathcal{Y}_s \in \{1, 2, 3, 4\}$  and  $\mathcal{Y}_t \in \{1, 2, 3, 4, 5\}$  where the labels are: undamaged class ‘1’ and labels ‘2’ to ‘5’ relate to damage at each degree-of-freedom numbered from the fixed end upwards (i.e. damage between the ground and first storey is labelled ‘2’, etc.). Labelling has been performed from the fixed end upwards as it is assumed that the ground boundary condition is most important for physical correspondence between the two structures. In order to keep the feature space consistent the first three damped natural frequencies are used as features, meaning  $X_s \in \mathbb{R}^{N_s \times 3}$  and  $X_t \in \mathbb{R}^{N_t \times 3}$ ,  $X_{test} \in \mathbb{R}^{N_{test} \times 3}$ ;  $N_s = 1200$ ,  $N_t = 1000$  and  $N_{test} = 1250$  (where each class has an equal weighting of data points). The (self-normalised) first two damped natural frequencies are presented in Fig. 13a, showing the inconsistent label spaces.

The MMD distances for each JDA mapping (using a linear kernel, where  $k = 2$  and  $\mu = 0.5$ ) are displayed in Fig. 13b; it can be seen that the scenario where data referring to label ‘5’ are removed from the target domain produces the smallest MMD distance. This JDA mapping is shown in Fig. 14a where positive transfer has occurred – again it is interesting to note that the JDA mapping places the target cluster relating to label ‘5’ near the origin, which is intuitive as this cluster was not used in learning the JDA mapping. Again, an unsupervised GMM (Fig. 14b) trained on the source and target transformed data correctly classifies the  $L$ -label sets from the training-target and test-target datasets (when assuming that the cluster label is obtained from the source domain data). As with the previous example, the extra target cluster (label ‘5’) is completely captured by its own Gaussian cluster.

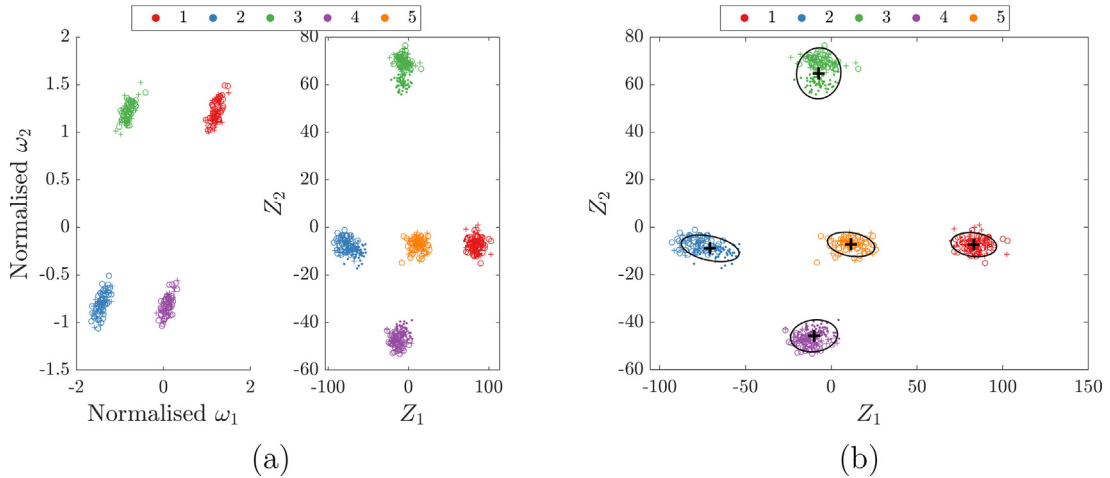
### 3.4.3. Four-storey to five-storey: labelling from the fixed end

Algorithm 1 has been demonstrated on two  $(L + 1)$ -problem case studies where the assumption holds, that there are geometric similarities in the cluster space for the  $L$ -labelled datasets from the source and target domains. In this next scenario, the algorithm is applied to a damage localisation problem between a four- (source) and five- (target) storey shear structure (with properties shown in Table 7). For the same reasons as Section 3.4.2 the first four damped natural frequencies are utilised as features such that  $X_s \in \mathbb{R}^{N_s \times 4}$  and  $X_t \in \mathbb{R}^{N_t \times 4}$ ,  $X_{test} \in \mathbb{R}^{N_{test} \times 4}$ ;  $N_s = 1500$ ,  $N_t = 1200$  and  $N_{test} = 1500$  (where each class has an equal weighting of data points). Again, the label space is numbered from the fixed end upwards, meaning  $\mathcal{Y}_s \in \{1, 2, 3, 4, 5\}$  and  $\mathcal{Y}_t \in \{1, 2, 3, 4, 5, 6\}$ , where, for example, ‘2’ refers to damage at  $k_1$ . The (self-normalised) source and target datasets are depicted in Fig. 15a, where it can be seen that the geometry in the datasets between the  $L$ -labels (i.e. ‘1’ to ‘5’) is not similar – i.e. rotations and translations of the source  $L$  clusters would not bring all the clusters equally close to the  $L$ -target clusters – meaning the algorithm is not expected to select the correct  $L$  clusters in the target domain, and therefore is unlikely to perform positive transfer.

The MMD distances for each JDA mapping (using a linear kernel, where  $k = 2$  and  $\mu = 10$ ), trained on each combination of target clusters, are shown in Fig. 15b. It is clear that the minimum MMD distance does not occur in the correct scenario, i.e. when cluster ‘6’ is removed from the target dataset. In fact, in this example, all of the JDA mappings produce negative transfer; Fig. 16 shows two JDA mappings, where cluster ‘3’ (corresponding to the smallest MMD distance) and ‘6’ (corresponding to the correct  $L$ -label set) are removed from the target dataset respectively. In both these cases JDA struggles to overcome the changing nature of the geometry of the clusters in the source and target domain, mainly due to the fact that no label knowl-



**Fig. 13.** Panel (a): Subset of normalised first and second natural frequencies for source (left) and target (right) domain (every fifth data point) for the three-storey to four-storey case study: labelling from the fixed end. Panel (b): MMD distances from JDA mappings where one cluster has been removed from the target domain.



**Fig. 14.** Panel (a): Subset of normalised first and second natural frequencies for the target domain (every fifth data point) where cluster '5' is removed (left) and JDA mapping (right). Panel (b): Unsupervised GMM trained on the 'optimal' JDA transfer components for the three-storey to four-storey case study: labelling from the fixed end.

