https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

WIND ENERGY SCIENCE DISCUSSIONS
Open Access
eawe european academy of wind energy

Probabilistic surrogate modeling of damage equivalent loads on onshore and offshore wind turbines using mixture density networks

Deepali Singh[1], Richard Dwight[1], and Axelle Viré[1]

[1]Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, 2629HS Delft, The Netherlands

Correspondence: Deepali Singh (d.singh-1@tudelft.nl)

Abstract. The use of load surrogates in offshore wind turbine site assessment has gained attention as a way to speed up the lengthy and costly siting process. We propose a novel probabilistic approach using mixture density networks to map 10-minute average site conditions to the corresponding load statistics. The probabilistic framework allows for the modeling of the uncertainty in the loads as a response to the stochastic inflow conditions. We train the data-driven model on the OpenFAST simulations of the IEA-10MW-RWT and compare the predictions to the widely used Gaussian process regression. We show that mixture density networks can recover the accurate mean response in all load channels with values for the coefficient of determination (R^2) greater than 0.95 on the test dataset. Mixture density networks completely outperform Gaussian process regression in predicting the quantiles, showing an excellent agreement with the reference. We compare onshore and offshore sites for training to conclude the need for a more extensive training dataset in offshore cases due to the larger feature space and more noise in the data.

1 Introduction

The selection of a suitable site for the installation of a wind farm plays an important role in limiting installation, maintenance, and operational costs, as well as in ensuring a safe operating lifetime of the structure. This process of *site assessment* or *site-suitability study* typically involves a thorough analysis of the structural integrity of the wind turbine at locations with different site-specific environmental inputs.

The International Electrotechnical Commission (IEC) defines a set of design standards IEC 61400-3-1 (2019) comprising of a set of design load cases (DLCs) corresponding to various loading, operating, and fault scenarios. Any combination of site conditions violating the IEC standard for the specific wind turbine class results in simulating the full design load basis to ensure a safe operational lifetime of the turbine. The load cases are typically simulated using time-domain coupled aero-servo-hydro-elastic tools to calculate the fatigue and ultimate loads. In the case of offshore wind turbines, the computational cost of each 10-minute simulation can range between tens of minutes to a few CPU hours, depending on the complexity of the model, computing framework and code efficiency. The total number of simulations that need to be evaluated at each site can be in the order of tens of thousands, significantly inflating the total computational overhead of a site assessment campaign. The cost is further exacerbated in the case of floating wind turbines where the cost per simulation is an order of magnitude higher, the initial transient longer, and the pool of load cases larger.

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

WIND ENERGY SCIENCE DISCUSSIONS

One of the approaches to expediting the site assessment process is to use *surrogates* to model the system. A surrogate model is a simpler and computationally inexpensive representation of the complex model that emulates the outputs as a function of the inputs. The data used by the surrogate as ground truth is often from a computational model but can also consist of real-life measurements. Surrogates are engineering tools developed for preliminary design calculations, optimization, or real-time control, where accuracy can be reasonably traded for computational efficiency. On a broad level, they can be categorized into physics-based and data-driven. On the one hand, physics-based models aim to reduce the system dynamics to the essential elements. In the context of floating wind turbines, several fast frequency-domain reduced-order models have been investigated by Lemmer et al. (2020, 2018); Smilden et al. (2016); Hall et al. (2022). On the other hand, data-driven surrogate models identify the system's behavior based on the observed input-output pairs. In this approach, the computer code is treated as a black box, and the physics is inaccessible to the user. They have the advantage of the ease of implementation in complex, non-linear systems where analytical closed-form solutions are intractable or identifying complex functional relationships between observations from experiments or field data where the physical properties cannot be easily modeled (Jiang et al., 2020).

Site-specific load surrogates are often designed using deterministic data-driven modeling approaches (Section 1.2). For a given training dataset $(\boldsymbol{X}, Y) = \{\boldsymbol{x}^q, y^q\}$, where $q = 1 \ldots n$, deterministic models map a set of K input features $\boldsymbol{x} \in \mathbb{R}^K$ to the corresponding output $y \in \mathbb{R}$. However, the assumption of a deterministic relationship between inputs and outputs does not hold in most real-life cases. This is because, despite having sufficient training data, the limited set of features defining the surrogate's input may not be sufficient to predict an accurate value of y for any $\boldsymbol{x}$. The uncertainty due to the presence of unknown or inexpressible features is called *aleatoric uncertainty*, and it can result in complex noise patterns in the data indicated by heteroscedasticity, non-Gaussianity, and multi-modality in the conditional response (Matthies, 2007; Der Kiureghian and Ditlevsen, 2009). By contrast, in a probabilistic approach, the input and output quantities are modeled as random vectors $\boldsymbol{X} \in \mathbb{R}^M$ and $Y \in \mathbb{R}$, allowing the propagation of uncertainties in the inflow to the corresponding load responses through the surrogate.

In this study, we develop a probabilistic data-driven surrogate that maps 10-minute averaged environmental conditions such as wind speed, turbulence intensity, and wave characteristics, like the significant wave height and period ($\boldsymbol{X} \in \mathbb{R}^6$), to the corresponding 10-minute load statistics ($Y \in \mathbb{R}$), calculated using an open-source, multi-fidelity, multi-physics solver called OpenFAST (NREL, 2022; Jonkman, 2013). The surrogate model learns to map $\boldsymbol{X}$ to the complete conditional probability density function (pdf) $p(Y \mid \boldsymbol{X} = \boldsymbol{x})$ of the load response.

1.1 The case for probabilistic reasoning in site-specific load surrogates

During the 10-minute period, wind turbines are subjected to randomly varying inflow turbulence and waves, regarding which the surrogate model has no information. For instance, for a given turbulence spectrum, average wind speed, and average turbulence intensity, there are unlimited variations of inflow turbulence patterns that result in equally varied load responses. OpenFAST takes as an input a frozen turbulence field generated by TurbSim, consisting of stochastic turbulence patterns created using pseudo-number generators initialized by random seeds. Thus, repeated simulations with the same input parameters but different seeds result in different values of output quantities, yielding a multi-valued mapping, highlighted in Figure 1.

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

On repeating the simulations with sufficient seeds, the load statistics converge towards a value of statistical moments that characterize a random variable, denoted $Y \mid \boldsymbol{X} = \boldsymbol{x}$. Furthermore, the shape of its pdf is a function of the controller actions, wind speed, wave period, and turbulence intensity, among other site conditions.

Deterministic regression models are generally of the type

$$y = f(\boldsymbol{x}) + \varepsilon, \tag{1}$$

where f is a deterministic function of the input features, or the conditional average of the target, and $\varepsilon \sim \mathcal{N}(0, \sigma)$ is the noise component. This framework can only accommodate the average conditional of the target. Therefore, when deterministic models are used, the common practice is to convert the multi-valued problem to a single-valued setting by averaging the response at each sample point over n unique random seeds to approximate $\mathbb{E}(Y|\boldsymbol{X} = \boldsymbol{x})$.

The main drawbacks of this approach are as follows.

– A finite sampling of input loading due to stochastic representations of wind and waves in time-domain simulations introduces an uncertainty in the turbine's load response. Zwick and Muskulus (2015) and Müller and Cheng (2018) show that the recommendation by the IEC61400-1 standard to average over a 60-minute long simulation or six 10-minute long simulations is insufficient to fully converge to the average fatigue loads. Liew and Larsen (2022) similarly concluded that some load channels are more sensitive to the number of seeds, and the average of the tower base moments can be off by around $3-4\%$, even with $n = 10$. For training surrogates, it is common to run 60-minute-long simulations and obtain the average response or perform four to ten 10-minute simulations over different random seeds to obtain the average response before training the surrogate (Dimitrov et al., 2018; Dimitrov and Natarajan, 2019; Shaler et al., 2022; Slot et al., 2020). The variability in the response with fewer seeds can be interpreted as *noise*, forcing the surrogate model to interpolate the noise in case of a small dataset or fit the mean of the response when the dataset is large. However, the mean may be wrongly inferred if the response is non-Gaussian. Seed repetitions add a significant computational cost to the data generation phase, especially when dealing with expensive simulations like in the case of floating wind turbines.

– Modern wind turbines are equipped with sophisticated controllers that can result in multi-modal responses in loads. Training the model on the expectation of a multi-modal distribution can misrepresent the actual load pattern, as the average of several correct target samples is not necessarily a meaningful target value.

– Most real-world learning tasks involve data sets with complex patterns of missing features that introduce aleatoric uncertainty. Unlike numerical simulations, seed repetitions cannot be performed on such datasets.

An alternate approach is to use probabilistic regression that models the targets as random variables, $Y : \Omega \to \mathbb{R}$, with an unknown conditional pdf, $p(Y|\boldsymbol{X} = \boldsymbol{x})$. Probabilistic models provide a framework for informed decision-making by predicting a confidence interval in addition to the most likely response. For instance, the uncertainty in the 10-minute damage equivalent loads (DELs) can be propagated to the lifetime DELs while also considering the distribution of the wind speeds to design less conservative, site-specific structures. Probabilistic surrogates such as conditional generative adversatial networks

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

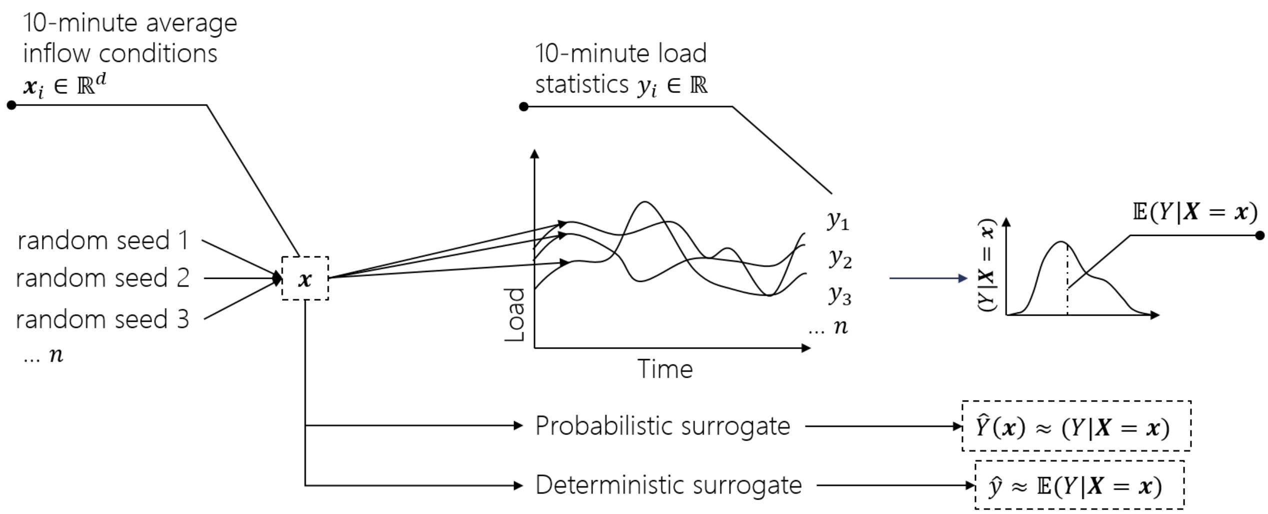


Figure 1. Schematic representation of the multi-valued mapping of the input-output pairs used to train the surrogate.

and conditional variational autoencoders use latent variables to infer meaningful quantities from data with complex noise distributions (Blei et al., 2017; Yang and Perdikaris, 2019; Kingma and Welling, 2014; Kneib, 2013). Other probabilistic regression approaches like mixture density networks (Bishop, 2006) and generalized lambda distributions (Zhu and Sudret, 2021) use maximum likelihood estimate to infer the conditional distributions in stochastic systems. Simulation repetitions before training are therefore unnecessary and significantly cut the training time short.

