

We thank the Associate Editor and both reviewers for their careful assessment of the manuscript and for the constructive comments provided throughout the review process. The feedback has helped us clarify the central comparison, strengthen the experimental design and interpretation, and improve the presentation of the manuscript. In the revised version, we have addressed all points raised in the Associate Editor's decision letter and have provided point-by-point responses below. We also respond separately to the comments of both reviewers.

Response to the Associate Editor

Comment 1

The manuscript needs a floor baseline. As it stands, the comparison runs between full-grid inputs, PCA-reduced inputs, and a flattened-grid MLP, with no configuration representing minimal spatial information. Without one, the paper cannot establish that flow-field representation influences predictive performance at all, which is the claim its title makes. Please add an input configuration using hub-height wind speed and variance, or their spatially averaged equivalents. This requires no new aeroelastic simulations and should be straightforward within your existing pipeline.

Response

We thank the Associate Editor for this suggestion. We have added the requested floor baseline, termed the **minimal baseline representation** in the revised manuscript. The baseline is constructed from two spatially aggregated quantities derived from each rotor-plane observation: (1) the spatial average of the temporal-mean wind-speed field and (2) the spatial average of the temporal-variance field. The latter is obtained by converting the released temporal standard-deviation field to variance before spatial averaging. Each observation is therefore represented by only two scalar input variables.

The baseline is evaluated using the same SimpleMLP architecture as the full flattened-grid and PCA-reduced representations. It produces substantially larger prediction errors than the distributed-input configurations for both electrical power and all twelve fatigue-load targets. These results provide a direct reference for assessing the predictive value of retaining distributed rotor-plane inflow information.

The baseline definition has been added to the *Input Representations* section, and the corresponding results have been incorporated into both the electrical-power and fatigue-load analyses. The abstract, discussion, and conclusions have also been revised to reflect this comparison.

Comment 2

The comparison of representations needs to be made at fixed architecture. At present the CNN and MLP arms differ in both input representation and model class, so differences cannot be attributed to representation, and a version of this objection has now been raised in both review rounds. I would suggest reporting the MLP applied to the flattened full grid, to the PCA-reduced inputs, and to the new floor baseline as the representation comparison, with the CNN reported separately as a topology-aware upper bound. I expect this to be largely a matter of reframing results you already have.

Response

We agree and have restructured the analysis accordingly.

The principal comparison of input representations is now performed at fixed architecture using the same SimpleMLP for the two-variable minimal baseline, the full 1800-dimensional flattened rotor-plane input, and the 128-component PCA representation.

The SimpleMLP architecture, training procedure, cross-validation folds, and optimisation settings are held fixed across these three configurations. Differences among these configurations are therefore interpreted primarily in terms of the retained input representation rather than differences in model class.

The CNN is now reported separately as a **topology-aware modelling configuration**. The revised manuscript explicitly states that the CNN differs from the fully connected configurations both in the two-dimensional structure of its input and in its architectural inductive bias. Its performance is therefore interpreted as the result of the **combined spatial representation and convolutional architecture**, rather than as an isolated effect of input representation.

The ResMLP results are retained as an additional architectural comparison and are not included in the principal fixed-architecture representation comparison.

This distinction has been made explicit throughout the manuscript, including the abstract, introduction, input-representation section, model description, results, discussion, limitations, and conclusions. For example, the revised electrical-power and fatigue-load results explicitly avoid attributing CNN–MLP differences to representation alone.

Comment 3

Please define the coordinate systems for all load channels. The bending load definitions for the main bearing are currently ambiguous, and tower-top and tower-bottom channels appear to use different conventions.

Response

We thank the Associate Editor for identifying this ambiguity. We have revised the dataset description and load-component table to clarify the physical interpretation of all twelve released load channels using the definitions provided in the DTU dataset documentation.