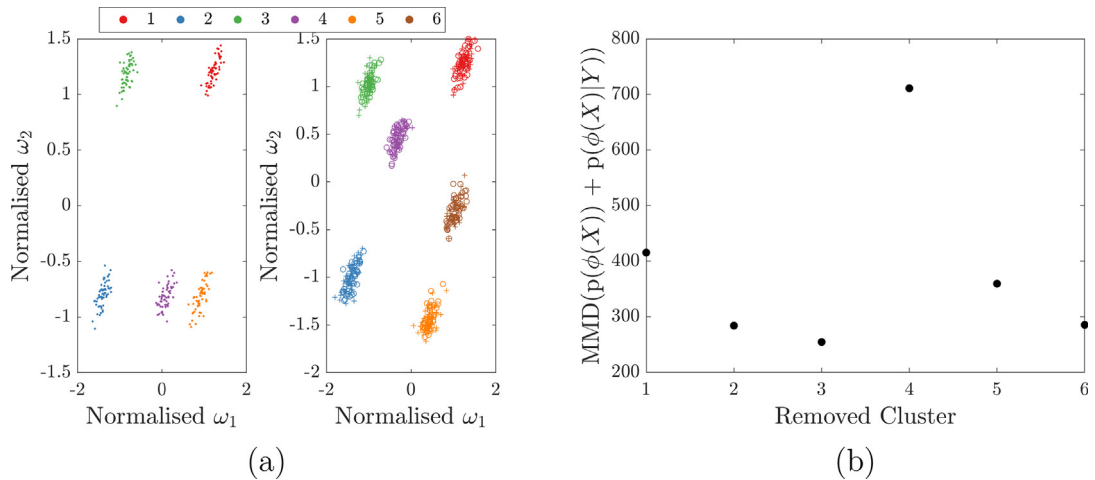
edge is known in the target domain, meaning that the inferred mapping cannot be pinned on the correct target clusters. For this example, Algorithm 1 has been demonstrated to fail; any classification result will be erroneous due to negative transfer.

This result indicates that the approach will not generalise for  $(L + 1)$ -problems across any combination of different storey shear structures, as the assumptions about the geometry of the cluster spaces is broken. This caveat means that care must be taking in applying Algorithm 1 to general SHM problems and highlights the issues with a fully black-box approach to transfer learning for PBSHM.

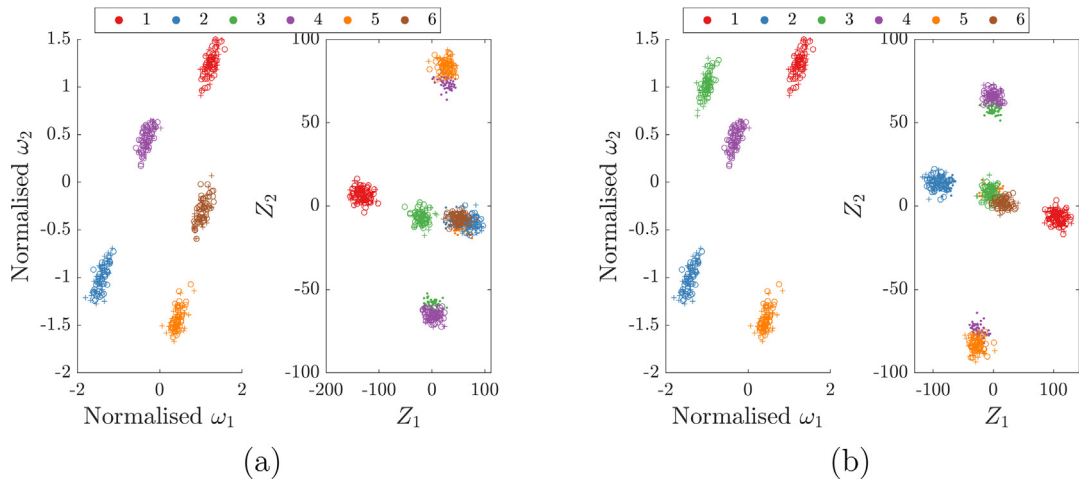
### 3.4.4. Two storey to three storey: labelling from the fixed end

Further investigation of the applicability of Algorithm 1 was performed with a two- (source) to three- (target) storey population (using the properties in Table 7). The motivation for this example is that physically, natural frequencies will become less sensitive to location specific damage (especially when solely affecting the stiffness at a particular degree-of-freedom), particularly at degrees-of-freedom near the free end – due to low amounts of strain energy – as the number of storeys increases.

The feature space represents the first two damped natural frequencies, meaning  $X_s \in \mathbb{R}^{N_s \times 2}$  and  $X_t \in \mathbb{R}^{N_t \times 2}$ ,  $X_{test} \in \mathbb{R}^{N_{test} \times 2}$ ;  $N_s = 900$ ,  $N_t = 800$  and  $N_{test} = 1000$  (where each class has an equal weighting of data points). Once more, the labels  $\mathcal{Y}_s \in \{1, 2, 3\}$  and  $\mathcal{Y}_t \in \{1, 2, 3, 4\}$ , are assigned from the fixed end upwards. Fig. 17a, again indicates



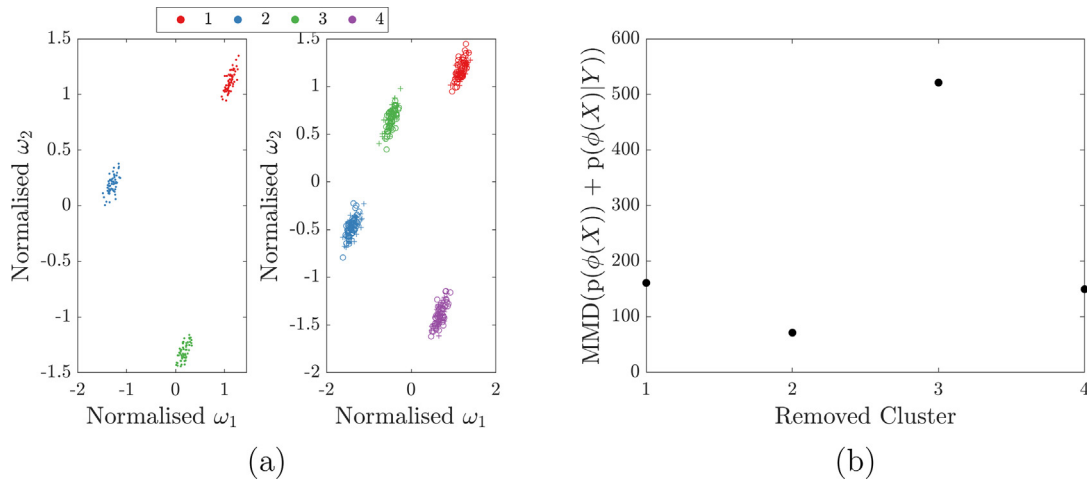
**Fig. 15.** Panel (a): Subset of normalised first and second natural frequencies for source (left) and target (right) domain (every fifth data point) for the four-storey to five-storey case study: labelling from the fixed end. Panel (b): MMD distances from JDA mappings where one cluster has been removed from the target domain.



