Population Based Structural Health Monitoring: Homogeneous Offshore Wind Model Development

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

[1]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@strath.ac.uk

Abstract. This is a development of the preceding paper that introduced the idea and methodology of population-based structural health monitoring (PBSHM). PBSHM involves transferring knowledge from one structure to a different structure so that predictions about the structural health on each of the members in 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 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), 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 which is common across all structures. Conversely, even if the models are heterogeneous, disparate 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 the way in which the structures are similar, and where the similarities lie. This determines what type of machine learning approach is necessary. For this study, the method analyzed the geometry, topology, operation, and material of the offshore structures in the prerequisite to this paper Innes Murdo Black. The main observation from this case study the structures for this 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 make 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 are focused on the area of renewable energy, the challenges and solutions in implementing transfer learning. Li et al. (2021b) proposes a strategy to tackle small data-sets 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 data for WT fault diagnosis. The main observation is TrAdaBoost shows its superior performance on dealing with data imbalance and different distributions. Gardner et al. (2022) focuses on the application of machine learning algorithms for structural health monitoring and highlights the importance of domain adaptation in improving the performance of these algorithms. The authors propose a hybrid machine learning model that shows improved performance on several populations of experimental and numerical structures. Schröder et al. (2022) introduces 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 study one month of raw SCADA data. Jamil et al. (2022) proposes a control strategy for deep transfer learning for fault detection on rotating machinery. The paper applies integrated signal processing on vibration signals. The main observation is the performance is significantly improved by reducing negative transfer and less data is required using this technique than standard deep learning.

The motivation for this study is to provide a solution for low-cost monitoring, where only a few wind turbine generators (WTG) are instrumented with sensors as apposed 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 data collected. The real-time data provided by a low-cost monitoring system can also aid in 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 there will have to be assumptions 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 from the transfer learning branch, and the last model utilises ensemble learning. Transfer learning is one approach with the purpose of improving the performance of the learner by transferring between different domains. Domain adaptation assumes that there is labeled data 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 taking the consensus.

2 Population-Based Structural health monitoring

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, the 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 - e.g., 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 Figure 1. This approach will highlight the four main sources of differences within a structure which are geometry, topology, material, and operation:

- Geometry links to the shape and size of the structure within the population.
- Topology depicts the construction, the connection, and location of the components in the structure.
- Material relates to the different, materials 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 too.

Adapting the definitions from graph theory for PBSHM, where a topologically homogenous 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 $p(\sigma_m)$, the probability mass of the distribution $p(\sigma_m)$, defines the small differences between the individuals within the population. A strong homogeneous population would have a uni-modal distribution with low dispersion for the geometrical, topological, and material properties. With the strictest, perfect, form 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 and apply this to another. Applying these conventions to the population and categorizing the individuals within the populations helps determine the difficulty of transferability.

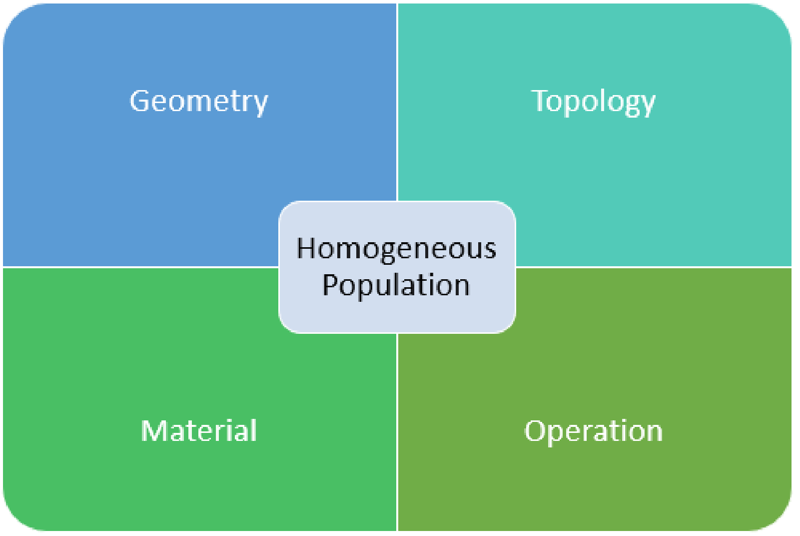

Figure 1. Categories of heterogeneous populations within the PBSHM framework. In the centre, all four categories are alike to a sufficient degree indicating a homogeneous population exists. It is noted that all four attributes can influence each other separately to create independent heterogeneous populations.

