Dear Referees

I have listed the accumulation of all the changes that have been made to the document and the compiled changes are supported in the Tracked Changes.pdf.

This section I have listed all the accumulated changes into a list, some of the changes from referee to another are consistent and I have addressed them together here. I will one by one breakdown the list of tracked changes in the manuscript for general changes I will not include all the changes. But when tackling specific problems, I will provide the specific changes:

1. We will revise the paper's organization to make it clearer and more concise.

   This is a general fix and can be noticed throughout the document

2. We will include a literature review in the introduction section.

   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.

3. We will emphasize the motivation for the study and highlight the contribution in both the introduction section.

   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.

4. This is a general fix and all references are now addressed the same. Clearly differentiate between feature data and feature space in our notation. Clearly define our set notation, including any conventions we are following.

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 $T_s \neq T_t$.

5. Correct the typo in "rewriting."

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

6. Add a definition for the WTG acronym.

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

7.