

The Influence of Flow-Field Representation on Electrical Power and Fatigue-Load Surrogate Modelling in Offshore Wind Farms

Pradhyumnaa Ganapathi Subramanian¹

¹Bauhaus-Universität Weimar, Geschwister-Scholl-Straße 8/15, 99423 Weimar, Thuringia, Germany

Correspondence: Pradhyumnaa Ganapathi Subramanian (pradhyumnaa.ganapathi.subramanian@uni-weimar.de)

Abstract. High-fidelity aeroelastic simulations provide accurate predictions of electrical power and structural fatigue loads in wind farms but remain computationally expensive for optimisation, control, and digital-twin applications. Machine-learning surrogate models offer an efficient alternative, and recent studies have shown that both surrogate architecture and inflow representation influence predictive performance. However, it remains unclear whether the most suitable representation of rotor-plane inflow information is the same for predicting electrical power and structural fatigue loads. This study compares three representations of identical temporally aggregated rotor-plane inflow data: a flattened vector, a principal component analysis (PCA) representation, and a two-dimensional spatial representation preserving the original rotor-plane topology. The representations are evaluated using fully connected and convolutional neural networks on the publicly available HAWC2Farm Lillgrund dataset under a common five-fold GroupKFold cross-validation framework. For electrical power prediction, the PCA representation achieved the highest predictive accuracy while reducing the input dimensionality from 1800 variables to 128 principal components. On the other hand, a convolutional neural network operating on the spatial representation consistently achieved the lowest prediction errors for all twelve fatigue-load components. These findings indicate that the preferred representation depends on the prediction target, with compact low-dimensional representations proving effective for electrical power prediction, whereas preserving the spatial topology of the inflow is advantageous for fatigue-load estimation.

1 Introduction

The increasing size and complexity of modern wind farms have intensified the need for computationally efficient tools capable of accurately predicting electrical power and structural fatigue loads. These quantities are fundamental to wind farm design, control, optimisation, and lifetime assessment, where large numbers of aeroelastic simulations are typically required to evaluate different operating conditions, control strategies, and environmental scenarios. Although high-fidelity aeroelastic tools provide accurate predictions of turbine response, their computational cost makes them impractical for many optimisation, control, and digital twin applications. Consequently, data-driven surrogate models have become an increasingly attractive alternative for approximating electrical power and structural loads at a fraction of the computational cost of conventional aeroelastic simulations (Gasparis et al., 2020; Haghi and Crawford, 2024; Singh et al., 2024).

A wide variety of surrogate modelling approaches have been proposed for wind energy applications. Early studies compared conventional regression techniques, artificial neural networks, and Gaussian process regression for predicting electrical power

and fatigue loads under varying operating conditions (Gasparis et al., 2020). More recent work has explored probabilistic surrogate models capable of quantifying prediction uncertainty (Singh et al., 2024) and graph neural networks for jointly predicting local inflow conditions, electrical power, and fatigue loads across arbitrary wind farm layouts (Duthé et al., 2023). Together, these studies demonstrate the growing use of machine learning methods for approximating aeroelastic responses across a range of wind-energy applications.

Several recent studies have investigated how different representations of rotor-plane inflow influence surrogate model performance. Shaler et al. (Shaler et al., 2022) demonstrated that temporal mean rotor-plane velocity measurements were sufficient for accurately predicting mean turbine quantities, whereas predicting fatigue loads and other higher-order statistics additionally required information describing the temporal variability of the inflow. Similarly, Liew et al. (Liew et al., 2023a) demonstrated that a surrogate model combining proper orthogonal decomposition (POD) of rotor-plane inflow fields with an artificial neural network substantially improved fatigue-load prediction. Guilloire et al. (Guilloire et al., 2024) proposed a surrogate model based on sector-averaged inflow quantities and demonstrated that this simplified representation remained sufficient to accurately predict damage equivalent loads under unwaked, waked, wake-steering, and induction-control operating conditions. More recently, Ramaswamy et al. (Ramaswamy et al., 2026) compared a hierarchy of local inflow parameterisations ranging from rotor-averaged quantities to POD-based and rotor-resolved representations, showing that rotor-averaged parameterisations exhibited the largest prediction errors, whereas higher-fidelity representations more reliably captured wake-induced loading. Taken together, these studies demonstrate that the representation of rotor-plane inflow can influence surrogate model performance for fatigue-load prediction.

The rotor-plane inflow representation studies discussed above all investigate structural fatigue-load prediction. In contrast, studies that jointly predict electrical power and structural loads have primarily focused on comparing surrogate modelling algorithms while employing a single inflow representation (Gasparis et al., 2020). Consequently, it remains unclear whether the representation that is most appropriate for fatigue-load estimation is also optimal for electrical power prediction when both prediction tasks are based on the same underlying rotor-plane inflow information.

The objective of this work is therefore to investigate how different representations of identical rotor-plane inflow data influence surrogate modelling performance for two fundamentally different prediction tasks: electrical power and fatigue-load estimation. To address this question, three representations of identical rotor-plane inflow information are considered: a flattened vector representation, a principal component representation obtained through principal component analysis (PCA), and a two-dimensional spatial representation that preserves the original rotor-plane topology. These representations are evaluated using fully connected and convolutional neural network architectures on the publicly available HAWC2Farm Lillgrund dataset (Liew et al., 2023b). By maintaining identical underlying inflow information across all experiments, this study isolates the influence of input representation and examines whether the most suitable representation differs between electrical power prediction and fatigue-load estimation.

2 Dataset Description

The dataset used in this study is derived from the large-scale HAWC2Farm simulations of the Lillgrund offshore wind farm introduced by Liew et al. (2023a) and is publicly available through the Technical University of Denmark Data Repository (Liew et al., 2023b). The wind farm consists of 48 wind turbines, each with a rated power of 2.3 MW.

The complete dataset comprises 1000 independent one-hour aeroelastic simulations generated under varying inflow conditions, of which 800 simulations are publicly available and used in this study. Each simulation is sampled at 100 Hz and is characterised by randomly generated atmospheric conditions, including wind direction, ambient wind speed ($4\text{--}20\text{ ms}^{-1}$), power-law shear exponent ($0.1\text{--}0.3$), and turbulence intensity ($2.5\text{--}10\%$). During each simulation, the rotor-plane inflow immediately upstream of every turbine is recorded on a uniform 30×30 Cartesian grid spanning an area of $1.1D \times 1.1D$, where the rotor diameter is $D = 92.6\text{ m}$.

The dataset is organised into six non-overlapping 10-minute intervals for each one-hour simulation. For every turbine and interval, the rotor-plane wind field is post-processed into two temporally aggregated statistics: the spatial mean and spatial standard deviation of the inflow velocity. Each turbine–interval pair constitutes a single observation, resulting in $48 \times 6 = 288$ observations per simulation and a total of 230,400 observations across the 800 simulations considered in this work.

Each observation comprises a single flattened 1800-dimensional input vector describing the temporally aggregated rotor-plane inflow over the corresponding 10-minute interval. This representation is formed by concatenating two 900-dimensional vectors containing the mean and standard deviation of the inflow, respectively. Since each 900-dimensional vector corresponds to a uniform 30×30 Cartesian sampling of the rotor plane, the original spatial fields can be recovered by reshaping the input into two 30×30 grids corresponding to the mean and standard deviation, respectively. The corresponding outputs comprise the electrical power together with twelve fatigue-relevant damage-equivalent load (DEL) quantities evaluated over the same turbine and 10-minute interval. The predicted fatigue load components are summarised in Table 1.

Figure 1 shows an example observation from the dataset after reshaping the flattened input vectors into their original 30×30 spatial representation. The left panel shows the temporally averaged rotor-plane inflow velocity over a representative 10-minute interval, while the right panel shows the corresponding temporal standard deviation.