1.2 Previous work

Wind turbine loads, for site-specific analysis and wind farm design, have commonly been approached with deterministic models like standard artificial neural networks (ANNs), (Schröder et al., 2018; Dimitrov, 2019; Shaler et al., 2022). ANNs are extremely powerful and emulate the loads well if the training data has been averaged over a set of inflow turbulence realizations. Shaler et al. (2022) compare the performance of inverse distance weighting, ANNs, radial basis functions, Kriging with a partial least squares dimension reduction, and regularized minimal-energy tensor-product b-splines in a wind farm array and observe the highest R^2 values for ANNs and the inverse distance weighting method. These approaches, however, do not aim to account for, or predict the variance of the load response.

The standard Gaussian process regression (GPR) (Rasmussen and Williams, 2006) is capable of uncertainty quantification but is restricted only to normally distributed homoscedastic responses. Nevertheless, due to its flexibility and ease of implementation, it is widely used as a load emulator to estimate the fatigue load response in wind turbines (Teixeira et al., 2017; Avendaño-Valencia et al., 2021; Li and Zhang, 2019, 2020). Gasparis et al. (2020) compare GPR to other data-driven methods like linear regression and artificial neural networks for modeling power and fatigue loads, showing a superior performance by the GPR. Similarly, Dimitrov et al. (2018) evaluate importance sampling, nearest-neighbor interpolation, polynomial chaos

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

expansion (PCE), GPR, and quadratic response surface (QRS), to conclude a better performance again by the GPR despite a computational penalty. Slot et al. (2020) provide a thorough comparison of the performance of PCE and GPR for the uncertainty quantification of fatigue loads on NREL's 5MW reference onshore wind turbine. They conclude the need for a minimum of four random seeds per training sample in the case of GPR to make high-accuracy predictions. They also note that GPR performs better per invested training simulation than PCE.

Further interest in quantifying the variability of the short-term fatigue loads as a function of the input parameters has initiated research into heteroscedastic surrogates. One of the ways to model heteroscedasticity is through replication-based approaches, wherein the simulations at each set of average input conditions are repeated for multiple realizations of the stochastic field to obtain statistical information about the response. Murcia et al. (2018) use 100 turbulent inflow realizations at each sample point to obtain the first two moments of the fatigue response. Thereafter, they create two independent surrogates using PCE to model the mean and standard deviation of the fatigue loads on the DTU 10MW reference wind turbine. Even though they use only 140 training samples for their model, the replications scale the computational cost by a factor of 100, eventually leading to a very expensive training database. Another replication-based approach is taken by Zhu and Sudret (2020) to model the load response using generalized lambda distributions. In this study, 50 TurbSim realizations are used at each input sample to estimate the four lambda parameters. Four PCE surrogates are then used to model the parameters independently. The main drawback of replication-based methods is the cost of generating the training database, which makes it difficult to apply them to computationally demanding applications such as floating wind turbines. Secondly, the goodness of fit relies heavily on the estimate of the statistical parameters in the first step.

Heteroscedasticity can also be modeled using statistical methods. Abdallah et al. (2019) use parametric hierarchical Kriging to predict blade-root-bending-moment extreme loads that are heteroscedastic on a 2MW onshore wind turbine. Their approach combines low- and high-fidelity observations, where the low-fidelity model informs the high-fidelity GPR. They show that introducing hierarchy helps make the model selection process more robust than the manual tuning of Kriging parameters. Singh et al. (2022) apply chained GPR that uses variational inference within a Bayesian framework to account for heteroscedasticity in the data and make predictions of site-specific load statistics on a more complex case of offshore wind turbines. The model can capture the heteroscedasticity in a small dataset but is not scalable to high dimensional problems or big data. In order to avoid replication prior to training, Zhu and Sudret (2021) extend the replication-based approach to derive a statistical method combining generalized least-squares with maximum conditional likelihood to estimate the lambda parameters without replications. The main advantage of this method is that it does not assume a Gaussian distribution. However, it can not handle multi-modality.

Only a few approaches attempt to model the uncertainty in the load response of the turbine and the tower, and of those that do, do not consider complex offshore conditions with heteroscedastic multi-modal responses. In this paper, we provide a methodology to build probabilistic data-driven surrogates using mixture density networks (MDN) (Bishop, 1994). The target is modeled as a mixture of $m \in \mathbb{N}$ Gaussians of varying proportions, capable of generating complex distributions when combined. MDN use feed-forward networks to learn the parameters of the mixture model. The performance of MDN is compared to the standard GPR since it is one of the more widely used and accurate load surrogate modeling approaches in the literature. We

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

train the wind turbine model on stochastic aerodynamic and hydrodynamic features to show the added difficulty in modeling offshore load surrogates.

The layout of the paper is as follows. MDN and GPR are introduced in Section 2. Section 3 presents the setup, including details on the wind turbine model, complex computational model, and the dataset generation methodology. Results are discussed in Section 4, followed by conclusions and future directions in Section 5.

2 Regression models

2.1 Gaussian process regression

Gaussian process regression (GPR) is a non-parametric, flexible, Bayesian machine learning framework. As mentioned in Section 1.1, the regression problem is defined as,

$$y = f(\boldsymbol{x}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2). \tag{2}$$

The standard GPR models the noise, ε, as a normally distributed quantity with a variance of σ^2. The function $f(\boldsymbol{x})$ is assigned a Gaussian process prior, that is, $f(\boldsymbol{x}) \sim \mathcal{GP}(\mu(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$. The covariance kernel k dictates the smoothness of the function. In this paper, we use the squared exponential kernel defined as,

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_h^2 \exp\left(-\frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}'\|_{\boldsymbol{l}}^2\right), \quad \|\boldsymbol{x} - \boldsymbol{x}'\|_{\boldsymbol{l}}^2 := \sum_{j=1}^{d} \frac{|x^{(j)} - x'^{(j)}|^2}{l^{(j)}}, \tag{3}$$

implying that the underlying function is smooth and infinitely differentiable, and where $x^{(j)}$ is the j-th component of $\boldsymbol{x}$. The characteristic length-scales $\boldsymbol{l} \in \mathbb{R}^d$ are defined per input parameter, and these and the variance σ_h^2 are *hyperparameters* that are tuned based on the training data. The aim is to make predictions $y^\star$ on unseen data points $\boldsymbol{x}^\star$. The observations $\boldsymbol{y}$ and prediction $y^\star$ are jointly Gaussian, as shown in Equation (4).

$$\begin{bmatrix} \boldsymbol{y} \\ y^\star \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(X) \\ \mu(\boldsymbol{x}^\star) \end{bmatrix}, \begin{bmatrix} K_{XX} + \sigma^2 I & K_{X\boldsymbol{x}^\star} \\ K_{\boldsymbol{x}^\star X} & K_{\boldsymbol{x}^\star\boldsymbol{x}^\star} + \sigma^2 I \end{bmatrix}\right) \tag{4}$$

The joint distribution is conditioned on the observed values to get the predictive distribution corresponding to a new input $\boldsymbol{x}^\star$ as,

$$y^\star \mid \boldsymbol{y}, X, \boldsymbol{x}^\star \sim \mathcal{N}(\hat{\mu}(x^\star), \hat{\Sigma}(x^\star)) \tag{5}$$

$$\hat{\mu}(\boldsymbol{x}^\star) = \mu(\boldsymbol{x}^\star) + K_{\boldsymbol{x}^\star X}(K_{XX} + \sigma^2 I)^{-1}(\boldsymbol{y} - \mu(X)) \tag{6}$$

$$\hat{\Sigma}(\boldsymbol{x}^\star) = K_{\boldsymbol{x}^\star\boldsymbol{x}^\star} - K_{\boldsymbol{x}^\star X}(K_{XX} + \sigma^2 I)^{-1} K_{X\boldsymbol{x}^\star} + \sigma^2 I \tag{7}$$

The hyperparameters σ_h and $\boldsymbol{l}$ may be fixed by the user, but an optimal value is often inferred from the data using type-II maximum likelihood (Rasmussen and Williams, 2006), wherein the negative log marginal likelihood is minimized with respect

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

to the hyperparameters. The negative log marginal likelihood is defined as,

$$-\log p(\boldsymbol{y}|X,\sigma_h,l_h) = \\ \frac{1}{2}(\boldsymbol{y}-\mu(X))^\top (K_{XX}+\sigma^2 I)^{-1}(\boldsymbol{y}-\mu(X)) \\ +\frac{1}{2}\log|K_{XX}+\sigma^2 I| + \frac{n}{2}\log 2\pi \tag{8}$$

The L-BFGS-B algorithm (Zhu et al., 1997) is used for optimization.

2.2 Mixture density networks

A mixture density network is a probabilistic regression method that combines Gaussian mixture models with artificial neural networks (Bishop, 1994). The conditional distribution of the target is represented by a mixture of Gaussian distributions,

$$p(y \mid \boldsymbol{x}) = \sum_{i=1}^{m} \alpha_i(\boldsymbol{x}) \mathcal{N}(y \mid \boldsymbol{\mu}_i(\boldsymbol{x}), \sigma_i^2(\boldsymbol{x})), \tag{9}$$

where $\alpha_i(\boldsymbol{x})$ are the weights or coefficients assigned to the i^{th} mixture component, and $\mathcal{N}(y \mid \boldsymbol{\mu}_i(\boldsymbol{x}), \sigma_i^2(\boldsymbol{x}))$ is a Gaussian kernel representing the conditional density of the i^{th} component of the target distribution, with parameters $\boldsymbol{\mu}_i(\boldsymbol{x})$ and $\sigma_i(\boldsymbol{x})$.

Instead of mapping the inflow features $\boldsymbol{x}$ to the load statistics y directly, the neural network is trained to predict the parameter vector, $\boldsymbol{z} \in \mathbb{R}$ consisting of α_j, μ_j and σ_j for $j = 1...m$.

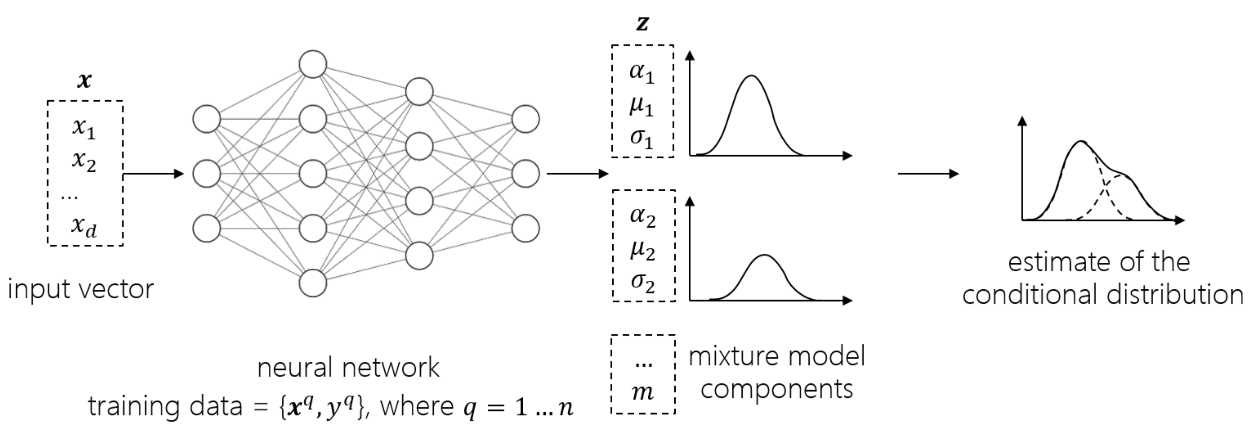


Figure 2. Schematic representation of Mixture Density Networks.

