
Article:
Gardner, P. orcid.org/0000-0002-1882-9728, Bull, L.A., Gosliga, J. et al. (3 more authors) (2022) A population-based SHM methodology for heterogeneous structures : transferring damage localisation knowledge between different aircraft wings. Mechanical Systems and Signal Processing, 172. 108918. ISSN 0888-3270

https://doi.org/10.1016/j.ymssp.2022.108918

A population-based SHM methodology for heterogeneous structures: transferring damage localisation knowledge between different aircraft wings

P. Gardner[1], L.A. Bull[2], J. Gosliga[1], J. Poole[1], N. Dervilis[1], K. Worden[1]

[1] Dynamics Research Group, Department of Mechanical Engineering
University of Sheffield, Mappin Street, Sheffield S1 3JD, UK

[2] The Alan Turing Institute, The British Library
London, NW1 2DB, UK

Abstract

Population-based structural health monitoring (PBSHM) offers a new viewpoint for structural health monitoring (SHM), allowing diagnostic information to be shared across populations of structures. By extending the set of available damage observations, a population-based approach can diagnose damage previously unseen on a structure of interest by leveraging damage information from other structures in the population. These technologies therefore provide significant benefits for making SHM practicable in a variety of industrial settings. It is proposed that PBSHM methodologies must be comprised of tools to *assess similarity* coupled with algorithms that perform *knowledge transfer*. The similarity tools are important for identifying whether an SHM task should be attempted for a given population by assessing both the structural similarities between members of a population and similarities in their data spaces. An abstract representation of structures in a graphical domain is presented as an objective way of assessing structural similarity, with distance metrics utilised for assessing data-space similarities. Knowledge transfer is performed using a branch of *transfer learning* called *domain adaptation*. By determining if members of a population are similar in a structural and data-space sense, the risk of *negative transfer* can be reduced; whereby domain adaptation reduces classification performance. This paper demonstrates a PBSHM methodology for transferring knowledge within a heterogeneous population (a group of non-identical structures). Specifically, the PBSHM methodology is shown to transfer localisation labels from a Gnat aircraft wing to an unlabelled Piper Tomahawk aircraft wing dataset, resulting in 100% classification accuracy.

Keywords: Population-based structural health monitoring, irreducible elements, attributed graphs, transfer learning, domain adaptation

1. Introduction

Obtaining damage state data from a structure of interest, prior to the start of a monitoring campaign, is often infeasible because of economic and safety concerns. This issue means that classification-based approaches to data-based structural health monitoring (SHM) are not practicable in the majority of industrial scenarios, limiting data-based methods to novelty detection. Population-based structural health monitoring (PBSHM), seeks to overcome this challenge by considering the available data from across a population of structures, which may include both physical and numerical systems

[1]. By expanding the available data in this way, observations of damage from previous monitoring campaigns on similar structures or from model-based simulations, can be utilised in diagnosing and classifying damage on a new structure of interest from the start of its monitoring campaign.

It is important to note that one cannot simply train a classifier using data from one structure and hope the classifier will generalise to data from another. This is because no two structures are ever structurally exactly identical (at least as a result of manufacturing tolerances), nor will they operate in exactly the same environmental conditions (even if located close to each other). These differences mean that the same feature data from any two structures will not be generated from the same underlying distribution, breaking the assumption of machine learning classifiers that training and testing data are generated from the same underlying distribution. This issue becomes worse when a heterogeneous population is considered (i.e. a population of non-identical structures), as the feature and label spaces may not even be consistent, let alone generated from the same underlying distribution. To overcome these issues, transfer learning in the form of domain adaptation is utilised, mapping labelled *source* datasets onto unlabelled *target* datasets, such that their distributions are harmonised [1].

Although there is a trend to move towards using more extensive sensor networks on operational aircraft for the purpose of health management, there is yet to be wide uptake in industry for various reasons. In addition, even if large quantities of data can be collected from in-service aircraft, much of that data will still relate to the aircraft operating in its normal, healthy condition. This issue means that very few observations from these large datasets will be associated with damage or structural failure, as aircraft are designed and regulated to high safety tolerances and reliability standards. As a result, observations of damage patterns will still be rare, meaning that a population-based view of SHM will still be required even into the future.

This paper proposes an overarching methodology for performing PBSHM on heterogeneous populations, with the ideas based on the foundations of PBSHM stated in a series of four papers [2–5]. Specifically, the proposed approach is demonstrated to perform damage localisation (of inspection panel removals) on a Piper Tomahawk aircraft wing using damage-state data from a Gnat aircraft wing. An abstract representation of structures in a graphical domain is utilised to determine where the two structures are structurally similar by identifying maximum common subgraphs. These maximum common subgraphs form potential candidate dataset pairings for the Gnat and Piper Tomahawk. Distance metrics are analysed for each of these pairings, identifying which is most similar from a data-space perspective, and least likely to lead to negative transfer. Domain adaptation is performed, mapping the Gnat and Piper Tomahawk datasets into a latent space where a classifier trained on the (labelled) Gnat dataset is shown to generalise and classify the (unlabelled) Piper Tomahawk dataset with 100% classification accuracy. A MATLAB implementation accompanies this paper: `https://github.com/pagard/EngineeringTransferLearning`.

1.1. Related work

PBSHM has also been termed *fleet-based monitoring* in the literature. For example, Fink *et al.* outline in [6], the challenges and future directions for deep learning in prognostics and health management (PHM), stating that fleet PHM is a potential solution to issues in applying deep learning for health management. They argue that transfer learning, in the form of domain adaptation, will be a key component for performing fleet PHM; however, this paper does not propose any particular methodologies itself.

Within fleet-based monitoring, there are approaches that are developed solely around a model or digital twin. For example. Zaccaria *et al.* perform fleet monitoring on a population of simulated

aircraft engines [7]. Their method utilises an adaptative physics-based model (as a digital twin) to simulate faults that can be compared with flight data. Fleet data are normalised based on the engine model's baseline normal condition, and novelty detection is performed. Given these results, fault identification is conducted by comparing the correlations of the fleet data signals to a digital twin over the average engine. Zaccaria *et al.* expanded this idea by using the physics-based model to train a Bayesian network classifier to diagnose fleets of gas turbines [8]. In a similar approach to [7], Moksadur *et al.* define a multi-level method for gas-turbine fleet monitoring, utilising nominal and updated physics-based models to generate baseline datasets for the population [9]. Although promising, these approaches are not experimentally validated, and may struggle to generalise when operational data are presented. The methods also differ to that proposed in this paper, as they rely on some baseline model, which must be valid for the complete population.

In contrast to using a model as a baseline for fleet monitoring, Moens *et al.* generate data from a 'smart maintenance living lab' of seven machines, which are used to train a more robust machine learner [10]. This is similar to the idea of a *form*, a machine learning model that captures a population's ideal response as well as its variations [2], and hence is more robust when making population-based inferences. Basora *et al.* develop a semi-supervised anomaly detection technique for aircraft fleets, demonstrated on real sensor data from a cooling-unit system on a modern wide-body aircraft from a major European airline [11]. At its core, the technique trains several deep neural network autoencoders on baseline data in order to perform novelty detection. A similar method by Michau *et al.* proposes a hierarchical extreme learning machine, constructed from a stack of autoencoders, as a way of determining similarity between datasets in order perform anomaly detection at a fleet level [12]. All of these methods differ from the approach proposed in this paper in that they only perform novelty detection, and all rely on some baseline that captures the normal condition of the population.

Some authors have pointed out that links and similarities between members of a population need to be captured as part of a fleet-based approach. Medina-Oliva *et al.* take this viewpoint, developing an ontological approach to fleet-level monitoring, helping to link and manage relevant knowledge about a population — particularly useful for a heterogeneous population [13]. Unlike the approach proposed here, the authors do not use the ontology to perform any inferences, and instead suggest how it might be used. Similarly, [3] outlines the abstract representation of structures as graphs but does not demonstrate the process of using the graphical space to aid inferences, as demonstrated in this paper.

Ideas of structural similarity have been adapted into algorithms for population-based inferences. For example, Hendrickx *et al.* propose an anomaly detection method for performing condition monitoring over a population of machines, demonstrated on a population of ten experimental electrical drivetrains [14]. Within their methodology, similar machines are grouped by hierarchically clustering the Euclidean distance or dynamic time-warping measure between any two machine's datasets in the fleet. The method expects that large groups of structures represent machines operating in their normal condition, with an anomaly score assigned to each machine based on the fraction of the total machines not in its cluster. The method proposed in this paper, is formed from similar ideas; however, the similarity between two structures' datasets are determined from a statistical-distance point of view. Physics-Informed Multi-source Domain Adversarial Networks (PhyMDAN) is another method that seeks to use structural similarity; however, in this case the method creates weights over datasets based on their structural attributes that are used to weight their importance in a deep learning-based domain adaptation approach [15]. The authors state that any structural attribute could be used to form the weight, and do not develop a principal to guide

other applications.

Other related literature involves the topic of transfer learning, and specifically domain adaptation [1, 16–22]. A large portion of the literature utilises domain adaptation in order to overcome changing loading conditions in condition monitoring [18–20, 22]. However, domain adaptation has also been utilised in a population-based setting [1, 4, 21, 23]. Two notable examples are Gardner *et al.*, who demonstrated the role of domain adaptation in transferring localisation labels between numerical and experimental shear-structures [1], and Bull *et al.*, where damage detectors were transferred between a population of six experimental tailplanes [23]. Neither of these approaches seek to develop methods for avoiding negative transfer by incorporating measures of similarity, as proposed in this paper.

The outline of this paper is as follows. Section 2 provides an overview of population-based structural health monitoring, outlining the key principles of assessing *similarity* between members of a population and *transferring* knowledge between them. The case study is introduced in Section 3, where the SHM task is to transfer knowledge about localisation of damage from a Gnat aircraft wing to a Piper Tomahawk aircraft wing (with the two wings forming a heterogeneous population). The proposed methodology is presented in Section 3.2, whereby localisation knowledge is successfully transferred, allowing damage to be localised on the Piper Tomahawk wing, with verification of the approach performed in Section 3.3. Lastly, conclusions are outlined in Section 4.

2. Population-based structural health monitoring

Population-based structural health monitoring (PBSHM) seeks to improve diagnostic capabilities, both at a population and/or individual level, by harnessing measured data from across individuals in a population [2–5]. By combining datasets from the complete population (which can include numerical simulations as well as real world datasets), a PBSHM methodology is able to expand the available observations of health-states of interest. As a result, a PBSHM approach can produce robust diagnostic predictions on structures in the population where no labelled data were available at the start of the monitoring campaign — achievable because labelled observations can be leveraged from other members of the population. The ability to diagnose health-states unobserved in training for a particular structure of interest, means that PBSHM overcomes a significant limitation of traditional data-based SHM approaches, where inferences are limited to novelty detection in the absence of labelled training data.

For datasets to be shared between members of a population, notions of *similarity* and *transferability* must be defined. The aim of any PBSHM method is to transfer knowledge between members of a population, forming some shared model that can diagnose them [4]. Before this process is performed, the engineer must determine if structures in the population, and their datasets, share similarities that can be harnessed at a population level [3, 24]. The more dissimilar any two structures are, the more at risk the PBSHM approach is of *negative transfer* [4] — a term, in a PBSHM context, used to describe the phenomenon where utilising a PBSHM approach reduces the performance when compared to an SHM approach on an individual structure. As a consequence, the engineer requires a set of PBSHM tools for determining when transfer between structures is advisable, given some notation of similarity between the structures and their datasets.

There are two population types that can be considered in a PBSHM approach. A *homogeneous population* is one where all the structures are nominally identical and are expected to exhibit similar behaviour; for example, a wind farm containing wind turbines of the same model type would be considered a homogeneous population. When members of a population cannot be classed as

nominally identical, the group are referred to as a *heterogeneous population*. There are a multitude of reasons for differences in a population, with significant factors being the topology of the structures, their material properties and their geometry [3, 4]. For example, a fleet of helicopters that were built to the same model specification, but with various custom modifications for clientele, would be classed as a heterogeneous population. Intuitively, PBSHM is generally easier to perform on homogeneous populations than heterogeneous populations, where there will be larger degrees of dissimilarity in the group of structures and their datasets. The following sections outline approaches that cover the two main aspects of a PBSHM methodology: assessing the similarities between structures in a population, and methods for transferring knowledge between structures.