The distribution of the electrical power output is shown in Figure 2. The observations span the complete operating range of the turbine, from low-power operating conditions to values approaching the rated power of 2.3 MW. This provides a diverse regression target across a wide range of inflow conditions.

Figure 3 summarises the distribution of the twelve fatigue damage-equivalent load (DEL) targets. The different structural components exhibit substantial differences in both their characteristic magnitudes and variability, with load values spanning more than three orders of magnitude. A logarithmic vertical axis is therefore used to facilitate comparison across all load components while preserving their physical units.

Table 1. Summary of predicted fatigue load components. All quantities are reported as 10-minute damage-equivalent loads (DELs) in kNm.

Abbreviation	Component	Load direction / type
MxBr	Blade root	Edgewise bending DEL
MyBr	Blade root	Flapwise bending DEL
MzBr	Blade root	Torsional DEL
MxTB	Tower base	Side–side bending DEL
MyTB	Tower base	Fore–aft bending DEL
MzTB	Tower base	Torsional DEL
MxTT	Tower top	Fore–aft bending DEL
MyTT	Tower top	Side–side bending DEL
MzTT	Tower top	Torsional DEL
MxMB	Main bearing	Bending DEL
MyMB	Main bearing	Bending DEL
MzMB	Main bearing	Torsional DEL

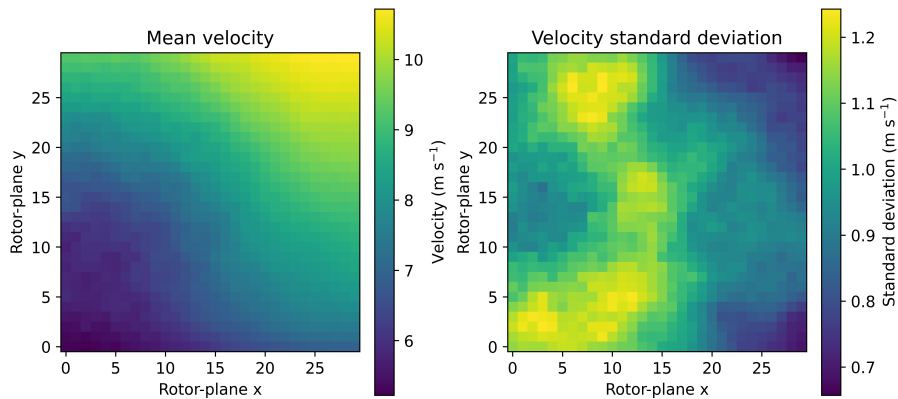


Figure 1. Example rotor-plane wind-field observation reconstructed from the released dataset. The left and right panels show the temporally aggregated mean and standard deviation of the inflow velocity, respectively, over a representative 10-minute interval.

3 Input Representations

90 The dataset provides the inflow conditions as two wind-field variables corresponding to the spatial distributions of the mean wind speed and the standard deviation over a 30×30 sampling grid. Although these fields are inherently two-dimensional, the released dataset stores them as two flattened vectors, resulting in an input dimensionality of 1800 features per observation.

This work considers three representations of the same underlying wind field. First, the released flattened representation is used directly. Second, the original spatial structure is restored by reshaping the vectors into two 30×30 fields. Finally, principal

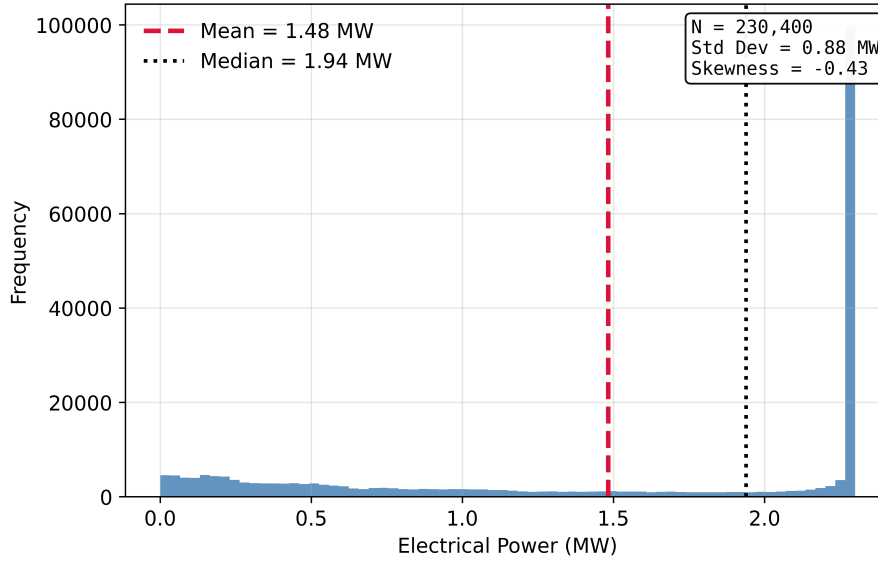


Figure 2. Distribution of electrical power across all observations in the dataset. The dashed red and dotted black lines denote the mean and median electrical power, respectively.

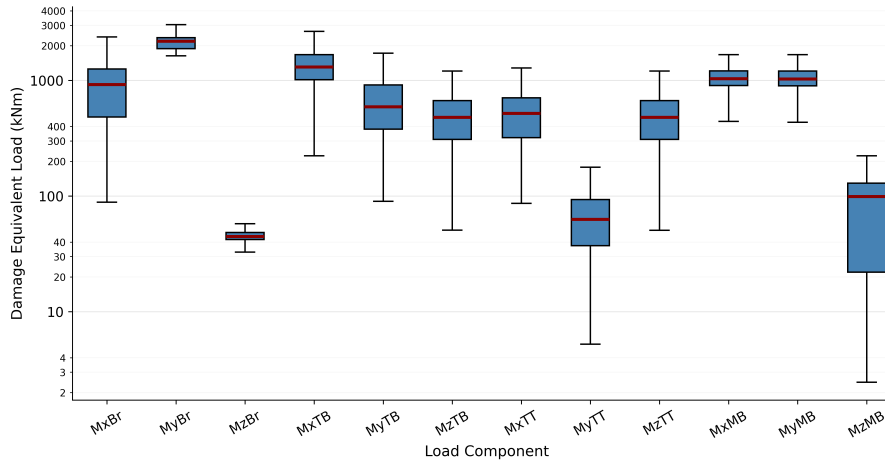


Figure 3. Distribution of the twelve fatigue damage-equivalent load (DEL) targets shown on a logarithmic scale. Boxes represent the interquartile range (IQR), the horizontal red line indicates the median, and whiskers extend to the most extreme observations within 1.5 times the IQR. Outliers beyond the whiskers are omitted for clarity.

95 component analysis (PCA) is employed to obtain a compact lower-dimensional representation. These representations contain identical physical information but differ in how that information is presented to the learning algorithm.

3.1 Flattened Representation

The flattened representation is used directly as provided by the dataset, resulting in an input vector of length 1800 for each observation. No reshaping or dimensionality reduction is applied to the wind field.

100 Representing the wind field as a one-dimensional feature vector removes the explicit spatial relationships between neighbouring grid locations. Consequently, the representation no longer preserves the geometric arrangement of the wind field, and the learning algorithm receives only an ordered collection of feature values. Any influence of the spatial organisation of the inflow on the prediction of electrical power or fatigue loads must therefore be learned implicitly, rather than being explicitly encoded within the input representation.

105 Because the input is represented as a feature vector rather than a two-dimensional field, fully-connected neural networks provide a natural modelling framework. Unlike convolutional neural networks, which employ local receptive fields and shared kernels to learn hierarchical representations from structured inputs (Lecun et al., 1998), fully-connected neural networks learn dense interactions between all input features without explicitly modelling spatial locality. The flattened representation therefore serves as the baseline against which the topology-preserving and dimensionality-reduced representations are evaluated.

