Wind Energ. Sci., 10, 2889–2901, 2025
https://doi.org/10.5194/wes-10-2889-2025
© Author(s) 2025. This work is distributed under
the Creative Commons Attribution 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE

Population-based structural health monitoring: homogeneous offshore wind model development

Innes Murdo Black[1], **Moritz Werther Häckell**[2], **and Athanasios Kolios**[1]

[1]Department of Naval Architecture, Ocean and Marine Engineering, University of Strathclyde, 16 Richmond St, Glasgow, G1 1XQ, Scotland
[2]Ramboll, Jürgen-Töpfer-Straße 48, 22763 Hamburg, Germany

Correspondence: Innes Murdo Black (innes.black@vattenfall.com)

Received: 4 October 2022 – Discussion started: 10 October 2022
Revised: 30 June 2023 – Accepted: 13 July 2023 – Published: 8 December 2025

Abstract. Population-based structural health monitoring (PBSHM) involves transferring knowledge from one structure to a different structure so that predictions about the structural health of each of the members of the population can be inferred. One of the most important aspects of PBSHM involves using the information on the source domain structure and the target domain structure to create an effective classifier. Domain adaptation is a subcategory of transfer learning that can create a general classifier using both the source and the target domain structures to create an enhanced overall classifier of the entire population. This paper presents a novel domain adaptation model for PBSHM in offshore wind.

1 Introduction

Moving beyond detecting damage on a single structure to diagnosing damage in an entire population raises the issue of acquiring data relating to each of the structures. One of the main concerns with this is the large cost associated with obtaining the information necessary to determine any damage to the structures. Population-based structural health monitoring (PBSHM) seeks to reduce this cost by developing methods that share the information between the structures. The concept of PBSHM is introduced in Bull et al. (2021), Gosliga et al. (2021), and Gardner et al. (2021). If the population of the structures is homogeneous, where the structures are nominally identical, then it may be possible to establish a general model that is common across all structures. Conversely, even if the models are heterogeneous, dissimilar structures, it may be possible to transfer select types of damage across the structures. The most promising technology that will allow for this transfer of information is found in the machine learning discipline of transfer learning.

The standard of similarity between the structures is indicative of the level of knowledge transfer between the structures. This can be achieved through quantifying how the structures are similar and where the similarities lie. This determines what type of machine learning approach is necessary. For this study, the geometry, topology, operation, and material of the offshore structures are the same. The main observation from this is that the population is of a strong homogeneous nature. Within a strongly homogeneous population, all the structures have the same material, geometry, and topology (this refers to the components for all the parts in the structure). This implies that all the structures are the same model and are made in the wind farm. The variation within the wind farm is due to the operational state, the location of the wind turbine, and manufacturing defects.

The five references provided focus on the area of renewable energy and the challenges in and solutions of implementing transfer learning. Y. Li et al. (2021b) propose a strategy to tackle small datasets using parameter-based transfer learning. The authors suggest a new model based on transfer learning for wind turbine diagnosis with small-scale data. The model can take the operational information from other wind turbines into account. Chen et al. (2021) propose a framework using unsupervised TrAdaBoost learning on SCADA (supervisory control and data acquisition) data for WT (wind turbine) fault diagnosis. The main observation is that TrAdaBoost shows its superior performance in dealing with data imbalance and different distributions. Gardner et al. (2022) focus on the application of machine learning algorithms for structural health monitoring and highlight the importance of domain adaptation in improving the performance of these algorithms. The authors propose a hybrid machine learning model that shows improved performance in several populations of experimental and numerical structures. Schröder et al. (2022) introduce a transfer-learning-based approach to include physics into data-driven normal behaviour monitoring models. An artificial neural network with an auto-encoder is used in this study to examine 1 month of raw SCADA data. Jamil et al. (2022) propose a control strategy for deep transfer learning for fault detection on rotating machinery. The paper applies integrated signal processing to vibration signals. The main observation is that the performance is significantly improved by reducing negative transfer, and less data are required using this technique instead of standard deep learning.

The motivation for this study is to provide a solution for low-cost monitoring, where only a few wind turbine generators (WTGs) are instrumented with sensors as opposed to the entire fleet. A low-cost monitoring strategy for offshore wind can provide numerous benefits to wind farm developers and operators. Implementing a cost-effective monitoring system can reduce the overall cost of monitoring and increase the reliability of the wind farm. The real-time data provided by a low-cost monitoring system can also aid in the early detection of performance issues and prompt maintenance, resulting in reduced downtime and improved maintenance practices. The monitoring data can also be used to better understand the wind resource and its variability, leading to improved wind farm design and operation. Furthermore, real-time monitoring can provide early warning of potential safety issues, improving safety for workers and maintenance personnel on offshore wind farms. The consequence of a low-cost monitoring strategy is that assumptions will have to be made on structures that are not instrumented, so the low-cost technique is developed based on a general classifier as apposed to individual models for a monitoring strategy that has the entire wind farm instrumented.

This work focuses on the case of strong homogeneous transfer, with four different machine learning models under consideration. Three models employ supervised domain adaptation techniques, a subcategory of the transfer learning branch, and the last model uses ensemble learning. Transfer learning is an approach with the purpose of improving the performance of the learner by transferring between different domains. Domain adaptation assumes that there are labelled data in the source domain that can be utilised to aid in the regression of the target domain by mapping the two domains into a common latent space on which the data distributions are coincident. There are assumptions of domain adaptation, where the input and output feature dimensions are consistent in the source and target domain. This means that structure one must have the same features as structure two in the wind farm. The former method of using an ensemble technique aims to improve the final prediction by grouping the views from the regression models and reaching a consensus.

2 Background

PBSHM involves mapping data and labels from different structures within the population so that a general classifier can be inferred across the entire population. As a result, asset management can potentially be performed digitally for any individual in the population. This section intends to define applicable forms of PBSHM.

For PBSHM it is pertinent to define the contextual difference between homogeneous and heterogeneous populations. This syntax is borrowed from graph theory where the names clearly explain how structures can be represented by attributes. To determine whether two systems are similar enough for knowledge transfer, it is unpracticable to consider every property or dimension of the structure; for example, comparing the geometric similarity of two structures using 3D, finite-element (FE), or computer-aided design (CAD) models of the structure directly would be computationally inefficient. For this study's desired goal, it is more efficient to consider only the properties and dimensions that have a significant effect on the transferability of knowledge.

Differences within the population occur for a magnitude of reasons, and structures are deemed different due to various properties. This can lead to groups of heterogeneous populations. The aspects of homogeneity are visualised in Fig. 1. This approach will highlight the four main sources of differences within a structure, namely geometry, topology, material, and operation:

- Geometry links to the shape and size of the structure within the population.
- Topology depicts the construction, connection, and location of the components in the structure.
- Material relates to the different material classes and specific materials with the associated properties for the structure in the population.
- Operation refers to the different states the operator can curtail the asset to.