2.1. Structures as graphs: assessing structural similarities

PBSHM requires the ability to assess the similarities between any two structures in a population, to determine whether knowledge transfer is advisable. As proposed by Gosliga *et al.* [3], an abstract representation of structures can be defined, meaning any structure can be converted into a graphical representation. In this graphical domain, metrics can be defined allowing the engineer to quantify the level of similarity between members of a population. More powerfully, graph-matching algorithms can be applied in this space of structures, meaning that substructures found in all members of a population can be identified. By locating these shared subsystems, it is possible to identify what knowledge can be shared in a population and what aspects of the dataset are most useful for PBSHM. Details of this abstract representation of structures are defined in [3], with the following section covering the main concepts.

An abstract representation of structures is formed in two stages. First, the engineer converts the structure into an *irreducible element* (IE) model. This process decomposes the structure into its essential structural elements, typically at the spatial resolution of the SHM problem. It is important to note that an IE model is not intended to capture every dimension and property of a structure, but only those that are significant for the PBSHM context. Elements are typically components with well-established dynamic behaviour, such as beams and plates, with some associated geometry and material properties. The *relationships* between elements describe both the physical relationships between these components e.g. bearing, bolts etc., and more abstract relationships such as perfect joints and boundary conditions. One of the main considerations when constructing an IE model is topology. The topology of the structure is determined by the physical relationships between each element in the IE model, and is particularly useful when performing localisation tasks, where corresponding topology is required between members of the population for localisation to be possible. An illustration of the IE models for the aircraft wings in this paper are shown in Figures 1 and 2 and are outlined in detail in Section 3.2.

The next stage transforms the IE model into an attributed graph (AG) such that a metric space can be defined and graph-matching algorithms utilised. Each element in the IE model becomes a node in the graph, with each relationship becoming an edge (and boundary conditions forming node-edge pairs). The AGs of the IE models in Figures 1 and 2 are presented in Figures 3 and 4 respectively, and details about their construction are provided in Section 3.2.

Once attributed graphs have been obtained for each member of the population, similarities can be quantified. Topological similarity can be measured by assessing the largest common substructure in the population; with the largest substructure for homogeneous populations being the entire structure itself. The graphical representation of structures means that substructures are represented as subgraphs (with structures represented as graphs). As a result, the problem of determining the largest common substructure between two structures can be posed as identifying the largest

common subgraph between two parent graphs. In order for the largest common subgraph to retain validity in a physical sense, the largest common subgraph must also be a connected and induced subgraph. An intuitive interpretation of these properties is that all the elements in a substructure must be connected (rather than two disconnected substructures) and that all the joints in the parent structure should appear in any substructure (for more mathematical definitions the reader is referred to [3]). The largest subgraphs that meet these conditions are defined as *maximum common subgraphs* (MCSs) and are said to be *topologically equivalent*.

One of the most widely-used methods for identifying all the connected, induced subgraphs between two parent graphs is the modified Bron-Kerbosch (BK) algorithm [3]. Once all these subgraphs are identified they can be sorted by size and the maximum common subgraph(s)[1] reported. Given the structural and physical importance of boundary conditions, the set of maximum common subgraphs can be reduced to the subset where the boundary condition nodes are aligned — this set of subgraphs are defined as *structurally equivalent*. A similarity measure can be defined over the set of maximum common subgraphs, between all pairwise combinations of structures in a population. In [3], the Jaccard distance is used, where a threshold on the Jaccard distance can be used to define a network of similar structures, which in turn can form a population for PBSHM. In addition to assessing the topological distance between a population of structures, the attributes can also be considered. Subgraphs can be identified where the attributes of the subgraph from each parent also match [3]. This condition can be used to rule out substructures that do not share key attributes, such as material properties, further enhancing the chance of positive transfer of knowledge between structures.

In summary, a graphical representation allows structures to be converted into a graphical domain, where maximum common subgraphs can be determined. In addition, distance metrics can be applied in this graphical domain providing a more objective interpretation of how similar structures are, helping to determine whether any set of structures should be considered as a population for a PBSHM methodology. As well as quantifying the level of topological (or structural) equivalence between structures, attributed graphs also aid in quantifying the differences between other key properties such as shared geometry and material properties. As a consequence, this process of identifying populations of similar structures via graphical representations helps form PBSHM problems that have high chances of positive knowledge transfer.

2.2. Transferring knowledge between structures

The main benefit of a PBSHM approach is the ability to transfer knowledge about health states between different members of the population, overcoming issues associated with a lack of available labelled data that hinder data-driven methods from being implemented in industrial contexts. Knowledge transfer may take on various forms in a PBSHM methodology, from sharing datasets in joint hierarchical statistical models of a population [25] to ideas of *forms* [2] and multi-task learning [26, 27]. In this section, the main emphasis will be *transfer learning* methodologies that enable knowledge from a set of *source* datasets to be used in improving inferences in some set of *target* datasets [4]. It is argued that the aims of transfer learning technologies are most closely aligned with the motivation behind PBSHM. That is, PBSHM seeks to overcome the issue associated with a lack of available labelled data for a particular structure of interest by leveraging and transferring knowledge in the form of labelled data from other structures in a population. More specifically, this aim is closest to a branch of transfer learning called *domain adaptation*, where datasets are adapted

[1] There may be several subgraphs of the same size between two graphs.

in a manner that allows a classifier constructed on source datasets to generalise to target datasets [1, 28–30].

There are many types of domain adaptation methodologies — from kernel-based approaches [31–34] to neural network-based algorithms [35–37]. The main concept behind all these approaches is the idea of harmonising datasets such that a classifier trained on the labelled source datasets will generalise to the unlabelled (or partially-labelled) target datasets, and hence labels can be *transferred* and assigned to the target datasets. The process of harmonisation usually takes the form of identifying a mapping (or potentially some set of mappings) that transforms the source and target datasets such that they can both be considered samples from the same underlying generative distribution. It is interesting to note that some of the simplest approaches to domain adaptation seek to align only the low-order statistical moments of the two datasets [30], which can be seen as being a form of standardisation for promoting transfer learning. Each domain adaptation approach is therefore designed around a variety of assumptions about the similarities between the source and target datasets, e.g. the source and target are derived from the same form of parametric distribution [38], the marginal [31] or joint distributions [32] are different, or that the datasets are observations from the same manifold [33].

In a PBSHM context, understanding the expected similarities between datasets from each structure is therefore a very important step, particularly in determining whether domain adaptation methods will work, and in deciding the most appropriate algorithm for a given scenario. As aforementioned, the abstract representation of graphs helps in identifying the structural similarities between members of the population. In addition to this process, metrics of the data space are also important for determining how similar any two feature spaces from two different structure may be [24]. It is intuitive that datasets that are 'closer' in terms of their statistical properties are more likely to be transferable. Any PBSHM approach must therefore perform transforms that maximise similarities between datasets before assessing how similar the source and target datasets are via some quantitative metric. If these metrics state that the datasets are similar enough, and all prior physical knowledge (particularly from the graph-matching process) suggests that knowledge transfer is possible, then the appropriate domain adaptation approach can be performed and a classifier trained on the (transformed) source datasets will generalise to the (transformed) target datasets. The process of classifying the target datasets allows label knowledge to be transferred from the source structures to the target structures.

3. Heterogeneous population of aircraft wings: Gnat and Piper Tomahawk

Heterogeneous populations pose a difficult set of challenges for any PBSHM methodology, as dissimilarity between structures increases the risk of negative transfer. The following case study seeks to demonstrate the applicability of PBSHM to heterogeneous populations of real world structures, namely a population of two aircraft wings from a Gnat trainer aircraft [39–41] and a Piper Tomahawk aircraft [42] (see Figure 1 and 2 respectively, for a schematic overview). The Gnat and Piper Tomahawk aircraft wings form two experimental datasets that were created to demonstrate data-driven approaches to SHM; for both datasets, the task was to perform localisation of damage in the form of inspection-panel removals. As such, neither monitoring campaign was designed with a population-based approach to SHM in mind. These datasets therefore represent an interesting case study, one in which a PBSHM approach is retrofitted to conventional monitoring campaigns.

The PBSHM scenario considered in this case study assumes that damage states have been observed

Table 1: Area of inspection panels on the Gnat aircraft wing [41].

Panel	1	2	3	4	5	6	7	8	9
Area (m^2)	0.0221	0.0496	0.00825	0.08	0.0176	0.00825	0.0392	0.0468	0.0234

Table 2: Area of inspection panels on the Piper Tomahawk aircraft wing [43].

Panel	1	2	3	4	5
Area (m^2)	0.0254	0.0254	0.0254	0.0254	0.0254

on the Gnat aircraft wing forming a labelled (source) dataset. The goal of the PBSHM methodology is to transfer this knowledge to the (target) Piper Tomahawk wing, where no labelled damage-state data are available. In this context, performing knowledge transfer means that damage localisation can be performed on the Piper Tomahawk using a classifier trained from the Gnat dataset without requiring any labelled (damage) observations on the Piper Tomahawk wing.

3.1. Datasets

The Gnat wing dataset [39–41] (collected in 2003) and Piper Tomahawk wing dataset [42] (collected in 2017) were both acquired with the aim of localising damage on their respective aircraft wing. In both cases, the damage type considered was the removal of inspection panels (as it was not possible to truly damage the wings); the area of the inspection panels for the Gnat and Piper Tomahawk aircraft wings are presented in Tables 1 and 2. In order to localise this form of damage, uniaxial accelerometers were positioned on each wing such that they formed transmissibility 'paths' that covered the panels of interest. The location of the inspection panels, accelerometers and the transmissibility paths formed by sensor pairs are presented in Figures 1 and 2. One key difference between the structures (besides the main geometric and therefore topological differences) was that the Gnat aircraft dataset was acquired when the wing was attached to the complete aircraft, whereas the Piper Tomahawk experiments involved mounting the wing (as an isolated component) to a substantial steel frame (herein referred to as the mount).

3.1.1. Gnat wing dataset: Source structure

The Gnat wing dataset has been well-studied within the SHM literature [39–41, 44–46]. The dataset comprises nine transmissibility paths, specified in Table 3, each of which were positioned in order to maximise changes in the feature, following removal of an inspection panel. The experiments involved exciting the structure with a broadband white-noise input applied on the bottom surface of the wing below Panel Four via an electrodynamic shaker. Each transmissibility covered a frequency range of 1024-2048Hz containing 1024 spectral lines, with the real and imaginary parts being converted into magnitude and phase and the phase then being discarded [41].

Table 3: Accelerometers used to form transmissibilities for the Gnat wing dataset and their associated inspection panel [41].

Transmissibility	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9
Reference accelerometer	AR	AR	AR	BR	BR	BR	CR	CR	CR
Response accelerometer	A1	A2	A3	B1	B2	B3	C1	C2	C3
Inspection panel	P1	P2	P3	P4	P5	P6	P7	P8	P9

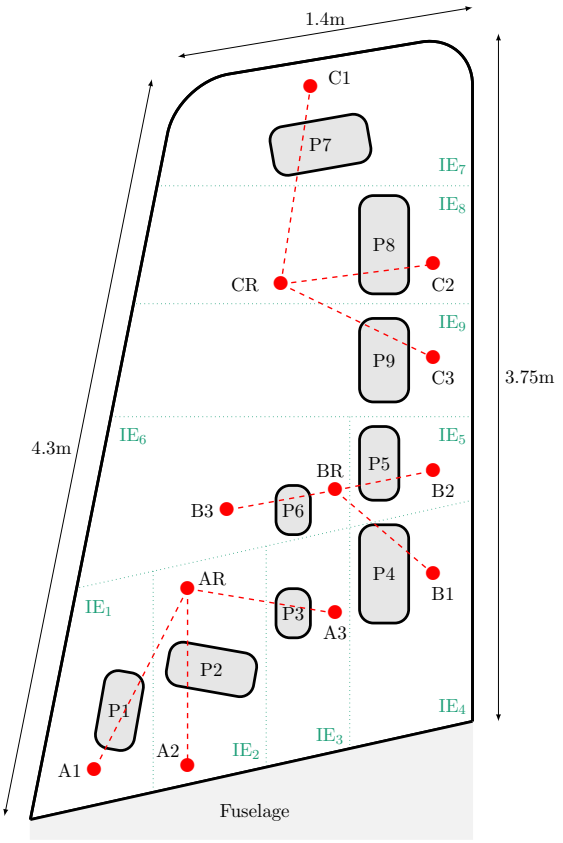

Figure 1: Schematic of the Gnat aircraft wing and associated irreducible element model (not to scale).