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 carried out. 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 form domains and distributions are different for each member when training and testing the model Pan and Yang (2009). Separate from multi-task learning, where the objective is to lean multiple tasks across different domains Zhang and Yang (2017), transfer learning utilises knowledge from the source to improve predictions on the target task, in our case, using the Damage Equivelant Moments (DEM) from two sperate 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 to another structure in the population. Formal definitions of transfer learning and transfer learning technologies are discussed in this section with domain adaptation having an entire subsection.

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 is a bounded sample from χ. The is the SCADA data for one structure in the context of this report.

Task – A task $T = (\Upsilon, f(\cdot))$, this would be 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, 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 build-up of finite samples sets X and Y. In the case of source domain data-sets $D_s = (\chi_{i,s}, y_{i,s})^N$ and with $x_{i,s} = \sum \chi_s$ and with $y_{i,s} = \sum \Upsilon_s$ and similarly for the target domain $D_t = (\chi_{i,t}, y_{i,t})^N$ and with $x_{i,t} = \sum \chi_t$ and $y_{i,t} = \sum \Upsilon_t$, Pan and Yang (2010). Given these artifacts, one can theoretically conduct transfer learning.

Transfer Learning – For transfer learning there must be a given source domain D_s and associated task T_s and a target domain D_t and task T_t. The objective is to improving the target predictive function $f_t(\cdot)$ in T_t by utilising the knowledge from the source, assuming $D_s \neq D_t$ and or $T_s \neq T_t$, Zhang and Yang (2018).

Homogeneous transfer – Homogeneous transfer learning assumes that $D_s = D_t$ and $T_s = T_t$ meaning the attributes are the exact same. A sub-category of this strong homogeneous transfer, where the domain and task are similar hence, $D_s \cong D_t$ and $T_s \cong T_t$.

Heterogeneous transfer – Heterogeneous transfer learning is when the domain, feature, and task space are non-identical hence,$D_s \neq D_t$, $T_s \neq T_t$, and $x_s \neq x_t$, respectively. It can also assume that $y_s \neq y_t$.

Domain adaptation – domain adaptation is relevant when the inference for the target domain D_t and T_t, and the target predictive function $f_t(\cdot)$ is improved given the source domain D_s and T_s. Assuming $x_s = x_t$ and $y_s = y_t$ but the distributions $p(x_s) \neq p(x_t)$.

To contextualize these definitions in the form of PBSHM for wind turbines, homogeneous transfer learning is a situation where both the source space and 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 exactly the same, but have different distribution due to sensor placement, location, to name a few. Hence $D_s \cong D_t$ and $T_s \cong T_t$. Heterogeneous transfer learning is applied when the features are dissimilar. A situation in the wind turbine industry would be when using data of two different wind turbine designs, e.g. a monopile foundation and a jacket structure. In this case the features will be dissimilar and the tasks dissimilar, hence $D_s \neq D_t$ and $T_s \neq T_t$.

3.2 Transfer Learning Technologies

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

Starting off 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 un-fixed layers are trained using target domain D_t. Examples of this are conducted in Innes Murdo Black (2022a) , 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 the 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).

Similar to knowledge graphs, the ontologies' goal is to give representations of entities that describe all the interdependences and interactions. Ontologies are useful in 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 most appropriate for one system to another. Ontologies have been explored in multiple industries, including Structural Health Monitoring (SHM) Li et al. (2021a), Tsialiamanis et al. (2020), 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 the target domains, based on the assumption that the marginal distributions of $p(x_s) \neq p(x_t)$ 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 which are parameter, feature, and instance based. Parameter-based domain adaptation takes the parameters of a pretrained model, is built using the source domain D_s data, and is then adapted to suit a model for the task domain D_t.

Feature-based domain adaptation techniques are designed on the research of common features which have similar attributes with respect to the source T_s 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 $p(x_s)$ and target $p(x_t)$ distributions. The task is then considered to be in an encoded space.

For instance-based domain adaptation the general principle is to redistribute the labeled training data to correct the differences between the source $p(x_s)$ and target $p(x_t)$ 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 directly used 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 most prominent when the source, D_s and the target, D_t domain is 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 contextualize when data is unlabeled, or the tasks are dissimilar.

Negative transfer raises the 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 Population Bases Structural Health Monitoring

This study uses data from the Wikinger wind farm, where there is only a select amount of the operational wind turbines that have CMS with strain gauges installed; to be specific, four out of 64. This section aims to develop a verity of models that can perform PBSHM using structures 64 and 45 as the source and target respectively 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 in determining the fatigue life. This section starts off with how the SHM is conducted, what the population form is for this study, then what the definition of the population is. Bases on the population form from the domain adaptation models are then described.