Adapting the definitions from graph theory for PBSHM, where a topologically homogeneous population is defined as a group of structures where the geometry σ_m and the material σ_m properties for the nodes and edges of the associated graph can be taken from the base distribution $p(\sigma_\mathrm{m})$, the probability mass of the distribution $p(\sigma_\mathrm{m})$ defines the small differences between the individuals within the population. A strong homogeneous population would have a unimodal distribution with low dispersion for the geometrical, topological, and material properties. With the strictest, perfect form

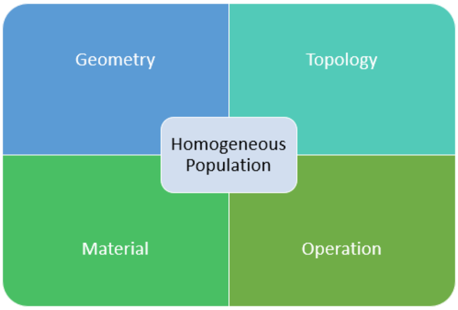

Figure 1. Categories of heterogeneous populations within the PBSHM framework. In the centre, all four categories are alike to a sufficient degree, indicating that a homogeneous population exists. It is noted that all four attributes can influence each other separately to create independent heterogeneous populations. Adapted from Gardner et al. (2021).

of homogeneous composition, the underlying distribution of the population is identical. The latter is uncommon, but this assumption can be made if you want to apply conventional machine learning (ML) methods trained on one structure to another structure. Applying these conventions to the population and categorising the individuals within the populations help to determine the difficulty of transferability.

Notable differences in the observable data may occur outside the structural properties of the individuals within the population beyond the categories previously discussed. These differences relate to how the data acquisition and any processing to obtain the features are conducted. A classic example of this would be sensor placement. This will lead to differences in the distribution of the data even though it is placed in the "same" position. Manufacturing and installation differences will also contribute to small variations in the homogeneity of the data distribution.

3 Transfer learning

Transfer learning technologies offer several opportunities for dealing with scenarios where the population forms domains, and distributions are different for each member when training and evaluating the model (Pan and Yang, 2010). Unlike multi-task learning, where the objective is to learn multiple tasks across different domains (Zhang and Yang, 2018), transfer learning utilises knowledge from the source to improve predictions on the target task, in our case using the damage-equivalent moments (DEMs) from two separate WTGs to create an improved general model. This type of learning is what makes PBSHM achievable. Even when performing in the homogeneous population scenario, variations in the structure, such as location, will lead to differences in the data distributions. Learners trained on one structure will not apply their training to another structure in the population. Formal definitions of transfer learning and transfer learning technologies are discussed in this section, with an entire subsection dedicated to domain adaptation.

3.1 Definitions

Domain. A domain $D = (\chi, p(X))$ is an object made up of a feature space χ and a marginal probability distribution $p(X)$ over feature data $X = x_{i=1}^{N}$, which are a bounded sample from χ. These are the SCADA data for one structure in the context of this report.

Task. $T = (\Upsilon, f(\cdot))$, the DEM for one structure, is an object made up of a label space Υ and predictive function $f(\cdot)(p(x|Y))$ in probabilistic terms and can be inferred from training data $X = x_{x=i,y=i}^{N}$, with $X_i = \sum \chi$ and $y_i = \sum \Upsilon$, noting that both χ and Υ are distributions, not individual observations, which are made up of finite sample sets X and Y. In the case of source domain datasets $D_\mathrm{s} = (\chi_{i,\mathrm{s}}, y_{i,s})^N$ with $x_{i,\mathrm{s}} = \sum \chi_\mathrm{s}$ and $y_{i,\mathrm{s}} = \sum \Upsilon_\mathrm{s}$, and, similarly for the target domain, $D_\mathrm{t} = (\chi_{i,\mathrm{t}}, y_{i,\mathrm{t}})^N$ with $x_{i,\mathrm{t}} = \sum \chi_\mathrm{t}$ and $y_{i,\mathrm{t}} = \sum \Upsilon_\mathrm{t}$ (Pan and Yang, 2010). Given these artefacts, one can theoretically conduct transfer learning.

Transfer learning. For transfer learning there must be a given source domain D_s with an associated task T_s and a target domain D_t with a task T_t. The objective is to improve the target predictive function $f_\mathrm{t}(\cdot)$ in T_t by utilising the knowledge from the source, assuming $D_\mathrm{s} \neq D_\mathrm{t}$ and/or $T_\mathrm{s} \neq T_\mathrm{t}$ (Zhang and Yang, 2018).

Homogeneous transfer. Homogeneous transfer learning assumes that $D_\mathrm{s} = D_\mathrm{t}$ and $T_\mathrm{s} = T_\mathrm{t}$, meaning the attributes are the same. A subcategory of homogeneous transfer is strong homogeneous transfer, where the domain and task are similar; hence $D_\mathrm{s} \cong D_\mathrm{t}$ and $T_\mathrm{s} \cong T_\mathrm{t}$.

Heterogeneous transfer. Heterogeneous transfer learning is when the domain, feature, and task space are non-identical; hence $D_\mathrm{s} \neq D_\mathrm{t}$, $T_\mathrm{s} \neq T_\mathrm{t}$, and $x_\mathrm{s} \neq x_\mathrm{t}$, respectively. It can also assume that $y_\mathrm{s} \neq y_\mathrm{t}$.

Domain adaptation. Domain adaptation is relevant when the inference for the target domain D_t, task T_t, and target predictive function $f_\mathrm{t}(\cdot)$ is improved given the source domain D_s and task T_s, assuming $x_\mathrm{s} = x_\mathrm{t}$ and $y_\mathrm{s} = y_\mathrm{t}$ but $p(x_\mathrm{s}) \neq p(x_\mathrm{t})$.

To contextualise these definitions in the form of PBSHM for wind turbines, homogeneous transfer learning is a situation where both the source space and the target space are the same. This is a situation where the context is the problem between similar assets. This could be where the wind turbines are the same but have different distributions due to sensor placement and location, to name a few. Hence $D_\mathrm{s} \cong D_\mathrm{t}$ and $T_\mathrm{s} \cong T_\mathrm{t}$. Heterogeneous transfer learning is applied when the features are dissimilar. An example in the wind turbine industry would be when using data from two different wind turbine designs, e.g. a monopile foundation and a jacket structure. In this case, the features and tasks will be dissimilar; hence $D_\mathrm{s} \neq D_\mathrm{t}$ and $T_\mathrm{s} \neq T_\mathrm{t}$.

3.2 Transfer learning technologies

There is a continually growing variety of transfer learning technologies. This section aims to briefly describe transfer learning, but the focus is on fundamental differences in the approaches of a subcategory called domain adaptation where parameter-, instance-, and feature-based approaches are described. See Friedjungová and Jiřina (2018) for a more comprehensive discussion on transfer learning.

Starting with a typical approach of deep learning and artificial neural networks, transfer learning technologies have been developed using fine-tuning. This methodology seeks to learn based on the parameter weights during a particular set of layers in the artificial neural network. The artificial neural network is trained on the domain D_s, and some of the layers are fixed. The remaining unfixed layers are trained using the target domain D_t. Examples of this are conducted in Black et al. (2022), Gao and Mosalam (2018), and Dorafshan et al. (2018).

Another approach to transfer learning is knowledge graphs, where the aim is to find objects that define specific entities and their interrelationships. This has been particularly successful in search engines incorporating semantic searches. Currently, knowledge graphs have been integrated as training data for machine learning models (Hamaguchi et al., 2018; Nickel et al., 2016).

Like knowledge graphs, the ontologies' goal is to give representations of entities that describe all the interdependencies and interactions. Ontologies are useful for outlining knowledge about specific domains. Most importantly, ontologies are helpful for explaining concepts and sharing information. If a new project is undertaken, ontologies can be reused or transferred to help identify more efficient processes. In the context of PBSHM, an ontology is knowing what types of techniques and methods are the most appropriate for one system to another. Ontologies have been explored in multiple industries, including structural health monitoring (SHM) (R. Li et al., 2021; Tsialiamanis et al., 2021; Anderlik et al., 2010).