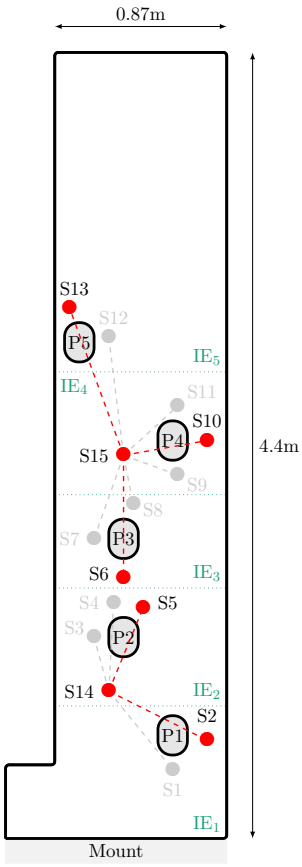

Figure 2: Schematic of the Piper Tomahawk aircraft wing and associated irreducible element model (not to scale).

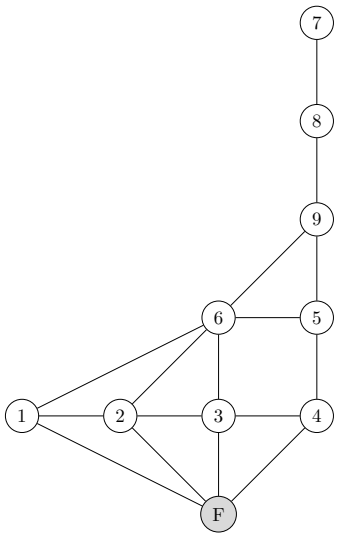

Figure 3: Attributed graph for the Gnat aircraft wing. White and grey circles denote structural and boundary nodes respectively.

Figure 4: Attributed graph for the Piper Tomahawk aircraft wing. White and grey circles denote structural and boundary nodes respectively.

Table 4: Overview of the test sequence for the Gnat dataset.

Damage	N	P1	P2	P3	N	P4	P5	P6
Observations	100	100	100	100	100	100	100	100
Label	$y_s = 0$	$y_s = 1$	$y_s = 2$	$y_s = 3$	$y_s = 0$	$y_s = 4$	$y_s = 5$	$y_s = 6$

Damage	N	P7	P8	P9
Observations	100	100	100	100
Label	$y_s = 0$	$y_s = 7$	$y_s = 8$	$y_s = 9$

The damage scenarios covered each inspection panel being removed and replaced sequentially, with the torque of the fasteners being monitored between configurations. The test strategy was designed around three transducer groups: A, B and C, each covering three inspection panels. The test sequence for each transducer group involved 100 observations of the normal condition, followed by 100 observations of each of the three inspection panels in the group. In the original Gnat dataset, variability was assessed by repeating the test sequence for each transducer group twice; however, this case study only considers one of the repeat test sequences for each transducer group as each repeat is a new observation of the underlying generative distribution and would require some form of domain adaptation before performing PBSHM (such analysis is presented in [46]). The test sequence of the Gnat dataset used in this case study is presented in Table 4, totalling 1200 observations.

The Gnat dataset forms the source dataset in the PBSHM (and transfer learning) context considered. The aim is to transfer a subset of five transmissibilities and their damage localisation labels onto the Piper Tomahawk dataset (which consists of five damage localisation states in addition to the normal condition). The complete Gnat dataset is formed from the transmissibility feature space $\{T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9\} \in \mathbb{R}^{1200\times 9216}$ and the label set $\boldsymbol{y}_s \in \mathbb{R}^{1200}$.

3.1.2. Piper Tomahawk wing dataset: Target structure

The Piper Tomahawk dataset is aquired from thirteen transmissibility paths, outlined in Table 5, where either two or three transmissibility paths cover each panel. In this case study, only five transmissibility paths are considered $\{T_2, T_5, T_6, T_{10}, T_{13}\}$, with the rest removed from the dataset. The five transmissibilities were chosen as each one had the most direct path over their respective inspection panel of interest. This step of reducing the transmissibility set down to one per panel aids the PBSHM process by bringing the Piper Tomahawk dataset closer in design to that of the Gnat dataset — a challenge for retrofitting PBSHM to existing conventional SHM datasets. An electrodynamic shaker was used to provide a broadband white-noise excitation to the underside of the wing approximately one metre from the root end of the wing, positioned on the main spar. The transmissibilities covered a frequency range of 0-2048Hz with 0.5Hz resolution; however, in this case study, the frequency range is truncated to 1024-2048Hz to match the Gnat dataset with the resolution downsampled to 1Hz (meaning there are 1024 spectral lines). The process of removing differences in the signal properties between the datasets maximises the chance of positive transfer and is an extra processing step required for retrofitting PBSHM to these datasets. As with the Gnat dataset [41], the transmissbilities were converted from real and imaginary into magnitude and phase, with the phase discarded.

The Piper Tomahawk dataset was initially designed to demonstrate the effectiveness of using single-site damage states as training data for data-driven methods that are required to locate multi-site damage scenarios [42]. However, given the aims of the PBSHM case study presented, only the single-site damage scenarios are used to form the dataset in this paper (Dataset A in [42]), as

Table 5: Accelerometers used to form transmissibilities for the Piper wing dataset and their associated inspection panel [42]. The transmissibilities used in this case study are denoted *.

Transmissibility	T_1	$T_2{}^*$	T_3	T_4	$T_5{}^*$	$T_6{}^*$	T_7	T_8	T_9	$T_{10}{}^*$	T_{11}	T_{12}	$T_{13}{}^*$
Reference accelerometer	S14	S14	S14	S14	S14	S15	S15	S15	S15	S15	S15	S15	S15
Response accelerometer	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
Inspection panel	P1	P1	P2	P2	P2	P3	P3	P3	P4	P4	P4	P5	P5

Table 6: Overview of the test sequence for the Piper Tomahawk dataset and corresponding labels [42].

Damage	N	P1	N	P2	N	P3	N	P4	N	P5
Observations	5×20	100	5×20	100	5×20	100	5×20	100	5×20	100
Label	$y_t = 0$	$y_t = 1$	$y_t = 0$	$y_t = 2$	$y_t = 0$	$y_t = 3$	$y_t = 0$	$y_t = 4$	$y_t = 0$	$y_t = 5$

multi-site experiments were not conducted on the Gnat aircraft.

The test sequence involved obtaining 100 observations of the normal condition before removing an inspection panel and obtaining 100 observation of that damage state (where an observation is comprised of a five-sample averaged response). This process was repeated, removing and reattaching all five panels in sequence, forming a dataset of 1000 observations (500 of the normal condition and 5×100 damage observations). The experiments also aimed to capture the variability in the normal condition caused by the eight screws used to attach each panel. As such, each block of 100 observations of the normal condition were comprised of five sets of twenty observations, where the inspection panel of interest was removed and reattached (where torque was monitored using a torque-controlled screwdriver). In addition, after each damage state, all five panels were removed and reattached. The overall test sequence is presented in Table 6. Once the set of 1000 observations had been obtained, the sequence of tests was repeated; however, this case study only utilises one of these repeats to avoid the problems associated with each repeat being a new observation of the underlying generative distribution, and in turn requiring some form of domain adaptation itself [46]. For more details about the dataset the interested reader is referred to [42].

In the context of PBSHM (and transfer learning), the Piper Tomahawk dataset is the target dataset. This means that the feature data are unlabelled and labels must be transferred from the source structure to perform localisation. The transmissibility feature space that forms the Piper Tomahawk dataset is $\{T_2, T_5, T_6, T_{10}, T_{13}\} \in \mathbb{R}^{1000 \times 5120}$, which are renumbered such that they align with their particular panel of interest i.e. T_2 is renumbered to T_1 given that it seeks to target panel 1, meaning $\{T_2, T_5, T_6, T_{10}, T_{13}\}$ becomes $\{T_1, T_2, T_3, T_4, T_5\}$. The dataset is considered unlabelled in the analysis, meaning that the label set $\boldsymbol{y}_t \in \mathbb{R}^{1000}$ is unknown during the PBSHM methodology and used only to validate the approach.

3.2. Methodology

The PBSHM methodology proposed in this paper seeks to transfer localisation labels in a heterogeneous population, namely from the Gnat aircraft wing (source) to the Piper Tomahawk wing (target). A particular challenge with these two structures is that the location labels for the two wings are not in one-to-one correspondence, because of topological differences between the two aircraft wings, i.e. there are nine localisation labels on the Gnat and only five localisation labels on the Piper Tomahawk, as there are a different number of inspection panels on each wing. This difference

in topology and label space means that the PBSHM approach must identify the most appropriate subset of label combinations from the Gnat that can be mapped into the Piper Tomahawk and produce positive transfer. Essentially, the PBSHM methodology must perform a 'nine-choose-five' problem where the order of the labels also matters. As a result, the PBSHM approach must select the most appropriate set of five localisation labels on the Gnat from over 15120 potential combinations[2]. Furthermore, this process must be performed without using any labelled observations from the target dataset. Performing domain adaptation in an exhaustive manner would be prohibitive because of the computational cost, and the approach would still require some method or metric for selecting which result out of the 15120 classifiers performed the best knowledge transfer, given that no target labels are available.

The proposed PBSHM methodology seeks to utilise an abstract representation of structures to identify sets of five localisation labels from the Gnat dataset that can be considered *structurally equivalent* using graph matching. Once this physical knowledge has been used to obtain a reduced subset of candidates from the source (Gnat) dataset, distance metrics are applied to determine which of these candidates are 'closest' to the target (Piper Tomahawk) dataset, with this candidate set being most likely to perform positive transfer. Pre-processing (that promotes transfer) and domain adaptation are subsequently performed, identifying a map into a latent space that harmonises the candidate source (Gnat) dataset and the target (Piper Tomahawk) dataset. In this latent space, a classifier is trained on the labelled source (Gnat) dataset and applied to the unlabelled target (Piper Tomahawk) dataset, transferring knowledge from the source to the target structure. A schematic overview of the process is outlined in Figure 5, with the following subsection providing detail on each particular process.

It is important to note that, although tailored to the case study in this paper, the general principles behind this approach are applicable across PBSHM scenarios. These principles are that an abstract representation of structures encodes physical understanding and can be used to simplify the transfer learning problem to only consider scenarios that maximise structural similarity. Distance metrics over the datasets in a population can also be used to further simplify the problem, helping determine if the datasets are similar enough that knowledge transfer should be attempted. After applying these two methods for assessing similarity between members of the population, transfer learning can be performed and label knowledge transferred between members.

3.2.1. Identifying structural similarity via maximum common subgraphs

The first stage of the PBSHM methodology is determining whether structural similarities exist between any two members of a population and locating substructures that can be considered structurally equivalent. This process is performed by creating irreducible element (IE) models and converting them into attributed graphs (AGs), such that graph matching can be performed. Details about constructing IE models and AGs for the aircraft-wing population are outlined below.

The Gnat and Piper Tomahawk aircraft wings are both constructed from sheet aluminium attached to riveted stiffeners along structural spars. From a macro-level, the main structural differences are that the Gnat aircraft has a sweptback wing with a RAE 102 modified aerofoil, and the Piper Tomahawk wing has straight leading and trailing edges with a NASA GA(W)-1 aerofoil; schematics of the wings with nominal dimensions are shown in Figures 1 and 2.

[2]There are 126 combinations from a nine choose five problem, where there are 120 permutations that are possible for each of these 126 combinations, meaning $126 \times 120 = 15120$.