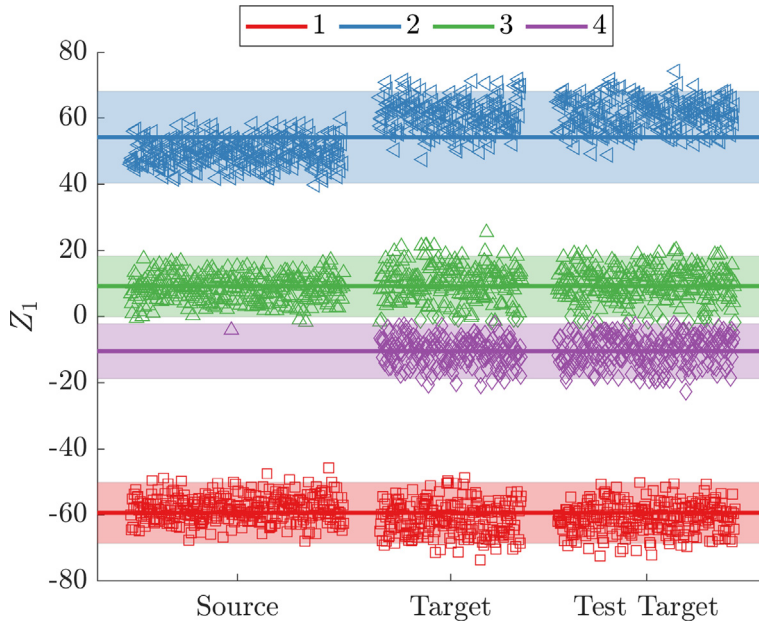
**Fig. 16.** Examples of JDA mapping for the four-storey to five-storey case study: labelling from the fixed end. Panel (a): Subset of normalised first and second natural frequencies for the target domain (every fifth data point) where cluster '3' is removed (left) and the JDA mapping (right). Panel (b): Subset of normalised first and second natural frequencies for the target domain (every fifth data point) where cluster '6' is removed (left) and the JDA mapping (right).

that the assumptions in Algorithm 1 are not applicable to this problem. Fig. 17b confirms this result, as the removal of cluster '2' from the target domain produces a closer mapping than removing cluster '4' (where all JDA mappings are inferred using a linear kernel, where  $k = 1$  and  $\mu = 0.5$ ). However, when cluster '4' is removed from the target domain, positive transfer occurs. In fact, when the JDA mapping, inferred from the target domain without cluster '4', is used to train a GMM (shown in Fig. 18), classification accuracies of 99.9%, 98.9% and 98.1% are achieved (for the source, training-target and test-target datasets respectively). This observation highlights the fact that, although the MMD distance has failed as a criterion for selecting the correct mapping, due to the invalidation of the assumptions in Algorithm 1, in theory another metric may work. The problem with using an alternative metric is that the target domain is unlabelled, and therefore very little is actually known about the structure of the target domain data beyond basic geometric assumptions – as assumed by Algorithm 1.

It is noted at this point that the labelling strategies utilised for the previous examples have all assumed that the most important information for consistently labelling the two topologically different structures is to begin from the ground boundary condition upwards. An alternative labelling strategy is also possible by taking a different view of the physics. This strategy assumes that damage located between the ground and first storey will be the most distinguishable cluster when using natural frequencies as a feature. In addition, the free end boundary condition on the top storey is consistent between any  $n$  and  $m$  storey shear structure. This observation means that a strategy that labels the damage locations from the free end downwards is equally valid, where the extra label will be given to damage between the ground and first floor in the target



**Fig. 17.** Panel (a): Subset of normalised first and second natural frequencies for source (left) and target (right) domain (every fifth data point) for the two-storey to three-storey case study: labelling from the fixed end. Panel (b): MMD distances from JDA mappings where one cluster has been removed from the target domain.

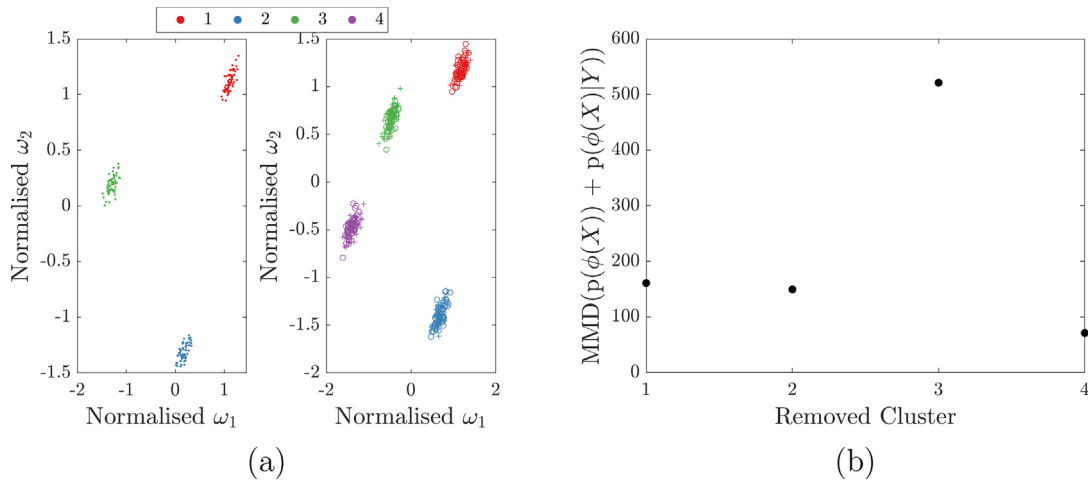


**Fig. 18.** GMM results trained on the JDA mapping when cluster ‘4’ is removed. The colours represent the predicted labels and the symbols represent the true class; where ( $\square$ ) is label ‘1’, ( $\triangleleft$ ) is label ‘2’, ( $\triangle$ ) is label ‘3’, ( $\blacklozenge$ ) is label ‘4’.

structure. This poses a challenge in determining which labelling strategy to use more generally for heterogeneous populations. The following examples consider the change in effectiveness of Algorithm 1 for the alternative labelling strategy, where damage locations are numbered from the free end downwards.