3.3 Domain adaptation

Domain adaptation is a subclass of transfer learning with the aim to transfer the feature space between the source and target domains, based on the assumption that the marginal distributions of $p(x_\mathrm{s}) \neq p(x_\mathrm{t})$ are not the same. This type of technique is primarily used in homogeneous transfer learning where the source domain and target domain are similar. There are three main approaches to domain adaptation, namely parameter, feature, and instance based.

Parameter-based domain adaptation involves transferring the parameters of a model trained on the source domain D_s and fine-tuning them to develop an equivalent model for the task domain D_t.

Feature-based domain adaptation techniques are designed taking into account the research of common features that have similar attributes concerning the source T_s and target T_t task. A new feature, often called the encoded feature space, is built with a projecting application, which aims to correct the difference between the source $p(x_\mathrm{s})$ and target $p(x_\mathrm{t})$ distributions. The task is then considered to be in an encoded space.

For instance-based domain adaptation, the general principle is to redistribute the labelled training data to correct the differences between the source $p(x_\mathrm{s})$ and target $p(x_\mathrm{t})$ distributions. This re-weighting consists of multiplying the individual loss of each training instance by a positive weight. The re-weighted training instances are then used directly to learn the task.

3.4 Negative transfer

One of the major drawbacks when performing transfer learning between WTGs is if the information is incorrectly detailed from one domain to another, as this can reduce the performance of the general learner when compared to the learning from the target domain alone. This phenomenon is known as negative transfer and is the most prominent when the source, D_s, and the target, D_t, domains are the most dissimilar, e.g. heterogeneous. The fundamental idea of transfer learning is that there must be some shared information across domains. This may be hard to contextualise when data are unlabelled or the tasks are dissimilar.

Negative transfer raises an important question: when is it right to transfer knowledge? This motivates the reasoning behind developing a measure of similarity of structures. The case study provides information on the heterogeneity of the data used in this work. This study reinforces our understanding of the data and helps mitigate the issue of negative transfer as we become aware of where differences in the distributions lie.

4 Methodology

This study uses data from the Wikinger wind farm, where there is only a select number of operational wind turbines that have CMS (condition-monitoring system) with strain gauges installed; to be specific, 4 out of 64. This section aims to develop a variety of models that can perform PBSHM using the structures of two structures as the source and target in the pursuit of a general classifier for the entire farm. The population form is the DEM on the foundation of the structure, which can be used to determine the fatigue life. This section starts with how the SHM data are gathered, what the population form is for this study, and what the definition of the population is. This is then used to detail the individual models, including how the data are pre-processed, the model development, and the error metrics used.

4.1 The data

The measurement data used as input for the low-cost monitoring technique comprise two different frequencies of data: 25 Hz CMS data and SCADA data at 10 min averages. The next subsection describes the process of determining the damage-equivalent moments (DEMs). The WTGs available for this study are equipped and unequipped with strain gauges (SGs). The three positions with SGs are wind turbines (WTs) 1, 2, and 3.

SCADA systems are provided on all WTGs, and, depending on the feature, the resolution varies. This encompasses meteorological information at the hub height, such as wind speed, wind direction, temperature, and pressure. The SCADA data also cover the operational signals, such as power production, the pitch angle of the individual blades, and the rotor rotational speed.

To increase the value of a low-cost monitoring programme, transferring knowledge that is unavailable in other wind turbines can provide insight into and confidence in other assets. If one can infer knowledge accurately on another WT, then one can save money by installing strain gauges on a fraction of the WTGs. Based on this principle, the population form is the DEM where only four WTGs have the CMS strain gauges installed.

There is a considerable number of model and feature spaces that can be applied to represent a population form. For a wind turbine, the form could range from wind turbine power curves to frequency responses. But, in this case, the form is fatigue-damage-equivalent moments for the jacket support structure. The entire population in this study has the same geometry and material, with small deviations in topology due to the location.

4.1.1 Fatigue-damage-equivalent moments

The condition-monitoring system calculates the forces from the strain gauges installed on the foundation. From these forces, the damage-equivalent loads are produced. The two-phase operation is as follows:
Phase 1 – calculation of forces from strain.

1. Run dynamic ROSA (Ramboll Offshore Structural Analysis) simulation (Ramboll, 2018).
2. Extract stress at selected element via Fatima.
3. Calculate strain using Hooke's law.
4. Calculate forces with internal functionality.
5. Compare forces with extracted forces.

Phase 2 – calculation of DEM from forces.

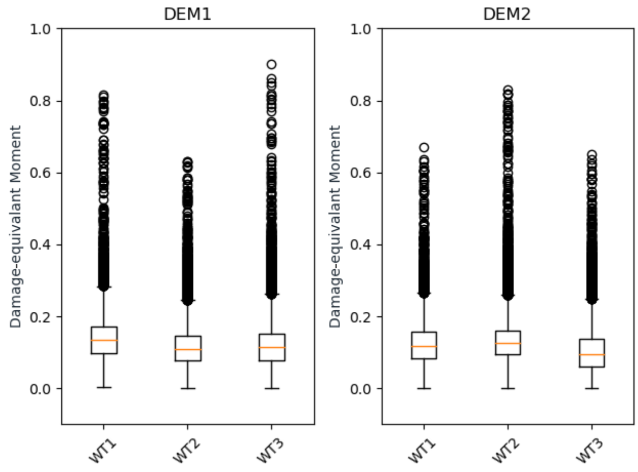

Figure 2. Normalised damage-equivalent moment boxplot of the two orthogonal directions for all three of the wind turbines.

1. Gather the applicable force location from the sensor location.
2. Calculate the cyclical forces at that sensor.
3. Apply the ASTM E1049-85 rainflow-cycle-counting algorithm (ASTM, 2017).
4. Apply a scale factor to force accumulation.
5. Sum the damage accumulation over the cycles to calculate the DEM.

4.1.2 Homogeneous population

A homogeneous population in the context of offshore wind is one where the task distributions are similar for all WTGs, which, depending on the features selected, may have similar feature spaces. This makes homogeneous populations ideal candidates for domain adaptation methods and for demonstrating the effectiveness of transfer learning for PBSHM. This section presents a homogeneous population of three wind turbines located in the Wikinger wind farm. All the structures are of the same design and capacity; hence, they have the same material and geometry. The damage-equivalent load histogram is presented in Fig. 2.

The three structures can be considered to be a homogeneous population, as they are structurally similar in their representation, and the material and geometry parameters can be described by a unimodal distribution with low deviation. The SHM problem presented here is due to small deviations in the DEM which arise from the location and the operational context of the individual wind turbines within the wind farm. Figure 3 displays two-dimensional heat maps of the DEM amplitudes for various operational features of the three wind turbines.