110 3.2 Spatial Representation

The spatial representation is obtained by reshaping the two flattened wind-field vectors into two 30×30 grids corresponding to the mean wind speed and the standard deviation. This transformation preserves all original measurements while recovering the original spatial arrangement of the sampling grid. Consequently, each observation is represented as a two-channel tensor with dimensions $2 \times 30 \times 30$.

115 Unlike the flattened representation, the spatial representation explicitly preserves the neighbourhood relationships between adjacent sampling locations. As a result, the geometric organisation of the wind field remains explicitly encoded within the input, making local neighbourhood information directly available to the learning algorithm when predicting electrical power and fatigue loads.

Because the input is represented as a structured two-dimensional field, convolutional neural networks provide a natural
120 modelling framework. Convolutional neural networks apply shared convolutional kernels over local receptive fields to learn hierarchical representations from structured inputs (Lecun et al., 1998). The spatial representation therefore provides a basis for assessing the influence of explicitly preserving the topology of the wind field on the prediction of electrical power and fatigue loads.

3.3 Principal Component Representation

125 Principal component analysis (PCA) provides a compact representation of the wind field by projecting the original high-dimensional input onto a lower-dimensional set of orthogonal components ordered according to the amount of variance explained (Jolliffe, 2002; Jolliffe and Cadima, 2016). Unlike the flattened and spatial representations, which preserve all original

measurements, the principal component representation expresses the wind field using a reduced set of variables that retain the dominant modes of variation present in the original wind field representation.

130 By concentrating the majority of the variability into a smaller number of orthogonal components, PCA reduces redundancy arising from the strong correlations between neighbouring wind-field measurements while retaining the most informative directions of variation (Jolliffe, 2002). However, because each principal component is formed as a linear combination of all original variables, the explicit spatial arrangement of the sampling grid is no longer preserved.

The principal component representation therefore provides an alternative means of representing the inflow conditions, al-
135 lowing the influence of dimensionality reduction to be assessed independently of preserving the explicit spatial topology of the wind field.

The subsequent sections introduce the surrogate models used in this study and describe the experimental framework adopted to evaluate the influence of each input representation on the prediction of electrical power and fatigue loads.

4 Experimental Setup

140 4.1 Implementation Environment

All experiments were conducted using Python 3.12.10 on a laptop equipped with an Intel(R) Core(TM) i7-14650HX processor (16 cores, 24 logical processors) and 32 GB of RAM. This hardware configuration was sufficient to train all surrogate models on the available dataset without the need for specialised accelerators.

4.2 Training Protocol

145 To obtain robust estimates of predictive performance, all surrogate models were evaluated using five-fold GroupKFold cross-validation (Hastie et al., 2009). Unlike conventional K -fold cross-validation, GroupKFold partitions the dataset according to predefined groups rather than individual samples. In the present work, each group corresponds to a single simulation realisation, with all 288 operating conditions generated from that simulation assigned exclusively to either the training or validation set within a given fold. This prevents information leakage arising from correlated samples and provides a more realistic assessment
150 of model generalisation to previously unseen inflow conditions.

Within each fold, the input and output normalization statistics were computed exclusively from the training partition and subsequently applied to both the training and corresponding validation partitions. This ensures that no information from the validation data influenced the preprocessing stage.

The use of GroupKFold also reduces the likelihood that the reported performance results arise from a favourable partitioning
155 of the data. Each surrogate model and input representation was therefore evaluated using the same five data partitions, ensuring that observed performance differences reflect the modelling approach rather than the particular choice of validation data.

The available dataset $\mathcal{D}$ of size N is partitioned into K non-overlapping folds $\{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_K\}$, where each fold contains complete simulation groups. For each fold k , the surrogate model is trained on the union of the remaining $K - 1$ folds and evaluated on the held-out fold $\mathcal{D}_k$, yielding K independent estimates of predictive performance.

160 For a generic performance metric M (e.g., RMSE, MAE or R^2), the cross-validated estimate is computed as

$$\bar{M} = \frac{1}{K} \sum_{k=1}^K M_k, \quad (1)$$

where M_k denotes the metric evaluated on fold k . The variability of model performance across folds can be quantified using the corresponding standard deviation,

$$\sigma_M = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (M_k - \bar{M})^2}. \quad (2)$$

165 Low variability across folds indicates that surrogate model performance is robust with respect to the particular train-validation partition and is therefore unlikely to result from a favourable split. Throughout this work, the reported performance corresponds to the mean and standard deviation across the five validation folds.

4.3 Evaluation Metrics

Surrogate model performance was evaluated using the root-mean-squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). RMSE and MAE quantify prediction error in the original physical units of the target variables, while R^2 measures the proportion of variance in the reference data explained by the surrogate model.

The RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (3)$$

where y_i and $\hat{y}_i$ denote the reference and predicted values, respectively, and N is the total number of samples.

175 The MAE is defined as

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (4)$$

which measures the average magnitude of the prediction error without placing greater emphasis on larger deviations.

The coefficient of determination is given by

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (5)$$

180 where $\bar{y}$ is the mean of the reference values. Values of R^2 closer to 1 indicate better predictive performance, whereas negative values indicate performance worse than predicting the mean reference value.

For all surrogate models, the reported performance metrics correspond to the mean and standard deviation across the five validation folds.

5 Surrogate Models

185 5.1 Common Training Configuration

All surrogate models were implemented in Python using the PyTorch deep learning framework (Paszke et al., 2019). Data preprocessing, principal component analysis, and cross-validation were performed using scikit-learn (Pedregosa et al., 2012), with NumPy (Harris et al., 2020) and Pandas (McKinney, 2010) used for numerical computation and data manipulation.

To ensure a fair comparison between surrogate architectures and input representations, all neural networks were trained
190 using a common optimisation strategy. Within each cross-validation fold, the input and output variables were standardised using statistics computed exclusively from the training partition. Specifically, each input and output feature x_j was transformed according to

$$x'_j = \frac{x_j - \mu_j}{\sigma_j}, \quad (6)$$

where x_j denotes the j -th input or output feature, and μ_j and σ_j are the corresponding mean and standard deviation com-
195 puted exclusively from the training partition. The resulting transformations were subsequently applied to both the training and validation data.

Models were trained using mini-batches of 256 samples, while validation employed a batch size of 512. Model parameters were optimised using the AdamW optimiser with an initial learning rate of 10^{-3} and a weight decay coefficient of 10^{-4} (Loshchilov and Hutter, 2017). The learning rate was updated after each training epoch using the CosineAnnealingLR
200 scheduler with $T_{\max} = 50$ epochs (Loshchilov and Hutter, 2016). The optimisation objective was the mean squared error (MSE) loss, implemented using the `MSELoss` criterion.

To reduce unnecessary computation and mitigate overfitting, early stopping with a patience of 20 epochs was employed based on the validation loss. Training was performed for a maximum of 200 epochs, with optimisation terminated if the validation loss did not improve for 20 consecutive epochs. The model parameters corresponding to the lowest validation loss
205 were retained for subsequent evaluation.

5.2 Simple Multilayer Perceptron

The simple multilayer perceptron (SimpleMLP) serves as the baseline fully-connected architecture for the flattened and principal component representations. As both representations are expressed as one-dimensional feature vectors, fully-connected

neural networks provide a natural modelling framework without assuming any underlying spatial structure. This provides a
 210 baseline for evaluating vector-based representations that do not explicitly preserve the spatial topology of the wind field.

The network consists of an input layer followed by two fully-connected hidden layers containing 256 neurons each, with rectified linear unit (ReLU) activation functions applied after each hidden layer. The resulting network architecture is given by

$$\text{Linear}(d_{\text{in}}, 256) \rightarrow \text{ReLU} \rightarrow \text{Linear}(256, 256) \rightarrow \text{ReLU} \rightarrow \text{Linear}(256, d_{\text{out}}),$$

where d_{in} denotes the input dimensionality and d_{out} the number of prediction targets. For the flattened representation,
 215 $d_{\text{in}} = 1800$, whereas for the principal component representation it corresponds to the retained number of principal components. The output dimensionality is $d_{\text{out}} = 1$ for electrical power prediction and $d_{\text{out}} = 12$ for fatigue-load prediction.