3.4.5. Two-storey to three-storey: labelling from the free end

An alternative labelling strategy, where the damage localisation labels begin from the free end downwards, is investigated for the two- (source) to three- (target) storey scenario. In this example the properties of the structure are the same as in Section 3.4.4, where the first two damped natural frequencies are used as features, with the same number of data points as Section 3.4.4. The only difference in this example, is that the label spaces  $\mathcal{Y}_s \in \{1, 2, 3\}$  and  $\mathcal{Y}_t \in \{1, 2, 3, 4\}$  are assigned from the free end downwards, i.e. undamaged class is ‘1’ and labels ‘2’ to ‘3’ refer to damage from the free end downwards, i.e. ‘2’ is damage at the top storey, where ‘4’ refers to damage located between the ground and first floor for the target structure. Fig. 19a presents the label space where the geometry of the clusters is now more consistent between the source and



**Fig. 19.** Panel (a): Subset of normalised first and second natural frequencies for source (left) and target (right) domain (every fifth data point) for the two-storey to three-storey case study: labelling from the free end. Panel (b): MMD distances from JDA mappings where one cluster has been removed from the target domain.

target domains, meaning Algorithm 1 should become applicable. Fig. 19b demonstrates that this is the case, with the smallest MMD distance coming from the scenario where the ‘4’ cluster is removed from the target domain (where the JDA mappings have been inferred using a linear kernel, where  $k = 1$  and  $\mu = 0.5$ ). In fact, it is interesting to note that Fig. 19b is a reordering of the distances from Fig. 17b, which is expected as they refer to the same JDA mappings but with the labels changed.

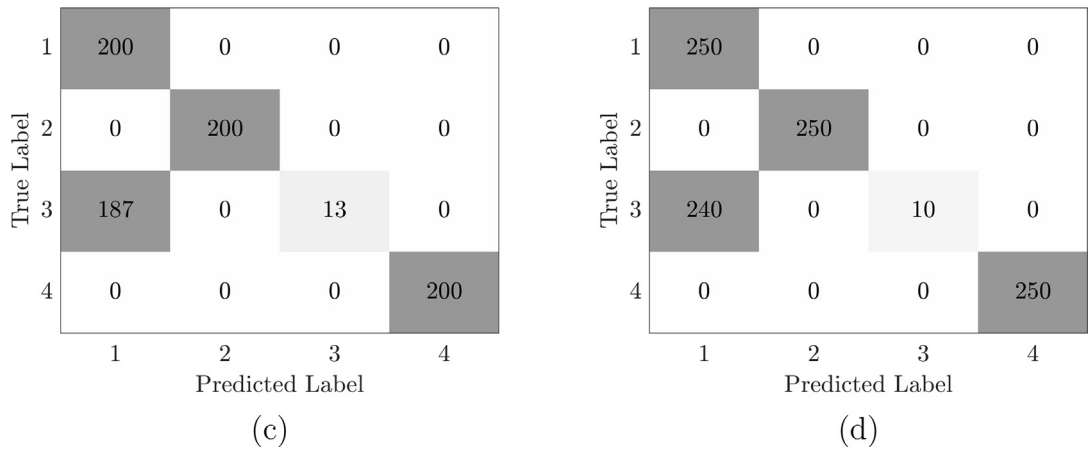
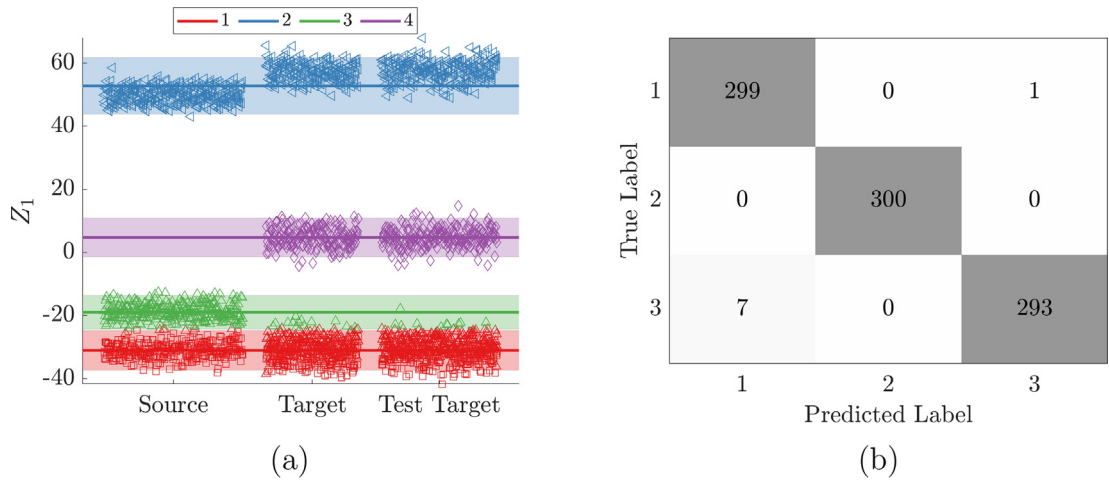
The JDA mapping – inferred when cluster ‘4’ is removed from the target domain – is shown in Fig. 20a, where a GMM has been trained on the transformed space. Fig. 20a illustrates that although positive transfer has occurred (the correct  $L$ -clusters are ‘close’ to each other), the mapping has not placed cluster ‘3’ from the source and target domains close enough together in the transformed space to make it separable from the undamaged cluster (labelled ‘1’) in the target domain. This is partly due to JDA being a dimensionality reduction tool, meaning the space reduces from two to one dimension, where Gaussian distributions in one-dimensional space will often be difficult to separate. For this example, the source domain has remained relatively separable in the transformed space, leading to a high classification accuracy of 99.1%, with relatively little confusion as shown in Fig. 20b. However, the transformed target domain is less separable, leading to classification accuracies of 76.6% and 76.0% (for the training-target and test-target datasets respectively), where confusion has occurred by labelling most of class ‘3’ as class ‘1’, as demonstrated in Fig. 20c and 20d. This example shows that, although the alternative labelling strategy does result in the correct combination of target clusters being used for transfer, and that positive transfer does occur, the mapping does not provide enough separability for classification. It is noted that class separability could be improved with additional machine learning tools. Alternatively, given a particular condition on separability of the clusters for learning, training sets could be drawn from the lower bound on damage extents for each location, which may potentially aid separability in low damage examples; however, the problem of class separability is not pursued here as it is not the focus of this paper.