The mixing coefficients $\alpha_i(\boldsymbol{x})$ must sum up to exactly 1. A *softmax* function is used to handle this constraint,

$$\alpha_i = \frac{\exp z_i^\alpha}{\sum_{j=1}^{m} \exp z_j^\alpha}, \tag{10}$$

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

where z_i^α are the network outputs predicting the mixture coefficients. Similarly, positive values of the standard deviation are ensured by representing them as exponential functions of the corresponding network outputs, z_i^σ,

$$\sigma_i = \exp(z_i^\sigma). \tag{11}$$

The likelihood $\mathcal{L}$ of the dataset is given by,

$$\mathcal{L} = \prod_{q=1}^{n} p(y^q \mid \boldsymbol{x}^q) p(\boldsymbol{x}^q) \tag{12}$$

A very commonly used error functions for probabilistic models is the negative log of the likelihood. From (9) and (12), the error can be written as,

$$E^q = -\ln\{\sum_{i=1}^{m} \alpha_i(\boldsymbol{x}^q) \mathcal{N}(y \mid \boldsymbol{\mu}_i(\boldsymbol{x}^q), \sigma_i^2(\boldsymbol{x}^q))\}, \tag{13}$$

where $p(\boldsymbol{x}^q)$ is not included as it is constant with respect to the parameters or weights. The derivative of the error function is calculated at the output layer and is back-propagated to get its gradient with respect to the network weights. Finally, we have everything we need to minimize the error function using a gradient descent optimization. In this study, we use the Adam optimizer (Kingma and Ba, 2017) to perform stochastic gradient descent. A 10-fold cross-validation set over 600 samples is performed at every training.

The hidden layers in our network use the rectified linear unit (ReLU), defined as,

$$ReLU(x) = \begin{cases} x & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases} \tag{14}$$

The output layer of the network does not have an activation function; therefore, the outputs are just linear combinations of the inputs from the previous layer.

Minimizing the error function is an ill-posed problem as there is a conflict between learning the function that fits the data perfectly and remaining robust under varying sets of training data. As the network size grows, the function space increases, and the tendency of the neural network is to overfit. Among several ways to avoid overfitting (Montavon et al., 2012), in this study, we implemented a combination of *early-stopping* (Yao et al., 2007) and *L1* and *L2 regularization* (Ng, 2004).

Early stopping

The error function measured on the cross-validation dataset first decreases, then starts increasing as the network begins overfitting the training data. This can be avoided by applying an early stopping mechanism that stops training as the validation loss stops decreasing over a certain number of iterations. We experimented with a range of early stopping iterations and found 100

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

to be sufficient for the negative log-likelihood on the validation dataset to converge, but not over-fit. That is, if the validation loss did not show any improvement after 100 iterations, we stopped training the model.

L1 and L2 regularization

L1-regularization penalizes the error function with the sum of the magnitude of the weights,

$$E_R^q = E^q + \lambda \sum |w_i| \tag{15}$$

It pushes the coefficients of uninformative features towards zero, effectively pruning the feature space.

Weight-decay or L2-regularization, on the other hand, penalizes the error function with a fraction of the squared magnitude of the weights,

$$E_R^q = E^q + \lambda \sum w_i^2 \tag{16}$$

L2-regularization encourages the weights to be small. In both approaches, λ is the *regularization parameter*. It is a hyperparameter that controls the complexity of the model, and the optimal value can be chosen using a search algorithm. We found that heavy regularization with $l1$ and $l2$ values of 0.1 deteriorated the results by over-smoothing the conditional response. On the other hand, no $l2$ regularization also resulted in relatively smaller R^2 values for the standard deviation prediction, likely due to some degree of over-fitting. On the basis of this hyperparameter study on one channel, we decided on a conservative value of $1e-3$ for both $l1$ and $l2$.

The main hyperparameters used in this study to train the models to obtain the results in Section 4 are summarized in Table 1. The features and targets are scaled with a standard scaler before training.

3 Setup of the OpenFAST engineering model

3.1 OpenFAST modeling approach

The surrogate is modeled on the responses of an aero-hydro-servo-elastic code, OpenFAST, which is used as ground truth in this study. It is a state-of-the-art, multi-physics numerical tool for modeling wind turbines. It combines analytical and empirical formulations with conservative assumptions to simplify the code and limit the computational cost. It can model environmental conditions like stochastic waves, currents, and a frozen wind turbulence field with randomized coherent turbulent structures superimposed on the random, homogeneous, background turbulence.

OpenFAST is used for setting up a numerical model of the real-world environment and system dynamics to produce the training data for the surrogate. All simulations are performed on the IEA-10MW (Bortolotti et al., 2019) offshore reference wind turbine. TurbSim (Jonkman and Buhl, 2007) simulations for inflow turbulence generation are performed with a grid resolution of 40 points in a $1.16\mathrm{D} \times 1.16\mathrm{D}$ domain, D being the rotor diameter. The total simulation duration is 900s, out of which the first 300s is discarded to exclude the initial transient. Based on the literature stated in Section 1.1, ten-minute

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

Table 1. Summary of the network hyperparameters

Network hyperparameter	Value
Number of mixture components	4
Hidden layers	2
Activation function (hidden layers)	ReLU
Activation function (output layer)	None
Learning rate	0.005
Maximum epochs	1000
Mini-batch size	100
Optimizer	Adam
Regularization	
λ for $L1-$regularization	10^{-3}
λ for $L2-$regularization	10^{-3}
Early-stopping	
Early-stopping patience	100
Early-stopping monitor	validation loss
Number of early-stopping validation samples	600

simulations on their own are insufficient for fatigue load estimations. However, the statistical variations in loads that one would expect over longer periods or multiple seed repetitions can be potentially inferred indirectly via probabilistic surrogates based on the variation in the quantity of interest at neighboring training samples. Therefore, ten-minute statistics are sufficient for a complete description of the load response as long as unique turbulence and wave seeds are used for each training simulation.

The loads are calculated with the ElastoDyn module in OpenFAST, which uses the Euler-Bernoulli beam theory to calculate the bending moments by assuming the structure to be straight and isotropic. The ServoDyn module is used to control the wind turbine. The controller settings differ from the DTU Wind Energy controller used in the HAWC2 simulations of the IEA-10MW reference document (Bortolotti et al., 2019). The main difference appears at low wind speeds, where the rotor RPM is not restricted to 6, and the collective blade pitch is zero until the rated wind speed. The OpenFAST simulation output with these controller settings at low wind speeds may interfere with the tower's natural frequencies; however, modification of the controller is beyond the scope of this study. The simulations are performed with single precision to limit file size and simulation times without any significant impact on the accuracy of the loads.

The implementation of the IEA-10MW-RWT in OpenFAST is relatively new (Bortolotti et al., 2019). As such, there continue to be constant updates to the public model based on user feedback. This study uses the IEA-10MW-RWT as ground truth for the surrogate modeling study. For that purpose, it need not be perfectly accurate but representative of the expected load response class. The machine learning methodology is expected to be easily transferable to different wind turbine types.

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

3.2 Definition and sampling of the input features

The IEA-10MW-RWT is designed for offshore conditions. However, in this study, we simulate it on both onshore (CASE-ONSHORE) and offshore (CASE-OFFSHORE) sites to be able to evaluate the additional training requirements in the offshore case against an onshore reference.

3.2.1 CASE-ONSHORE

Aero-servo-elastic: The wind turbine is subjected only to aerodynamic loading. Wind speed, power-law exponent, and turbulence intensity are selected as the input parameters for the aerodynamic simulations as they have been shown to have the highest impact on the load response in previous studies (Dimitrov et al., 2018). The variable bounds are also the same as the ones defined in (Dimitrov et al., 2018), listed in Table 2. The power-law exponent and turbulence intensity are functions of the wind speed. R and z are the rotor radius and the hub height, respectively. The samples are drawn from a three-dimensional Sobol sequence to ensure an even spread of points in the sample space. The random seed for turbulence generation is not included as a training variable. Each sample is therefore associated with a unique random seed, there are no repetitions.

3.2.2 CASE-OFFSHORE

Aero-servo-hydro-elastic: The offshore wind turbine is placed on a monopile foundation at 30m water depth and is subject to both aerodynamic and hydrodynamic loading. Along with the aerodynamic parameters of CASE-ONSHORE, there are additional wave parameters in this case as listed in Table 2. For designing load surrogates suitable for multiple sites, ideally the $H_s - T_p$ diagrams from several sites should be combined to define conservative ranges for the two variables. In this study, as an example, we sample the values from a joint $H_s - T_p$ kernel from a representative distribution. Additionally, the minimum and maximum range of H_s may also be defined as a function of wind speed in order to sample from the joint $u - H_s - T_p$ distribution. The first order waves are modeled using the JONSWAP spectrum in HydroDyn. The values of the aerodynamic variables are the same as in CASE-ONSHORE. In particular, the Turbsim solution files are therefore shared between CASE-ONSHORE and CASE-OFFSHORE. Similar to CASE-ONSHORE, the wave and turbulence seeds are not included in training the model. Each sample is therefore associated with a unique set of random seeds, there are no repetitions.

3.3 Responses

We assessed load statistics like mean, max, and fatigue primarily at the tower base (TwrBs) and tower top (YawBr) fore-aft moments. MDN surrogates can, in theory, also be used to model blade loads, blade aerodynamics, gearbox loads, maximum blade displacement, power output, or nacelle acceleration.

Since the direction of the incoming flow is always aligned with the rotor, the fore-aft direction at the tower base is defined in the local coordinate system of the inflow wind. The tower bottom loads must be projected appropriately in the global coordinate system of the wind turbine before integrating to calculate the lifetime fatigue damage in the global coordinate system. In this study, we only calculate the short-term damage in the local coordinate system. The 10-minute fatigue is calculated using short-

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

Table 2. Variables and variable bounds for CASE-ONSHORE and CASE-OFFSHORE

Variable	Lower Bound	Upper Bound	Sampling
Wind Speed (u) [ms^{-1}]	4	25	Uniform, Sobol
Turbulence Intensity (ti) [%]	2.5	$\frac{18}{u}(6.8+0.75u+3(\frac{10}{u})^2)$	Uniform, Sobol
Power Law Exponent (α) [$-$]	$0.15-0.23(\frac{u_{max}}{u})(1-(0.4\log\frac{R}{z})^2)$	$0.22+0.4(\frac{R}{z})(\frac{u_{max}}{u})$	Uniform, Sobol
Significant Wave Height (H_s) [m]	0	6	Kernel density estimate, pseudo-random
Spectral Peak Period (T_p) [s]	1	21	Kernel density estimate, pseudo-random
Wave Direction ($wdir$) [deg]	-180	180	Uniform, pseudo-random
Turbulence Random Seed [$-$]	-50000	50000	Uniform, pseudo-random
Wave Random Seed 1 [$-$]	-50000	50000	Uniform, pseudo-random
Wave Random Seed 2 [$-$]	-50000	50000	Uniform, pseudo-random

term damage equivalent loads (DEL^{ST}). DEL^{ST} converts the irregular load time series to a constant amplitude and frequency signal that produces an equivalent amount of fatigue damage loads. Rainflow counting (Matsuishi and Endo, 1968) algorithm is used to obtain the load ranges S_i and the number of load cycles n_i needed to calculate the DEL^{ST} as,

$$DEL^{ST} := \left(\frac{n_i S_i^m}{n^{ref}}\right)^{1/m}, \tag{17}$$

where n_{ref} is 600 for 1Hz DELs over 10 minutes. m is the Wöhler coefficient with values 3.5 for the tower, 10 for blade flapwise, and 8 for blade edgewise moments.