A relatively shallow architecture was deliberately adopted to provide a strong baseline while avoiding unnecessary model complexity. The use of two hidden layers allows nonlinear relationships between the input features and target variables to be learned while maintaining a relatively simple network architecture. Consequently, differences in predictive performance can
 220 be attributed more directly to the input representation rather than increased architectural complexity.

5.3 Residual Multilayer Perceptron

The residual multilayer perceptron (ResMLP) extends the SimpleMLP architecture by incorporating residual learning, facilitating the training of deeper fully-connected networks while alleviating optimisation difficulties associated with increasing network depth (He et al., 2015; Touvron et al., 2021). Similar to the SimpleMLP, the ResMLP operates directly on one-
 225 dimensional feature vectors and is therefore compatible with vector-based input representations.

The network first projects the input feature vector into a 512-dimensional latent feature space through a fully-connected layer followed by the SiLU activation function. This latent representation is subsequently processed by a sequence of twelve residual blocks before being mapped to the prediction targets using a final linear output layer. The resulting network architecture is given by

$$\text{Linear}(d_{\text{in}}, 512) \rightarrow \text{SiLU} \rightarrow 12 \times \text{Residual Block} \rightarrow \text{LayerNorm} \rightarrow \text{Linear}(512, d_{\text{out}}),$$

where d_{in} denotes the input dimensionality and d_{out} the number of prediction targets. As with the SimpleMLP, $d_{\text{in}} = 1800$ for the flattened representation, while it corresponds to the retained number of principal components for the principal component representation. The output dimensionality is $d_{\text{out}} = 1$ for electrical power prediction and $d_{\text{out}} = 12$ for fatigue-load prediction.

235 Each residual block employs layer normalisation followed by a two-layer feed-forward network using the SwiGLU activation function (Shazeer, 2020). The transformed features are combined with the block input through a residual skip connection of the form

$$\mathbf{y} = \mathbf{x} + \mathcal{F}(\mathbf{x}), \quad (7)$$

where $\mathbf{x}$ denotes the block input and $\mathcal{F}(\cdot)$ denotes the nonlinear transformation comprising layer normalisation, linear transformations, and the SwiGLU activation function within the residual block. Additionally, dropout and stochastic depth (Drop-Path) regularisation are incorporated within each block to improve generalisation, while learnable layer-scaling parameters are employed to stabilise optimisation during training (Huang et al., 2016; Touvron et al., 2021).

Compared with the SimpleMLP, the ResMLP substantially increases the representational capacity of the fully-connected architecture while preserving the same vector-based input representation. Consequently, comparing the two architectures isolates the influence of architectural complexity while controlling for the input representation.

5.4 Convolutional Neural Network

The convolutional neural network (CNN) is employed exclusively with the spatial representation of the wind field, where each observation is represented as a two-channel tensor containing the mean wind speed and standard deviation over the original 30×30 sampling grid. By preserving the spatial topology of the inflow conditions, convolutional neural networks are able to exploit local neighbourhood information through shared convolutional kernels, making them well suited to learning spatial patterns from structured two-dimensional inputs (Lecun et al., 1998).

The network consists of three convolutional stages with progressively increasing numbers of feature maps. Each stage comprises two convolutional layers with 3×3 kernels followed by batch normalisation and rectified linear unit (ReLU) activation functions. Spatial downsampling is performed using max-pooling after the first and second convolutional stages, while an adaptive average pooling layer aggregates the final feature maps into a compact latent representation. The resulting network architecture is given by

Conv(2, 32) $\rightarrow$ BN $\rightarrow$ ReLU $\rightarrow$ Conv(32, 32) $\rightarrow$ BN $\rightarrow$ ReLU $\rightarrow$ MaxPool
 $\rightarrow$ Conv(32, 64) $\rightarrow$ BN $\rightarrow$ ReLU $\rightarrow$ Conv(64, 64) $\rightarrow$ BN $\rightarrow$ ReLU $\rightarrow$ MaxPool
 $\rightarrow$ Conv(64, 128) $\rightarrow$ BN $\rightarrow$ ReLU $\rightarrow$ Conv(128, 128) $\rightarrow$ BN $\rightarrow$ ReLU $\rightarrow$ AdaptiveAvgPool
 $\rightarrow$ Linear(128, 256) $\rightarrow$ ReLU $\rightarrow$ Dropout(0.2) $\rightarrow$ Linear(256, d_{out}),

where $d_{\text{out}} = 1$ for electrical power prediction and $d_{\text{out}} = 12$ for fatigue-load prediction.

The convolutional layers learn hierarchical spatial features from the two-channel wind field by combining information from neighbouring sampling locations using shared convolutional kernels. This weight-sharing mechanism substantially reduces the number of trainable parameters compared with fully-connected architectures operating on the same spatial input while maintaining translation-equivariant feature extraction (Lecun et al., 1998). Batch normalisation is employed to improve optimisation stability, while max-pooling progressively reduces the spatial resolution and increases the receptive field. Adaptive

Table 2. Performance comparison of the evaluated surrogate models for electrical power prediction. Reported values correspond to the mean and standard deviation across the five GroupKFold validation folds. The final column reports the relative change in RMSE with respect to the SimpleMLP baseline. Lower RMSE and MAE indicate better predictive performance, whereas higher R^2 values are preferred.

Model	RMSE	MAE	R^2	Relative RMSE Change (%)
SimpleMLP	6756.06 ± 243.45	4176.41 ± 196.73	0.9999 ± 0.0000	0.00
ResMLP	19145.74 ± 2312.49	14315.16 ± 1991.36	0.9995 ± 0.0001	+183.37
CNN	7635.97 ± 566.00	4795.58 ± 543.23	0.9999 ± 0.0000	+13.02
PCA	5534.45 ± 378.49	2790.91 ± 133.26	1.0000 ± 0.0000	−18.08

average pooling produces a fixed-length latent representation that is subsequently processed by a fully-connected prediction head.

Compared with the fully-connected architectures, the CNN explicitly exploits the spatial organisation of the wind field while maintaining the same underlying physical measurements. Consequently, comparisons with the SimpleMLP and ResMLP isolate the influence of explicitly preserving the spatial topology of the wind field while controlling for the prediction task.

5.5 Principal Component Analysis Configuration

Principal component analysis was applied as a preprocessing step to obtain a lower-dimensional representation of the flattened wind-field vectors. Within each cross-validation fold, PCA was fitted exclusively to the standardised training partition, after which the learned transformation was applied to both the training and corresponding validation data. This ensured that the principal component basis was determined solely from the training data, preventing information leakage during preprocessing.

For the primary comparison of surrogate models, the principal component representation retained 128 components. This configuration was selected based on the sensitivity analysis presented in Section 8, which evaluated the influence of the retained dimensionality on predictive performance.

This analysis enables the effect of dimensionality reduction on surrogate modelling performance to be assessed independently of the network architecture while identifying the level of compression that provides the most favourable balance between information retention and predictive accuracy.

6 Results for Electrical Power Prediction

This section presents the predictive performance of the evaluated surrogate models for electrical power prediction. Performance is reported using the root-mean-squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2), averaged across the five GroupKFold validation folds. The corresponding mean and standard deviation values are summarised in Table 2, while Fig. 4 provides a visual comparison of the RMSE and MAE obtained by each surrogate model.