Specifically, the revised manuscript now states that blade-root M_x , M_y , and M_z correspond to edgewise bending, flapwise bending, and torsion; tower-base M_x , M_y , and M_z correspond to side–side bending, fore–aft bending, and torsion; tower-top M_x , M_y , and M_z correspond to fore–aft bending, side–side bending, and torsion; and the main-bearing M_x and M_y channels are identified in the released dataset as bending components, with M_z denoting torsion.

We have also clarified why the x - and y -labels should not be interpreted as one universal turbine-fixed convention. HAWC2 permits structural moment outputs to be expressed in user-selected main-body, global, or local beam coordinate systems, whose orientations may differ between structural components. The generation input files specifying the exact output frame used for each channel are not included in the public dataset release. We therefore report the physical load directions supplied with the released dataset and explicitly avoid inferring an unreported common coordinate transformation or assigning more specific directions to the two main-bearing bending channels than are supported by the available metadata.

The table caption and accompanying explanatory paragraph have been revised accordingly.

Comment 4

Please reorder the sections so that the parameter study precedes the sections whose results depend on it.

Response

We have reordered the manuscript accordingly. The PCA dimensionality sensitivity analysis is now presented in the *PCA Analysis* section before the electrical-power and fatigue-load comparative results.

The parameter study evaluates retained dimensionalities from 2 to 256 principal components and identifies 128 components as the configuration used for the subsequent comparative evaluation. The revised manuscript therefore establishes and justifies the PCA dimensionality before presenting any results that depend on this choice.

Comment 5

Please add a discussion on limitations and relevance. This should state explicitly which applications the approach targets and where the rotor-plane inflow would come from in those applications.

Response

We have added a dedicated subsection entitled “Limitations and practical relevance.” The revised discussion now explicitly states that the framework primarily targets simulation-based applications requiring repeated evaluation of turbine power or structural loads, including wind-farm layout optimisation, flow-control studies, fatigue-aware optimisation, and large-scale parametric investigations.

For these applications, we clarify that the required rotor-plane inflow would be supplied by wake or flow-field simulations evaluated at the turbine location before the corresponding aeroelastic turbine response is computed. In this setting, the trained surrogate can replace repeated high-fidelity structural-response calculations while retaining inflow information supplied by a lower-cost flow model.

We additionally discuss the requirements for potential operational or measurement-informed use. Nacelle-mounted lidar and state-estimation approaches are identified as possible sources of incoming-flow information, while the limitations associated with sparse measurements, spatial filtering, reconstruction assumptions, line-of-sight effects, noise, missing observations, and wake contamination are explicitly acknowledged. The revised section also discusses the limitations associated with temporal aggregation, the set of surrogate architectures considered, the combined representation–architecture interpretation of the CNN, and the use of a single simulated wind-farm dataset.

We thank the Associate Editor again for the clear guidance provided in the decision letter. We believe that these revisions have clarified the experimental comparison and the scope of the conclusions, and that all points identified in the closed list have now been addressed in the revised manuscript.

Response to Reviewer 1

We thank Reviewer 1 for the positive assessment of the revised manuscript and for recommending acceptance. As no additional substantive comments were raised in this review round, no further point-by-point response is required.

Response to Reviewer 2

We thank Reviewer 2 for the detailed assessment of the revised manuscript and for highlighting several issues regarding the interpretation of the representation comparison, practical applicability, load-channel definitions, and methodological scope. The revised manuscript has been further updated in response to these concerns and to the Associate Editor's decision letter. Our point-by-point responses are provided below.

Comment 1

Since different surrogate models are used (CNN and MLP), it is not clear whether the spatial representation is the actual reason for the differences. In my opinion, you mainly compare different surrogates and not different spatial representations.

Response

We agree that differences between the CNN and MLP configurations cannot be attributed to input representation alone because both the input structure and the model architecture change simultaneously.

To address this, the principal representation comparison in the revised manuscript is now performed at fixed architecture. The minimal baseline representation, the full 1800-dimensional flattened grid, and the 128-component PCA representation are all evaluated using the same SimpleMLP architecture and the same training and cross-validation procedure.