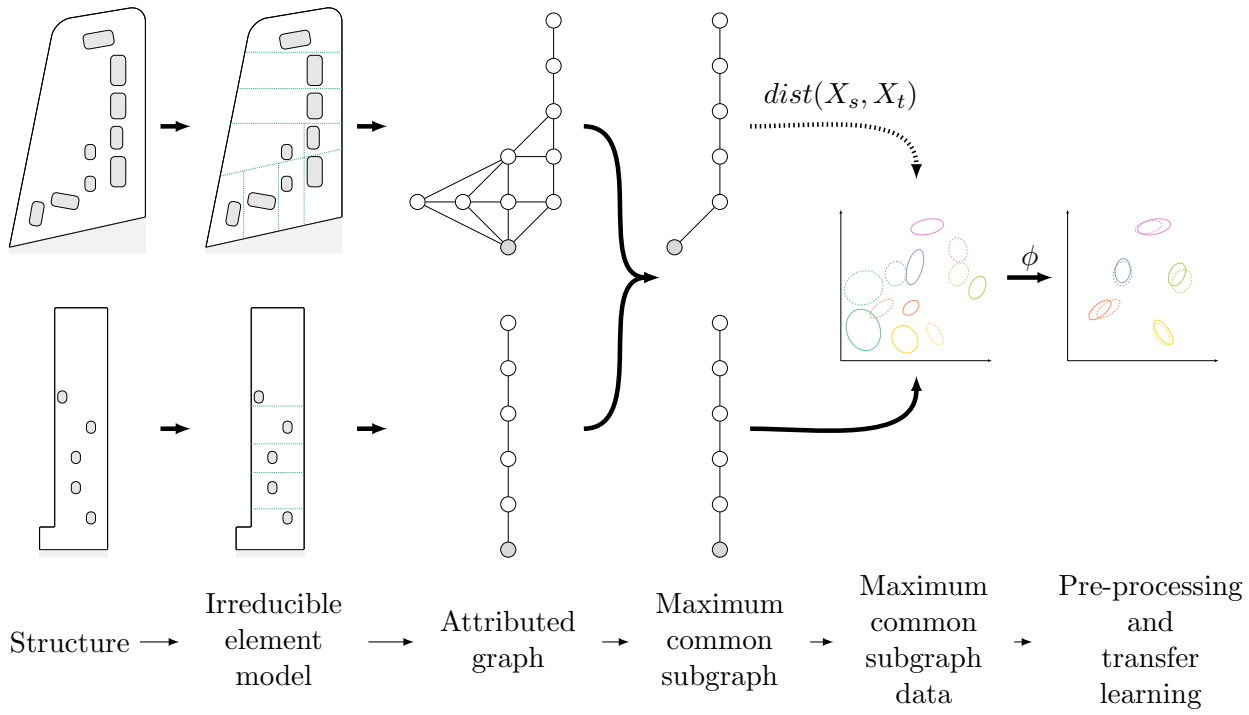

Figure 5: Schematic overview of the population-based methodology.

Before developing an IE model for a PBSHM task, it is important to define the level of resolution required. In this context, the aim of the PBSHM method is to transfer localisation labels from the Gnat to the Piper Tomahawk wing using data from transmissibility paths. The key structural details that must be encoded in the IE models for this context are the topology of the panels on the wing, as these form the main structural differences in the localisation label spaces between the members of the population. As a result, each wing was divided into IEs that focus on defining an IE for each inspection panel as shown in Figures 1 and 2. For both IE models, the relationships between each IE are defined as 'perfect', as further resolution about rivet and stiffener locations are not required for obtaining the key differences between the inspection panel locations. While the physical boundary conditions for each structure are slightly different — the Gnat wing is attached to the fuselage and the Piper Tomahawk wing is positioned on a mount — they are both represented as 'ground' in these IE models as, for this particular SHM problem, the particular details of these boundary conditions was not deemed important for performing the localisation task. The IE models for both structures are outlined in tabular form in Appendix A.

Once constructed, the IE models can be converted to AGs as shown in Figures 3 and 4; in this case study, each node has been assigned a number in relation to the inspection panel that the IE encompasses. The white nodes define structural elements, whilst the grey nodes (and their associated edges) define boundary conditions. The AGs are now representations of the structure in a graphical domain, where similarities can now be quantified (in relation to the PBSHM context). In this case study the important attributes for the localisation problem are the topology of the inspection panels, and therefore further geometric and material attributes were not incorporated into the AGs, and the graph matching was performed solely on the topology of the AGs. The modified BK algorithm identified four connected and induced subgraph pairings outlined in Table 7, with Figure 6 showing

Piper		M	1	2	3	4	5
Gnat	(a)	F	4	5	9	8	7
	(b)	F	1	6	9	8	7
	(c)	F	2	6	9	8	7
	(d)	F	3	6	9	8	7

Table 7: The node combinations for the maximum common subgraphs using the modified BK algorithm.

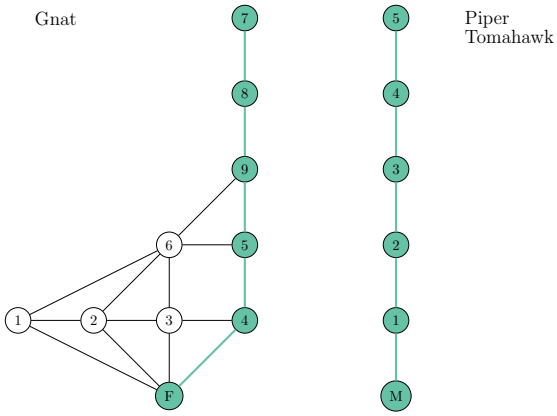

Figure 6: Maximum common subgraph (a).

maximum common subgraph (a) as an example. It is noted that the Jacard distance between the Gnat and Piper Tomahawk structures is 0.6. The results from the graph-matching stage mean that there are four combinations of structural elements in the Gnat that can be mapped onto the Piper Tomahawk whilst maintaining the level of structural equivalence encoded in their respective IE models and AGs. This process has ruled out over 99.97% of inspection panel combinations that could be mapped from the Gnat onto the Piper Tomahawk wing, significantly reducing the computational load of the problem from 15120 potential mappings to just four. The remaining part of the PBSHM methodology is therefore selecting which of these four maximum common subgraphs are most likely to produce positive transfer and allow the localisation labels from the Gnat to be used to diagnose the Piper Tomahawk.

It is interesting to note that the maximum common subgraphs identified in Table 7 effectively define 'cantilever-like' combinations of IEs in the Gnat wing. These subgraphs have been identified, as the Piper Tomahawk wing can effectively be seen as a complex cantilevered structure. The identified subgraphs therefore match a level of physical intuition. Furthermore, it might be expected that maximum common subgraph (a) is the most similar 'cantilever-like' combination of IEs that map onto the Piper Tomahawk, as (a) defines elements along the trailing edge of the Gnat wing that will be stiffened by the spars along that edge in a similar manner to the Piper Tomahawk elements that lie close to the main spar.

3.2.2. Assessing transferability between data spaces

The 'closer' two datasets are in the feature space, the more likely that domain adaptation will lead to positive transfer; this is because the mapping needed to harmonise the two datasets will be less complex. The second key stage of the proposed PBSHM methodology is to perform any pre-processing that might aid transfer by removing differences in the data space. After pre-processing has been performed, the distance between the proposed feature spaces is calculated, with the 'closest' feature spaces being most likely to produce positive transfer.

Pre-processing

Pre-processing is important for any machine learning approach to SHM, helping identify the most optimal features for a given SHM task. The same principle applies, and arguably is of increased importance, in a PBSHM context. For a PBSHM scenario, any pre-processing that will help align feature spaces prior to applying domain adaptation algorithms should be performed. In this paper statistic alignment [30] is conducted before applying principal component analysis to each feature

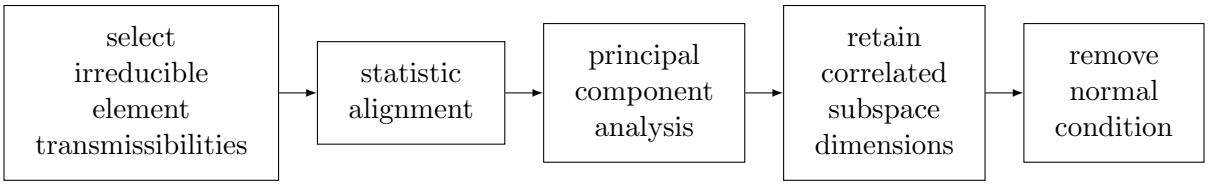

Figure 7: Pre-processing steps for the source and target datasets.

space; an overview of the steps is shown in Figure 7.

The first pre-processing step is to select the set of transmissbility paths that will form the candidate source and target data spaces. For the j^{th} pair of subgraphs (in the set of 15120 potential subgraphs) the structural node sets G_s^j and G_t^j are used to form the transmissibility spaces T_s^j and T_t^j, where only transmissibility paths that are associated with a structural IE in the graph are kept, i.e. for maximum common subgraph pair (a), the transmissibility spaces $T_s^{(a)} = \{T_i : T_i \in G_s^{(a)} \wedge y_s(T_i) \in G_s^{(a)}\} \in \mathbb{R}^{800\times 5120}$ where $G_s^{(a)} = \{4\ 5\ 9\ 8\ 7\}$ and $T_t^{(a)} = \{T_i : T_i \in G_t^{(a)} \wedge y_t(T_i) \in G_t^{(a)}\} \in \mathbb{R}^{1000\times 5120}$ where $G_t^{(a)} = \{1\ 2\ 3\ 4\ 5\}$ (where $\wedge$ is the logical AND symbol). As a result, the label spaces for the j^{th} pair of subgraphs are truncated such that only the normal condition and localisation labels associated with the j^{th} subgraph set are kept, $\boldsymbol{y}_s^j = \{y_{s,i}^j : y_{s,i}^j \in \{0, G_s^j\}\} \in \mathbb{R}^{800}$ and $\boldsymbol{y}_t^j = \{y_{t,i}^j : y_{t,i}^j \in \{0, G_t^j\}\} \in \mathbb{R}^{1000}$, where '0' defines the label associated with the normal condition.

Statistic alignment — the second pre-processing step — is a set of methods that align some set of statistics about two data spaces [30]. This group of approaches can be seen as analogous to performing standardisation in conventional machine learning methods, aiding numerical stability and convergence of algorithms and promoting equal treatment of features. The proposed statistic alignment step in this paper seeks to align the datasets by standardising to their respective normal condition statistics i.e. both dataset's normal condition statistics are aligned to have zero mean and unit variance. The reason for aligning the source and target data by the first two moments of their normal condition is that the complete training datasets for each structure may not cover the same range of behaviours (i.e. data from some classes may be missing) and each dataset may have different levels of class imbalance. This problem means that the global statistics of the dataset are more likely to be affected by sample bias than the statistics of the normal condition. Furthermore, in most typical SHM applications, the initial measured data points are assumed to come from the structure in its normal operating condition, meaning these labels are 'known', and the majority of data points from a system will be from the structure in their normal condition. Formally, the approach can be defined as,

$$\bar{T}_s^j = \frac{T_s^j - \mu(\{t_{s,i}^j : y_s^j(t_{s,i}^j) = 0 \wedge i \in N_s^{tr}\})}{\sigma(\{t_{s,i}^j : y_s^j(t_{s,i}^j) = 0 \wedge i \in N_s^{tr}\})} \tag{1a}$$

$$\bar{T}_t^j = \frac{T_t^j - \mu(\{t_{t,i}^j : y_t^j(t_{t,i}^j) = 0 \wedge i \in N_t^{tr}\})}{\sigma(\{t_{t,i}^j : y_t^j(t_{t,i}^j) = 0 \wedge i \in N_t^{tr}\})} \tag{1b}$$

where $\bar{T}_s^j$ and $\bar{T}_t^j$ denote the statistic-aligned transmissibilities for the j^{th} subgraph for the source and target. The functions $\mu(\cdot)$ and $\sigma(\cdot)$ define the empirical mean and standard deviations of their arguments. The sets N_s^{tr} and N_t^{tr} define the sets of training indices for the source and target datasets, where a 60:40 training-to-testing data split has been used (with the testing dataset treated

as a held-out set for validation). Studies have shown this type of subset statistic alignment to be effective for PBSHM scenarios [47].