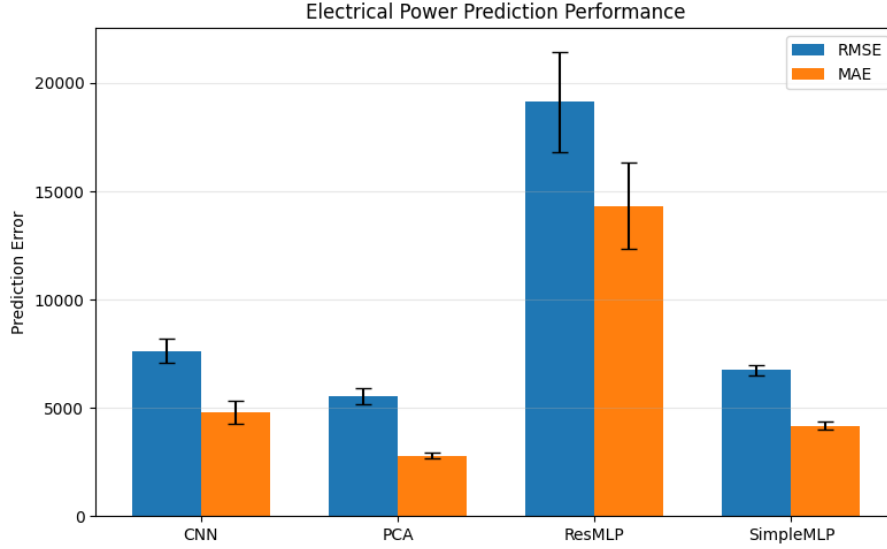


Figure 4. Comparison of electrical power prediction performance for the evaluated surrogate models. Error bars denote one standard deviation across the five GroupKFold validation folds.

285 The PCA-based representation achieved the best overall predictive performance. Compared with the SimpleMLP baseline, the RMSE was reduced by approximately 18.1%, demonstrating that substantial dimensionality reduction can improve predictive accuracy. This suggests that the retained principal components preserve the dominant predictive information required for accurate electrical power prediction.

The CNN achieved lower prediction errors than the ResMLP but was outperformed by both the SimpleMLP and the PCA-
 290 based representation. Despite explicitly preserving the spatial topology of the wind field, this additional structural information did not translate into superior predictive performance for electrical power prediction. These results indicate that preserving the spatial organisation of the inflow field does not appear to be a prerequisite for accurate electrical power prediction and that compact vector representations are capable of capturing the dominant predictive relationships required for this prediction task.

The ResMLP consistently produced the highest prediction errors despite its substantially greater architectural complexity.
 295 Relative to the SimpleMLP, the increased network depth and residual learning strategy did not improve predictive performance, suggesting that the electrical power prediction task does not benefit from the additional representational capacity provided by the deeper fully-connected architecture.

All surrogate models achieved coefficients of determination exceeding $R^2 = 0.9995$, demonstrating that each architecture accurately captures the overall variability of the reference solutions. However, the substantially lower RMSE and MAE achieved
 300 by the PCA representation highlight that these error-based metrics provide greater discrimination between competing surrogate models than R^2 alone.

Table 3. Comparison of RMSE for the twelve fatigue load components. Reported values correspond to the mean and standard deviation across the five GroupKFold validation folds. The lowest RMSE for each component is highlighted in bold.

Component	SimpleMLP	ResMLP	CNN	PCA
MxBr	42.67 ± 1.35	47.68 ± 2.11	40.83 ± 1.35	42.76 ± 1.11
MyBr	32.72 ± 2.40	35.15 ± 1.65	32.12 ± 1.90	33.33 ± 2.31
MzBr	0.985 ± 0.052	1.048 ± 0.048	0.935 ± 0.042	0.988 ± 0.052
MxTB	116.94 ± 3.18	123.12 ± 4.20	103.79 ± 4.38	113.68 ± 3.07
MyTB	110.58 ± 5.85	112.43 ± 5.65	105.33 ± 5.33	110.37 ± 5.93
MzTB	32.22 ± 1.47	34.51 ± 1.16	26.96 ± 0.70	31.12 ± 1.29
MxTT	29.97 ± 1.14	33.52 ± 2.12	25.97 ± 0.41	28.67 ± 1.05
MyTT	5.857 ± 0.231	6.013 ± 0.181	5.316 ± 0.231	5.843 ± 0.254
MzTT	32.21 ± 1.47	34.42 ± 1.02	26.95 ± 0.69	31.12 ± 1.28
MxMB	30.99 ± 1.47	34.61 ± 1.66	27.05 ± 1.12	30.53 ± 1.60
MyMB	30.91 ± 1.57	34.50 ± 1.72	26.94 ± 0.97	30.57 ± 1.59
MzMB	7.210 ± 0.293	7.470 ± 0.296	7.043 ± 0.217	7.472 ± 0.188

7 Results for Fatigue Prediction

This section presents the predictive performance of the evaluated surrogate models for the twelve fatigue load components. Performance is primarily assessed using the root-mean-squared error (RMSE), averaged across the five GroupKFold validation folds. The corresponding mean and standard deviation values are summarised in Table 3, while Fig. 5 provides a visual comparison of the RMSE obtained for each fatigue load component.

The CNN achieved the lowest RMSE for all twelve fatigue load components, consistently outperforming the vector-based representations. Unlike electrical power prediction, these results indicate that preserving the spatial topology of the wind field is advantageous for fatigue-load estimation.

Among the vector-based representations, the PCA representation achieved lower prediction errors for the majority of fatigue load components, although the improvements over the SimpleMLP were generally modest. In contrast to electrical power prediction, where the PCA representation noticeably outperformed the SimpleMLP, its advantage was considerably smaller for fatigue-load estimation.

Consistent with the electrical power prediction results, the ResMLP exhibited the largest prediction errors across nearly all fatigue load components despite its substantially greater architectural complexity. Increasing the depth of the fully-connected network therefore did not compensate for the absence of explicit spatial feature extraction.

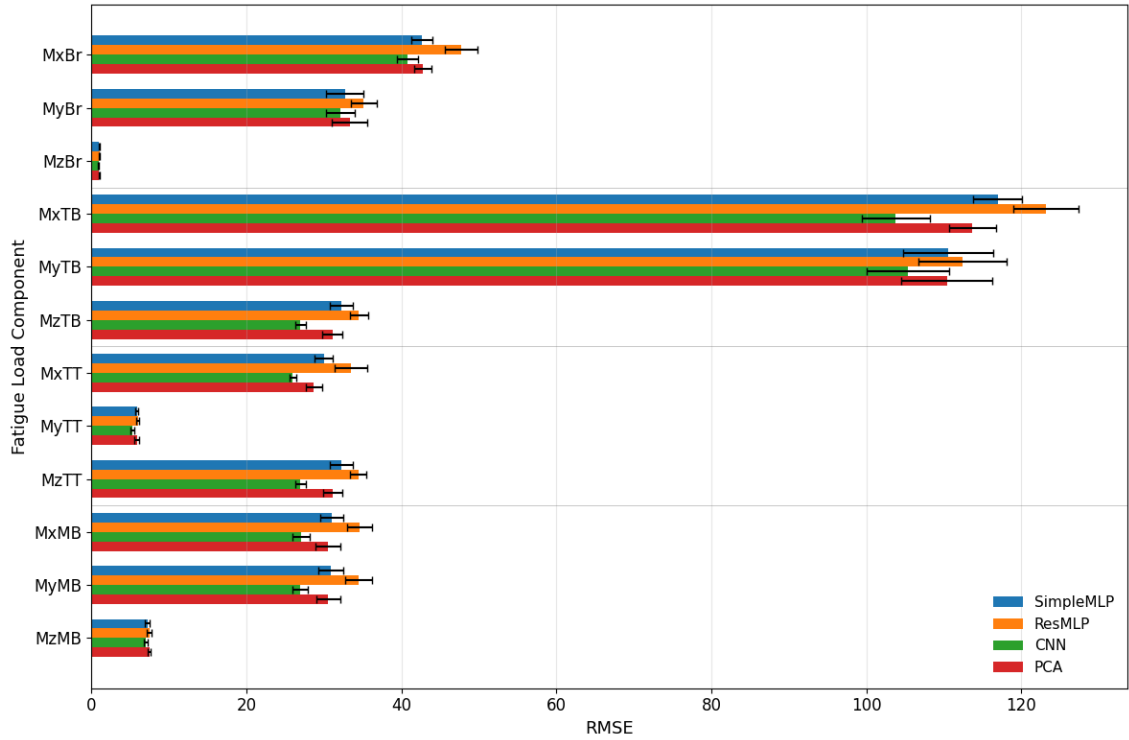


Figure 5. Comparison of RMSE across the twelve fatigue load components for the evaluated surrogate models. Error bars denote one standard deviation across the five GroupKFold validation folds. Horizontal separators distinguish blade-root, tower-base, tower-top, and main-bearing load components.