Small deviations in the overall distributions are highlighted in Fig. 2. One of the contributors to these results is the

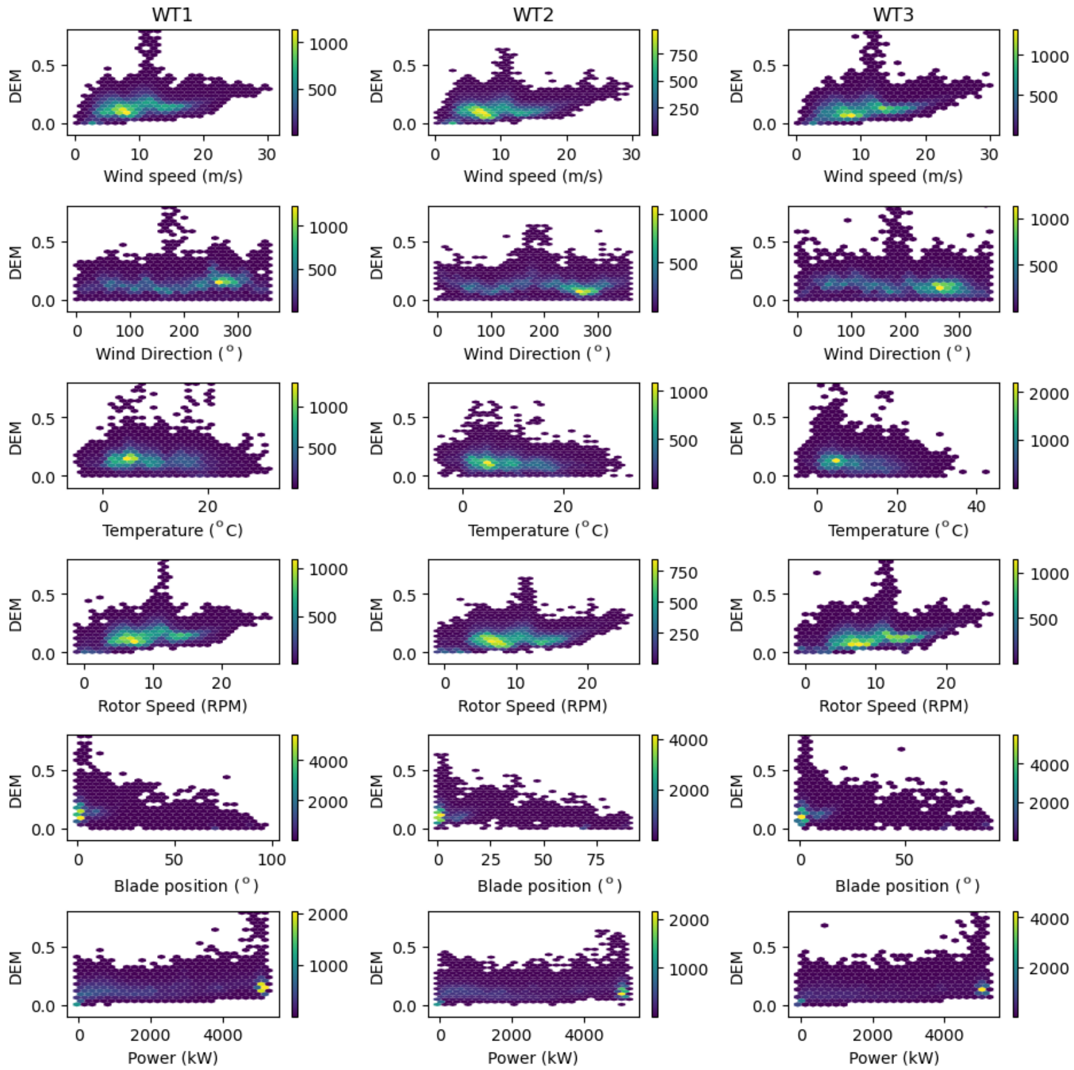

Figure 3. This series of figures highlights the operational and environmental effects on the population of the wind farm. This includes the power, blade position, wind speed, rotor speed, and temperature against the fatigue-damage-equivalent moments in direction 1.

location, as the entire wind farm has deviations in the water depth. The design of the individual WTs does not require the topology and geometry to be altered for single locations but rather in design clusters spanning a range of water depths. One consistency is the height of the transition piece. It needs to be at a constant height across the population, meaning that there are variations in the length of the foundation which influence the dynamics. Another aspect is the operation and metrological differences. Figure 3 displays a heat map of these features for each of the WTs, which are influenced by the location and the degree of turbulence intensity based on the direction of the wind. Navigating these deviations in the distribution is the aim of the general model development.

The Fréchet number is a numerical method of determining the similarity of a population form for a given domain in a metric space. The Fréchet distance is a popular measure of similarity between the two domains and is calculated by

$$\Delta = \|\mu_\mathrm{s} - \mu_\mathrm{t}\|_2^2 + T_\mathrm{r}\left(|\boldsymbol{\Sigma}_\mathrm{s} + \boldsymbol{\Sigma}_\mathrm{t} - 2(\boldsymbol{\Sigma}_\mathrm{s} \cdot \boldsymbol{\Sigma}_\mathrm{t})^{1/2}\right), \tag{1}$$

where μ_s and μ_t are the mean along the source and target along the first axis, respectively, and $\boldsymbol{\Sigma}_\mathrm{s}$ and $\boldsymbol{\Sigma}_\mathrm{t}$ are the covariance matrices of the source and target domain datasets, respectively. This considers the location and the ordering of the points along the points of both domains. For homogeneous populations, the value of the Fréchet distance will be 0, for strong homogeneous distributions the value will be small, and for heterogeneous populations the value will be large. The value is dependent on the length of the instances in the domain and the magnitude of the values.

By applying this to the DEM for each of the wind turbines, the results in Table 1 indicate that the population form is homogeneous with values close to 0. This is a numerical rep-

Table 1. Fréchet distance for all three wind turbines.

Fréchet distance			
WT 1	0.00		
WT 2	13.48	0.00	
WT 3	14.48	14.92	0.00

Table 2. Statistical values of both DEMs for all three wind turbines.

	DEM 1			DEM 2		
	WT 1	WT 2	WT 3	WT 1	WT 2	WT 3
Mean	0.1382	0.1503	0.1193	0.1233	0.1092	0.1049
SD	0.0634	0.0576	0.0629	0.0588	0.0623	0.0623
Min	0.0015	0.0015	0.0015	0.0017	0.0015	0.0017
Max	0.8158	0.6304	0.9024	0.6669	0.8286	0.6494

resentation, and it is limited in the scope of interpretation on the homogeneity. This is a technique that mathematically reinforces the decision after observation of the data. Nevertheless, in the following section, the location is the main source of heterogeneity for this population form, and breaking this down into more detail will provide greater insight into the small deviations in heterogeneity observed in Table 1.

To contextualise this, some statistical measures have been calculated based on the normalised DEM, and these are represented in Table 2. This includes the mean, standard deviation (SD), and minimum and maximum of the DEM in both orthogonal directions. There is an almost identical nature to the three WTGs; WT 2 does have higher and lower maximums and minimums, but both the mean and the standard deviations are identical to four significant figures.

4.2 Models

An artificial neural network (ANN) is an optimal base model suited for stochastic problems to estimate the DEM without direct measurement. Sensors often cannot be placed on the sub-structure, and SCADA data are usually gathered in the tower and nacelle; hence, this does not provide direct information on important parameters of the foundation's structural behaviour. Therefore, a general model using the SCADA data to determine the DEM would provide excellent potential in the application on wind turbine foundation monitoring. This subsection briefly describes an ANN and the differences that the domain adaptation models make to the original ANN model.

Three domain adaptation methods are implemented in this study that all use a base model of an ANN and are altered based on their specific procedures. However, before these procedures are explained, the original model is developed using the data from one wind turbine only. No transfer learning is carried out to generate this model, only hyperparameter optimisation using WT 1.

The architecture of an ANN is made up of hidden layers, where each layer has a density of neurons attached to it. Techniques such as dropout can be introduced to aid in removing bias within the architecture, increasing weights (w) to 0 and 1 such that the sum of the weights remains constant. The individual neurons have a synaptic weight associated with them, which can be represented as an activation function. For this type of problem, the rectified linear unit (RELU) is used. Weights w are associated with each activation function, and the entire network is curated using the Adam optimiser (Kingma and Ba, 2015).