Principal component analysis (PCA) has been demonstrated to aid domain adaptation as a pre-processing step [48] and even forms the basis of several transfer learning technologies [33, 49, 50]. The idea is that identifying and sorting both datasets by components that maximise variance the two spaces are brought 'closer' into alignment. In the PCA pre-processing step, the statistic-aligned transmissibility spaces are transformed via PCA (calculated on the training data) into linear subspaces X_s^j and X_t^j.

Gong *et al.* propose a method for promoting transfer when using linear subspaces. The approach seeks to select an optimal number of subspace dimensions such that the source and target subspaces are 'well-correlated'. Gong *et al.* propose using a subspace disagreement measure (SDM), outlined in Appendix B, that seeks to select only the principal components that are 'well-correlated' (on the Grassmannian, the set of linear subspaces) between the source and target PCA subspaces whilst balancing variance preservation of each dataset. In this case study, a dimension of ten is used[3].

Once the PCA subspaces are identified, the normal condition data are removed from the feature space i.e. $X_s^j = \{x_{s,i}^j : y_{s,i} \in G_s^j\} \in \mathbb{R}^{500\times 10}$ and $X_t^j = \{x_{t,i}^j : y_{t,i} \in G_t^j\} \in \mathbb{R}^{500\times 10}$ and label space $\boldsymbol{y}_s^j = \{y_{s,i} : y_{s,i} \in G_s^j\} \in \mathbb{R}^{500}$ and $\boldsymbol{y}_t^j = \{y_{t,i} : y_{t,i} \in G_t^j\} \in \mathbb{R}^{500}$. The normal condition is crucial in learning the PCA embeddings as it anchors the mappings to the feature-aligned space. However, the normal condition is not needed at the classification or domain adaptation stage, as novelty detectors have been shown to adequately perform damage detection in isolation [40, 41] and the PBSHM task is performing localisation.

Distance metrics over the feature space

Domain adaptation techniques seek to harmonise source and target data spaces; to do this, most algorithms identify a mapping that minimises the distance between the two datasets. There is therefore a prior assumption that the 'closer' the generative data distributions are for any two datasets, the more similar the datasets are, and the more likely transfer is possible when compared to datasets which are 'far' away from each other. This notion that 'closer' datasets are more likely to produce positive transfer can be used to select the most optimal source and target datasets for a PBSHM problem. A distance metric-based approach is proposed in order to determine which particular source-target pair of datasets from the maximum common subgraph set (Table 7) should be used to perform transfer for the aircraft wing population. Specifically, the two metrics, the *rank of domain* and the *maximum mean discrepancy* are discussed.

The rank of domain (ROD) was proposed by Gong *et al.* specifically for determining which source datasets would produce the best performance for target datasets without training a domain adaptation technique or classifier [33]. The metric is designed for scenarios where the feature sets are linear subspaces, as is the case in this paper. Mathematically, the ROD is a weighted symmetrised KL-divergence, where the weights are the principal angles between the source and target. As a result, the ROD assesses the statistical distance between the source and target, whilst also considering

[3]A dimension of ten was selected as it was the maximum dimension from performing the SDM analysis for all 15120 combinations of source-target datasets (with dimensions varying from two to ten). By selecting a dimension of ten, all candidate feature sets contain all their 'well-aligned' components and objective distance metric comparisons can be performed between all the potential feature sets. In addition, a study was performed investigating the sensitivity of the PBSHM methodology to the number of PCA dimensions, where dimensions around ten all produced similar outcomes.

the geometric alignment of the two subspaces. The symmetrised KL-divergence used in the ROD assumes that each dimension of the feature space for both the source and target subspaces can be approximated with one-dimensional Gaussian distributions. This assumption is reasonable, given that the two subspaces are linear subspaces identified using PCA. Formally, the ROD is defined as,

$$D_{ROD}(\mathcal{S}, \mathcal{T}) = \frac{1}{d} \sum_{i}^{d} \theta_i \left(KL(\mathcal{S}_i || \mathcal{T}_i) + KL(\mathcal{T}_i || \mathcal{S}_i) \right) \tag{2}$$

where $\mathcal{S}_i$ and $\mathcal{T}_i$ are both one-dimensional Gaussian distribution approximations of the data distributions projected onto the subspace (as mentioned above), θ_i is the i^{th} principal angle between the subspaces, and d is the dimension of the subspaces (e.g. in this paper $d = 10$). For further mathematical details, see Appendix C.

The second metric utilised in this section is the maximum mean discrepancy — a metric typically used as part of the cost function in many domain adaptation algorithms [28, 32, 51, 52]. The MMD is a popular choice of distance metric, as it is a non-parametric distance, meaning that the distribution forms do not need to be known (or approximated). The MMD quantifies the difference between the mean embeddings for two datasets (using a universal reproducing kernel Hilbert space), meaning that it is a metric on distributions [51]. A Gaussian kernel is utilised in this paper (as it is characteristic on $\mathbb{R}^d$), $k(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\left|\left|x - x'\right|\right|^2 / 2l^2)$, where l is the scale parameter which is determined here using the median heuristic [51]. The MMD can be formulated to assess the distance between the marginal distributions (i.e. $P(X_s)$ and $P(X_t)$), or the joint distributions (i.e. $P(\boldsymbol{y}_s, X_s)$ and $P(\boldsymbol{y}_t, X_t)$)[4]. These two forms of the distance are useful, as the marginal MMD can be considered an unsupervised metric, whereas the joint MMD more accurately assesses the differences between the two feature spaces, but requires labels. The squared MMD distances[5] (both marginal and joint), can be defined as,

$$D^2_{MMD}(\mathcal{D}_s, \mathcal{D}_t) = \operatorname{tr}\left(K \sum_{c=0}^{C} M_c \right) \tag{3}$$

where $k(X, X') = K \in \mathbb{R}^{(N_s+N_t)\times(N_s+N_t)}$, which is the kernel matrix, given that $X = X_s \cup X_t \in \mathbb{R}^{(N_s+N_t)\times d}$, and M_c defines the empirical mean such that,

$$(M_c)_{i,j} = \begin{cases} \frac{1}{N_s^{(c)} N_s^{(c)}}, & x_i, x_j \in \mathcal{D}_s^{(c)} \\ \frac{1}{N_t^{(c)} N_t^{(c)}}, & x_i, x_j \in \mathcal{D}_t^{(c)} \\ \frac{-1}{N_s^{(c)} N_t^{(c)}}, & \begin{cases} x_i \in \mathcal{D}_s^{(c)},\ x_j \in \mathcal{D}_t^{(c)} \\ x_j \in \mathcal{D}_s^{(c)},\ x_i \in \mathcal{D}_t^{(c)} \end{cases} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

[4]The distance between the joint distribution is approximated by considering the distance between the marginal (i.e. $P(X_s)$ and $P(X_t)$) and class-conditional distributions (i.e. $P(X_s \,|\, \boldsymbol{y}_s)$ and $P(X_t \,|\, \boldsymbol{y}_t)$), removing the need for a model.

[5]Throughout the paper any squared MMD distance is referred to as an MMD distance.

Table 8: Distance metrics for the four sets of candidate feature spaces from the maximum common subgraphs. Lowest metrics are highlighted in bold.

Subgraph set	(a)	(b)	(c)	(d)
ROD	**0.62**	3.16	1.07	1.51
Marginal MMD	0.13	**0.08**	0.12	0.10
Joint MMD	**3.02**	4.41	6.37	4.18

Table 9: F1 scores for the four sets of candidate feature spaces from the maximum common subgraphs. Highest F1 scores are highlighted in bold.

Subgraph set			(a)	(b)	(c)	(d)
F1	Source	Train	**1.0**	**1.0**	**1.0**	**1.0**
		Test	**1.0**	**1.0**	**1.0**	**1.0**
	Target	Train	**1.0**	0.4	0.2	0.6
		Test	**1.0**	0.4	0.2	0.6

where, when $C = 0$ the MMD assesses the distance between the marginal distributions[6], and when $c = \{0, 1 \ldots, C\}$ the MMD approximates the distance between the joint distribution (where C is the total number of classes). The notation $(M_c)_{i,j}$ denotes the i^{th} row and j^{th} column of the matrix M_c. Furthermore, $\mathcal{D}_s^{(c)} = \{\boldsymbol{x}_i : \boldsymbol{x}_i \in \mathcal{D}_s \wedge y(\boldsymbol{x}_i) = c\}$ denotes the source instances that belong in class c and $\mathcal{D}_t^{(c)} = \{\boldsymbol{x}_i : \boldsymbol{x}_i \in \mathcal{D}_t \wedge y(\boldsymbol{x}_i) = c\}$ are the target instances that belong in class c, where $N_s^{(c)} = |\mathcal{D}_s^{(c)}|$ and $N_t^{(c)} = |\mathcal{D}_t^{(c)}|$.

The ROD, marginal and joint MMD distances were quantified, as shown in Table 8, for the four (pre-processed) sets of candidate feature spaces, each of which are associated with a maximum common subgraph in Table 7. The values from the two unsupervised metrics, the ROD and marginal MMD, show conflicting results, with the rank order of the subgraph sets reversed for the two metrics. The ROD, which takes into account the alignment of the subspaces as well as the statistical distance, shows the smallest distance for set (a) and largest for set (b), whereas the marginal MMD distance is smallest for set (b) and largest for set (a). However, the MMD distances for all four sets are more similar than the ROD values, which indicates that set (a) is 'closer' than (b), (c) and (d). Consulting the (supervised) joint MMD metric, which requires knowledge of the source and target labels (something not possible in practice), set (a) comes out as the most 'similar', producing the smallest distance, and (b) the most dissimilar. Given this analysis, set (a) is selected as the feature spaces for performing transfer learning, having performed best on two out of the three metrics, (i.e. of the two metrics that really separate the options, both clearly choose (a)) with the expectation that these datasets are most likely to lead to positive transfer (out of the set of four candidates). Exploration of the link between data-space distances and transferability is discussed further in Section 3.3.2.

3.2.3. Transferring localisation labels

Domain adaptation algorithms allow label information from source datasets to be utilised in classifying (unlabelled) target datasets. Once a pair of candidate source and target datasets have been selected (via structural and data similarity measures), domain adaptation is performed, identifying a latent space where the source and target datasets are harmonised. In this latent space a classifier is trained on the source dataset and used to diagnose the target dataset.

Multiple domain adaptation algorithms exist in the literature [29–31], each of which has particular assumptions and strengths for specific transfer learning problems. In this paper, balanced distribution adaptation (BDA) [34] is utilised as the algorithm[7] for performing transfer, with details stated

[6]Here $c = 0$ is used to refer to the complete dataset, i.e. $\mathcal{D}_s^{(0)} = \{\boldsymbol{x}_i : \boldsymbol{x}_i \in \mathcal{D}_s\}$, $\mathcal{D}_t^{(0)} = \{\boldsymbol{x}_i : \boldsymbol{x}_i \in \mathcal{D}_t\}$, $N_s^{(0)} = N_s$ and $N_t^{(0)} = N_t$; this should not be confused with the use of '0' as the normal condition label earlier in the manuscript.

[7]It is noted that the novelty of this paper is in presenting the complete PBSHM methodology (visually outlined in

in Appendix D. BDA was applied with a Gaussian kernel (where l was chosen using the median heuristic), five iterations, $k = 2$ transfer components, regularisation parameter $\mu = 0.1$, and the balance factor $\lambda = 1$; meaning the datasets are similar and matching the conditional distribution is most important [34], something that can be assumed given the preprocessing (that mostly aligns the marginal distributions) and graph-matching steps. The BDA mappings were inferred using the source (labelled) and target (unlabelled) training datasets with generalisation of the mapping validated on the testing datasets. Finally, a k-nearest neighbour (kNN) classifier is used to transfer knowledge from the source to the target in the latent space. A kNN is chosen, as BDA seeks to minimise the distance between these source and target datasets, therefore making them 'close' in Euclidean space. A kNN was trained on the latent space for the labelled source training data, where the number of neighbours was selected using cross-validation. This classifier was applied to the source testing and target training and testing datasets.