4.1 Fatigue Damage Equivalent Moments

There are a great number of model and feature spaces that can be applied to represent a population form. For a wind turbine, the form could be 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.

The condition monitoring system calculates the forces from the strain gauges on the foundation of the structure. 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 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

1. Gather applicable force location from the sensor location
2. Calculate the cyclical forces at that sensor

3. Apply ASTM E1049-85 rain-flow 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

4.2 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 to demonstrate 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 loads histogram is presented in Figure 3.

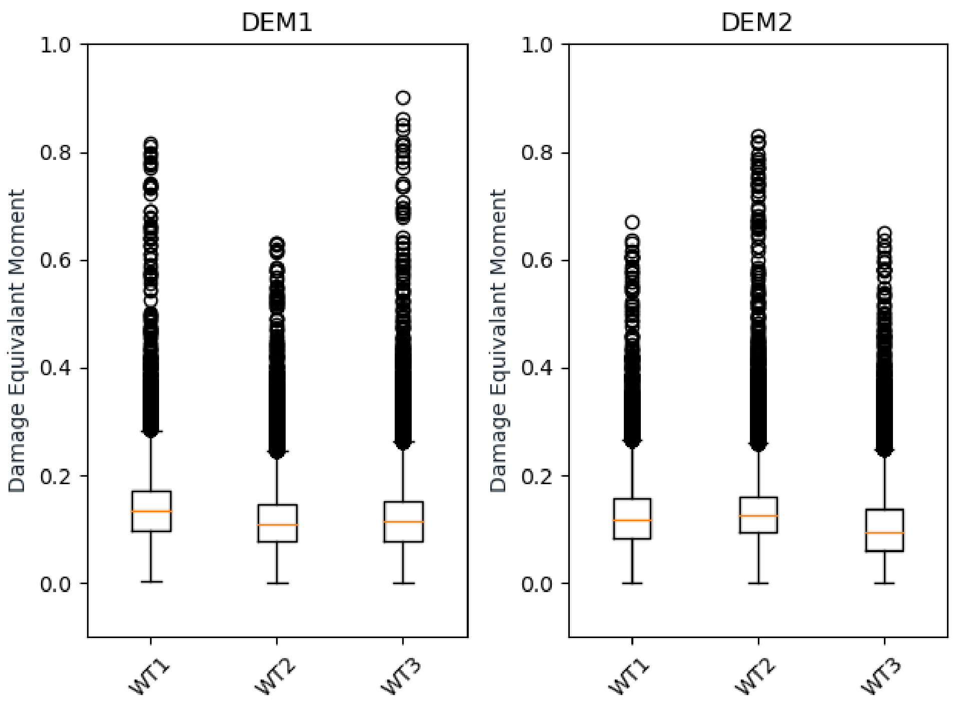

Figure 2. Normalized damage equivalent moments box plot of the two orthogonal directions for all three of the wind turbines

The three structures can be considered as a homogeneous population, as they are structurally similar in their representation and the material and geometry parameters can be described by a uni-modal 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 4 displays two-dimensional heatmaps of the DEM amplitudes for various operational features for the three wind turbines.

Small deviations in the overall distributions are highlighted in Figure 4. One of the contributors to these results is the location as the entire wind farm has deviations on the water depth. The design of the individual WTs does not require the topology and geomoetry 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 influences the dynamics. Another aspect is the operation and metrological differences. Figure 5 displays a heatmap 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 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:

$$\Delta = \|\nu_s - \mu_t|\,{}^2_2 + T_r(|\Sigma_s + \Sigma_T - 2(\Sigma_s \cdot \Sigma_T)^{1/2}) \tag{1}$$

Where μ_s, $\mu_T$are the mean along the source and target along the first axis, Σ_s and $\Sigma_t$are the covariance matrix of the source and target domain datasets. 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 heterogeneous will be large. The value is dependent on the length of the instances in the domain and the magnitude of the values.

