Response to Reviewer 2

We sincerely thank Reviewer 2 for the positive assessment of our work and for recommending acceptance after minor revisions. We appreciate the careful reading of the manuscript and the constructive comments, which helped us clarify methodological choices and improve the presentation. Each comment is addressed below. All corrections and clarifications have been incorporated into the revised manuscript.

Major Comment 1

Justification for using Deep Embedded Clustering (DEC)

Reviewer comment: The motivation for using DEC instead of simpler methods such as k-means is not sufficiently justified. Are the cluster boundaries stable when the number of requested clusters is varied?

Response

We agree that a clearer justification was needed. DEC is not only used for clustering but its actually forcing the latent dimension to be clusterable, it is part of the training. If we rely only on K-means we would not be sure that the latent space would have clusters.

In the revised manuscript, we expanded the discussion in Section \ref{sec:regimes} to explicitly contrast the latent structure obtained without DEC and with DEC, and to explain why post-hoc clustering alone is insufficient in this context.

Changes made:

When no clustering constraint is applied (i.e., DEC is disabled and $\beta=0$), the latent-space is shaped only by reconstruction and domain-adversarial objectives. In this setting, there is no explicit geometric incentive for the encoder to form multiple compact, well-separated operational regimes. The embedding therefore organizes primarily according to the largest, most separable dynamical differences in the data. Empirically, this yields three dominant groups, as illustrated in Fig.~\ref{fig:umap_wo_dec}: two small clusters corresponding to non-producing conditions (standstill and parked), and one large cluster that aggregates the full continuum of producing operation. The separation between the two non-producing clusters is consistent with a control-driven distinction (mainly pitch angle differences under low or zero rotor speed), which produces distinct low-frequency spectral signatures. In contrast, within the producing regime, sub-rated, rated, and curtailed behavior form a smooth progression in spectral space (driven by gradual changes in rotor speed, aerodynamic loading, and control action), and therefore remain embedded as a single connected manifold rather than splitting into discrete clusters. This behavior is visible in Fig.~\ref{fig:umap_wo_dec}, where producing states form a single connected manifold the splitting is mainly driven by the pitch angle.

By introducing DEC, the latent structure is explicitly encouraged to become clusterable. DEC adds a set of learnable centroids $\{\mu_j\}_{j=1}^{K}$ and optimizes the encoder such that embeddings are pulled toward these centroids via a KL-divergence objective between soft assignments and a sharpened target distribution. This explicitly trades a purely continuous representation for one that partitions the operating manifold into K compact regions. With DEC enabled and $K=5$ (a user-defined choice), the previously broad operating manifold is refined into multiple regimes that distinguish different levels of power production and control action. This five-cluster configuration aligns with the canonical division of turbine behavior used in SCADA-based classification, while being inferred directly from vibrations and at higher temporal resolution.

Additionally, DEC integrates clustering into the training objective: centroids are part of the model and are refined jointly with the encoder during optimization. In that sense, regime discovery is learned end-to-end rather than imposed only as a purely post-hoc clustering step on the final embeddings (although, as standard in DEC, centroids are initialized from a preliminary clustering such as k-means before being refined during training).

Major Comment 2

Clarification of "train-on-4 / test-on-1" and fleet-wide generalization

Reviewer comment: The meaning of "train-on-4 / test-on-1" is unclear. Is this a split within each fleet-leader turbine, or across turbines? This raises questions regarding fleet-wide deployment claims.

Response

We thank the reviewer for pointing out this ambiguity. The fatigue-estimation experiment follows a leave-one-fleet-leader-out strategy across the five strain-instrumented turbines. In the reported setup, the regression head is trained using DEM data from FL1–FL4 and evaluated on FL1-FL5. FL5 is entirely unseen during training. The encoder remains fixed during this stage.

This turbine-level split is intentional and necessary: DEM can only be computed on turbines equipped with strain gauges. As a result, metrics such as R^2 and MSE cannot be evaluated on non-instrumented turbines, because no ground-truth fatigue reference exists for those units. Holding out one fleet-leader turbine is therefore the only practical way to evaluate generalization to unseen turbines while still allowing quantitative validation.

Change made in manuscript

We clarified the evaluation protocol in Section \ref{sec:fatigue_protocol}, and explicitly stated that FL5 is excluded from training and used only for evaluation. The following text was added (as in the revised manuscript):

Acceleration data are stored in one-hour files, whereas DEM values are available at 10-minute resolution. Each acceleration file is segmented into six non-overlapping 10-minute intervals, each associated with the corresponding DEM value computed over the same time window.

The encoder is kept fixed, and only an LSTM-based regression head is trained to map sequences of embeddings to DEM values. Training and evaluation follow a train-on-4 / test-on-1 cross-validation strategy across the five strain-instrumented fleet-leader turbines; in the reported setup, the model is trained on FL1–FL4 and evaluated on FL5, which is entirely unseen during training. Evaluation is performed over the full summer period, which includes a wide range of operational conditions and numerous stop–start events.

Major Comment 3

Effect of domain-adversarial training on DEM prediction

Reviewer comment: What is the DEM prediction performance with and without domain-adversarial training? Is there a risk of over-suppressing physically meaningful variability?

Response

The primary focus of this paper is operational-state inference and the learning of turbine-invariant operational embeddings. The DEM experiment is included as an auxiliary validation to demonstrate that the learned embeddings preserve load-related information, not as a dedicated study of fatigue modeling or of the interaction between domain adaptation and fatigue prediction.

For this reason, we did not perform a systematic ablation study of DEM prediction with and without domain-adversarial training. Nevertheless, the reported results indicate that domain-adversarial regularization does not prevent fatigue estimation: good predictive performance is obtained on the held-out fleet leader FL5, while turbine identity is largely suppressed in the embeddings (Fig.~\ref{fig:mi_dann}). This suggests that fatigue-relevant variability is retained despite suppressing turbine identity.

We agree that a focused study of the trade-off between turbine invariance and preservation of fatigue-relevant, site-specific variability (e.g., soil–structure interaction) is valuable and should be investigated in future work.

Change made in manuscript

We added an explicit statement in the Conclusions identifying this as future work. The revised manuscript now includes:

While fatigue estimation was not the primary focus of this work, these results indicate that the operational embeddings retain physically meaningful variability beyond regime identification. A dedicated investigation of the interaction between domain-adversarial regularization and fatigue prediction—quantifying the trade-off between turbine invariance and preservation of load- and site-specific effects such as soil–structure interaction— remains an important direction for future research. Such a study would require a broader set of strain-instrumented turbines and is therefore left for future work.

Minor comments

Minor Comment 1

Lines 230–235: meaning of MMM vs. FFF

Response

We agree this was unclear.

Change made in manuscript

Clarified in Section \ref{sec:arch}: MMM now explicitly denotes the full input dimensionality and FFF the number of retained frequency bins.

Minor Comment 2

Line 420: "R2" typo

Response

Correct.

Change made in manuscript

Corrected to R^2.

Minor Comment 3

Line 500: missing "there is"

Response

Correct.

Change made in manuscript

Sentence rewritten to correct the missing copula.

Minor Comment 4

Line 540: paragraph ends suddenly

Response

We agree.

Change made in manuscript

Paragraph rewritten to answer reviewer 1