#### 3.4.6. Three-storey to four-storey: labelling from the free end

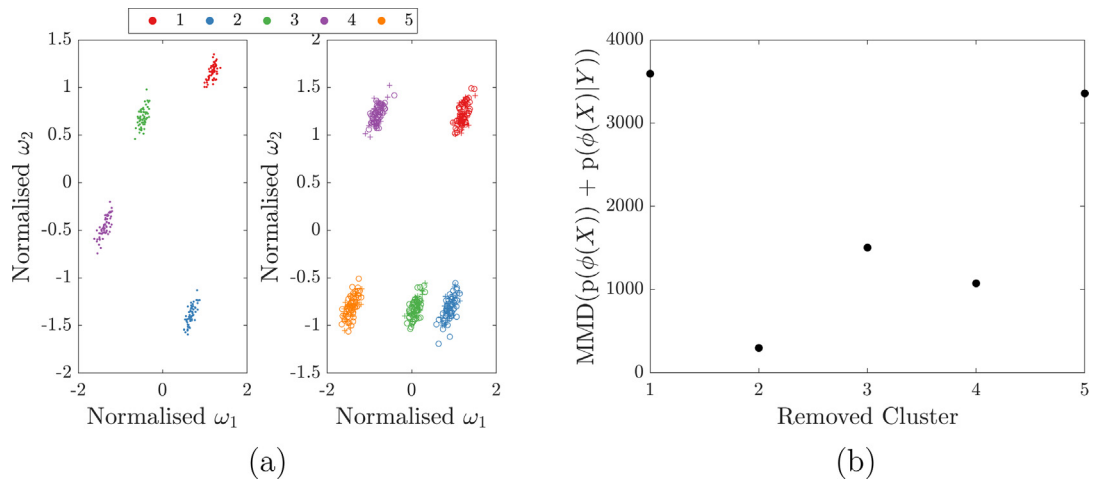
In order to assess the effectiveness of the alternative labelling strategy the same approach as in Section 3.4.5 was applied to the three- (source) to four- (target) storey scenario from Section 3.4.2, i.e. label ‘1’ is undamaged and label ‘2’ corresponds to damage at the top storey etc. The source and target datasets are depicted in Fig. 21a. Here the geometry of the  $L$ -clusters changes between the source and target domain, invalidating the assumption in Algorithm 1. This results in the algorithm producing a minimum MMD distance for the JDA mapping (using a linear kernel, where  $k = 2$  and  $\mu = 0.5$ ) inferred from the incorrect set of target clusters, as shown in Fig. 21b (where again Fig. 21b is a reordering of Fig. 13b). As expected even when the correct  $L$ -clusters are selected from the target domain, negative transfer occurs. This example provides additional evidence that the algorithm’s assumptions are too strong for the general  $n$ - to  $m$ -storey shear structure problem. It also highlights that the labelling strategy is important in creating a successful JDA mapping, and that labelling from the free and fixed end both produce problems as physical correspondence of localisation labels is difficult to definitely determine.

#### 3.4.7. Four-storey to five-storey: labelling from the free end

The last example in Section 3.4 utilises the free end labelling strategy on the four- (source) to five- (target) storey ( $L + 1$ )-problem from Section 3.4.3; labelling from the free end downwards i.e. label ‘1’ is undamaged and label ‘2’ corresponds to damage at the top storey etc., where the target domain has an extra label ‘6’ that corresponds to damage between the ground



**Fig. 20.** Panel (a): GMM results trained on the JDA mapping where the colours represent the predicted labels and the symbols represent the true class; where ( $\square$ ) is label '1', ( $\circ$ ) is label '2', ( $\Delta$ ) is label '3', ( $\blacklozenge$ ) is label '4'. Panel (b), (c) and (d) are the resulting confusion matrices for the source (b), target (c) and test target (d) datasets.



**Fig. 21.** Panel (a): Subset of normalised first and second natural frequencies for source (left) and target (right) domain (every fifth data point) for the three-storey to four-storey case study; labelling from the free end. Panel (b): MMD distances from JDA mappings where one cluster has been removed from the target domain.

and first floor. The label space is shown in Fig. 22a where once more the geometry of the  $L$ -clusters in the source and target domain invalidate the assumption in Algorithm 1. This results in Fig. 22b, showing that the incorrect target cluster combination is selected by Algorithm 1 (where the JDA mapping uses a linear kernel, where  $k = 2$  and  $\mu = 0.5$ ). As with the example in Section 3.4.4, it is interesting to note that if the correct combination of clusters from the target domain is selected, positive transfer occurs, as demonstrated in Fig. 23b, when compared to the negative transfer that occurs in Fig. 23a; mainly in source cluster '5' being paired with target cluster '6'. This reiterates the argument that the MMD distance may not be the best criterion in selecting the 'optimal' mapping. However, due to the lack of information about the target domain (i.e. it is unlabelled) it is difficult to go beyond a distance criterion and the assumptions in Algorithm 1 without incorporating more physical knowledge. These results therefore demonstrate the limitations of a black-box approach to transfer learning for PBSHM, and motivate the need for leveraging as much physical information as possible in performing transfer learning. Physics might be introduced in the form of prior constraints on the mapping, better manifold assumptions on the data for each domain elicited from the physics, or could involve utilising the physics to map the data onto a space where the geometric properties of the clusters are consistent (making Algorithm 1 applicable after that stage).

### 3.5. Heterogeneous populations: from physics-based models to experimental structures

The final case study considers a population in which one member is formed from a numerical model and the other an experimental structure. This highlights the potential within PBSHM to utilise computer models in labelling other members of a population, potentially overcoming a significant SHM challenge, namely that damage-state data often have limited availability, as it is not feasible nor economically viable to damage a structure in various manners in order to collect data for a machine learner. In this case study the labelled source domain is generated from a numerical simulation of a three-storey shear structure (in the same manner as the previous case studies) where the target domain is an unlabelled experimental structure, as depicted in Fig. 24.

The properties of the numerical model are presented in Table 9, where the dimensions are those measured from the structure, and the material properties and damping coefficients are estimated, given that the structure is made from aluminium 6082. The SHM problem  $\mathcal{SP}$  is an extent problem<sup>6</sup> where a midpoint saw cut (modelled as an open crack) is applied to a single beam between the ground and first storey in 5 mm increments from 0 mm to 20 mm, meaning label '1' corresponds to undamaged, '2' to a 5 mm saw cut, '3' to a 10 mm saw cut etc. The numerical model is deliberately oversimplified such that the case study demonstrates the effectiveness of domain adaptation in utilising physics-based models, which are often challenging to validate and may not fully agree with observational data due to model form-errors, in labelling real world structures. Consequently, the population of the numerical model and experimental structure can be thought of as a heterogeneous population. The feature spaces represent the first three bending natural frequencies meaning  $X_s \in \mathbb{R}^{N_s \times 3}$  and  $X_t \in \mathbb{R}^{N_t \times 3}$ . The number of data points in each class for the simulation were 250, meaning  $N_s = 1250$ . The target domain data were obtained from modal testing using an electrodynamic shaker attached to the first floor of the structure, where the acceleration response was measured from three uni-axial accelerometers attached to each floor. For each damage extent, the structure was excited five times with a 6553.6 Hz broadband white-noise excitation containing 16384 spectral lines (0.2 Hz resolution) with a Hanning window on the force excitation and acceleration response. The size of the complete target dataset was therefore  $N_t = 25$ . The feature spaces for the source and target domains are presented in Fig. 25, where the numerical model's natural frequencies are around half those of the structure, highlighting the need for transfer learning. JDA was applied to the source and target feature data, where the inferred mapping and classification results are shown in Fig. 26 (where  $k = 2$  and  $\mu = 1$  identified from fivefold cross validation). A  $k$ NN classifier (where  $k = 4$ ) trained on the source domain, was utilised in classifying the target domain in Fig. 26. The improvement on not performing JDA (i.e. training a  $k$ NN on the source domain and applying it to the target domain) is shown in Table 10, where accuracy increases from 48.0% to 88.0%. This highlights the effectiveness of utilising both physics-based models and observational datasets within a PBSHM framework.