Table 1. Fréchet Distance for all three wind turbines

Fréchet Distance			
WT1	0.00		
WT2	13.48	0.00	
WT3	14.48	14.92	0.00
	WT1	WT2	WT3

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 representation, 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 contextualize this, some statistical measures have been calculated based on the normalized DEM and these are represented in Table 2. This includes the mean, standard deviation (Std), and minimum and maximum of the DEM in both orthogonal

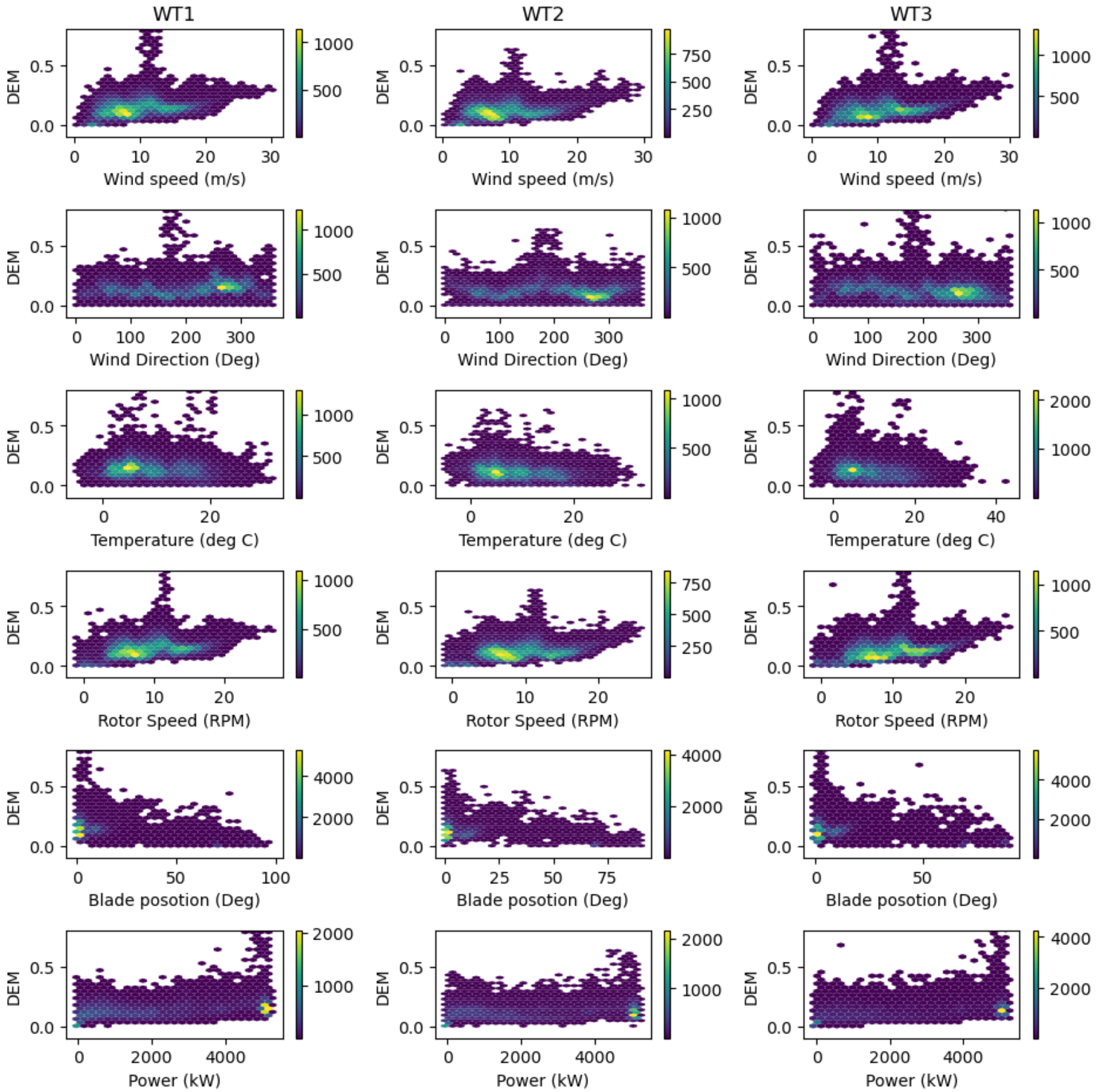

Figure 3. This series of figures highlights the operational and environmental effects on the population of the wind farm. This includes the power, blade position, wind speed, rotor speed, and temperature against the fatigue damage equivalent moments in direction 1.

directions. There is an almost identical nature to the three WTGs; WT2 does have higher and lower maximums and minimums, but both the mean and standard deviations are identical to four significant figures.

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

DEM 1				DEM 2		
	WT1	WT2	WT3	WT1	WT2	WT3
Mean	0.1193	0.1193	0.1193	0.1049	0.1049	0.1049
Std	0.0629	0.0629	0.0629	0.0623	0.0623	0.0623
Min	0.0015	0.0015	0.0015	0.0017	0.0015	0.0017
Max	0.9024	0.9024	0.9024	0.6494	0.6494	0.6494

4.3 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 structures, and SCADA does not provide direct information on important parameters on the 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 will briefly describe 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 hyper parameter optimization using WT1.