4.2.1 CORrelation ALignment (CORAL)

CORAL (Sun et al., 2015) is a feature-based domain adaptation method with the aim of minimising the domain shift from the source D_s to the target D_t by aligning the second-order statistics of the source and target distributions. The method transforms the source features to minimise the Frobenius norm (Lord, 1999) between the correlation matrix, the input target data, and the transformed input source data. The transformation is described by the following optimisation:

$$\min_A \| A^\mathrm{T} \mathbf{C}_\mathrm{s} A - \mathbf{C}_\mathrm{t} \|_F^2, \tag{2}$$

where A is the feature transformation matrix such that C_s and $\mathbf{C}_\mathrm{t}$ are the correlation matrices of the source and target data, respectively. The solution of this operation can be written in explicit form, and the feature transformation is computed in four steps:

$$\mathbf{C}_\mathrm{s} = \mathrm{Cov}(X_\mathrm{S}) + \lambda I_\mathrm{P}, \tag{3}$$

$$\mathbf{C}_\mathrm{t} = \mathrm{Cov}(X_\mathrm{T}) + \lambda I_\mathrm{P}, \tag{4}$$

$$X_\mathrm{s} = X_\mathrm{S} C S_\mathrm{S}^{-1/2}, \tag{5}$$

$$X_\mathrm{s} = X_\mathrm{S} C S_\mathrm{T}^{1/2}, \tag{6}$$

where λ is the regularisation parameter.

4.2.2 TwoStageTrAdaBoostR2

The TwoStageTrAdaBoostR2 algorithm is an instance-based domain adaptation method suited to regression tasks (Pardoe and Stone, 2010). This method is characterised by the "reverse boosting" principle where the weights of the source instances predicted a decrease at each boosting iteration, and one of the instances increases. The "two-stage" version of the TrAdaBoostR2 algorithm is where the weights of the source and target instances are carried out separately (Pardoe and Stone, 2010). In the first stage, the weights of the source instances are frozen, but the ones of the target instances are updated according to the classical AdaBoost.R2 (Dai et al., 2007). In the second stage, the weights of the target instance are now fixed, whereas the ones of the source are updated according to TrAdaBoostR2. During each first stage, a cross-validation score is computed with the labelled target data.

The cross-validation score is used to determine the most effective estimator within all boosting iterations. This algorithm performs the following steps:

- Normalise the weights $\sum w_s + \sum w_t = 1$.
- Fit an AdaBoost.R2 estimator $f_s()$ on the source, and target labelled data (x_s, y_s) and (x_t, y_t) with the respective importance initial weights w_s and w_t. During the training of AdaBoost.R2, the weights of w_s are frozen.
- Compute the cross-validation score on (x_t, y_t).
- Compute the error vectors

$$e_s = L(f(X_s), y_s), \tag{7}$$
$$e_s = L(f(X_s), y_s). \tag{8}$$

- Normalise the vectors

$$e_s = e_s / \max_{e \in e_s \cup e_t}, \tag{9}$$
$$e_s = e_s \max_{e \in e_s \cup e_t}. \tag{10}$$

- Update the source and target weight

$$w_s = w_s \beta_s^{e_s} / Z, \tag{11}$$
$$w_t = w_t / Z, \tag{12}$$

where Z is the normalising constant, and B_s is chosen so that the sum of the weights is equal to $\frac{n_t}{n_t + n_s} + \frac{t}{N-1}\left(1 - \frac{n_t}{n_t + n_s}\right)$ with t being the current boosting iteration number. B_s is located with a binary search.

- Return to the first step, and loop until the number of boosting iterations is reached.

The general model is selected by the best estimator according to the cross-validation.

4.2.3 RegulartransferANN

RegulartransferANN (Chelba and Acero, 2004) is a parameter-based domain adaptation method. This assumes that an effective global estimator can be obtained using ladled target data. The aim consists of fitting the neural network on the target data based on the objective function which is regularised by the Euclidean distance of both the source and the target parameters:

$$\beta_t {=}_{\beta_1, \ldots \beta_D} \| f(X_t, \beta) - y_t \|^2 + \sum_{i=1}^{D} \lambda_i \| \beta_i - \beta_{Si} ||^2, \tag{13}$$

where the estimation function is f with D network layers. B_t is related to the target parameters, and β_s is the source neural network parameters:

$$\beta_s {=}_{\beta} \| f(X_s, \beta) - y_s \|^2. \tag{14}$$

The trade-off parameter is λ_i, where training is biased towards source or target domains depending on the associated weighting.

4.3 Pre-processing

The task of training models for predictions that involve multiple different data streams from different structures requires coordination so that effective ML modelling can take place. Several issues arose when working with different data streams in ML. This section discusses the process of dealing with these issues by using feature, selection, projection, and data cleaning. The first barrier that prevents effective ML modelling in this problem is that sensors tend to break, just like most industrial components. This meant identifying a suitable time frame, where the maximum period of operational uptime should be met. The optimal period was between 2 October 2018 and 2 July 2019 (9 months). Secondly, synchronising the CMS data to the SCADA data had to be conducted, and matching the instances of 10 min intervals was the next step.

After synchronising the data-matching instances for all WTs, the next stage was to perform data cleaning. This procedure involved implementing previous value imputation in the place of missing values and NaNs. At this point, there were 0 missing values and over 300 features. The next step was to reduce this to suit the needs of the ML task.

For the feature selection process, the inputs were taken from the SCADA data since these are provided on all WTGs. The final features comprised hand-picked features and statistically relevant features. The outputs were the DEMs in the two orthogonal directions from the CMS systems' strain gauge rings. Now that both the input and the output data have been established, the feature projection is implemented where the features are normalised from 0–1.

Lastly, the datasets were split into training and test data to a ratio of 80 % to 20 %, respectively. The source domain was WT 1, and the target domain was WT 2 using 80 % of the datasets for the domain adaptation model. Testing was conducted using the subsequent 20 % of both the source and the domain datasets and the entirety of WT 3. In summary, SCADA data features were used as inputs, with WT 1 as the source domain and WT 2 as the target domain, where the general model makes estimations on the DEM in the two orthogonal directions. Testing takes the remaining 20 % of the source and target domains and the entire dataset of WT 3 to validate the results using the same input and output features for all WTGs.

4.4 Model development

The method implemented involves three stages. Stage one involved defining the most suitable ANN and source domain, and this became the basis for all the subsequent experiments as the architecture is mimicked. Stage two entailed hyperparameter tuning the remaining experiments using the same architecture from the ANN and application of the three domain adaptation algorithms. Stage three investigated the optimal model further by altering the source and domain data.

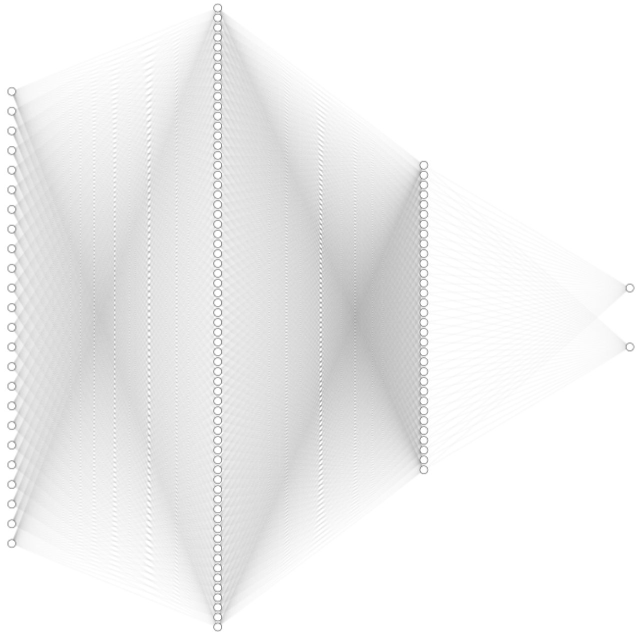

Figure 4. An artificial neural network architecture with 2 hidden layers with a density of 64 and 32, a batch size of 1, and a dropout ratio of 0.2.