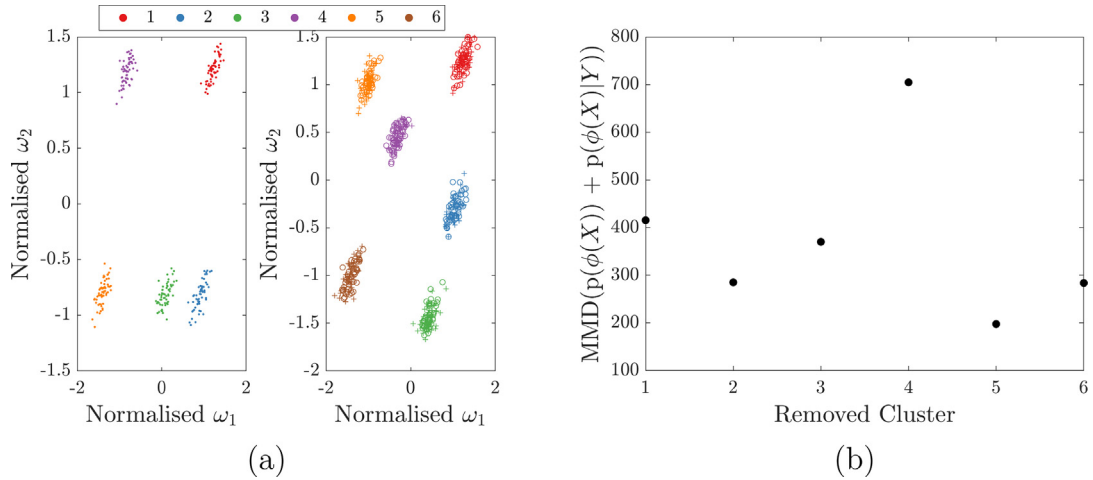
## 4. Discussion and conclusions

Knowledge transfer and mapping are important processes in developing a population-based approach to SHM. The idea of mapping data from different structures within a population such that knowledge about health states is transferred, can be achieved via transfer learning. It is important when applying transfer learning to determine what similarities exist between structures within a population, and what information should be transferred, such that negative transfer is avoided. With this aim in mind, it is helpful to categorise structures based on their differences. Here two main types of population within PBSHM have been considered: *homogenous populations* and *heterogeneous populations*. Further to these two main categories, three more subdivisions have been discussed for heterogeneous populations: *geometry*, *material* and *topology*. These sources of dissimilarity form part of an Irreducible Element and Attributed Graph representation of structures [2].

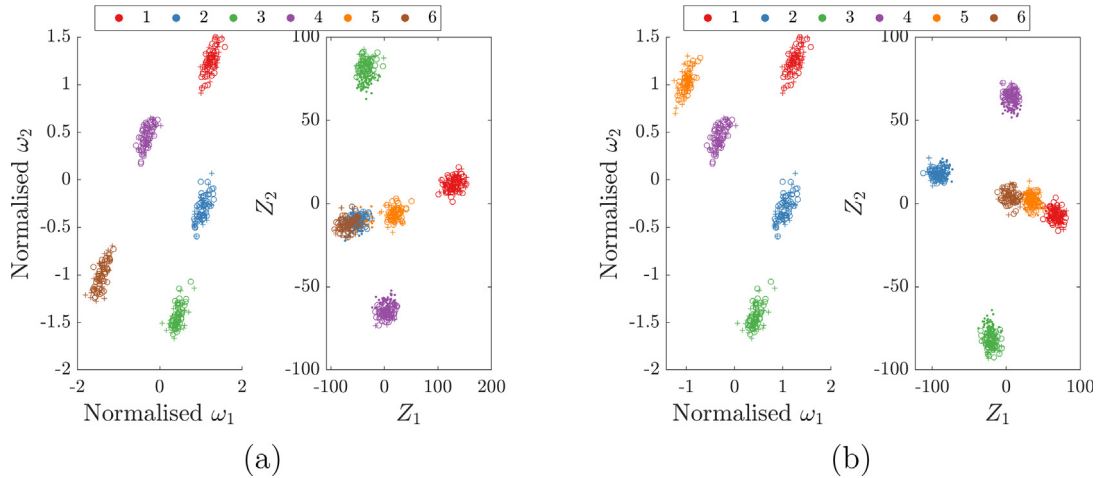
By starting from what type of SHM problem is required, it is possible to categorise the level of consistency that exists for each type of population. This helps define what assumptions are required for a given transfer learning method, and whether

<sup>6</sup> It is noted that typically extent problems should be posed as regression problems, rather than classification problems. The extent problem here is solved as a classification problem in order to demonstrate transfer learning in PBSHM and not intended to indicate best practice.





**Fig. 22.** Panel (a): Subset of normalised first and second natural frequencies for source (left) and target (right) domain (every fifth data point) for the four-storey to five-storey case study: labelling from the free end. Panel (b): MMD distances from JDA mappings where one cluster has been removed from the target domain.



**Fig. 23.** Examples of JDA mapping for the four-storey to five-storey case study: labelling from the free end. Panel (a): Subset of normalised first and second natural frequencies for target domain (every fifth data point) where cluster ‘3’ is removed (left) and the first two components of the JDA mapping (right). Panel (b): Subset of normalised first and second natural frequencies for target domain (every fifth data point) where cluster ‘6’ is removed (left) and the first two components of the JDA mapping (right).

homogenous or heterogeneous transfer learning should be implemented. Furthermore, label inconsistency between members of the population will increase the likelihood of negative transfer. This issue shows the power of an Irreducible Element (IE) and Attributed Graph (AG) representation of structures, as this approach means that similarities between members of the population can be extracted, reducing the transfer learning problem to one in which labels are considered consistent.

Transfer learning, in the form of domain adaptation, has been demonstrated to be applicable for problems when both feature and label spaces are consistent. By defining the mapping problem within an SHM and population type context, several PBSHM scenarios can be considered to fulfil these requirements. This paper has demonstrated that domain adaptation is applicable for both a homogeneous population, and heterogeneous population context; where the geometric differences are changes in dimensions and two different metallic materials are used.