A visualisation of the BDA mapping identified for feature set (a) is shown in Figure 8 (where the source labels are renumbered to match their corresponding target label). It can be seen that all five classes have been harmonised correctly, and hence lie on top of each other in the latent space. In this latent space the kNN classifier trained on the source training dataset generalises to the target datasets with 100% classification accuracy. The PBSHM methodology shows an increase on the performance of the two multi-class SVM classifiers in [42], where a classification performance of the single-site damage states was 98.29% for Classifier 1 (trained using a range of single- and multi-site damage states) and 98.34% using Classifier 2 (trained using only the five single-site damage states).

It is interesting to note that the pre-processing and domain adaptation steps have reduced the feature space dimensionality from 5120 to just two. The latent space for set (b) is shown in Figure 9 for comparison, showing that negative transfer has occurred and the incorrect sets of classes have been aligned. A comparison of performance for the kNN classifiers trained on the BDA latent spaces for the four candidate feature sets is shown in Table 9. These results show that the PBSHM methodology has successfully allowed the Gnat labelled dataset to localise damage on the Piper Tomahawk wing using candidate set (a) — the set chosen by both physical intuition and by distance metrics. Analysis is shown in Section 3.3 that verifies the methodology, showing comparison to all 15120 combinations of source-target pairs.

3.3. Verification of the methodology

The PBSHM approach outlined in Section 3.2 has been demonstrated as successful in diagnosing damage on the Piper Tomahawk using information from the Gnat dataset. The following section seeks to further verify that the proposed methodology provides an efficient solution to performing PBSHM in heterogeneous populations. Section 3.3.1 demonstrates the benefits of generating abstract representations of structures as graphs, with Section 3.3.2 exploring the relationship between distance metrics and transferability.

3.3.1. Verification of the abstract representation of structures

The PBSHM task was to transfer localisation labels from the Gnat (source) dataset to the Piper Tomahawk (target) dataset. A challenge with this problem is that there are over 15120 combinations of localisation label sets that could be transferred from the Gnat onto the Piper Tomahawk dataset. The PBSHM methodology has shown that by utilising abstract representations of structures in a

Figure 5), of which domain adaptation is one (important) step; BDA could therefore be substituted for any appropriate alternative domain adaptation algorithm.

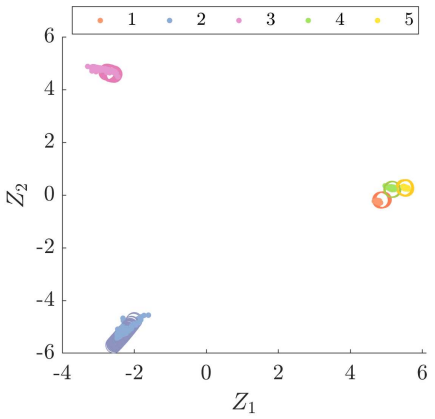

Figure 8: A visualisation of the BDA latent space for candidate set (a), where • denotes the source and ◯ denotes the target datasets.

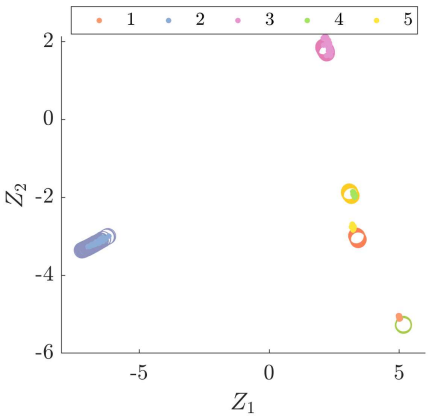

Figure 9: A visualisation of the BDA latent space for candidate set (b), where • denotes the source and ◯ denotes the target datasets.

graphical domain, the problem can be reduced to four candidate sets. The same pre-processing, domain adaptation and classification steps were performed for all 15120 combinations. It was found that only 92 combinations lead to classification performances with an F1 score of 1 on both the target training and testing dataset (around 0.6% of the complete set). A box and whisker plot of the F1 scores on target testing data is presented in Figure 10, comparing those feature combinations that were not MCSs to those that were[8]. It is interesting to note that the worst possible effect of negative transfer has been removed from the MCS set (i.e. no MCSs produced an F1 score of 0).

These results show that without some prior knowledge, the PBSHM problem would be challenging to perform in an exhaustive manner, and even more challenging to select positive transfer scenarios without access to target labels. By generating AGs and performing graph matching, this problem was simplified, improving the chance of identifying a set that produced positive transfer to one in four. Before reduction, the probability that the F1 score equalled 1 was 0.006; after, it was 0.25. This demonstrates the power of utilising knowledge about the structural similarities between members of a population.

The classification results can be further grouped into those feature sets that form no path on the AGs (denoted 'Random'); those that have some topological equivalence, being valid paths of five structural nodes on the AGs (denoted 'Cat. 1'); those with some structural equivalence, being valid paths of five structural nodes and one boundary condition node on the AGs (denoted 'Cat. 2'); and MCSs. The chance of identifying a feature set in a particular group that produced a classifier with 100% accuracy on the target dataset were: Random: 0.6%, Cat. 1: 0%, Cat. 2: 3.4%, and MCS: 25%. Box and whisker plots for these four groupings are presented in Figure 11 showing F1 scores on the target testing data. It is noted that the performance of the set defined by Cat. 1 is much worse than the Random set (see Figure 11), demonstrating the importance of boundary conditions in defining structural similarity. In fact, the set defined by Cat. 2 has a mean and interquartile range that are worse than the Random set and Cat. 1; however, there is a higher probability of finding a combination that produced an F1 score of 1 within Cat. 2 than Cat 1. or the Random set. The results from these groupings demonstrate that, given the AG representations, subgraphs that were connected and induced (i.e. MCSs) were much more likely to produce positive transfer

[8]The same trends were seen for the target training dataset as well.

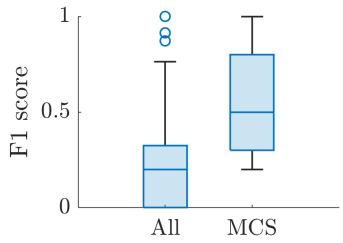

Figure 10: Box and whisker plots of testing F1 scores from kNN classifiers trained on BDA latent spaces. 'All' denotes feature set combinations that are not MCSs.

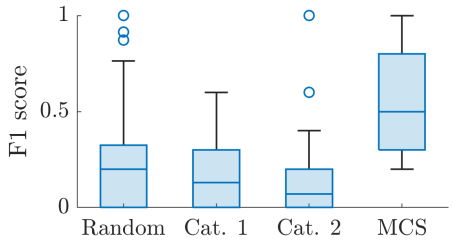

Figure 11: Box and whisker plots of testing F1 scores from kNN classifiers trained on BDA latent spaces. 'Random' denotes feature set combinations that are not valid paths in the AGs, 'Cat. 1' are combinations that are valid paths in the AGs and 'Cat. 2' combinations that are valid paths and include the boundary condition nodes in the AGs.

than other valid subgraphs, verifying the assumption that these criteria better reflect structural similarities.

3.3.2. Feature space distances and classification performance

One of the key challenges when performing transfer learning is determining the risk of negative transfer. It is difficult to quantify the risk of negative transfer, as it is affected by many different factors, such as whether two datasets have some causal mechanism that is linked between them, or whether the nonlinear mapping and the assumptions in the domain adaptation algorithm are valid. One factor investigated in this paper is whether transferability can be determined by assessing the distance between source and target feature spaces, an idea that has been utilised in the transfer learning literature [33].

The ROD, marginal MMD and joint MMD distances were assessed for all 15120 (pre-processed) feature sets and compared to the target testing F1 scores, presented in Figures 12, 13 and 14. It can be seen that the two unsupervised metrics, the ROD and marginal MMD, both show almost no correlation to F1 score. The reason for the lack of correlation is that the ROD and marginal MMD only depend on the unlabelled feature space, meaning that these distances are the same for any permutations of the labels for a given set of features in the n-choose-k problem (effectively there are 126 different ROD and marginal MMD distances from the n-choose-k problem). Given that the ordering of how the labels are paired between the Gnat and Piper Tomahawk data matters, both these metrics fail to fully capture the notion of distance between the datasets, and hence there is a large amount of scatter for a given ROD and marginal MMD distance. The joint MMD does incorporate knowledge of the labels, and hence is different for all 15120 feature set combinations. This metric shows a clear negative correlation (the Pearson correlation coefficient is -0.69), although there is large scatter of joint MMD distances that produce the same F1 scores. Comparisons of marginal and joint MMD and ROD and joint MMD are presented in Figures 15 and 16. These figures show that generally the lowest joint MMD values for a given marginal MMD or ROD distance lead to the best F1 scores. Knowledge of the difference in joint distribution between datasets therefore provides more information about transferability. These results show the challenge of determining the risk of negative transfer from a data-based viewpoint *a priori*, and demonstrates why the process of forming AGs and identifying MCSs is so important in a PBSHM context. Further research will investigate whether there is an unsupervised metric that has a clear link to transferability.

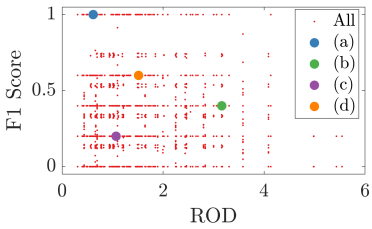

Figure 12: Rank of domain (ROD) against F1 score on the target testing dataset. 'All' denotes feature set combinations that are not MCSs and (a), (b), (c) and (d) denote the MCS set.

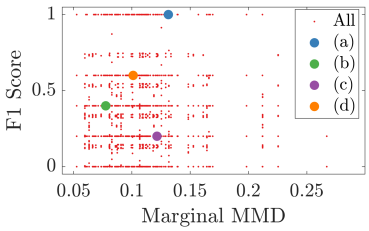

Figure 13: Marginal maximum mean discrepancy (MMD) against F1 score on the target testing dataset. 'All' denotes feature set combinations that are not MCSs and (a), (b), (c) and (d) denote the MCS set.

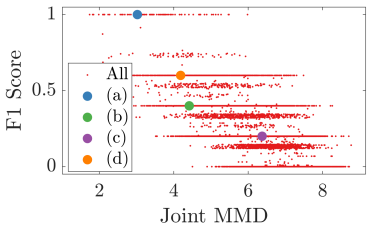

Figure 14: Joint maximum mean discrepancy (MMD) against F1 score on the target testing dataset. 'All' denotes feature set combinations that are not MCSs and (a), (b), (c) and (d) denote the MCS set.

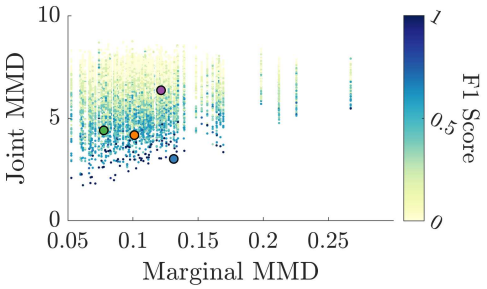

Figure 15: Marginal maximum mean discrepancy (MMD) against Joint MMD, coloured by the F1 score on the target testing dataset. The coloured (•) denote the MCSs with blue for set (a), green for (b), purple for (c) and orange for (d).

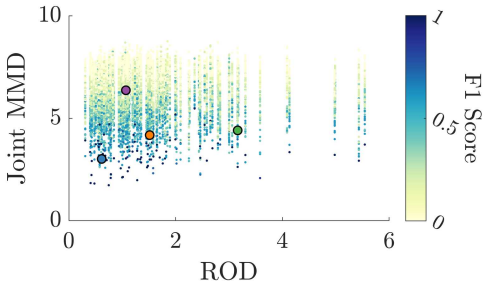

Figure 16: Rank of domain (ROD) against Joint MMD, coloured by the F1 score on the target testing dataset. The coloured (•) denote the MCSs with blue for set (a), green for (b), purple for (c) and orange for (d).