Taken together with the electrical power prediction results, these findings demonstrate that the most suitable wind-field representation depends on the prediction task. While compact vector representations are sufficient for accurately predicting electrical power, preserving the spatial organisation of the inflow field is advantageous for estimating fatigue loads.

320 8 PCA Analysis

To investigate the influence of the retained PCA dimensionality on electrical power prediction, a sensitivity analysis was performed using the SimpleMLP architecture. The objective of this experiment was to determine an appropriate PCA representation for the subsequent comparative evaluation rather than to establish the final benchmark performance. Consequently, only the first GroupKFold validation split was considered, substantially reducing the computational cost while providing a
325 consistent basis for comparing different PCA configurations.

The number of retained principal components was varied from 2 to 256 using powers of two. Figure 6 illustrates the corresponding prediction error together with the cumulative explained variance of the retained principal components.

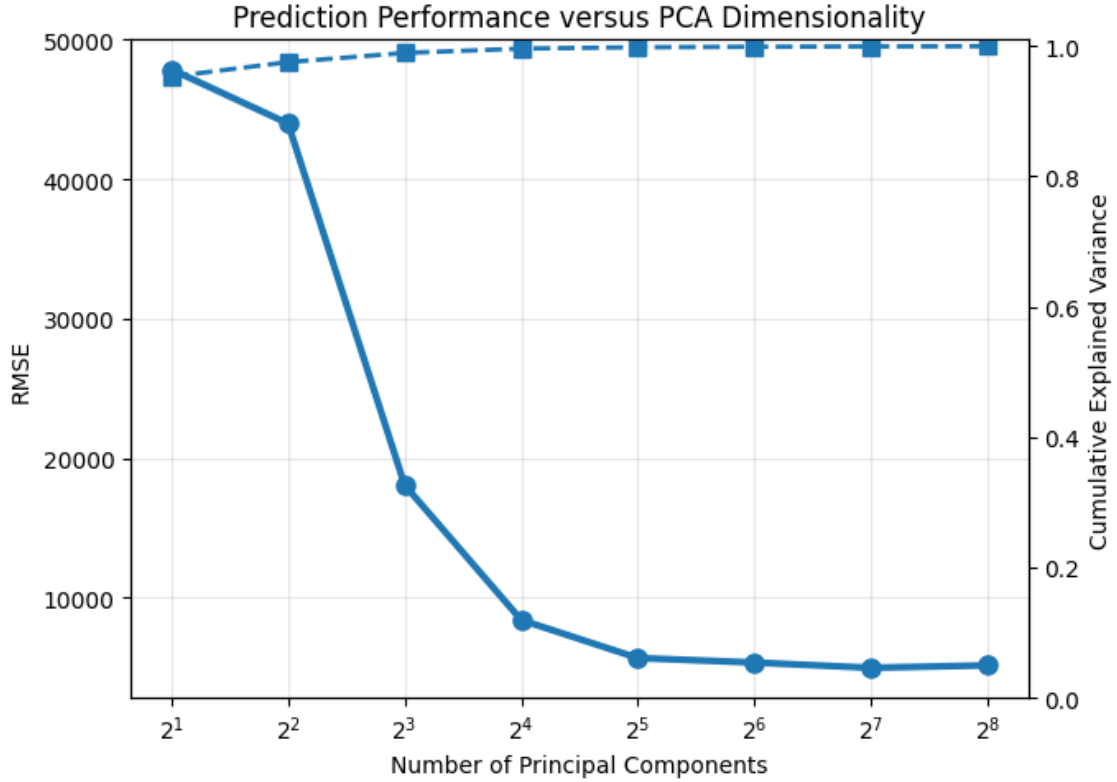


Figure 6. Sensitivity analysis of the retained PCA dimensionality for electrical power prediction using the SimpleMLP architecture. The left axis shows the root-mean-squared error (RMSE), while the right axis shows the cumulative explained variance as a function of the number of retained principal components. Results are reported for the first GroupKFold validation split.

The prediction error decreases rapidly as additional principal components are retained, particularly between 4 and 32 components. Although the cumulative explained variance already exceeds 95% with only two principal components, the corresponding prediction error remains comparatively large. This demonstrates that retaining a high proportion of the statistical variance alone is insufficient to ensure accurate surrogate predictions.

Beyond approximately 64 principal components, the reduction in RMSE becomes progressively smaller despite the cumulative explained variance approaching 100%. The lowest RMSE was obtained using 128 principal components, while increasing the dimensionality to 256 components did not provide any further improvement. These results indicate that the information most relevant for electrical power prediction is captured by a relatively small subset of the principal components, with additional low-variance components providing little benefit.

To complement the quantitative sensitivity analysis, Figs. 7 and 8 illustrate representative reconstructions of the rotor-plane inflow using increasing numbers of retained principal components. Each reconstruction is obtained by projecting the original