When label spaces are considered inconsistent, such as damage localisation problems in heterogeneous populations where topological differences occur, the risk of negative transfer increases, posing a significant challenge. This paper has outlined particular types of label inconsistency, namely an  $(L + N)$ - and  $(L + 1)$ -problem; where the source label space is a subset of the target label space with  $N$  or 1 extra labels in the target label space respectively. For the  $(L + 1)$ -problem a novel MMD-based algorithm has been proposed, based on an assumption of geometric similarities in the source and target datasets. This approach has been demonstrated on several PBSHM examples involving a population of  $n$ - and  $m$ -storey structures. These



Fig. 24. Target domain structure: Experimental setup of three-storey building structure [4].

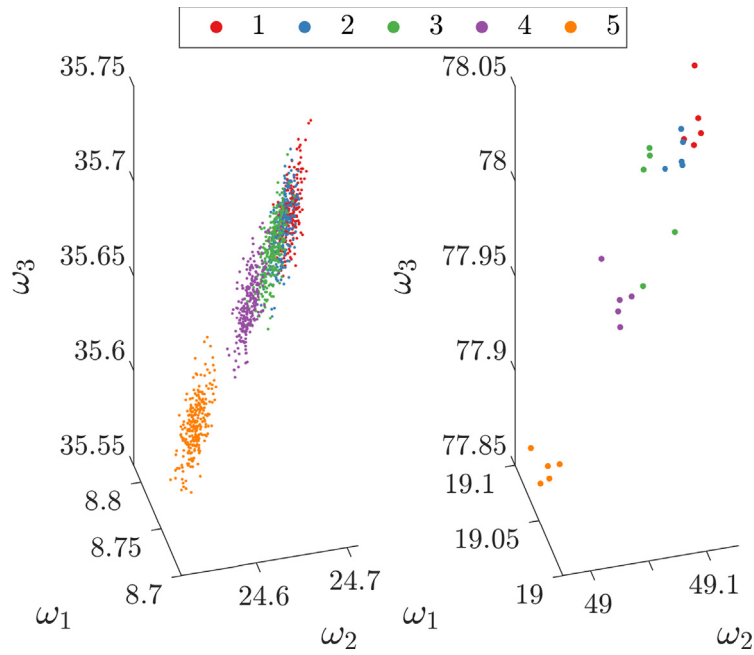
examples have demonstrated the applicability of the novel approach in cases where the geometric similarities between the source and target datasets occur. Unfortunately, outside of specific cases within a shear structure context, this assumption breaks down, meaning a more complex approach is required. The examples shown in this paper motivate the need to go beyond a black-box approach, as without more knowledge of the target space, domain adaptation is limited to assumptions like that proposed in the novel approach. As a consequence, future research will investigate the potential for leveraging physical knowledge of structures in mapping the observational feature sets from the source and target domains onto a consistent manifold, where domain adaptation then becomes applicable. Without incorporating physical knowledge, there will always be a danger of black-box domain adaptation performing negative transfer. In addition, these case studies highlighted that labelling strategies will effect the outcome of any knowledge transfer. These issues should be pursued as part of further research, such that any PBSHM approach is robust to the labelling strategy employed.

The final case study presented the potential for utilising physics-based models as part of a PBSHM dataset. Specifically, an example has been shown where an unvalidated numerical model has been used to label an experimental structure. This

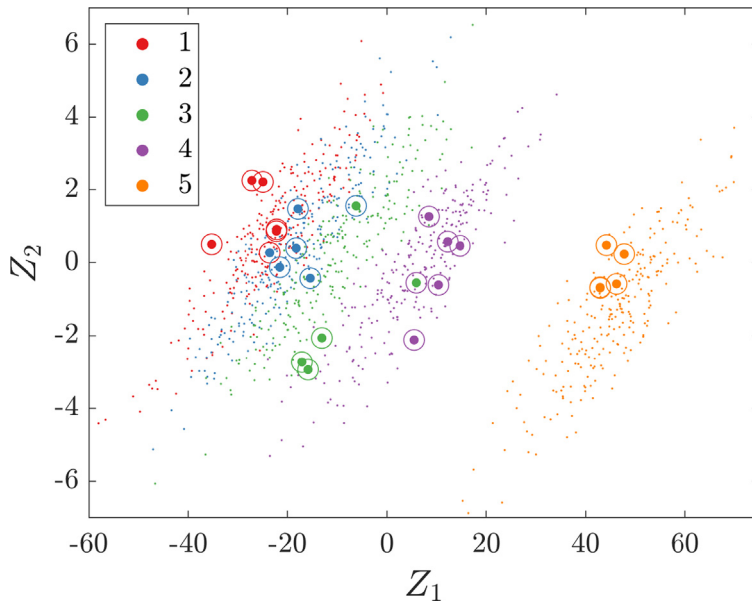
**Table 9**

Properties of the source structure for the heterogeneous population case study mapping from a physics-based model to experimental structure.

Property	Unit	Source
Beam geometry, $\{l_b, w_b, t_b\}$	mm	$\{177.8, 25.4, 6.35\}$
Mass geometry, $\{l_m, w_m, t_m\}$	mm	$\{304.8, 254.0, 25.4\}$
Crack geometry, $\{l_{cr}, l_{loc}\}$	mm	$\{17.5, 88.9\}$
Elastic modulus, $E$	GPa	$\mathcal{N}(71, 1 \times 10^{-9})$
Density, $\rho$	kg/m <sup>3</sup>	$\mathcal{N}(2700, 50)$
Damping coefficient, $c$	Ns/m	$\mathcal{G}(9, 0.5)$



**Fig. 25.** Source (left) and target (right) domain features for the heterogeneous population case study mapping from a physics-based model to experimental structure (where damped natural frequencies are in Hz).



**Fig. 26.** Transfer components and classification results for the source and target domains in the heterogeneous population case study mapping from a physics-based model to experimental structure. The labelled source domain transfer components are denoted by (•), the target-domain transfer components are denoted by (•) for the true label and (○) for the predicted label respectively.

result means that physics-based models can be leveraged in providing label knowledge not currently obtainable from real world structures.

Finally, critical foundations of population-based SHM, namely in the use of transfer learning in performing knowledge transfer between homogeneous and heterogeneous populations, have been defined. These foundations, show that for a population-based approach to SHM, understanding the required SHM problem is crucial for determining the level of structural similarity required and in selecting the most effective transfer learning technology. The integration of an IE and AG-

**Table 10**

Classification results for the heterogeneous population case study mapping from a physics-based model to experimental structure trained on the labelled source domain and applied to the unlabelled target domain.

	Method	kNN	JDA
Target	Accuracy	48.0%	88.0%

based approach to quantifying knowledge about structures is therefore vital in understanding the effectiveness of transfer learning in a PBSHM setting.

### CRedit authorship contribution statement

**P. Gardner:** Conceptualization, Formal analysis, Investigation, Methodology, Software. **L.A. Bull:** Methodology, Writing - review & editing. **J. Gosliga:** Methodology, Writing - review & editing. **N. Dervilis:** Methodology, Writing - review & editing, Supervision. **K. Worden:** Conceptualization, Methodology, Funding acquisition, Writing - review & editing, Supervision.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

The authors would like to acknowledge the support of the UK Engineering and Physical Sciences Research Council via grants EP/R006768/1, EP/R003645/1 and EP/R004900/1.