4. Conclusions

Population-based structural health monitoring seeks to improve SHM inferences by overcoming issues associated with a lack of labelled damage observations by expanding the available dataset to a population of structures. PBSHM becomes more difficult as structures in a population become increasingly different (groups called heterogeneous populations). The challenge arises as populations include structures that are increasingly dissimilar, with data spaces that are further apart and less aligned. This paper seeks to demonstrate that it is possible to perform PBSHM for hetergeneous populations, demonstrating a novel PBSHM methodology for a population of two real-world structures, a Gnat aircraft wing and a Piper Tomahawk aircraft wing. Specifically, this paper demonstrates that localisation labels can be transferred from the Gnat wing dataset onto the Piper Tomahawk dataset, meaning localisation can be performed on the Piper Tomahawk wing without the need for damage observations. It should be noted that this problem is a severe stress test for PBSHM. Ideally, sensor systems on multiple structures would be aligned in ways that promote transfer. In the situation here, the sensor sets were chosen for the individual problem only, demonstrating the power of retrofitting a PBSHM approach to exisiting SHM datasets.

The proposed PBSHM methodology utilises abstract representations of structures in a graphical domain via the use of irreducible element models and attributed graphs. This process of converting structures into attributed graphs allows for objective comparisons of structural similarities, as metrics can be defined over the space of structures. In addition, maximum common subgraphs can

be extracted across a population of structures, locating sets of structural elements that exist in any two members of a population. This process was applied to the heterogeneous population of aircraft wings, where the AGs were utilised to reduce the potential candidate feature spaces for transfer learning down from 15120 to just four, by identifying four maximum common subgraphs. This use of AGs as a method for encoding structural similarities was shown to be extremely effective with one of the four MCSs leading to F1 scores of 1. The process of utilising an abstract representation of structures in a graphical domain meant that the chance of selecting a set of candidate features that lead to 100% classification accuracy increased from around 0.6% to 25%.

The second main stage of the PBSHM methodology sought to perform pre-transfer alignment via pre-processing before assessing the distances between the source-target dataset pairs. Three distance metrics were used to assess the distance between the source-target pairs, namely the rank of domain, marginal maximum mean discrepancy and the joint maximum mean discrepancy. It was found that one of the MCS feature sets led to the smallest distance in two out of the three metrics and was selected as the most likely to perform positive transfer. However, analysis of the relationship between these distance metrics and transferability was not clear, with this being an area for further research. In particular, the two metrics that were unsupervised (i.e. did not require labels) showed no clear correlation to the target F1 score for all 15120 potential source-target feature sets. This is because transferability depends on the joint distribution of the dataset, with every class cluster needing to be 'close' for transfer to work. This was reflected in the joint MMD that provided a clearer correlation with testing F1 score, as opposed to the marginal MMD and ROD, but required knowledge of the source and target labels.

The final step in the PBSHM methodology was to perform domain adaptation for set (a) from the MCSs group. This lead to 100% classification accuracy on the target dataset, showing that localisation labels can be transferred from the Gnat wing to the Piper Tomahawk wing.

5. 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. The authors would also like to thank Dr Graeme Manson and Dr Rob Barthorpe for the Gnat and Piper Tomahawk datasets respectively, and Dan Brennan for his help with the Irreducible Element modelling.

Appendix A. Irreducible element models

Appendix A.1. Gnat aircraft wing irreducible element model

The two tables below define the complete IE model for the Gnat aircraft wing, where Table A.10 states the structural elements and Table A.11 the relationships between those elements.

Table A.10: List of elements and their properties for the Gnat aircraft wing.

Gnat aircraft wing				
Element Name	**Description**	**Material Type**	**Geometry Type**	**Contextual Type**
1	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
2	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
3	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
4	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
5	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
6	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
7	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
8	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
9	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
Element Name	**Description**	**Boundary**	-	-
A	Fuselage	Ground	-	-

Table A.11: List of relationships and their properties for the Gnat aircraft wing.

Gnat aircraft wing		
Relationship Name	**Element set**	**Type**
A	{1, 2}	Perfect
B	{2, 3}	Perfect
C	{3, 4}	Perfect
D	{1, 6}	Perfect
E	{2, 6}	Perfect
F	{3, 6}	Perfect
G	{4, 5}	Perfect
H	{5, 6}	Perfect
I	{5, 9}	Perfect
J	{6, 9}	Perfect
K	{9, 8}	Perfect
L	{8, 7}	Perfect
M	**{1, A}**	**Boundary**
N	**{2, A}**	**Boundary**
O	**{3, A}**	**Boundary**
P	**{4, A}**	**Boundary**

Appendix A.2. Piper Tomahawk aircraft wing irreducible element model

The two tables below define the complete IE model for the Piper Tomahawk aircraft wing, where Table A.12 states the structural elements and Table A.13 the relationships between those elements.

Table A.12: List of elements and their properties for the Piper Tomahawk aircraft wing.

Piper Tomahawk aircraft wing				
Element Name	**Description**	**Material Type**	**Geometry Type**	**Contextual Type**
1	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
2	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
3	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
4	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
5	Wing Panel	Metal → Aluminium	Plate → Other	Aerofoil
Element Name	**Description**	**Boundary**	-	-
A	Mount	Ground	-	-

Table A.13: List of relationships and their properties for the Piper Tomahawk aircraft wing.

Piper Tomahawk aircraft wing		
Relationship Name	**Element set**	**Type**
A	{1, 2}	Perfect
B	{2, 3}	Perfect
C	{3, 4}	Perfect
D	{4, 5}	Perfect
E	**{1, A}**	**Boundary**

Appendix B. Subspace disagreement measure

The subspace disagreement measure (SDM) was introduced by Gong *et al.* [33] as an unsupervised method for determining the size of subspace (both source and target), that would be optimal for performing knowledge transfer using the geodesic flow kernel algorithm [33]. The intuition behind this idea was that the size of subspace should be determined by selecting the maximum number of dimensions that are 'well-correlated' between the source and target subspaces in an unsupervised way. In order to achieve this idea, the SDM harnesses the fact that PCA subspaces are linear subspaces, and therefore, for a set dimension d, can be defined on the Grassmannian [33]. If the source and target PCA subspaces are similar, then these subspaces, and their joint subspace — here referred to as the source-target subspace (calculated using the combined source and target datasets) — should all be 'close' on the Grassmannian. Mathematically, this is performed by calculating the averaged minimum correlation distance on the Grassmannian for a given dimension d,

$$SDM(d) = \frac{1}{2}\left(\sin\alpha_d + \sin\beta_d\right) \tag{B.1}$$

where α_d and β_d are the d^{th} principal angles between the source PCA and source-target PCA subspaces, and target PCA and source-target PCA subspaces respectively (where $\sin\alpha_d$ and $\sin\beta_d$ are the respective minimum correlation distances).

Equation (B.1) is useful as it is bounded on [0 1], meaning that small SDM values indicate that the source and target PCA subspaces are well-aligned at the d^{th} dimension with small principal angles, and values of one indicate that the source and target subspaces are orthogonal with $\alpha_d = \beta_d = \pi/2$. Gong *et al.* therefore proposed to use the SDM in a greedy manner in order to identify the optimal subspace dimension $\hat{d}$,

$$\hat{d} = \min\left(d \mid SDM(d) = 1\right) \tag{B.2}$$

where ideally $\hat{d}$ is not so high that the subspaces have orthogonal directions, but high enough that the dimensions preserve the variances in the datasets.

Appendix C. Rank of domain

The rank of domain (ROD) seeks to calculate the weighted symmetrised KL-divergences between two datasets, a source dataset[9] $X_s \in \mathbb{R}^{N_s \times D}$ and a target dataset $X_t \in \mathbb{R}^{N_t \times D}$, projected onto d-dimensional linear subspaces. If P_s and P_t are the basis of the two PCA subspaces then the principal angles and vectors can be calculated via singular value decomposition,

$$P_s^\mathsf{T} P_t = U \Gamma V^\mathsf{T} \tag{C.1}$$

where,

$$\theta_i = \arccos \gamma_i, \quad \boldsymbol{s}_i = (P_s U)_{\cdot,i}, \quad \boldsymbol{t}_i = (P_t V)_{\cdot,i} \tag{C.2}$$

where $\boldsymbol{\gamma} = \text{diag}(\Gamma)$ and the notation $(A)_{\cdot,i}$ returns the i^{th} column of the matrix A. The data distributions are projected using the principal vectors via $X_s^\mathsf{T} \boldsymbol{s}_i$ and $X_t^\mathsf{T} \boldsymbol{t}_i$. Once projected, the distributions are approximated as one-dimensional Gaussian distributions, an assumption that holds if PCA has identified fully independent components. Given that X_s and X_t have been centred, they will both have zero means, meaning that only the variances are required for the KL-divergences, defined as,

$$\sigma_{i,s}^2 = \frac{1}{N_s} \boldsymbol{s}_i^\mathsf{T} X_s^\mathsf{T} X_s \boldsymbol{s}_i, \quad \sigma_{i,t}^2 = \frac{1}{N_t} \boldsymbol{t}_i^\mathsf{T} X_t^\mathsf{T} X_t \boldsymbol{t}_i \tag{C.3}$$

so that with the Gaussian assumption, the ROD can be calculated in closed-form as,

$$ROD(S,T) = \frac{1}{d} \sum_i^d \theta_i \left(\frac{1}{2} \frac{\sigma_{i,s}^2}{\sigma_{i,t}^2} + \frac{1}{2} \frac{\sigma_{i,t}^2}{\sigma_{i,s}^2} - 1 \right) \tag{C.4}$$

Appendix D. Balanced distribution adaptation

Balance distribution adaptation (BDA), proposed by Wang *et al.* [34], performs domain adaptation via an MMD-based cost function that seeks to 'balance' the effect of minimising the MMD between the marginal (i.e. $P(X)$) and class conditional distributions (i.e. $P(X|\, y)$) of a source and target dataset. In addition, weightings are used to reduce the effect of class imbalance when minimising the distance between the class-conditional distributions. This distance is formed as,

$$D(\mathcal{D}_s, \mathcal{D}_t) \approx (1-\lambda) D(P(X_s), P(X_t)) + \lambda D\left(\alpha_s P(X_s \mid \boldsymbol{y}_s), \alpha_t P(X_t \mid \boldsymbol{y}_t)\right) \tag{D.1}$$

[9] In order to avoid notational clutter, the notation used in this appendix differs from the rest of the manuscript.

where $\lambda \in [0,\ 1]$ is a balance factor, where $\lambda \to 1$ assumes the datasets are more similar and the conditional distributions are more important to adapt [34]. Approximations of $\alpha_s = P(\boldsymbol{y}_s)/P(X_s)$ and $\alpha_t = P(\boldsymbol{y}_t)/P(X_t)$ are formed using the class prior of both datasets, leading to a weight matrix for each class W_c,

$$(W_c)_{i,j} = \begin{cases} \frac{P\left(\boldsymbol{y}_s^{(c)}\right)}{N_s^{(c)} N_s^{(c)}}, & x_i, x_j \in \mathcal{D}_s^{(c)} \\ \frac{P\left(\boldsymbol{y}_t^{(c)}\right)}{N_t^{(c)} N_t^{(c)}}, & x_i, x_j \in \mathcal{D}_t^{(c)} \\ \frac{-\sqrt{P\left(\boldsymbol{y}_s^{(c)}\right) P\left(\boldsymbol{y}_t^{(c)}\right)}}{N_s^{(c)} N_t^{(c)}}, & \begin{cases} x_i \in \mathcal{D}_s^{(c)},\ x_j \in \mathcal{D}_t^{(c)} \\ x_j \in \mathcal{D}_s^{(c)},\ x_i \in \mathcal{D}_t^{(c)} \end{cases} \\ 0, & \text{otherwise} \end{cases} \tag{D.2}$$

where $P\left(\boldsymbol{y}_s^{(c)}\right)$ and $P\left(\boldsymbol{y}_t^{(c)}\right)$ are the class priors on class c. The MMD-based cost function is constructed by separating out the marginal MMD distance (when $c = 0$ in equations (3) and (4)), and the conditional MMD distance (with M_c for $c = \{1, \ldots, C\}$ replaced with the weightings in equation (D.2)). By utilising the low-rank empirical kernel embedding $\tilde{K} = KAA^{\mathsf{T}}K$ [53] and regularisation constraints (a Frobenius-norm on the mapping weight A and subject to kernel PCA) the optimisation problem becomes,