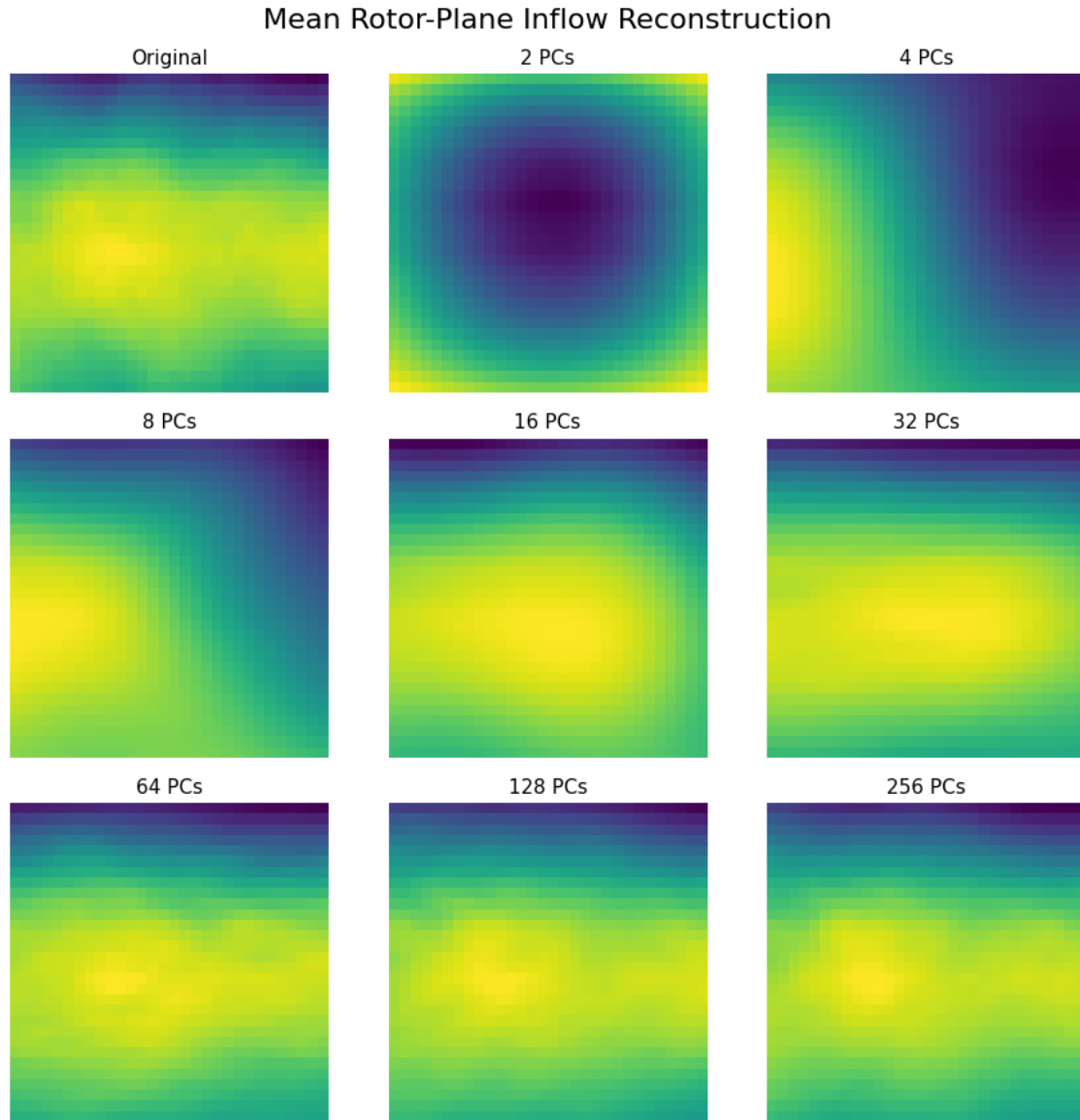


Figure 7. Reconstruction of a representative temporally aggregated mean rotor-plane inflow field using increasing numbers of retained principal components.

1800-dimensional input onto the PCA basis, retaining only the first n principal components, and reconstructing the original
 340 rotor-plane inflow representation through the inverse PCA transformation.

As additional principal components are retained, the reconstructed fields progressively recover the dominant spatial characteristics of the original inflow. For both the mean and standard-deviation fields, the largest-scale spatial structures are recovered using relatively few principal components, while increasing the retained dimensionality gradually restores smaller-scale spa-

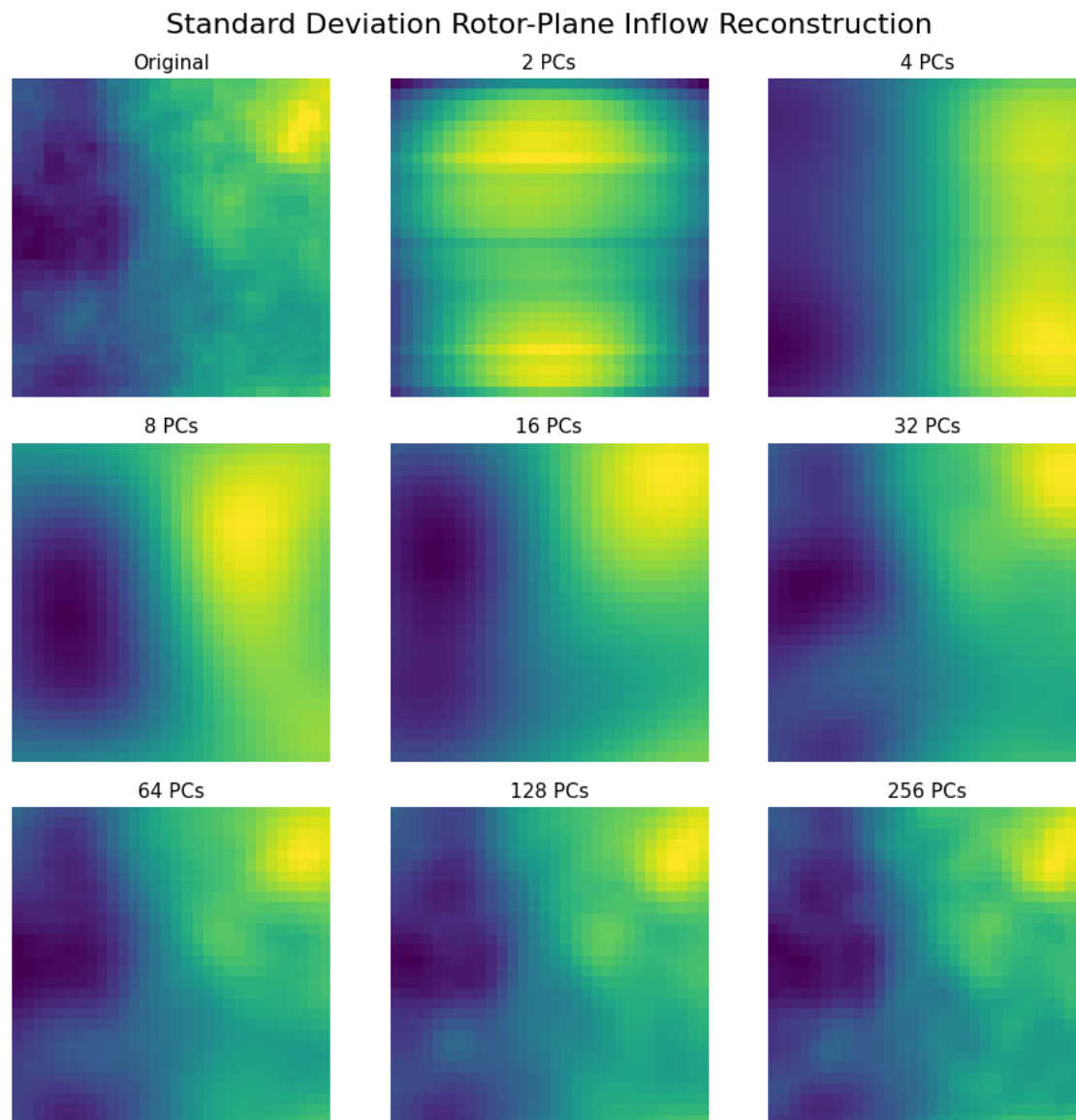


Figure 8. Reconstruction of a representative temporally aggregated standard-deviation rotor-plane inflow field using increasing numbers of retained principal components.

tial variations. The reconstructions obtained using 128 principal components are visually very similar to the original sample, whereas retaining 256 principal components produces only comparatively minor qualitative changes.

This qualitative behaviour is consistent with the sensitivity analysis presented in Fig. 6, where prediction accuracy improved rapidly as additional principal components were retained before reaching a plateau at approximately 128 components. The

reconstruction therefore provides an intuitive visual interpretation of why retaining additional principal components beyond this point yields only marginal improvements in predictive performance.

350 Based on these results, the 128-component representation was selected for the comparative evaluation of surrogate models presented in this work. Although the detailed sensitivity analysis presented in this section was performed for electrical power prediction, an equivalent analysis was also conducted for fatigue-load prediction. Similar saturation behaviour was observed, with improvements becoming marginal beyond 128 principal components. Comparing 128 and 256 retained principal components, six of the twelve fatigue-load components exhibited slightly lower prediction errors using 128 principal components, 355 while the remaining six exhibited slightly lower prediction errors using 256 principal components. As neither dimensionality demonstrated a consistent advantage for fatigue-load prediction, the 128-component representation was adopted for both prediction tasks to provide a common experimental framework for comparing the different input representations.

9 Discussion

This study investigated how different representations of the rotor-plane inflow influence surrogate modelling performance for 360 electrical power and fatigue load estimation. By systematically comparing flattened, principal component, and spatial input representations across multiple neural network architectures, several consistent observations emerge.