3.4 Test datasets

The prediction accuracy of the conditional pdf by MDNs is tested on two independently sampled test datasets, that have not been used in the training procedure, referred to as TEST1 and TEST2.

TEST1 consists of 50 pseudo-randomly-selected points spanning the entire sampling domain, with parameter bounds the same as in Table 2. At each test point, engineering simulations with OpenFAST are repeated 300 times to get a reference pdf by keeping the input features constant but changing the turbulence and wave random seeds, resulting in a total of 15000 TurbSim and OpenFAST simulations. In TEST2, we alter only the wind speed and turbulence intensity, keeping the other inflow parameters constant. The values are listed in Table 3.

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

Table 3. Variables and their corresponding values in TEST2 dataset.

Variable parameters	[Min : Max : Δ]
Wind Speed (u) [ms^{-1}]	[6 : 21 : 3]
Turbulence Intensity (ti) [%]	$\begin{cases} [10:40:10] & \text{for } u = 6 \\ [6:24:6] & \text{for } u = 9 \\ [4:16:4] & \text{for } u > 10 \end{cases}$
Fixed parameters	**Value**
Power Law Exponent (α) [−]	0.08
Significant Wave Height (H_s) [m]	1
Spectral Peak Period (T_p) [s]	7
Wave Direction ($wdir$) [deg]	0
Random seeds	**Value**
Turbulence Random Seed [−]	$\mathcal{U}(-50000, 50000)$
Wave Random Seed 1 [−]	$\mathcal{U}(-50000, 50000)$
Wave Random Seed 2 [−]	$\mathcal{U}(-50000, 50000)$

The test and training points for CASE-OFFSHORE are shown in Figure 3. The test points for CASE-ONSHORE are identical, but only for the turbulence inflow features, namely, u, ti, and α.

Figure 4 shows, as an example, the 10-minute average tower bottom fore-aft moment as a function of wind speed, along with the conditional distributions at two wind speeds from the TEST1 dataset. The surrogate models aim to predict this kind of conditional variation in the loads due to the stochastic inflow without the need for seed repetitions during training.

3.5 Accuracy metric

The qualitative assessment of the performance of the surrogate model is based on two criteria: the coefficient of determination and the Wasserstein distance, as further described hereafter.

3.5.1 Coefficient of determination R^2

The coefficient of determination, also known as the R^2, is a common measure of the goodness of fit of a model. It is defined as,

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)}{\sum(y_i - \bar{y})}, \tag{18}$$

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

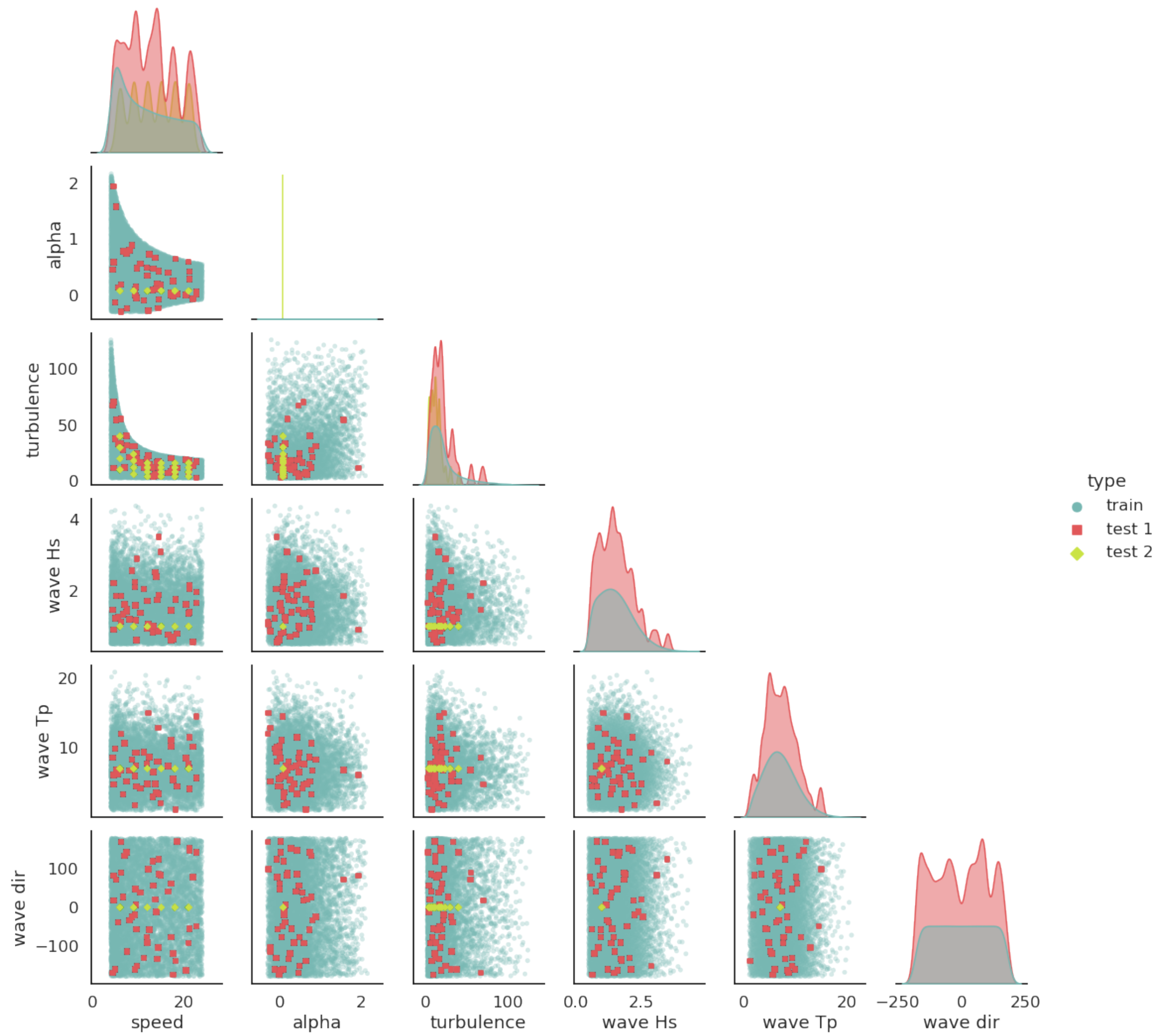


Figure 3. Pairplot of the input features for CASE-OFFSHORE showing the training samples along with the test datasets.

where $\hat{y}_i$ is the predicted output, y_i is the observed value and $\bar{y}$ is the mean of the observed values. R^2 is interpreted as the linear correlation between the predicted and observed values of the output vector. To assess the accuracy of the predicted conditional distribution of the response compared to the OpenFAST reference (Figure 4), we calculate the R^2 value for the conditional pdf's mean, standard deviation, 5%, and 95% quantiles.

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

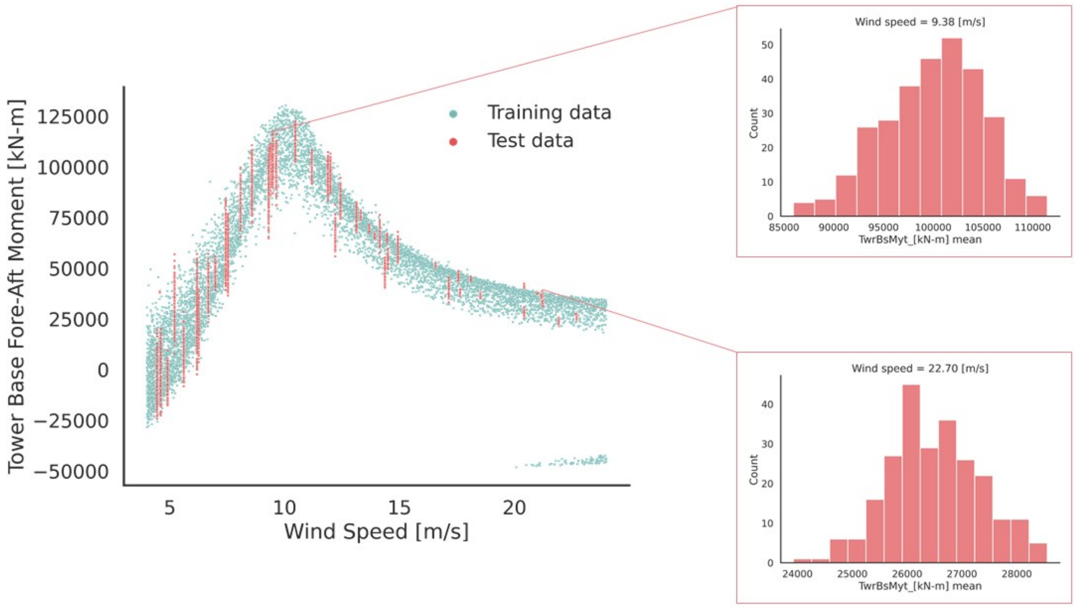


Figure 4. The left plot shows the 10-minute average tower bottom fore-aft moment as a function of wind speed. The samples belong to the training database of CASE-OFFSHORE. On the right, two examples of reference histograms generated using 300 turbulence seed repetitions in OpenFAST at wind speeds of $9.38\mathrm{ms}^{-1}$ and $22.7\mathrm{ms}^{-1}$ are shown from the TEST1 dataset.

3.5.2 Wasserstein distance

The Wasserstein metric is a distance function to compare the difference between the pdfs of any two random variables. It is symmetric, non-negative, and satisfies the triangle inequality, making it a proper distance. The normalized 2-Wasserstein distance (Villani, 2009; Peyré and Cuturi, 2019; Ramdas et al., 2015) between two pdfs Y and $\hat{Y}$ is defined as,

$$d_{W2}(Y,\hat{Y}) = \frac{(\int_0^1 |F^{-1}(t) - G^{-1}(t)|^2 dt)^{1/2}}{\sigma(Y)}, \tag{19}$$

where F^{-1} and G^{-1} are the quantile functions of Y and $\hat{Y}$ respectively. A value of $d_{W2} = 1$ is, therefore, the distance between a distribution with mean $\mu(Y)$, scale $\sigma(Y)$, and a degenerate distribution with the same mean.