The CNN is now reported separately as a topology-aware modelling configuration. The manuscript explicitly states that its performance reflects the combined effect of preserving the two-dimensional spatial structure and using a convolutional architecture designed to exploit that structure. CNN–MLP differences are therefore no longer interpreted as the isolated effect of representation.

Comment 2

Why is it important to analyze surrogates for fatigue loads and electrical power jointly?

Response

The purpose of considering both prediction tasks is not simply to evaluate two outputs within the same study, but to investigate whether the same inflow representation is equally appropriate for two physically different turbine-response quantities.

The revised results show that the relative effectiveness of the evaluated representations is target-dependent. For electrical power prediction, the PCA

representation provided the lowest prediction error under the fixed SimpleMLP architecture. For fatigue-load prediction, the full-grid and PCA representations produced broadly similar errors under the same SimpleMLP, while the separately evaluated topology-aware CNN achieved the lowest errors across all twelve fatigue-load components.

The joint analysis therefore allows the study to assess whether conclusions regarding inflow representation generalise across different prediction objectives rather than assuming that a representation effective for one response quantity is necessarily optimal for another.

Comment 3

Considered loads are not fully clear, e.g., the coordinate systems are not defined, so that bending loads for the main bearing are not defined, and tower top and tower bottom seem to have different coordinate systems.

Response

We thank the reviewer for identifying this ambiguity. The dataset description and load-component table have been revised to clarify the physical interpretation of all twelve released load channels using the definitions provided in the DTU dataset documentation.

The revised manuscript now states that blade-root M_x , M_y , and M_z correspond to edgewise bending, flapwise bending, and torsion; tower-base M_x , M_y , and M_z correspond to side–side bending, fore–aft bending, and torsion; tower-top M_x , M_y , and M_z correspond to fore–aft bending, side–side bending, and torsion; and the main-bearing M_x and M_y channels are reported as bending components, with M_z denoting torsion.

The revised text also clarifies that the x , y , and z labels should be interpreted as channel-specific HAWC2 output axes rather than as a single universal turbine-fixed coordinate convention. The generation input files specifying the exact output frame for each channel are not included with the public dataset release, and the manuscript therefore avoids assigning more specific directions to the main-bearing bending channels than are supported by the available metadata.

Comment 4

If I am not mistaken, you train your surrogates in each fold with $8004/56 \times 48 = 184.320$ samples. This is a lot of data. As each sample is based on an aero-elastic simulation, such a high amount of training data is not practical. For standard surrogates (without spatial information), normally only a few hundred to a few

thousand samples are used. Furthermore, it is not investigated how much training data is required for each surrogate. This convergence behavior is more important than the final accuracy for a high number of training samples.

Response

We agree that sample efficiency and learning-curve behaviour are important practical considerations for surrogate modelling.

The present study, however, is focused specifically on isolating the influence of inflow representation under a common training-data regime rather than on determining the minimum number of training observations required for each model. All configurations are therefore evaluated using the same GroupKFold partitions so that the representation comparison is not confounded by differences in the amount of available training data.

We have expanded the discussion of practical relevance to clarify that the proposed surrogates are intended primarily for simulation-based settings in which a training database has already been generated and repeated turbine-response evaluations are subsequently required. A systematic learning-curve analysis would provide valuable additional insight into data efficiency, but such an analysis is outside the scope of the present study.

Comment 5

Section 5 requires more explanations and reasoning. For most of the section, it is not clear why you have chosen which surrogate/hyperparameter/architecture/etc. Some choices seem arbitrary.

Response

We thank the reviewer for this comment. We have clarified the methodological role of the different surrogate architectures in the revised manuscript. The SimpleMLP is now explicitly used as the common fully connected architecture for the principal fixed-architecture comparison of the baseline, full-grid, and PCA representations. A relatively shallow architecture was selected to provide a consistent modelling framework while avoiding unnecessary architectural complexity. The ResMLP is retained as an additional deeper fully connected comparison, while the CNN is evaluated separately with the topology-preserving spatial input because convolutional kernels are specifically suited to exploiting local neighbourhood structure.