To establish a standardised process for all the experiments, the pre-processing procedure was applied to all three WTGs, as this provides a standard platform for model development and increases the performance of the model. This also provides a consistent feature for training and testing, since all three WTG domains consist of the same input SCADA features and output DEMs based on CMS data. To train all the domain adaptation models, the source domain dataset and features are made up of the data from WT 1 and the target domain dataset of features from WT 2. WT 3 is not used in any of the training and is only used for validating the model.

In stage one, an optimal model was produced by conducting an exhaustive search iterating the number of hidden layers, the density of neurons on each layer, the ratio of dropout, batch size, and the degree of convolution. The final form of the ANN is displayed in Fig. 4.

In stage two, the experiments were set up to firstly investigate the performance of each of the techniques and secondly to establish what the most suitable method would be to carry forward for further testing. Techniques from all the main approaches of domain adaptation are implemented, which include parameter-, feature-, and instance-based methods; these are RegularTransferNN (Chelba and Acero, 2004), CORAL (Sun et al., 2015), and TwoStageTrAdaBoostR2 (Pardoe and Stone, 2010), respectively. In each instance, the model is tuned using its distinct hyperparameters for each of the methodologies.

The goal of the experiments is to transfer knowledge between all three WTGs using a single general model by taking SCADA data and making inferences on the DEM such that a low-cost monitoring methodology can be applied to the entire wind farm. The RegularTransferNN is the most accurate model on all three metrics for this architecture, with input and output data for WT 1 as the source domain and WT 3 as the target domain. Further testing was conducted to investigate what the most suitable target and source domains are by altering them.

4.5 Error assessment

The performance of the regression algorithms is based on how the general classifier can make predictions on DEM for all three wind turbines. In this case, common key performance indicators, which provide a percentage of the performance, are implemented.

The mean absolute error (MAE) is a measure of the errors between the paired observations. This is the arithmetic average of the absolute error, where $\widehat{y_i}$ is the prediction, and y_i is the true value.

$$\mathrm{MAE} = \sum_{i=1}^{n} \frac{|y_i - \widehat{y_i}|}{n} \tag{15}$$

The coefficient of determination (R^2) represents the proportion of the variation from the predicate value to the actual value, and μ is the arithmetic mean.

$$R^2 = \frac{\sum\limits_{i=1}^{n} (y_i - \widehat{y_i})^2}{\sum\limits_{i=1}^{n} (y_i - \mu)^2} \tag{16}$$

Cumulative fatigue prediction error (CFPE) encompasses the total error for all instances used in the model, where a conservative result would be a negative percentage value and an underestimate would have a positive percentage value. Here m is the power factor, and in this case it is set to 4.

$$\mathrm{CFPE} = \frac{\sum\limits_{i=1}^{n} \left(\sqrt[m]{y_i}^{\,m} \right) - \sum\limits_{i=1}^{n} \left(\sqrt[m]{\widehat{y_i}}^{\,m} \right)}{\sum\limits_{i=1}^{n} \left(\sqrt[m]{y_i}^{\,m} \right)} \tag{17}$$

5 Results

The Results section is broken down into three subsections in the pursuit of an optimal model. Stage one aims to determine the most suitable ANN architecture for traditional SHM, training on one structure only and then applying the remaining WTGs to this model for comparison. Stage two takes the model from stage one and applies it to all three of the domain adaptation techniques. Stage three alternates the source and target domain adaptation techniques in pursuit of the optimal model.

Table 3. Standard SHM approach to PBSHM, with the test results from the ANN trained on one WTG and tested on the resulting two WTGs.

	MAE (%)			CFPE (%)			R^2 (%)		
WTG	WT 3	WT 2	WT 1	WT 3	WT 2	WT 1	WT 3	WT 2	WT 1
WT3	0.025	0.031	0.029	−16.12	−29.43	−18.71	0.65	0.51	0.54
WT2	0.037	0.030	0.042	−7.37	−22.80	−31.56	0.31	0.45	0.15
WT1	0.034	0.030	0.029	−17.04	−29.06	−21.81	0.42	0.45	0.55

Table 4. Cumulative error for the optimal models during the model development stage. The test datasets are used to determine the error for both DEMs.

	Accumulative error (%)			
Model	WT 1	WT 2	WT 3	Avg
Artificial neural network	−21.806	−31.558	−18.712	−24.025
CORR	−23.400	−34.475	−23.239	−27.038
RegularTransferNN	−2.225	−18.216	−5.639	−8.693
TwoStageTrAdaBoostR2	−8.626	−17.240	−4.338	−10.068

Table 5. Mean absolute error for the optimal models during the model development stage. The test datasets are used to determine the error for both DEMs.

	MAE (%)			
Model	WT 1	WT 2	WT 3	Avg
Artificial neural network	0.026	0.043	0.030	0.033
CORR	0.037	0.029	0.034	0.034
RegularTransferNN	0.033	0.030	0.031	0.031
TwoStageTrAdaBoostR2	0.034	0.050	0.039	0.041

5.1 Stage one

To demonstrate the applied form of strong homogeneous populations, three structurally equivalent WTGs are applied to an ANN using standard SHM techniques. The ANN is trained on one structure, and then the trained model is applied to the other WTGs. The data used are the test datasets during normal operation for all three WTGs. The results are summarised in Table 3, with all three metrics presented.

The normal operation test using the standard approach of training on one structure and then applying the resultant WTGs provides varying degrees of accuracy. All the tests in this case fall below the threshold for the CFPE of ±10 %. This is expected as there are small perturbations in label space due to manufacturing tolerances and location-specific effects.

5.2 Stage two

One of the main challenges of PBSHM is performing damage identification on the population with different label spaces $y_s \neq y_t$. However, in this case, the label spaces are strongly homogeneous, but we have identified that the general classifier using the standard approach to SHM does not provide adequate inferences from Sect. 1. The results of applying the three domain adaptation techniques using WT 1 as the source domain and WT 3 as the target domain are displayed in Tables 4–6.

Table 6. R^2 score for the optimal models during the model development stage. The test datasets are used to determine the error for both DEMs.

	R^2 score (%)			
Model	WT 1	WT 2	WT 3	Avg
Artificial neural network	0.546	0.151	0.538	0.412
CORR	0.637	0.191	0.564	0.464
RegularTransferNN	0.594	0.338	0.553	0.495
TwoStageTrAdaBoostR2	0.431	0.158	0.310	0.300

The goal of the experiments is to transfer knowledge between all three WTGs using a single general model by taking SCADA data and making inferences on the DEM such that a low-cost monitoring methodology can be applied to the entire wind farm. Tables 4–6 highlight the variation in the accuracy of the ANN in the three domain adaptation techniques. The RegularTransferNN is the most accurate model in all three metrics for this architecture, input data, and output data and is the only model that reaches the target CFPE of ±10 %.