### References

- [1] L.A. Bull, P. Gardner, J. Gosliga, N. Dervilis, E. Papatheou, A.E. Maguire, C. Campos, T.J. Rogers, E.J. Cross, and K. Worden. Foundations of population-based structural health monitoring, Part I: Homogeneous populations and forms. *Mech. Syst. Signal Process.*, Submitted..
- [2] J. Gosliga, P. Gardner, L.A. Bull, N. Dervilis, and K. Worden. Foundations of population-based structural health monitoring, Part II: Heterogeneous populations and structures as graphs, networks, and communities. *Mech. Syst. Signal Process.*, Submitted..
- [3] E. Papatheou, N. Dervilis, A.E. Maguire, I. Antoniadou, and K. Worden. Population-based SHM: a case study on an offshore wind farm, in *Proceeding of the 10<sup>th</sup> International Workshop on Structural Health Monitoring*, 2015..
- [4] P. Gardner, X. Liu, and K. Worden. On the application of domain adaptation in structural health monitoring. *Mech. Syst. Signal Process.*, 2019, Accepted..
- [5] S.J. Pan, Q. Yang, A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22 (2010) 1345–1359.
- [6] Y. Zhang, Q. Yang, An overview of multi-task learning. *Natl. Sci. Rev.* 5 (2018) 30–43.
- [7] O. Day, T.M. Khoshgoftaar, A survey on heterogeneous transfer learning. *J. Big Data* 4 (2017) 29.
- [8] K. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning. *J. Big Data* 3 (2017) 29.
- [9] S. Dorafshan, R.J. Thomas, M. Maguire, Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. *Constr. Build. Mater.* 186 (2018) 1031–1045.
- [10] Y. Gao, M. Yuqing, K.M. Mosalam, Deep transfer learning for image-based structural damage recognition. *Comput.-Aided Civil Infrastr. Eng.* 33 (9) (2018) 748–768.
- [11] K. Jang, N. Kim, Y. An, Deep learning-based autonomous concrete crack evaluation through hybrid image scanning. *Struct. Health Monit.* (2019).
- [12] J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer, N.D. Lawrence, *Dataset Shift in Machine Learning*. MIT Press, 2009.
- [13] S.J. Pan, I.W. Tsang, J.T. Kwok, Q. Yang, Domain adaptation via transfer component analysis. *IEEE Trans. Neural Networks* 22 (2011) 199–210.
- [14] L. Duan, D. Xu, I.W. Tsang, Learning with augmented features for heterogeneous domain adaptation, in: *Proceedings of the 29<sup>th</sup> International Conference on Machine Learning, ICML, 2012*.
- [15] M. Long, J. Wang, G. Ding, J. Sun, P.S. Yu, Transfer feature learning with joint distribution adaptation, in: *In 2013 IEEE International Conference on Computer Vision, 2013*, pp. 2200–2207.
- [16] W. Li, L. Duan, D. Xu, I.W. Tsang, Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* 36 (2014) 1134–1148.
- [17] M. Long, J. Wang, G. Ding, S.J. Pan, P.S. Yu, Adaptation regularization: a general framework for transfer learning. *IEEE Trans. Knowl. Data Eng.* 26 (2014) 1076–1089.
- [18] E. Kodirov, T. Xiang, Z. Fu, S. Gong, Unsupervised domain adaptation for zero-shot learning, in: *In 2015 IEEE International Conference on Computer Vision (ICCV), 2015*, pp. 2452–2460.
- [19] T.R. Gruber, A translation approach to portable ontology specifications. *Knowledge Acquisition* 5 (1993) 199–220.
- [20] S. Anderlik, R. Stumpfner, B. Freudenthaler, M. Fritz, A proposal for ontology-based integration of heterogeneous decision support systems for structural health monitoring, in: *In Proceedings of the 12th International Conference on Information Integration and Web-Based Applications and Services, 2010*, pp. 168–175.
- [21] G. Tsialiamanis, D.J. Wagg, I. Antoniadou, K. Worden, An ontological approach to structural health monitoring. *Proceeding of IMAC XXXVIII* (2020).
- [22] R. Li, T. Mo, J. Yang, S. Jiang, T. Li, Y. Liu, Ontologies-based domain knowledge modeling and heterogeneous sensor data integration for bridge health monitoring systems, in: *IEEE Trans. Industr. Inf.*, 2020, p. 1.
- [23] M. Nickel, K. Murphy, V. Tresp, E. Gabrielovich, A review of relational machine learning for knowledge graphs. *Proc. IEEE* 104 (1) (2016) 11–33.
- [24] T. Hamaguchi, H. Oiwa, M. Shimbo, Y. Matsumoto, Knowledge transfer for out-of-knowledge-base entities: a graph neural network approach, in: *In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017*, pp. 1802–1808.
- [25] A. Rytter. *Vibrational Based Inspection of Civil Engineering Structures* (Ph.D. thesis), Aalborg University, Denmark, 1993..
- [26] C.R. Farrar, N.A.J. Lieven, Damage prognosis: the future of structural health monitoring. *Philos. Trans. R. Soc. A: Math., Phys. Eng. Sci.* 365 (2007) 623–632.

- [27] O. Chapelle, B. Schölkopf, A. Zien, *Semi-Supervised Learning*, MIT Press, 2006..
- [28] H. Sohn, Effects of environmental and operational variability on structural health monitoring, *Philos. Trans. R. Soc. A* 365 (2007) 539–560.
- [29] M.T. Rosenstein, Z. Marx, L.P. Kaelbling, T.G. Dietterich, To transfer or not to transfer, *NIPS 2005 Workshop on Transfer Learning* (2005).
- [30] E. Eaton, M. Desjardins, T. Lane, Modeling transfer relationships between learning tasks for improved inductive transfer, *Lect. Notes Comput. Sci.* (2008).
- [31] L. Ge, J. Gao, H. Ngo, K. Li, A. Zhang, On handling negative transfer and imbalanced distributions in multiple source transfer learning, in: *In Proceedings of the 2013 SIAM International Conference on Data Mining SDM*, 2013, 2013..
- [32] Z. Wang, Z. Dai, B. Póczos, J. Carbonell, Characterizing and avoiding negative transfer, in: *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11293–11302.
- [33] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf, A. Smola, A kernel two-sample test, *J. Mach. Learn. Res.* 13 (2012) 723–773.
- [34] B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Comput.* 10 (1998) 1299–1319.
- [35] S. Christides, A.D.S. Barr, One-dimensional theory of cracked Bernoulli-Euler beams, *Int. J. Mech. Sci.* (1984).
- [36] K.P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.