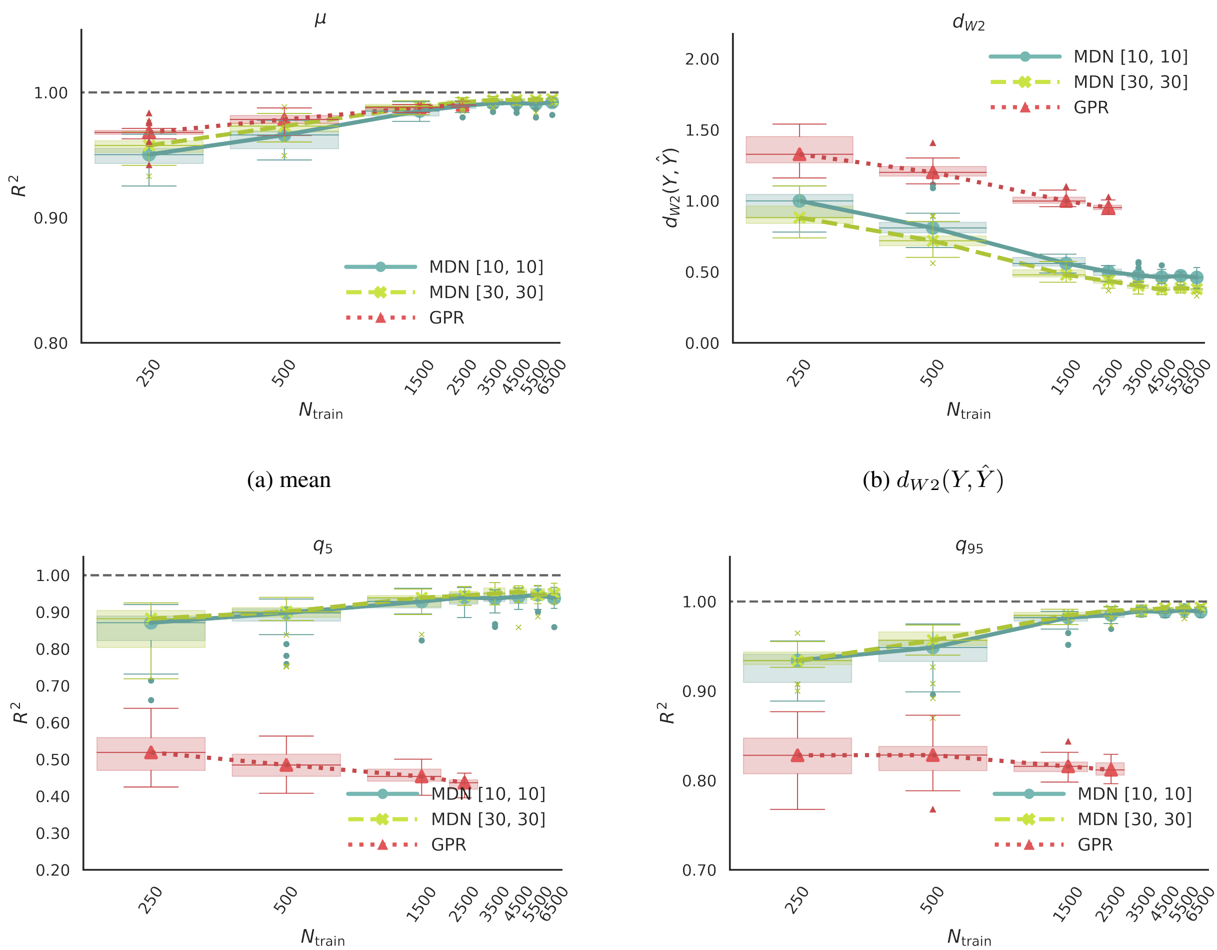
4 Results and discussion

In this section, we assess how well the surrogate models predict the conditional load distribution on the TEST1 and TEST2 datasets mentioned in Section 3.4. The first part focuses on convergence studies, specifically the impact of training data size on the prediction of the average 10-minute standard deviation of the tower bottom fore-aft moment. The goal is to measure how

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

the accuracy of the predictions varies based on the hyperparameters and initialization of the optimization algorithm. Once the network architecture and training sample size are fixed, we do a rigorous analysis of the model's performance in Section 4.2.

4.1 Convergence



(a) mean

(b) $d_{W2}(Y,\hat{Y})$

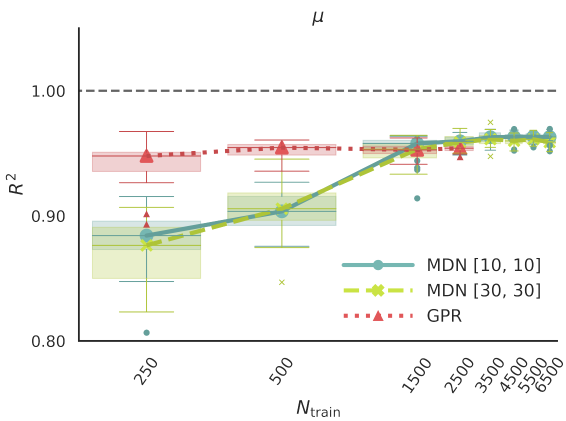
(c) 0.05 quantile

(d) 0.95 quantile

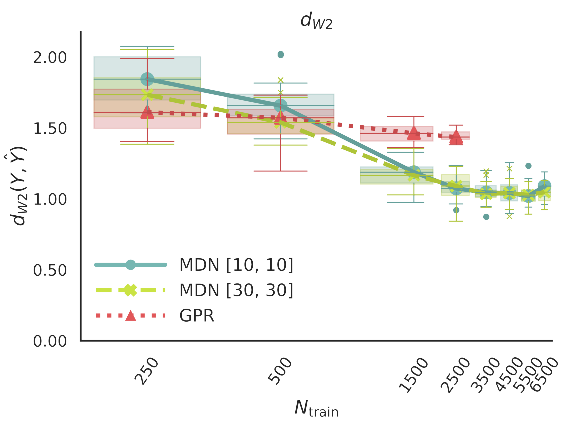
Figure 5. CASE-ONSHORE: Figures showing the change in the normalized 2-Wasserstein distance, R^2 value of the mean, 0.05 quantile and 0.95 quantile of the predicted pdf as a function of the training samples. The study is performed on the tower base fore-aft moment standard deviation (TwrBsMyt stddev).

330 Generally speaking, more data translates to better accuracy. However, an increase in data after a certain point gives diminishing returns in accuracy. Given the computational cost of generating the training database, we want to ensure a good model fit with as little training data as possible.

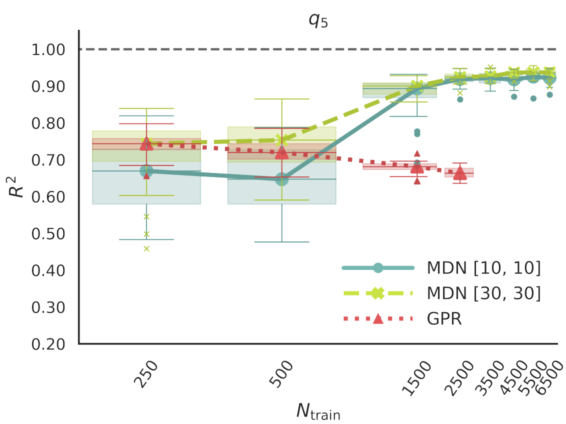
https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024



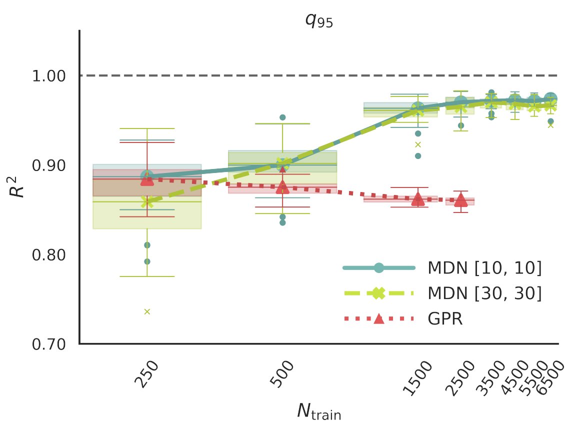
(a) mean



(b) $d_{W2}(Y, \hat{Y})$



(c) 0.05 quantile



(d) 0.95 quantile

Figure 6. CASE-OFFSHORE: Figures showing the change in the normalized 2-Wasserstein distance, R^2 value of the mean, 0.05 quantile and 0.95 quantile of the predicted pdf as a function of the training samples. The study is performed on the tower base fore-aft moment standard deviation (TwrBsMyt stddev).

In this section, we look at the convergence of the model with respect to the number of training samples in two two-layer networks with 10 ($[10, 10]$) and 30 ($[30, 30]$) units in each layer. Two different networks are chosen to comment on the robustness of the approach with respect to the network architecture. The convergence study is performed both on CASE-ONSHORE in Figure 5 and CASE-OFFSHORE in Figure 6. The network is trained on the tower base fore-aft moment standard deviation (TwrBsMyt [kN-m] stddev), as it is found to be the most difficult to fit. At every N_{train}, the model training is repeated on 25 uniquely sampled subsets of the data with 10-fold cross-validation. The plots in Figure 6 show the convergence of the model in terms of predicting the normalized 2-Wasserstein distance and statistics, including the response pdf's mean, 5% quantile and 95% quantile. The metrics are averaged over the validation dataset formed by combining TEST1 and TEST2.

Figure 5 and Figure 6 also show the Gaussian process regression predictions with 25 repetitions. We expect GPR to only predict the right estimate of the mean of the response. Since it is based on Bayesian inference, which is very different from the back-propagation mechanism used in MDNs, it can infer the response estimate with a much smaller training dataset. The GPR model is not trained on datasets larger than 2500 samples because it scales poorly and the training expense grows exponentially.

In both CASE-ONSHORE and CASE-OFFSHORE, the difference between the predictions of the two MDN architectures $[10,10]$ and $[30,30]$ is negligible for μ, $q_{5\%}$ and $q_{95\%}$. An improvement of roughly 15% in terms of d_{W2} is seen in CASE-ONSHORE, whereas a negligible difference is observed in CASE-OFFSHORE. As we do not have an infinite pool of data, the uncertainty bounds concerning the choice of the training samples progressively reduce as we approach the total available training samples. At smaller N_{train} values, the uncertainty is also driven by the missing information in the training data and the choice of the initial conditions used by the stochastic gradient descent optimizer. Significantly better GPR estimates of the response mean for less than 1500 training samples can be attributed to the Bayesian formulation. Beyond that point, however, MDNs and GPR are comparable, with $R^2 > 0.99$ in CASE-ONSHORE and $R^2 > 0.95$ in CASE-OFFSHORE. MDN significantly better estimates all other quantities.

The estimates of the tails of the pdf, quantified by the lower 5% quantile, are extremely well captured by MDN in both onshore and offshore case studies. Overall, the model's accuracy in terms of the statistical quantities is approximately 5% better in CASE-ONSHORE for the same number of training points. However, the average 2-Wasserstein distance is 50% smaller in CASE-ONSHORE than in CASE-OFFSHORE, signifying, overall, a much better inference of the latent pdf in the onshore conditions than offshore.

Figure 7 shows the training and validation losses plotted against the number of epochs for CASE-ONSHORE. MDN overfits the data at $N_{train} = 500$ because, as the training loss decreases, the validation loss increases, indicating that the model cannot handle previously unseen data. Figure 7b is well-fitted as the training and validation losses decrease at the same rate. The plot also shows the auto-stop algorithm at work, which halts training after 100 epochs of approximately zero-gradient loss to avoid overfitting.

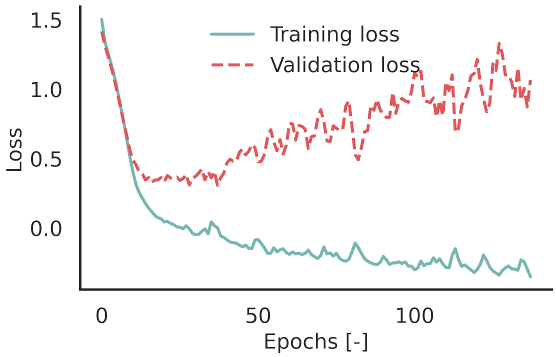
For the remainder of this study, we will use a two-layer MDN with ten activation units in each layer trained on 4500 samples. It offers a good balance between training time, model complexity, and accuracy. For GPR, a training set of 500 samples will be used as it is found to be sufficiently accurate.