The revised manuscript also reports the complete common training configuration, including preprocessing, optimisation, learning-rate scheduling, batch sizes, regularisation, early stopping, and loss function. These settings are applied

consistently across the relevant comparisons to reduce confounding from differences in training procedure. The present study does not aim to perform an exhaustive hyperparameter optimisation for each architecture; rather, the emphasis is on maintaining a consistent experimental framework for comparing the evaluated input representations and model configurations.

Comment 6

If I see it correctly, the errors in power production are really low (NRMSE of 1% or less). Hence, arguing that one surrogate is better than another is not expedient. Here, more relevant is the amount of required training data.

Response

We agree that the absolute electrical-power prediction errors are small for the richer distributed-input configurations, and the revised manuscript avoids interpreting relatively small numerical differences as evidence of universal superiority.

Instead, the electrical-power results are now framed within the fixed-architecture SimpleMLP comparison. Within that controlled comparison, PCA reduced RMSE relative to the full flattened-grid input while reducing the input dimensionality from 1800 variables to 128 components.

The manuscript also explicitly notes that the distributed-input configurations all achieve very high R^2 values and that error-based metrics provide more discrimination between them. We agree that training-data efficiency is an important complementary criterion, but a systematic sample-efficiency study is beyond the scope of the present work.

Comment 7

A comparison with results of a “simple” surrogate using only fully aggregated spatial information, e.g., only mean wind speed and variance (both spatially averaged) or only the wind speed and variance at hub height, is missing and would be very beneficial to demonstrate the relevance of spatial information.

Response

We agree and have added the requested minimal-information baseline.

The baseline uses two spatially aggregated quantities: the spatial average of the temporal-mean wind-speed field and the spatial average of the temporal-variance field. This two-variable representation is evaluated using the same SimpleMLP architecture as the full flattened-grid and PCA configurations.

The baseline produces substantially larger prediction errors for both electrical power and all twelve fatigue-load components, providing direct evidence that retaining distributed rotor-plane information contributes predictive value beyond these two aggregated quantities.

Comment 8

The results of the parameter study of Section 8 are used in Section 6 and 7. Hence, the order of the sections is not logical.

Response

We have reordered the manuscript accordingly.

The PCA sensitivity analysis is now presented before the final comparative results. The selected dimensionality of 128 principal components is therefore established and justified before it is used in the electrical-power and fatigue-load evaluations.

Comment 9

Limitations of this work and a critical discussion regarding the relevance are missing.

Response

We have added a dedicated subsection entitled “**Limitations and practical relevance.**”

The revised discussion now explicitly states the intended application scope of the surrogate framework, including wind-farm layout optimisation, flow-control studies, fatigue-aware optimisation, and large-scale parametric investigations.

For simulation-based use, the rotor-plane inflow would be obtained from wake or flow-field simulations evaluated at the turbine location before the corresponding aeroelastic turbine response is computed. The trained surrogate can then replace repeated high-fidelity response evaluations.

The revised manuscript also discusses the possibility of measurement-informed use through nacelle-mounted lidar or state-estimation approaches, while explicitly acknowledging the additional challenges associated with sparse measurements, noise, line-of-sight effects, spatial filtering, reconstruction assumptions, missing observations, and wake contamination.

The limitations associated with temporal aggregation, the evaluated model classes, the CNN interpretation, and the use of a single simulated wind-farm dataset are also discussed explicitly.

Closing

We thank Reviewer 2 again for the detailed comments. The revised manuscript now separates representation effects from architecture effects more clearly, includes the requested minimal baseline, clarifies the load-channel definitions, improves the section ordering, and substantially expands the discussion of practical relevance and limitations.