CORAL focuses on aligning second-order statistics of the source and target domains by transforming feature representations. TwoStageTrAdaBoostR2 adapts the AdaBoost algorithm for domain adaptation by assigning higher weights to unclassified target domain instances during boosting. RegularTransferNN utilises neural networks for transfer learning, involving pre-training on a source domain and fine-tuning on the target domain. CORAL and RegularTransferNN have higher potentials through the alignment or fine-tuning, while TransferAdaBoost is specifically tailored to adapt the AdaBoost algorithm. In this particular case, the RegularTransferNN was best suited to this particular dataset. The fine-tuning of the trained ANN provided a better general regression model than the other two. However, this may not always be the case – it is always dependent on the dataset.

Table 7. Comparison of the test results from the RegulartransferANN using different source and target datasets.

	Training data (source and target)								
	MAE (%)			CFPE (%)			R^2 (%)		
WTG	64/45	45/10	64/10	64/45	45/10	64/10	64/45	45/10	64/10
WT 3	0.031	0.029	0.023	−5.64	−18.02	−20.24	0.55	0.49	0.63
WT 2	0.030	0.029	0.041	−18.22	−25.48	−10.36	0.34	0.57	0.17
WT 1	0.033	0.034	0.025	−2.23	−26.10	−18.58	0.59	0.42	0.69
Average	0.031	0.031	0.030	−8.69	−23.20	−16.39	0.49	0.49	0.50

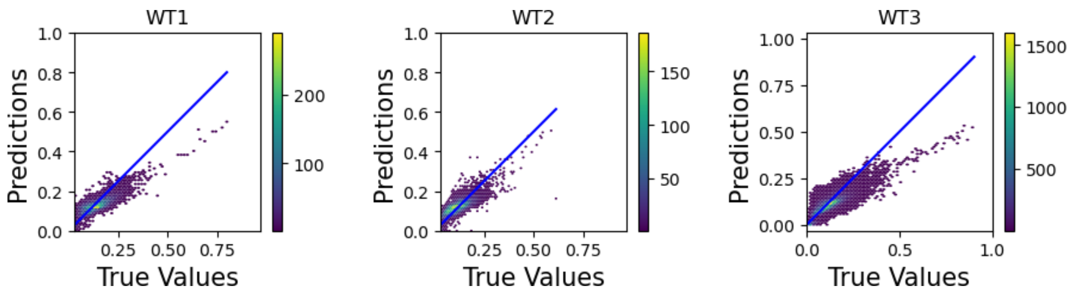

Figure 5. Error plot showing the predictions from the general model against the actual value for the DEM 1 of all three WTGs.

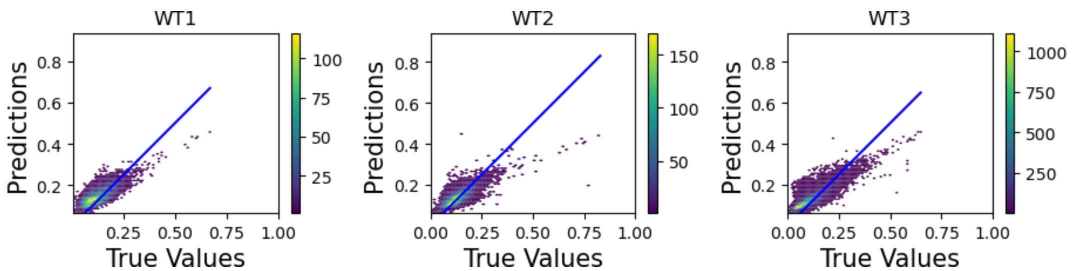

Figure 6. Error plot showing the predictions from the general model against the actual value for the DEM 2 of all three WTGs.

5.3 Stage three

In the pursuit of an optimal general classifier for the Wikinger wind farm, the first two stages are set up to determine the optimal choice of model. Stage three investigates what model data sources are best suited for achieving an average optimal score in all three metrics. The aim of this is to determine the sensitivity of the model's accuracy with the input data. If there is a large difference in the results, this may contribute to a large degree of variance by the general model for an entire wind farm.

The selection process for determining the optimal model is to have the lowest number of CFPEs and the highest R^2 and MAE. A threshold is placed on the CFPE of $\pm 10\,\%$ since the purpose of this type of model is to predict the DEM, which in turn is used to determine the fatigue life estimations of the structure. A high or low estimation of the accumulation of DEM is linked to poor fatigue life estimations. Ideally, one should be conservative in predicting the accumulated DEM to avoid overestimating the fatigue life of the structure. There is only one set-up in this entire process that achieves this; unfortunately it does not have the highest average R^2 or MAE as seen in Table 7 with the WT 1 and WT 3 source and target domain set-up. However, this particular set-up achieves higher results when making inferences on both the source and the target domain but is less accurate in the inferences on WT 2. Thus, it does not achieve the desired goal of a general classifier, in contrast to the high consistency of the MAE of around 3 %.

Figures 5 and 6 display the evaluation of the general model estimations. The hyperparameter λ was altered from 0.1 to 0.99, varying the bias general model training from the source domain data to the target. The optimal model presented has a 50/50 split with $\lambda = 0.5$. One aspect of the final general model is the lack of accuracy at high DEM values, where one sees the highest deviation from the real value, which is the main contributing factor to the reduced performance of the CFPE.

6 Conclusions

Knowledge transfer is an important process in PBSHM. The benefit of transferring knowledge about structural health from one structure to another within the population is imperative to the progress of low-cost digitally enabled asset management for WTGs. It is essential when applying this technique that a process is conducted to determine which similarities exist within the population so that negative transfer can be avoided. In this paper two categories of structures have been discussed: homogeneous and heterogeneous.

Implementing transfer learning in the form of domain adaptation has been demonstrated to effectively mitigate problems where both the features and the label spaces are consistent. This paper has shown that domain adaptation applies to homogeneous populations where there are small deviations in the geometry due to the water depth, manufacturing tolerances, and sensor placement.

CORAL focuses on aligning second-order statistics of the source and target domains by transforming feature representations. TwoStageTrAdaBoostR2 adapts the AdaBoost algorithm for domain adaptation by assigning higher weights to unclassified target domain instances during boosting. RegularTransferNN utilises neural networks for transfer learning, involving pre-training on a source domain and fine-tuning on the target domain. CORAL and RegularTransferNN have higher potentials through alignment or fine-tuning, while TransferAdaBoost is specifically tailored to adapt the AdaBoost algorithm. In this particular case, RegularTransferNN proved to be the most suitable for the dataset, as fine-tuning the pre-trained artificial neural network resulted in superior predictive performance. It is important to note that the effectiveness of these techniques is highly dependent on the characteristics of the dataset, and the choice of the most suitable approach may vary accordingly.

If a higher accuracy of the model is required to determine the remaining fatigue life, the following further optimisation measures can be taken: (1) separating the general model into discrete models for operational modes, (2) discrete model development based on the wind direction, (3) further studies on the feature selection and hyperparameter tuning, and (4) implementation of high-frequency SCADA data for higher-order statistics. Measures 1 and 2 may lead to developing specific models and may increase bias within the estimation. As such, a less accurate but more general model may offer improved robustness and better capability to capture unforeseen events.

Code availability. The software code developed and used in this study is not publicly available due to confidentiality agreements. However, the authors are open to sharing the code on a case-by-case basis for research purposes, as they are keen to support further development of this work.

Data availability. The data used throughout this study are not publicly available due to confidentiality agreements. However, the authors are open to sharing the dataset on a case-by-case basis for research purposes, as they are keen to support further development of this work.

Author contributions. Innes Murdo Black conceptualised the study, developed the methodology, implemented the modelling framework, and conducted the analysis. Moritz Werther Häckell contributed to data provision, technical validation, and interpretation of the results. Athanasios Kolios supervised the research, contributed to the conceptual framing, and provided critical revisions to the manuscript. All authors discussed the results and approved the final version of the paper.