4.2 Load prediction

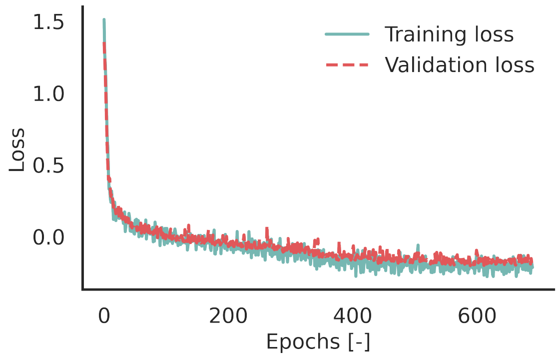
In this section, we evaluate the prediction of the 10-minute damage equivalent loads on the wind turbine for CASE-ONSHORE and CASE-OFFSHORE. Table 4 summarizes the predictions of various statistical properties of the response pdf for both cases in terms of the R^2 values. The absolute magnitude of R^2 is sensitive to the optimization initialization, choice of the test samples, and the choice of the training subset as seen in Figure 6. It is, therefore, important to note that the absolute R^2 values

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS



(a) $N_{train} = 500$



(b) $N_{train} = 4500$

Figure 7. Training and validation losses for the [10, 10] MDN with 500 (a) and 4500 (b) training samples, CASE-ONSHORE.

do not carry much objective meaning on their own. They are only used here for comparing the performance of models relative to one another. Figure 8 shows the corresponding plots for CASE-OFFSHORE.

Table 4. Comparison of the prediction of the statistical properties of the response pdf for the tower base fore-aft loads

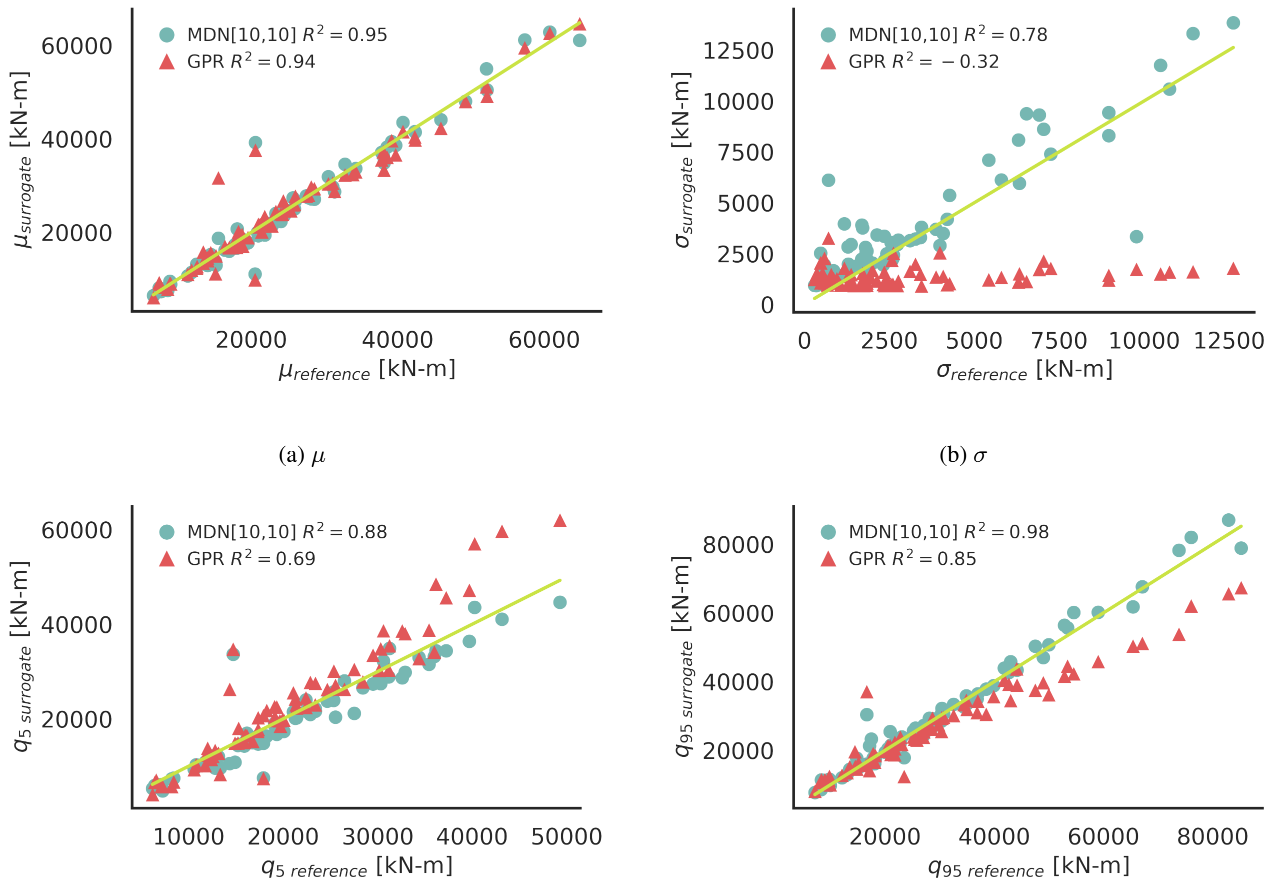
Tower base fore-aft 10-minute DEL R^2

	CASE-ONSHORE				**CASE-OFFSHORE**			
	μ	σ	q_5	q_{95}	μ	σ	q_5	q_{95}
MDN	0.994	0.942	0.962	0.991	0.955	0.782	0.879	0.977
GPR	0.943	-0.85	0.488	0.819	0.936	-0.322	0.693	0.852

Overall, MDN performs better than GPR across all the metrics listed in Table 4 for CASE-ONSHORE and CASE-OFFSHORE. The conditional average, μ, is well estimated by both models. The standard deviation, σ, is constant in the case of the GPR model, as it is a homoscedastic formulation. The minor variations in Figure 8b in $\sigma_{surrogate}$ can be attributed to a combination of model uncertainty and Monte Carlo sampling. Estimates of the standard deviation by MDN are excellent in CASE-ONSHORE, but the performance drops in CASE-OFFSHORE. However, the results are encouraging compared to GPR and show that MDN can handle heteroscedastic datasets. The 5% and 95% quantiles, which are essential for design considerations, are exceptionally well predicted by MDN. In Figure 8c and Figure 8d, GPR shows a bias in the quantile estimate, increasing with the quantity's magnitude, which can be directly ascribed to the underestimation of the standard deviation of the response.

The R^2 values show that the model fit for CASE-ONSHORE is relatively better than for CASE-OFFSHORE. As noted in the previous section, it appears to be much easier to train a case with only aerodynamic features for the same number of training samples and network architecture.

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024



(a) μ (b) σ (c) q_5 (d) q_{95}

Figure 8. CASE-OFFSHORE: Prediction of the statistics of the response pdf of the 10-minute damage equivalent loads for the tower base fore-aft moment.

On taking a closer look at the conditional pdfs, it becomes clear why the predictions made by MDN are superior. The 10-minute DELs for the tower base fore-aft moment (Figure 9), tower top fore-aft moment (Figure 10), blade root flapwise moment (Figure 11) and blade root edgewise moment (Figure 12) are plotted at three operational conditions falling in low, medium and high wind speed blocks. The values of the input features are noted in Table 5.

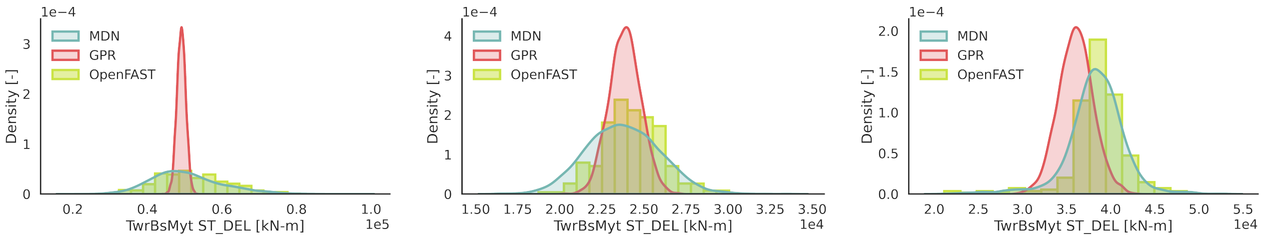
Clearly, the responses are not always normally distributed. The variance of the response is not constant across wind speeds. Near the cut-out wind speed, we also notice a multi-modal response, as the wind turbine switches between idling and power production, depending on the local variations in the inflow wind patterns. Here, MDN is shown to leverage the flexibility of learning complex noise patterns to then approximate the full picture of the response that deterministic models would otherwise miss.

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 5. Values of the input features for the pdfs in Figures 9 to 12.

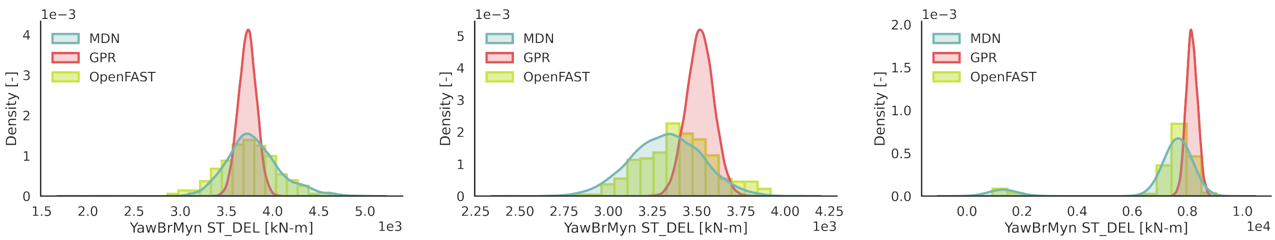
Wind condition	u [ms^{-1}]	ti [-]	α [-]	H_s [m]	T_p [s]	$wdir$ [deg]
Low wind speed	6	40	0.08	1.0	7.0	0
Medium wind speed	12	16	0.08	1.0	7.0	0
High wind speed	21.2	18.5	0.42	2.2	11.5	148.9



(a) Low wind speed (b) Medium wind speed (c) High wind speed

Figure 9. CASE-OFFSHORE: Predicted and reference (OpenFAST) conditional pdf for the tower base fore-aft moment 10-minute DEL.

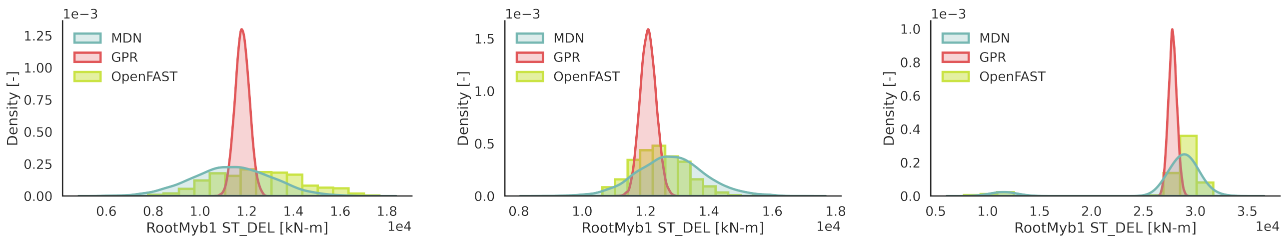
In Figure 13, d_{W2} is plotted on TEST2 dataset. We notice that there are certain test samples where neither MDN nor GPR is successful in inferring the underlying function. From the figure, it appears that the model is consistently unable to detect the correct patterns at very low turbulence intensities across different wind speeds and load channels. The poor performance at low turbulence could result from insufficient training data, causing the model to regress to the mean, non-plausible operating conditions that introduce an unexpected gradient in the response surface or a modeling error such as mode collapse. Application-wise, these regions are not the most critical, as fatigue is primarily driven by larger turbulent disturbances.