The architecture of an ANN is built up of hidden layers where each layer has a density of neurons attached to that. 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 on 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 to each activation function and the entire network is curated based using the Adam Kingma and Ba (2015) Optimizer.

4.3.1 CORrelation Alignment (CORAL)

CORAL Sun et al. (2015) is a feature-based domain adaptation method, with the aim of minimizing 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 minimisze the Frobenius norm Lord (1999) between the correlation matrix o the input target data and the transformed input source data. The transformation is described by the following optimization:

$$\min_{A} ||A^T C_S A - CT||_F^2 \qquad (2)$$

Where, A is the feature transformation matrix such that C_s and C_t is 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:

$$C_S = Cov(X_S) + \lambda I_P \tag{3}$$

$$C_T = Cov(X_T) + \lambda I_P \tag{4}$$

$$X_s = X_S CS_S^{-1/2} \tag{5}$$

$$X_s = X_S CS_T^{1/2} \tag{6}$$

Where λ is the regularization parameter.

4.3.2 TwoStageTrAdaBoostR2

TwoStageTrAdaBoostR2 algorithm, Pardoe and Stone (2010), is an instance-based domain adaptation method suited to regression tasks. This method is characterized by the 'reverse boosting' principle where the weights of the source instances predicted decrease at each boosting iteration, and one of the instances increase. The 'two stages' version of TRAdaBoostR2, Pardoe and Stone (2010), algorithm is where the weights of the source and target instances are carried out separately. In the first stage, the weights from the source instances are frozen, but the ones on the target instances are updated according to the classical AdaBoostR2 Dai et al. (2007). In the second stage the weights of the target instance are now fixed whereas the ones on the source are updated according to TrAdaBoostR2. During each first stage, a cross-validation score is computed with the labeled 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 AdaBoostR2 estimator $f_s(\,)$ on the source and target labelled data (x_s, y_s), (x_t, y_t) with the respective importance initial weights w_s, w_t. During the training of AdaBoostR2 the weights of w_s is 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 normalizing constant B_s is chosen so that the sum of the weights is equal to $\frac{n_t}{n_t+n_s} + \frac{t}{N-1}(1 - \frac{n_t}{n_t+n_s})$with t 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.

4.3.3 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 regularized by the Euclidean distance of both the source and target parameters:

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

Where the estimation function is f with D network layers. B_t is related to the target parameters, 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 is, where training is biased towards source or target domains depending on the associated weighting.

5 Methodology

The following section discusses the process of creating a PBSHM model and the comparison of domain adaptation models used. This process is about coordinating the data stream and effective use of metrics to determine an optimal model out of the technologies utilized. This section includes how the data is pre-processed, the model development and the error metrics used.

5.1 The Data

The measurement data used as input for the low-cost monitoring technique takes high frequency 25Hz CMS data that is processed into 10-minute averages. The process of determining the damage equivalent moments (DEM) is in section 4.1. The WTGs available for this study that are both equipped and unequipped with strain gauges (SG). The three positions with SG are WT 1, 2 and 3.

SCADA systems are equipped 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 covers 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 program, transferring knowledge that is unavailable in other wind turbines can provide insight and confidence on 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 WTG. Based on this principle, the population form is the DEM where only four WTGs have the CMS strain gauges installed.

5.2 Pre-Processing

The task of training models for predictions that involve multiple different data streams from different structures requires coordination so that effective ML modeling 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 modeling in this problem is that sensors tend to break, just like most industrial components. This meant identifying a suitable timeframe, where the maximum period of operational uptime should be met. The optimal period took place between 02/10/2018 and 02/07/2019 (nine months). Secondly, synchronizing the CMS data to the SCADA data had to be conducted and matching the instances of 10-minute intervals was the next step.

After synchronizing the data matching instances for all WTs, the next stage was to perform data cleaning. This procedure involved implementing previous value imputation in place of missing values and NaNs. At this point there were zero 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 this is equipped on all WTGs. The final features comprised hand-picked features and statistically relevant features. The outputs are the DEM in the two orthogonal directions from the CMS systems' strain gauge rings. Now that both the input and output data have been established, the feature projection is implemented where the features are normalized 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 WT1, and the target domain was WT2 using 80% of the datasets for the domain adaptation model. Testing was carried out using the subsequent 20% of both the source and domain datasets and the entirety of WT3. In summary, SCADA data features were used as inputs, with WT1 as the source domain and WT2 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 domain and the entire dataset of WT3 using the same input and output features for all WTG.