Competing interests. At least one of the (co-)authors is a member of the editorial board of *Wind Energy Science*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. The authors acknowledge the financial support provided by the Engineering and Physical Sciences Research Council (EPSRC) through the Wind and Marine Energy Systems Centre for Doctoral Training under grant number EP/S023801/1.

Financial support. The authors have received financial support provided by the Engineering and Physical Sciences Research Council (EPSRC) through the Wind and Marine Energy Systems Centre for Doctoral Training under grant number EP/S023801/1.

Review statement. This paper was edited by Amir R. Nejad and reviewed by four anonymous referees.

References

Anderlik, S., Stumptner, R., Freudenthaler, B., and Fritz, M.: A proposal for ontology-based integration of heterogeneous decision support systems for structural health monitoring, iiWAS, https://doi.org/10.1145/1967486.1967515, 2010.

ASTM: Standard Practices for Cycle Counting in Fatigue Analysis, https://store.astm.org/e1049-85r17.html (last access: 10 October 2023), 2017.

Black, I. M., Cevasco, D., and Kolios, A.: Deep Neural Network Hard Parameter Multi-Task Learning for Condition Monitoring of an Offshore Wind Turbine, Journal of Physics: Conference Series, 2265, 032091, https://doi.org/10.1088/1742-6596/2265/3/032091, 2022.

Bull, L. A., Gardner, P. A., Gosliga, J., Rogers, T. J., Dervilis, N., Cross, E. J., Papatheou, E., Maguire, A. E., Campos, C., and Worden, K.: Foundations of population-based SHM, Part I: Homogeneous populations and forms, Mech. Syst. Signal Process., 148, 107141, https://doi.org/10.1016/j.ymssp.2020.107141, 2021.

Chelba, C. and Acero, A.: Adaptation of Maximum Entropy Capitalizer: Little Data Can Help a Lo, in: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Barcelona, Spain, 285–292, https://aclanthology.org/W04-3237 (last access: 10 October 2023), 2004.

Chen, W., Qiu, Y., Feng, Y., Li, Y., and Kusiak, A.: Diagnosis of wind turbine faults with transfer learning algorithms, Renew. Energy, 163, 2053–2067, https://doi.org/10.1016/j.renene.2020.10.121, 2021.

Dai, W., Yang, Q., Xue, G.-R., and Yu, Y.: Boosting for transfer learning, in: Proceedings of the 24th International Conference on Machine Learning (ICML '07), Corvallis, OR, USA, 2007, 193–200, https://doi.org/10.1145/1273496.1273521, 2007.

Dorafshan, S., Thomas, R. J., and Maguire, M.: Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete, Construction and Building Materials, 186, 1031–1045, https://doi.org/10.1016/j.conbuildmat.2018.08.011, 2018.

Friedjungová, M. and Jiřina, M.: An Overview of Transfer Learning Focused on Asymmetric Heterogeneous Approaches, in: Data Management Technologies and Applications, edited by: Filipe, J., Bernardino, J., and Quix, C., Springer International Publishing, Cham, ISBN 978-3-319-94809-6, 2018.

Gao, Y. and Mosalam, K. M.: Deep Transfer Learning for Image-Based Structural Damage Recognition, Comput.-Aid. Civ. Infrastruct. Eng., 33, 748–768, https://doi.org/10.1111/mice.12363, 2018.

Gardner, P., Bull, L. A., Gosliga, J., Dervilis, N., and Worden, K.: Foundations of population-based SHM, Part III: Heterogeneous populations – Mapping and transfer, Mech. Syst. Signal Process., 149, 107142, https://doi.org/10.1016/j.ymssp.2020.107142, 2021.

Gardner, P., Bull, L., Dervilis, N., and Worden, K.: On the application of kernelised Bayesian transfer learning to population-based structural health monitoring, Mech. Syst. Signal Process., 167, 108519, https://doi.org/10.1016/j.ymssp.2021.108519, 2022.

Gosliga, J., Gardner, P. A., Bull, L. A., Dervilis, N., and Worden, K.: Foundations of Population-based SHM, Part II: Heterogeneous populations – Graphs, networks, and communities, Mech. Systems Signal Process., 148, 107144, https://doi.org/10.1016/j.ymssp.2020.107144, 2021.

Hamaguchi, T., Oiwa, H., Shimbo, M., and Matsumoto, Y.: Knowledge Base Completion with Out-of-Knowledge-Base Entities: A Graph Neural Network Approach, T. Jpn. Soc. Artific. Intel., 33, F-H72_1–10, https://doi.org/10.1527/tjsai.f-h72, 2018.

Jamil, F., Verstraeten, T., Nowé, A., Peeters, C., and Helsen, J.: A deep boosted transfer learning method for wind turbine gearbox fault detection, Renew. Energy, 197, 331–341, https://doi.org/10.1016/j.renene.2022.07.117, 2022.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, arXiv [preprint], https://doi.org/10.48550/arXiv.1412.6980, 2015.

Li, R., Mo, T., Yang, J., Jiang, S., Li, T., and Liu, Y.: Ontologies-Based Domain Knowledge Modeling and Heterogeneous Sensor Data Integration for Bridge Health Monitoring Systems, IEEE T. Indust. Inform., 17, 321–332, 2021.

Li, Y., Jiang, W., Zhang, G., and Shu, L.: Wind turbine fault diagnosis based on transfer learning and convolutional autoencoder with small-scale data, Renew. Energy, 171, 103–115, https://doi.org/10.1016/j.renene.2021.01.143, 2021.

Lord, N.: Matrix computations, in: The Mathematical Gazette, 3rd Edn., edited by: Golub, G. H. and Van Loan, C. F., Johns Hopkins University Press, 556–557, ISBN 0801854148, ISBN 080185413X, https://doi.org/10.2307/3621013, 1999.

Nickel, M., Murphy, K., Tresp, V., and Gabrilovich, E.: A Review of Relational Machine Learning for Knowledge Graphs, Proc. IEEE, 104, 11–33, https://doi.org/10.1109/JPROC.2015.2483592, 2016.

Pan, S. J. and Yang, Q.: A survey on transfer learning, IEEE T. Knowl. Data Eng., 22, 1345–1359, 2010.

Pardoe, D. and Stone, P.: Boosting for Regression Transfer, in: Proceedings of the 27th International Conference on Machine Learning (ICML 2010), Haifa, Israel, 21–24 June 2010, 863–870, https://doi.org/10.5555/3104322.3104432, 2010.

Ramboll: ROSAP-Ramboll Offshore Structural Analysis Package, Version 53, 2018.

Schröder, L., Dimitrov, N. K., Verelst, D. R., and Sørensen, J. A.: Using Transfer Learning to Build Physics-Informed Machine Learning Models for Improved Wind Farm Monitoring, Energies, 15, 558, https://doi.org/10.3390/en15020558, 2022.

Sun, B., Feng, J., and Saenko, K.: Return of Frustratingly Easy Domain Adaptation, CoRR, abs/1511.05547, arxiv [preprint], https://doi.org/10.48550/arXiv.1511.05547, 2015.

Tsialiamanis, G., Wagg, D., Antoniadou, I., and Worden, K.: An ontological approach to structural health monitoring, in: Topics in Modal Analysis & Testing, Volume 8, edited by: Dilworth, B. and Mains, M., 51–59, Springer International Publishing, https://doi.org/10.1007/978-3-030-47717-2_6, 2021.

Zhang, Y. and Yang, Q.: An overview of multi-task learning, Nat. Sci. Rev., 5, 30–43, 2018.