(a) Low wind speed (b) Medium wind speed (c) High wind speed

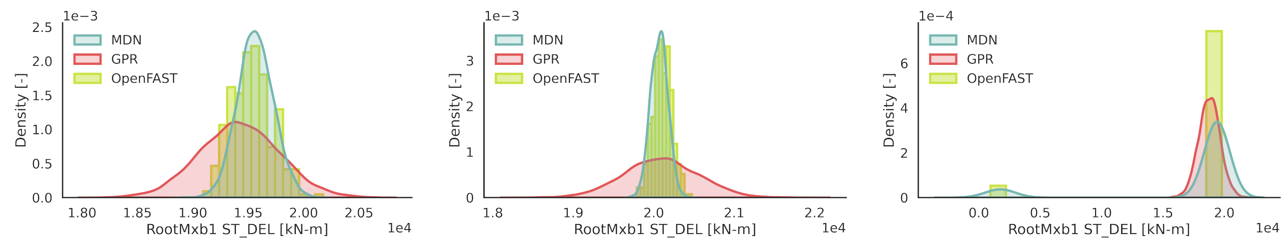
Figure 10. CASE-OFFSHORE: Predicted and reference (OpenFAST) conditional pdf for the tower top fore-aft moment 10-minute DEL.

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024



(a) Low wind speed (b) Medium wind speed (c) High wind speed

Figure 11. CASE-OFFSHORE: Predicted and reference (OpenFAST) conditional pdf for the blade root flapwise moment 10-minute DEL.



(a) Low wind speed (b) Medium wind speed (c) High wind speed

Figure 12. CASE-OFFSHORE: Predicted and reference (OpenFAST) conditional pdf for the blade root edgewise moment 10-minute DEL.

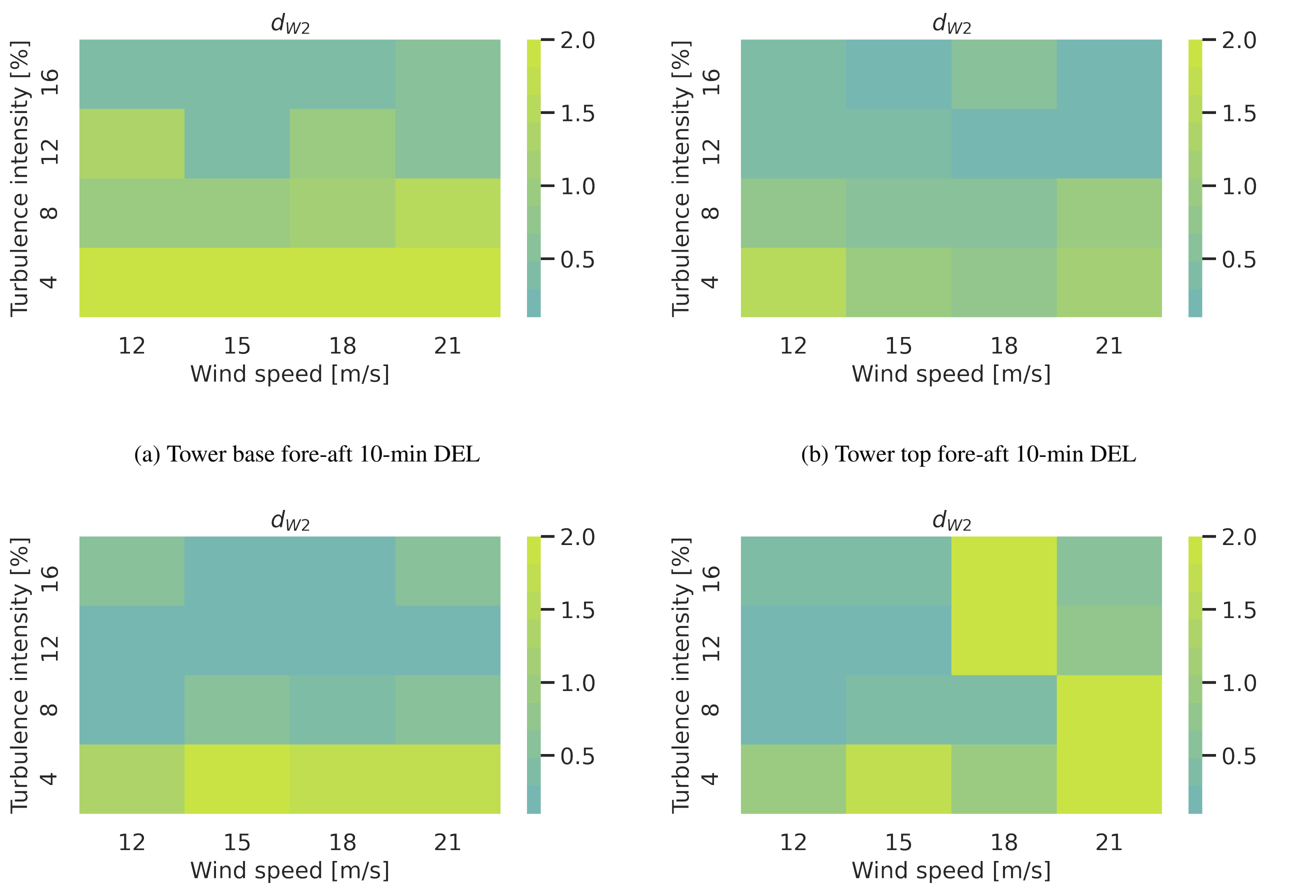
5 Conclusions

This paper presents a novel probabilistic approach based on mixture density networks to make efficient and flexible load surrogates for offshore siting. The data-driven surrogate uses aero-servo-hydro-elastic OpenFAST simulations of the 10-MW reference wind turbine for training. We compare the performance of MDN to the widely used Gaussian process regression model and show an improvement in the estimation of the load uncertainty associated with the stochastic representation of inflow turbulence and waves.

The surrogate is trained on a wind turbine subject to aerodynamic (CASE-ONSHORE) and aero-hydrodynamic (CASE-OFFSHORE) loading with the intent of comparing the difficulty in designing load surrogates for the two cases. The reference conditional pdfs for validating the models' performance are produced using 300 random seeds at each of the 50 combinations of inflow conditions. A convergence study is performed to assess the accuracy of the surrogate as a function of the number of training samples. Two different MDN architectures and the standard Gaussian process regression are evaluated. It is shown that the surrogate is more accurate for the same number of training samples in CASE-ONSHORE (three features) as opposed to CASE-OFFSHORE (six features), based on the 2-Wasserstein distance between the predicted and the reference conditional pdf of the response. A minimum of 2500 samples are required by MDN to surpass a R^2 value of 0.95 for the prediction of the mean and quantiles in CASE-OFFSHORE. The GPR model is shown to be more accurate in predicting the mean of the

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

WIND ENERGY SCIENCE DISCUSSIONS



(a) Tower base fore-aft 10-min DEL

(b) Tower top fore-aft 10-min DEL

(c) Blade root flapwise 10-min DEL

(d) Blade root edgewise 10-min DEL

Figure 13. Normalized 2-Wasserstein distance computed on CASE-OFFSHORE validation dataset for the MDN model. The performance is plotted on a turbulence intensity and wind speed grid.

response even with a small dataset of 250 samples. However, beyond 1500 samples, MDN predictions are consistently better. The quantiles are well captured by MDN in both cases.

The conditional pdfs from the validation dataset are evaluated for low, medium, and high wind speed cases to demonstrate the ability of MDN to capture heteroscedastic, multi-modal responses with high accuracy, even with limited training data. We note a poor performance of the MDN model at low turbulence intensity conditions across all load channels, indicating either the need for a higher sampling rate in those regions in the training dataset or the presence of a sharp gradient in the response surface that the model could not appropriately capture.

The probabilistic modeling of the loads, although shown to have a slight improvement in the prediction of the expectation of the response compared to the state-of-the-art Gaussian process regression, can capture the variances and quantiles of the

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

response far better. With the added benefit of not needing seed repetitions prior to training, we show that this approach also
425 cuts down significantly on the computational cost associated with generating the training database. This work shows promising results for using MDN as a surrogate in site assessment of onshore and offshore wind turbines. Work is currently in progress to determine how the information on the uncertainty in the short-term load response can be propagated to the lifetime loads to help inform engineering decisions.

Data availability. Datasets related to this article, described in Section 3, can be found at https://doi.org/10.4121/21939995.v1, hosted at
430 4TU.ResearchData (Singh, 2023).

Appendix A: Machine learning framework

Figure A1 shows the basic framework used for model calibration and load estimation. All data generated from OpenFAST is included in the training base without any repetition or pre-filtering step involved.

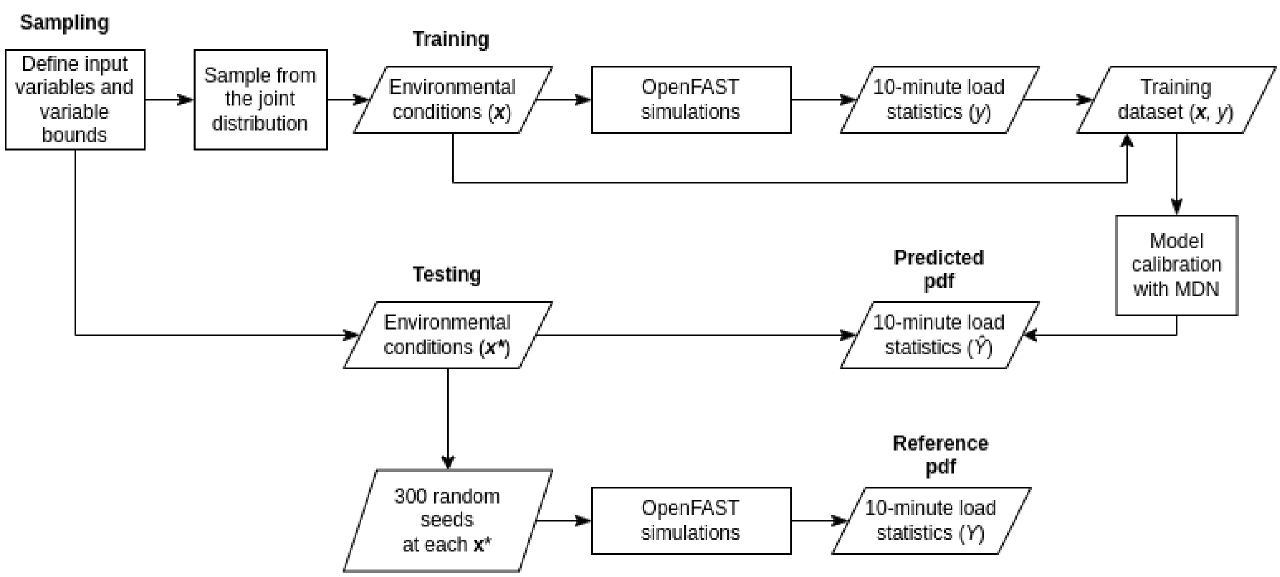


Figure A1. Schematic of the machine learning framework.

Author contributions. **Deepali Singh:** Conceptualization, Methodology, Software, Validation, Data curation, Investigation, Writing - original
435 draft, Visualization. **Richard P. Dwight:** Conceptualization, Supervision, Methodology, Writing - review & editing, Project administration. **Axelle Viré:** Supervision, Writing - review & editing, Project administration, Funding acquisition.