For electrical power prediction, the principal component representation achieved the highest predictive performance despite reducing the input dimensionality from 1800 variables to 128 principal components. This indicates that the dominant information required for accurate electrical power prediction is captured by a relatively low-dimensional representation of the inflow 365 field. The accompanying PCA sensitivity analysis further demonstrated that retaining a larger proportion of the statistical variance does not necessarily translate into improved predictive performance, emphasising that principal component selection should be guided by predictive accuracy rather than explained variance alone.

In contrast, fatigue load prediction exhibited a different pattern. The convolutional neural network consistently achieved the lowest prediction errors across all twelve fatigue load components, indicating that preserving the spatial organisation of 370 the rotor-plane inflow provides a clear advantage for estimating fatigue loads. Unlike electrical power prediction, estimating fatigue loads appears to benefit more strongly from preserving the spatial organisation of the inflow field, indicating that local spatial information plays a more important role in this prediction task.

The comparison between the SimpleMLP and ResMLP further indicates that increasing architectural complexity alone does not necessarily improve surrogate modelling performance. Despite incorporating substantially greater representational capacity 375 through deeper residual learning, the ResMLP consistently produced higher prediction errors than the simpler fully-connected baseline across both prediction tasks.

Overall, the results demonstrate that no single representation is universally optimal for wind farm surrogate modelling. Instead, the most suitable representation depends on the prediction target. Compact vector representations, particularly when combined with dimensionality reduction, appear beneficial for predicting aggregated quantities such as electrical power, 380 whereas preserving the spatial topology of the inflow becomes increasingly important for predicting local fatigue loads. These

findings suggest that input representation should be selected according to the intended prediction task rather than assuming that a single representation is equally suitable for different prediction targets. This highlights the importance of considering the prediction target alongside the characteristics of the available input information when designing data-driven surrogate models for wind farm applications.

385 10 Conclusions

This work compared flattened, principal component, and spatial rotor-plane inflow representations for surrogate modelling of wind farm electrical power and fatigue loads. The results indicate that the most suitable wind-field representation depends on the prediction objective. While a PCA-based representation combined with a SimpleMLP achieved the highest predictive accuracy for electrical power prediction, a convolutional neural network operating on the spatial representation of the inflow
390 appears to be advantageous for fatigue-load estimation. Increasing architectural complexity through residual fully-connected networks did not improve predictive performance for the prediction tasks considered in this study.

These findings suggest that selecting an appropriate wind-field representation may have a greater influence on surrogate modelling performance than increasing model complexity alone. Furthermore, they indicate that different wind farm prediction tasks can benefit from different representations of the underlying inflow field. Future work could extend the present compar-
395 ison to graph neural networks, transformer-based architectures, and multi-task surrogate models capable of jointly predicting electrical power and structural loads.

Data availability. The data used in this study are publicly available from the Technical University of Denmark (DTU) data repository at <https://doi.org/10.11583/DTU.21399645>.

Author contributions. PGS was solely responsible for the conceptualization, methodology, software development, analysis, and writing of
400 this manuscript.

Competing interests. The author declares no conflict of interest.

References

- Duthé, G., Santos, F. d. N., Abdallah, I., Réthore, P., Weijtjens, W., Chatzi, E., and Devriendt, C.: Local flow and loads estimation on wake-affected wind turbines using graph neural networks and PyWake, *Journal of Physics: Conference Series*, 2505, 012014, <https://doi.org/10.1088/1742-6596/2505/1/012014>, 2023.
- Gasparis, G., Lio, W. H., and Meng, F.: Surrogate Models for Wind Turbine Electrical Power and Fatigue Loads in Wind Farm, *Energies*, 13, <https://doi.org/10.3390/en13236360>, 2020.
- Guilloré, A., Campagnolo, F., and Bottasso, C. L.: A control-oriented load surrogate model based on sector-averaged inflow quantities: capturing damage for unwaked, waked, wake-steering and curtailed wind turbines, *Journal of Physics: Conference Series*, 2767, 032019, <https://doi.org/10.1088/1742-6596/2767/3/032019>, 2024.
- Haghi, R. and Crawford, C.: Data-driven surrogate model for wind turbine damage equivalent load, *Wind Energy Science*, 9, 2039–2062, <https://doi.org/10.5194/wes-9-2039-2024>, 2024.
- Harris, C. R., Millman, K. J., van der Walt, S., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E.: Array Programming with NumPy, *CoRR*, abs/2006.10256, <https://arxiv.org/abs/2006.10256>, 2020.
- Hastie, T., Tibshirani, R., and Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, New York, USA, 2 edn., <https://doi.org/https://doi.org/10.1007/978-0-387-84858-7>, 2009.
- He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, <https://arxiv.org/abs/1512.03385>, 2015.
- Huang, G., Sun, Y., Liu, Z., Sedra, D., and Weinberger, K.: Deep Networks with Stochastic Depth, <https://arxiv.org/abs/1603.09382>, 2016.
- Jolliffe, I. T.: *Principal Component Analysis*, Springer, 2 edn., 2002.
- Jolliffe, I. T. and Cadima, J.: Principal component analysis: a review and recent developments, *Philosophical Transactions of the Royal Society A*, 374, 20150202, 2016.
- Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 86, 2278–2324, <https://doi.org/10.1109/5.726791>, 1998.
- Liew, J., Riva, R., and Göçmen, T.: Efficient Mann turbulence generation for offshore wind farms with applications in fatigue load surrogate modelling, *Journal of Physics: Conference Series*, 2626, 012050, <https://doi.org/10.1088/1742-6596/2626/1/012050>, 2023a.
- Liew, J. Y., Riva, R., and Göçmen, T.: Simulated fatigue load data set of the Lillgrund wind farm, <https://doi.org/10.11583/DTU.21399645.v1>, 2023b.
- Loshchilov, I. and Hutter, F.: SGDR: Stochastic Gradient Descent with Restarts, *CoRR*, abs/1608.03983, <http://arxiv.org/abs/1608.03983>, 2016.
- Loshchilov, I. and Hutter, F.: Fixing Weight Decay Regularization in Adam, *CoRR*, abs/1711.05101, <http://arxiv.org/abs/1711.05101>, 2017.
- McKinney, W.: *Data Structures for Statistical Computing in Python*, *SciPy 2010*, <https://doi.org/10.25080/Majora-92bf1922-00a>, 2010.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E. Z., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, *CoRR*, abs/1912.01703, <http://arxiv.org/abs/1912.01703>, 2019.

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., VanderPlas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, CoRR, abs/1201.0490, <http://arxiv.org/abs/1201.0490>, 2012.
- 440 Ramaswamy, H., Guilloire, A., Riva, R., Quick, J., Pettas, V., Murcia, J., Campagnolo, F., Bottasso, C., and Réthoré, P.: Comparison of wind turbine load surrogate model performances with multi-scale local inflow parameterization, *Journal of Physics: Conference Series*, 3224, 032 073, <https://doi.org/10.1088/1742-6596/3224/3/032073>, 2026.
- Shaler, K., Jasa, J., and Barter, G. E.: Efficient Loads Surrogates for Waked Turbines in an Array, *Journal of Physics: Conference Series*, 2265, 032 095, <https://doi.org/10.1088/1742-6596/2265/3/032095>, 2022.
- 445 Shazeer, N.: GLU Variants Improve Transformer, <https://arxiv.org/abs/2002.05202>, 2020.
- Singh, D., Dwight, R., and Viré, A.: Probabilistic surrogate modeling of damage equivalent loads on onshore and offshore wind turbines using mixture density networks, *Wind Energy Science*, 9, 1885–1904, <https://doi.org/10.5194/wes-9-1885-2024>, 2024.
- Touvron, H., Bojanowski, P., Caron, M., Cord, M., El-Nouby, A., Grave, E., Izacard, G., Joulin, A., Synnaeve, G., Verbeek, J., and Jégou, H.: ResMLP: Feedforward networks for image classification with data-efficient training, <https://arxiv.org/abs/2105.03404>, 2021.