$$\min_{A^{\mathsf{T}}KHKA = \mathbb{I}} = \operatorname{tr}\left(A^{\mathsf{T}} K \left((1-\lambda) M_0 + \lambda \sum_{c=1}^{C} W_c \right) KA \right) + \mu \operatorname{tr}\left(A^{\mathsf{T}} A \right) \tag{D.3}$$

where $A \in \mathbb{R}^{(N_s+N_t)\times k}$ are the mapping weights that map onto a k-dimensional latent space $Z = KA \in \mathbb{R}^{(N_s+N_t)\times k}$. From the regularisation constraints, μ controls the complexity of the mapping, $H = \mathbb{I} - 1/(N_s + N_t)\mathbf{1}$ is a centring matrix, $\mathbb{I}$ is an identify matrix and $\mathbf{1}$ a matrix of ones. By using a Lagrange multiplier approach, the optimisation problem can be solved as a generalised eigendecomposition,

$$\left(K \left((1-\lambda) M_0 + \lambda \sum_{c=1}^{C} W_c \right) K + \mu \mathbb{I} \right) A = KHKA\Phi. \tag{D.4}$$

where Φ are the Lagrange multipliers, and A the optimal mapping weights. It is noted that BDA seeks to perform domain adaptation on an unlabelled target dataset. This restriction means that in order to approximate the target conditional distribution the algorithm utilises psuedo-labels, i.e. $\mathcal{D}_t^{(c)} = \{\boldsymbol{x}_i : \boldsymbol{x}_i \in \mathcal{D}_t \wedge \hat{y}(\boldsymbol{x}_i) = c\}$ are the instances that belong in class c given the pseudo-target label $\hat{y}(\boldsymbol{x}_i)$. The psuedo-labels are initialised using a classifier trained on the source dataset and applied to the target dataset, and are updated iteratively by a classifier trained on the source data in the latent space Z_s which is applied to the target data in the latent space Z_t [34].

References

[1] P. Gardner, X. Liu, and K. Worden. On the application of domain adaptation in structural health monitoring. *Mechanical Systems and Signal Processing*, 138:106550, 2020.

[2] 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. *Mechanical Systems and Signal Processing*, 148:107141, 2021.

[3] 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. *Mechanical Systems and Signal Processing*, 148:107144, 2021.

[4] P. Gardner, L.A. Bull, J. Gosliga, N. Dervilis, and K. Worden. Foundations of population-based structural health monitoring, Part III: Heterogeneous populations, transfer and mapping. *Mechanical Systems and Signal Processing*, 149:107142, 2021.

[5] G. Tsialiamanis, C. Mylonas, E. Chatzi, N. Dervilis, D.J. Wagg, and K. Worden. Foundations of population-based structural health monitoring, Part IV: The geometry of spaces of structures and their feature spaces. *Mechanical Systems and Signal Processing*, 157:107692, 2021.

[6] O. Fink, Q. Wang, M. Svensén, P. Dersin, W.-J. Lee, and M. Ducoffe. Potential, challenges and future directions for deep learning in prognostics and health management applications. *Engineering Applications of Artificial Intelligence*, 92:103678, 2020.

[7] V. Zaccaria, M. Stenfelt, I. Aslanidou, and K.G. Kyprianidis. Fleet monitoring and diagnostics framework based on digital twin of aero-engines. volume 6: Ceramics; Controls, Diagnostics, and Instrumentation; Education; Manufacturing Materials and Metallurgy of *Turbo Expo: Power for Land, Sea, and Air*, 2018. V006T05A021.

[8] V. Zaccaria, A.D. Fentaye, M. Stenfelt, and K.G. Kyprianidis. Probabilistic model for aero-engines fleet condition monitoring. *Aerospace*, 7(6), 2020.

[9] M. Rahman, V. Zaccaria, X. Zhao, and K. Kyprianidis. Diagnostics-oriented modelling of micro gas turbines for fleet monitoring and maintenance optimization. *Processes*, 6(11), 2018.

[10] P. Moens, V. Bracke, C. Soete, S. Vanden Hautte, D. Nieves Avendano, T. Ooijevaar, S. Devos, B. Volckaert, and S. Van Hoecke. Scalable fleet monitoring and visualization for smart machine maintenance and industrial IoT applications. *Sensors*, 20(15), 2020.

[11] L. Basora, P. Bry, X. Olive, and F. Freeman. Aircraft fleet health monitoring with anomaly detection techniques. *Aerospace*, 8(4), 2021.

[12] G. Michau, T. Palmé, and O. Fink. Fleet PHM for critical systems: Bi-level deep learning approach for fault detection. In *PHM Society European Conference*, volume 4, 2018.

[13] G. Medina-Oliva, A. Voisin, M. Monnin, and J-B. Leger. Predictive diagnosis based on a fleet-wide ontology approach. *Knowledge-Based Systems*, 68:40–57, 2014.

[14] K. Hendrickx, W. Meert, Y. Mollet, J. Gyselinck, B. Cornelis, K. Gryllias, and J. Davis. A general anomaly detection framework for fleet-based condition monitoring of machines. *Mechanical Systems and Signal Processing*, 139:106585, 2020.

[15] S. Xu and H.-Y. Noh. PhyMDAN: Physics-informed knowledge transfer between buildings for seismic damage diagnosis through adversarial learning. *Mechanical Systems and Signal Processing*, 151:107374, 2021.

[16] D. Chakraborty, N. Kovvali, B. Chakraborty, A. Papandreou-Suppappola, and A. Chattopadhyay. Structural damage detection with insufficient data using transfer learning techniques. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, page 798147, 2011.

[17] J. Ye, T. Kobayashi, H. Tsuda, and M. Murakawa. Robust hammering echo analysis for concrete assessment with transfer learning. In *Proceedings of the the 11th International Workshop on Structural Health Monitoring*, pages 943–949, 2017.

[18] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*, 17(2), 2017.

[19] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun. Multi-layer domain adaptation method for rolling bearing fault diagnosis. *Signal Processing*, 157:180 – 197, 2019.

[20] Z. Wang, Z. Dai, B. Póczos, and J. Carbonell. Characterizing and avoiding negative transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11293–11302, 2019.

[21] P. Gardner and K. Worden. On the application of domain adaptation for aiding supervised SHM methods. In *Proceedings of the 12th International Workshop on Structural Health Monitoring*, pages 3347–3357, Stanford, USA, 2019.

[22] S.-X. Chen, L. Zhou, Y.-Q. Ni, and X.-Z. Liu. An acoustic-homologous transfer learning approach for acoustic emission–based rail condition evaluation. *Structural Health Monitoring*, page 1475921720976941, 2020.

[23] L.A. Bull, P. Gardner, N. Dervilis, E. Papatheou, M. Haywood-Alexander, R.S. Mills, and K. Worden. On the transfer of damage detectors between structures: An experimental case study. *Journal of Sound and Vibration*, 501:116072, 2021.

[24] C.T. Wickramarachchi, W. Leahy, K. Worden, and E.J. Cross. On metrics assessing the information content of datasets for population-based structural health monitoring. In *Proceedings of the European Workshop on Structural Health Monitoring*, 2021.

[25] M. Dhada, M. Girolami, and A.K. Parlikad. Anomaly detection in a fleet of industrial assets with hierarchical statistical modeling. *Data-Centric Engineering*, 1:e21, 2020.

[26] P. Gardner, L.A. Bull, N. Dervilis, and K. Worden. Kernelised Bayesian transfer learning for population-based structural health monitoring. In *Model Validation and Uncertainty Quantification, Volume 3*, pages 209–215, Cham, 2020. Springer International Publishing.

[27] P. Gardner, L.A. Bull, N. Dervilis, and K. Worden. On the application of kernelised bayesian transfer learning to population-based structural health monitoring. *Mechanical Systems and Signal Processing*, 167:108519, 2022.

[28] S.J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22:1345–1359, 2010.

[29] K. Weiss, T.M. Khoshgoftaar, and D. Wang. A survey of transfer learning. *Journal of Big Data*, 3:29, 2017.

[30] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2021.

[31] S.J. Pan, I.W. Tsang, J.T. Kwok, and Q. Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22:199–210, 2011.

[32] M. Long, J. Wang, G. Ding, J. Sun, and P.S. Yu. Transfer feature learning with joint distribution adaptation. In *2013 IEEE International Conference on Computer Vision*, pages 2200–2207, 2013.

[33] B. Gong, Y. Shi, F. Sha, and K. Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2066–2073, 2012.

[34] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen. Balanced distribution adaptation for transfer learning. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 1129–1134, 2017.

[35] Y. Ganin and V. Lempitsky. Unsupervised domain adaptation by backpropagation. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, pages 1180–1189, 2015.

[36] M. Long, Y. Cao, J. Wang, and M.I. Jordan. Learning transferable features with deep adaptation networks. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, page 97–105, 2015.

[37] M. Long, Y. Cao, J. Wang, and M.I. Jordan. Conditional adversarial domain adaptation. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

[38] A. Karbalayghareh, X. Qian, and E.R. Dougherty. Optimal Bayesian transfer learning. *IEEE Transactions on Signal Processing*, 66(14):3724–3739, 2018.

[39] K. Worden, G. Manson, and D.J. Allman. Experimental validation of a structural health monitoring methodology: Part I. Novelty detection on a laboratory structure. *Journal of Sound and Vibration*, 259(2):323 – 343, 2003.

[40] G. Manson, K. Worden, and D.J. Allman. Experimental validation of a structural health monitoring methodology: Part II. Novelty detection on a gnat aircraft. *Journal of Sound and Vibration*, 259(2):345 – 363, 2003.

[41] G. Manson, K. Worden, and D.J. Allman. Experimental validation of a structural health monitoring methodology: Part III. Damage location on an aircraft wing. *Journal of Sound and Vibration*, 259(2):365 – 385, 2003.

[42] R.J. Barthorpe, G. Manson, and K. Worden. On multi-site damage identification using single-site training data. *Journal of Sound and Vibration*, 409:43–64, 2017.

[43] R.J. Barthorpe. *On model- and data-based approaches to structural health monitoring*. PhD thesis, University of Sheffield, 2011.

[44] L.A. Bull, K. Worden, R. Fuentes, G. Manson, E.J. Cross, and N. Dervilis. Outlier ensembles: A robust method for damage detection and unsupervised feature extraction from high-dimensional data. *Journal of Sound and Vibration*, 453:126 – 150, 2019.

[45] G. Tsialiamanis, D.J. Wagg, P. Gardner, N. Dervilis, and K. Worden. On partitioning of an SHM problem and parallels with transfer learning. In *Proceedings of IMAC XXXVIII International Conference on Modal Analysis*, Houston, USA, 2020.

[46] P. Gardner, L.A. Bull, N. Dervilis, and K. Worden. Overcoming the problem of repair in structural health monitoring: Metric-informed transfer learning. *Journal of Sound and Vibration*, 510:116245, 2021.

[47] J. Poole, P. Gardner, N. Dervilis, L.A. Bull, and K. Worden. On normalisation for domain adaptation in population-based structural health monitoring. In *Proceedings of the 13th International Workshop on Structural Health Monitoring*, 2021.

[48] G.M.Y. Dauphin, X. Glorot, S. Rifai, Y. Bengio, I. Goodfellow, E. Lavoie, X. Muller, G. Desjardins, D. Warde-Farley, P. Vincent, A. Courville, and J. Bergstra. Unsupervised and transfer learning challenge: a deep learning approach. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, volume 27 of *Proceedings of Machine Learning Research*, pages 97–110, Bellevue, Washington, USA, 2012.

[49] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In *2013 IEEE International Conference on Computer Vision*, pages 2960–2967, 2013.

[50] X. He. Quantum subspace alignment for domain adaptation. *Phys. Rev. A*, 102:062403, 2020.

[51] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schöolkopf, and A. Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, 2012.

[52] M. Long, J. Wang, G. Ding, S.J. Pan, and P.S. Yu. Adaptation regularization: a general framework for transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 26: 1076–1089, 2014.

[53] B. Schölkopf, A. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299–1319, 1998.