5.3 Model Development

The method implemented involves three stages. Stage one involves 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 hyper-parameter tuning the remaining experiments using the same architecture from the ANN and application of the three domain adaptation algorithms. Stage three will investigate the optimal model further by altering the source and domain data.

To establish a standardized 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 WT1, the target domain dataset with features from WT2. WT3 is not used in any of the training and is only used for validating the model.

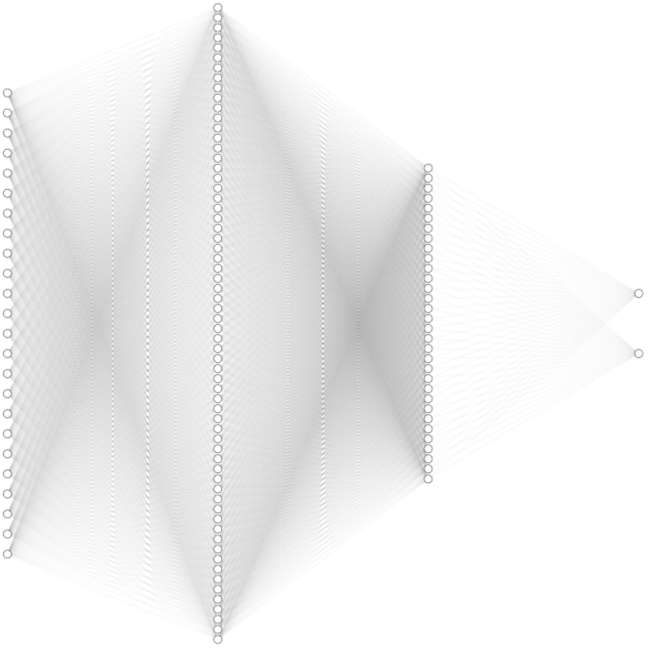

Figure 4. An artificial neural network architecture with 2 hidden layers of density 64 and 32 respectively, batch size of 1, and a dropout ratio of 0.2.

Stage one, an optimal model was produced by carrying out an exhaustive search iterating the number of hidden layers, density of neurons on each layer, ratio of drop out, batch size, and the degree of convolution. The final form of the ANN is displayed in figure 4.

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 their distinct hyper parameters for each of the methodologies.

The goal for 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 data and output data for WT1 as the source domain and WT3 as the target domain. Further testing was conducted to investigate what is the most suitable target and source domain by altering them.

5.4 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 KPI's are implemented which provide a percentage of the performance.

Mean absolute error (MAE) this is a measure of the errors between the paired observations. This is the arithmetic average of the absolute error where $\hat{y}_i$ is the prediction, and y_i is the true value.

$$MAE = \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{n} \tag{15}$$

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

$$R_2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \mu)^2} \tag{16}$$

Cumulative error ($CFPE$) this encompasses the total error for all instances used in the model. Where a conservative result would be a negative % value and an underestimate would have a positive % value. Where m is the power factor, in this case its set to 4.

$$CFPE = \frac{\sum_{i=1}^{n}(\sqrt[m]{y_i}^{m}) - \sum_{i=1}^{n}\sqrt[m]{\hat{y}_i}^{m})}{\sum_{i=1}^{n}\sqrt[m]{y_i}^{m})} \tag{17}$$

6 Results

The results section is broken down into three sections 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, then applying the remaining WTG 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.

6.1 Stage 1

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 is the test dataset during normal operation for all three WTGs. The results are summarized in Table 3 with all three metrics presented.

Table 3. Standard SHM approach to PBSHM, table indicating the test results from the ANN trained on one WTGand tested on the resulting two WTGs.

	MAE			CFPE			R2		
WTG	WT3	WT2	WT1	WT3	WT2	WT1	WT3	WT2	WT1
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

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

6.2 Stage 2

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 section 1. The results of applying the three domain adaptation techniques using WT1 as the source domain and WT3 as the target domain are displayed in Tables 4, 5, and 6.

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

Accumulative Error				
Model	WT1	WT2	WT3	avg
Artificual 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

The goal for 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

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

MAE				
Model	WT1	WT2	WT3	avg
Artificual 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

Table 6. R2 Score for the optimal models during the model development stage. The test data sets are used to determine the error for both DEMs.