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements. The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 860737 (STEP4WIND project, step4wind.eu). The authors are grateful to Kasper Laugesen and Erik Haugen (Siemens Gamesa Renewable Energy, Denmark), for their valuable feedback.

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

References

Abdallah, I., Lataniotis, C., and Sudret, B.: Parametric hierarchical kriging for multi-fidelity aero-servo-elastic simulators — Application to extreme loads on wind turbines, Probabilistic Engineering Mechanics, 55, 67–77, https://doi.org/10.1016/j.probengmech.2018.10.001, 2019.

Avendaño-Valencia, L. D., Abdallah, I., and Chatzi, E.: Virtual fatigue diagnostics of wake-affected wind turbine via Gaussian process regression, Renewable Energy, 170, 539–561, https://doi.org/10.1016/j.renene.2021.02.003, 2021.

Bishop, C. M.: Mixture density networks, Tech. rep., Aston University, ISBN NCRG/94/004, 1994.

Bishop, C. M.: Pattern Recognition and Machine Learning, Springer New York, ISBN 9780387310732, https://doi.org/10.1007/978-3-030-57077-4_11, 2006.

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D.: Variational Inference: A Review for Statisticians, Journal of the American Statistical Association, 112, 859–877, https://doi.org/10.1080/01621459.2017.1285773, 2017.

Bortolotti, P., Tarres, H. C., Dykes, K. L., Merz, K., Sethuraman, L., Verelst, D., and Zahle, F.: IEA Wind TCP Task 37: Systems engineering in wind energy - WP2.1 Reference Wind Turbines, Tech. Rep. NREL/TP-5000-73492, National Renewable Energy Lab., Colorado, 2019.

Der Kiureghian, A. and Ditlevsen, O.: Aleatoric or epistemic? Does it matter?, Structural Safety, 31, 105–112, https://doi.org/10.1016/j.strusafe.2008.06.020, 2009.

Dimitrov, N.: Surrogate models for parameterized representation of wake-induced loads in wind farms, Wind Energy, 22, 1371–1389, https://doi.org/10.1002/we.2362, 2019.

Dimitrov, N. and Natarajan, A.: From SCADA to lifetime assessment and performance optimization: How to use models and machine learning to extract useful insights from limited data, Journal of Physics: Conference Series, 1222, https://doi.org/10.1088/1742-6596/1222/1/012032, 2019.

Dimitrov, N., Kelly, M. C., Vignaroli, A., and Berg, J.: From wind to loads: Wind turbine site-specific load estimation with surrogate models trained on high-fidelity load databases, Wind Energy Science, 3, 767–790, https://doi.org/10.5194/wes-3-767-2018, 2018.

Gasparis, G., Lio, W. H., and Meng, F.: Surrogate models for wind turbine electrical power and fatigue loads in wind farm, Energies, 13, 6360, https://doi.org/10.3390/en13236360, 2020.

Hall, M., Housner, S., Zalkind, D., Bortolotti, P., Ogden, D., and Barter, G.: An open-source frequency-domain model for floating wind turbine design optimization, Journal of Physics: Conference Series, 2265, 042 020, https://doi.org/10.1088/1742-6596/2265/4/042020, 2022.

IEC 61400-3-1: Wind energy generation systems - Part 3-1: Design requirements for fixed offshore wind turbines, Standard, IEC, Geneva, CH, 2019.

Jiang, P., Zhou, Q., and Shao, X.: Surrogate-model-based design and optimization, Springer Nature Singapore Pte Ltd, ISBN 9789811507304, https://doi.org/10.1007/978-981-15-0731-1_7, 2020.

Jonkman, B. J. and Buhl, M. L., J.: TurbSim User's Guide: Revised February 2007 for Version 1.21, https://doi.org/10.2172/903075, 2007.

Jonkman, J.: The New Modularization Framework for the FAST Wind Turbine CAE Tool, AIAA, https://doi.org/10.2514/6.2013-202, 2013.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, 2017.

Kingma, D. P. and Welling, M.: Auto-encoding variational bayes, 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings, pp. 1–14, 2014.

Kneib, T.: Beyond mean regression, Statistical Modelling, 13, 275–303, https://doi.org/10.1177/1471082X13494159, 2013.

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Lemmer, F., Yu, W., Cheng, P. W., Pegalajar-Jurado, A., Borg, M., Mikkelsen, R. F., and Bredmose, H.: The TripleSpar campaign: Validation of a reduced-order simulation model for floating wind turbines, https://doi.org/10.1115/OMAE2018-78119, 2018.

Lemmer, F., Yu, W., Luhmann, B., Schlipf, D., and Cheng, P. W.: Multibody modeling for concept-level floating offshore wind turbine design, Multibody System Dynamics, 49, 203–236, https://doi.org/10.1007/s11044-020-09729-x, 2020.

Li, X. and Zhang, W.: Probabilistic Fatigue Evaluation of Floating Wind Turbine using Combination of Surrogate Model and Copula Model, AIAA, https://doi.org/10.2514/6.2019-0247, 2019.

Li, X. and Zhang, W.: Long-term fatigue damage assessment for a floating offshore wind turbine under realistic environmental conditions, Renewable Energy, 159, 570–584, https://doi.org/10.1016/j.renene.2020.06.043, 2020.

Liew, J. and Larsen, G. C.: How does the quantity, resolution, and scaling of turbulence boxes affect aeroelastic simulation convergence?, Journal of Physics: Conference Series, 2265, 032 049, https://doi.org/10.1088/1742-6596/2265/3/032049, 2022.

Matsuishi, M. and Endo, T.: Fatigue of metals subjected to varying stress, Japan Society of Mechanical Engineers, Fukuoka, Japan, 68, 37–40, 1968.

Matthies, H. G.: Quantifying uncertainty: Modern computational representation of probability and application, in: Extreme Man-Made and Natural Hazards in Dynamics of Structures, pp. 105–135, Springer Netherlands, Dordrecht, ISBN 978-1-4020-5656-7, 2007.

Montavon, G., Orr, G. B., and Müller, K.-R., eds.: Neural networks: Tricks of the trade, Springer Berlin, Heidelberg, 2 edn., ISBN 978-3-642-35288-1, https://doi.org/10.1007/978-3-642-35289-8, 2012.

Müller, K. and Cheng, P. W.: Application of a Monte Carlo procedure for probabilistic fatigue design of floating offshore wind turbines, Wind Energy Science, 3, 149–162, https://doi.org/10.5194/wes-3-149-2018, 2018.

Murcia, J. P., Réthoré, P.-E., Dimitrov, N., Natarajan, A., Sørensen, J. D., Graf, P., and Kim, T.: Uncertainty propagation through an aeroelastic wind turbine model using polynomial surrogates, Renewable Energy, 119, 910–922, https://doi.org/10.1016/j.renene.2017.07.070, 2018.

Ng, A. Y.: Feature Selection, L1 vs. L2 Regularization, and Rotational Invariance, in: Proceedings of the Twenty-First International Conference on Machine Learning, ICML '04, p. 78, Association for Computing Machinery, New York, NY, USA, ISBN 1581138385, https://doi.org/10.1145/1015330.1015435, 2004.

NREL: OpenFAST/openfast: Main repository for the NREL-supported OpenFAST whole-turbine and FAST.Farm wind farm simulation codes., https://github.com/OpenFAST/openfast, accessed: 2022-09-29, 2022.

Peyré, G. and Cuturi, M.: Computational optimal transport: With applications to data Science, Foundations and Trends in Machine Learning, 11, 355–607, https://doi.org/10.1561/2200000073, 2019.

Ramdas, A., Garcia, N., and Cuturi, M.: On Wasserstein two sample testing and related families of nonparametric tests, https://doi.org/10.48550/arxiv.1509.02237, 2015.

Rasmussen, C. E. and Williams, C. K. I.: Gaussian processes for machine learning, Adaptive computation and machine learning, MIT Press, Cambridge, Mass., 2006.

Schröder, L., Dimitrov, N. K., Verelst, D. R., and Sørensen, J. A.: Wind turbine site-specific load estimation using artificial neural networks calibrated by means of high-fidelity load simulations, Journal of Physics: Conference Series, 1037, 062 027, https://doi.org/10.1088/1742-6596/1037/6/062027, 2018.

Shaler, K., Jasa, J., and Barter, G. E.: Efficient Loads Surrogates for Waked Turbines in an Array, Journal of Physics: Conference Series, 2265, https://doi.org/10.1088/1742-6596/2265/3/032095, 2022.

Singh, D.: Training and validation datasets for training probabilistic machine learning models on NREL's 10-MW reference wind turbine, https://doi.org/10.4121/21939995.v1, 2023.

https://doi.org/10.5194/wes-2024-20
Preprint. Discussion started: 6 March 2024

Singh, D., Dwight, R. P., Laugesen, K., Beaudet, L., and Viré, A.: Probabilistic surrogate modeling of offshore wind-turbine loads with chained Gaussian processes, Journal of Physics: Conference Series, 2265, 032 070, https://doi.org/10.1088/1742-6596/2265/3/032070, 2022.

Slot, R. M., Sørensen, J. D., Sudret, B., Svenningsen, L., and Thøgersen, M. L.: Surrogate model uncertainty in wind turbine reliability assessment, Renewable Energy, 151, 1150–1162, https://doi.org/10.1016/j.renene.2019.11.101, 2020.

Smilden, E., Horn, J.-T. H., Sørensen, A. J., and Amdahl, J.: Reduced order model for control applications in offshore wind turbines, IFAC-PapersOnLine, 49, 386–393, https://doi.org/10.1016/j.ifacol.2016.10.435, 2016.

Teixeira, R., O'Connor, A., Nogal, M., Krishnan, N., and Nichols, J.: Analysis of the design of experiments of offshore wind turbine fatigue reliability design with Kriging surfaces, Procedia Structural Integrity, 5, 951–958, https://doi.org/10.1016/j.prostr.2017.07.132, 2nd International Conference on Structural Integrity, ICSI 2017, 4-7 September 2017, Funchal, Madeira, Portugal, 2017.

Villani, C.: The Wasserstein distances, in: Optimal Transport: Old and New, pp. 93–111, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-540-71050-9, https://doi.org/10.1007/978-3-540-71050-9_6, 2009.

Yang, Y. and Perdikaris, P.: Conditional deep surrogate models for stochastic, high-dimensional, and multi-fidelity systems, Computational Mechanics, 64, 417–434, https://doi.org/10.1007/s00466-019-01718-y, 2019.

Yao, Y., Rosasco, L., and Caponnetto, A.: On Early Stopping in Gradient Descent Learning, Constructive Approximation, 26, 289–315, https://doi.org/10.1007/s00365-006-0663-2, 2007.

Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization, ACM Trans. Math. Softw., 23, 550–560, https://doi.org/10.1145/279232.279236, 1997.

Zhu, X. and Sudret, B.: Replication-based emulation of the response distribution of stochastic simulators using generalized lambda distributions, International Journal for Uncertainty Quantification, 10, 249–275, https://doi.org/10.1615/Int.J.UncertaintyQuantification.2020033029, 2020.

Zhu, X. and Sudret, B.: Emulation of Stochastic Simulators Using Generalized Lambda Models, SIAM/ASA Journal on Uncertainty Quantification, 9, 1345–1380, https://doi.org/10.1137/20M1337302, 2021.

Zwick, D. and Muskulus, M.: The simulation error caused by input loading variability in offshore wind turbine structural analysis, Wind Energy, 18, 1421–1432, https://doi.org/10.1002/we.1767, 2015.