R2 Score				
Model	WT1	WT2	WT3	avg
Artificual 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

wind farm. Tables 4, 5, and 6 highlight the variation in the accuracy from the ANN to the three domain adaptation techniques. The RegularTransferNN is the most accurate model on all three metrics for this architecture, input data and output data, and is the only model that reaches the target CFPE of $\pm 10\%$.

6.3 Stage 3

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 to achieve 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 then this may constitute 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 least $CFPE$ and the highest $R2$ and MAE. A threshold is places 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. With a high or low estimation of the accumulation of DEM will directly link to poor fatigue life estimations. Ideally one would e conservative in the prediction of the accumulated DEM as one would not like to over sell. There is only one setup in this entire process that achieves this, unfortunately it does not have the highest average $R2$ or MAE as seen in Table 7 with the WT1 and 3 source domain and target domain setup. However, this particular set-up achieves higher results when making inferences on both the source and target domain but is less accurate in the inferences on WT2. Thus, not achieving the desired goal of a general classifier.

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

	MAE			CFPE			R2		
	Training Data WK(Source/Target)								
WTG	64/45	45/10	64/10	64/45	45/10	64/10	64/45	45/10	64/10
WT3	0.031	0.029	0.023	-5.64	-18.02	-20.24	0.55	0.49	0.63
WT2	0.030	0.029	0.041	-18.22	-25.48	-10.36	0.34	0.57	0.17
WT1	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

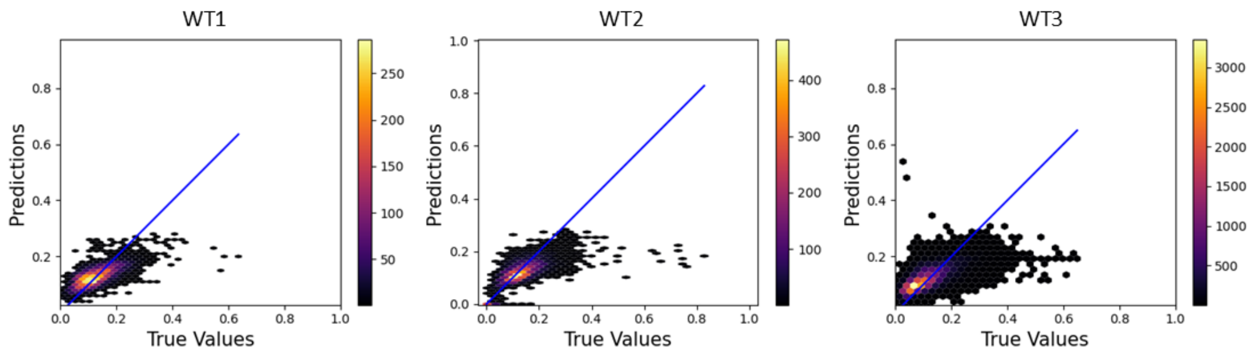

Figure 5. Error plot showing the predictions from the general model against the actual value for the DEM 1 of all three WTGs.

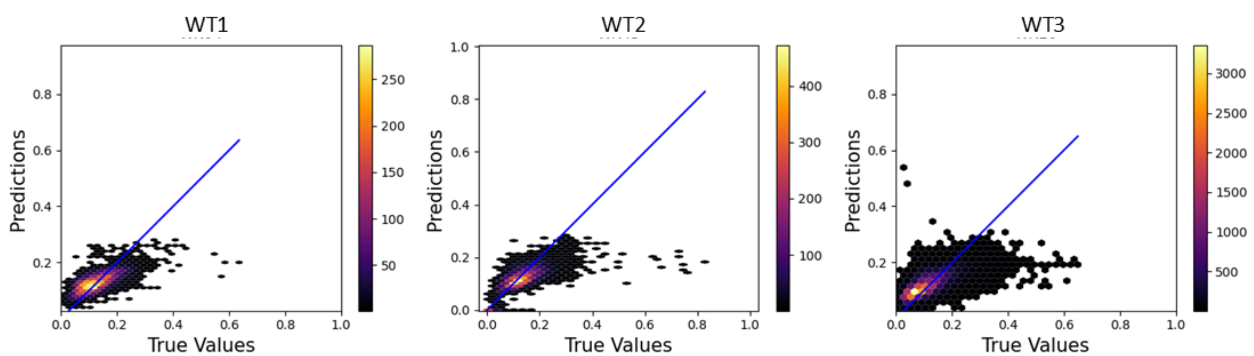

Figure 6. Error plot showing the predictions from the general model against the actual value for the DEM 2 of all three WTGs.

In contrast to the high consistency of the MAE of around 3% Figure 5-6 display the evaluation of the general model estimations. The hyper parameter λ was altered from 0.1 – 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 and is the main constituting factor to the reduced performance of the $CFPE$.

7 Conclusions

Knowledge transfer is an important process in PBSHM. The benefit of transferring knowledge about the structural health from one structure to another within the population is imperative to the progress of low-cost digital enabled asset management for WTGs. It is important when applying this technique that a process is conducted to determine what similarities exist within the population so that negative transfer can be avoided. In this text 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 label spaces are consistent. This paper has demonstrated that domain adaptation is applicable to homogeneous populations where there are small deviations in the geometry due to the water depth, manufacturing tolerances and sensor placement.

If a higher accuracy of model is required to determine the remaining fatigue life, further optimization measures can be taken: These would include: 1 - separating the general model into discrete models for operational modes; 2 - discreet model development based on the wind direction; 3 - further studies on the feature selection and hyper parameter tuning; 4 - implementation of high frequency SCADA data for higher order statistics. Measures 1 and 2 may lead to developing specific models and increase bias within the estimation. As such a less accurate but more general model may produce greater estimations on unforeseen events.

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, in: iiWAS, 2010.

ASTM: Standard Practices for Cycle Counting in Fatigue Analysis, 2017.

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, Mechanical systems and signal processing, 148, 107 141, 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, pp. 285–292, Association for Computational Linguistics, Barcelona, Spain, https://aclanthology.org/W04-3237, 2004.

Chen, W., Qiu, Y., Feng, Y., Li, Y., and Kusiak, A.: Diagnosis of wind turbine faults with transfer learning algorithms, Renewable Energy, 163, 2053–2067, https://doi.org/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: ICML '07, 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, 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, 2018.

Gao, Y. and Mosalam, K. M.: Deep Transfer Learning for Image-Based Structural Damage Recognition, Computer-Aided Civil and Infrastructure Engineering, 33, 748–768, https://doi.org/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, Mechanical systems and signal processing, 149, 107 142, 2021.

Gardner, P., Bull, L., Dervilis, N., and Worden, K.: On the application of kernelised Bayesian transfer learning to population-based structural health monitoring, Mechanical Systems and Signal Processing, 167, 108 519, https://doi.org/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, Mechanical systems and signal processing, 148, 107 144, 2021.

Hamaguchi, T., Oiwa, H., Shimbo, M., and Matsumoto, Y.: Knowledge Base Completion with Out-of-Knowledge-Base Entities: A Graph Neural Network Approach, Transactions of the Japanese Society for Artificial Intelligence, 33, F–H72_1–10, https://doi.org/10.1527/tjsai.f-h72, 2018.

Innes Murdo Black, Debora Cevasco, A. K.: Deep Neural Network Hard Parameter Multi-Task Learning for Condition Monitoring of an Offshore Wind Turbine, in: Torque 2022, 2022a.

Innes Murdo Black, Moritz Werther Häckell, A. K.: Population-Based Structural Health Monitoring: Investigation into the Heterogeneity of an Offshore Wind Farm, Wind Energy, 2022b.

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

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, CoRR, abs/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 Transactions on Industrial Informatics, 17, 321–332, 2021a.

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

Lord, N.: Matrix computations, 3rd edition, by G. H. Golub and C. F. Van Loan. Pp. 694. 1996. £25 (paper), £54 (hard). ISBN 0 8018 5414 8, 0 8018 5413 X. (Johns Hopkins University Press)., The Mathematical Gazette, 83, 556–557, 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, Proceedings of the 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 Transactions on knowledge and data engineering, 22, 1345–1359, 2009.

Pan, S. J. and Yang, Q.: A survey on transfer learning, IEEE Transactions on knowledge and data engineering, 22, 1345–1359, 2010.

Pardoe, D. and Stone, P.: Boosting for Regression Transfer, in: ICML, 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, https://www.mdpi.com/1996-1073/15/2/558, 2022.

Sun, B., Feng, J., and Saenko, K.: Return of Frustratingly Easy Domain Adaptation, CoRR, abs/1511.05547, http://arxiv.org/abs/1511.05547, 2015.

Tsialiamanis, G., Wagg, D., Antoniadou, I., and Worden, K.: An ontological approach to structural health monitoring, 2020.

Zhang, Y. and Yang, Q.: An overview of multi-task learning, National Science Review, 5, 30–43, https://doi.org/10.1093/nsr/nwx105, 2017.

Zhang, Y. and Yang, Q.: An overview of multi-task learning, National Science Review, 5, 30–43, 2018.