Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Farm-wide fatigue loads estimation: a data-driven approach

Francisco de N Santos[1,*] | **Nymfa Noppe**[1] | **Wout Weijtjens**[1] | **Christof Devriendt**[1]

[1]OWI-Lab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

Correspondence
*F de N Santos, Email: francisco.de.nolasco.santos@vub.be

Abstract

Fatigue has become a major consideration factor in modern offshore wind farms as optimized design codes and a lack of lifetime reserve have made continuous fatigue life monitoring become an operational concern. In this contribution we discuss a data-driven methodology for farm-wide tower-transition piece fatigue load estimation. We specifically tackle the employment of this methodology in a real-world farm-wide setting and the implications of continuous monitoring. With reliable nacelle-installed accelerometer data at all locations, along with the customary ten-minute SCADA statistics and three strain gauge-instrumented *'fleet-leaders'* we discuss the value of two distinct approaches: use of fleet-leader or population-based data for training a physics-guided neural network model with a built-in conservative bias, with the latter taking precedence. In the context of continuous monitoring, we touch on the importance of data imputation, working under the assumption that if data is missing, then its fatigue loads should be modelled as under idling. With this knowledge at hand, we analyzed the errors of the trained model over a period of nine months, with monthly accumulated errors always kept below $\pm 5\%$. A particular focus was given to performance under high loads, where higher errors were found. The cause for this error was identified as being inherent to the use of ten-minute statistics, but mitigation strategies have been identified. Finally, the farm-wide results are presented on fatigue load estimation, which allowed to identify outliers, whose behaviour we correlated with the operational conditions. Finally, the continuous data-driven, population-based approach here presented can serve as a springboard for further lifetime-based decision-making.

KEYWORDS:
Offshore wind turbine fatigue; Farm-wide fatigue estimation; Population-based SHM

1 | INTRODUCTION

1.1 | Motivation

Despite the rapid growth seen in recent years with a current estimated 63,2 GW of installed capacity in 2022 [1], wind energy as an industry and a research field retains several considerable challenges. One key challenge faced by the different stakeholders relates to the fatigue lifetime of offshore wind turbines. Increasingly, a greater focus has been given to sustainable asset management: if wind energy is to realize its potential in supplying sustainable and cheap energy, one is required to know more precisely the current lifetime consumption of each turbine within a farm. This is because operation and maintenance (O&M) amounts to nearly a third of global costs [2], making operators keenly look for solutions that minimize these. An accurate account of the fatigue lifetime can allow a better scheduling of inspections and suppress unnecessary actions, further lowering costs and resource usage.

Furthermore, fatigue has remained a key design driver [3], with wind turbine structures being designed for dynamics rather than for bearing capacity [4]. This means that the substructure dimensions have been optimized during design for fatigue life, in order to match the intended lifetime of a project (typically 20-25 years) as closely as possible. In some older offshore wind farms, real-world observations through the use of structural health monitoring (SHM) appear to have identified an additional structural reserve, suggesting that fatigue life consumption was less than as per design [5]. Conversely, this structural reserve

seems to be absent in more recent projects which employ updated design procedures which allow a more accurate description of structural behaviour (*e.g.* PISA[6]). This has meant that operators of newer farms, instead of looking for a possible lifetime extension, need to keep tabs on the fatigue progression of their assets to ensure that the intended project life can be reached. Particularly, external factors which interfere with normal operating conditions and potentially affect lifetime consumption need to be tracked. *E.g.*, if a turbine is shut down/curtailed to meet market needs or requirements from the transmission system operator. Looking specifically to standstill, its high impact on lifetime occurs because, with turbines' foundations progressively increasing size, the natural frequencies are closer to the dominant wave frequencies[7]. This, combined with a larger surface area for hydrodynamic loading and deeper locations, has led to a greater impact of wave loads on fatigue[8], which is felt more heavily when there is no aerodynamic damping (*i.e.*, non-operational conditions as parked and idling). Furthermore, curtailment introduces additional transitional load cycles which may impact lifetime consumption[9]. The farm operator needs therefore to ensure that the complex interplay of requirements leading to downtime (*e.g.* compliance with transmission system operator or energy market requirements, component malfunction, etc.) does not jeopardize achieving the expected fatigue life. Thus, in absence of any reserve, continuous fatigue life monitoring has become an operational concern.

1.2 | Fatigue monitoring

The analysis of the fatigue life can broadly be approached from two distinct avenues[10]: (i) the cumulative fatigue damage is used to predict fatigue life, based on the assumption that failure occurs after a number of loading cycles for a particular tension/stress range; and (ii) examining fracture behaviour of mechanical elements under dynamic loads, considering that, once cracks have grown to a critical length, failure will occur. Due to the complexity of crack behaviour models, the first approach (i) is more commonly applied in practice.

Thus, a process that initially involved physical inspections (which can be dangerous, time consuming and costly)[11] has evolved to include an increasing use of remote, non-intrusive health monitoring systems[12]. This remote sensing has been principally undertook through strain gauge instrumentation at the substructures' interface level with the turbine[13,14,15], with this sensing being selected[16] due to its ability of capturing the history of long-term types of damage[17]. The interface between turbine and substructure is an interface in design for load exchanges. At least for a first assessment a comparison between the measured and as-designed loads offers insight in the residual life of the asset.

Based on substructure strain measurements, the fatigue damage might then be extrapolated into the future and open the doors for lifetime quantification[18,19,20]. However, to measure mechanical stress by instrumenting all turbines of a given wind farm with strain gauges and maintaining these is not economically viable. The authors have observed that, as a result, in real-world scenarios, fewer than 10% of turbines in a farm are equipped with a SHM system that incorporates strain gauges. Therefore, a number of researchers have suggested alternatives involving mainly the use of supervisory control and data acquisition (SCADA), available at every turbine[21,22]. The farm-wide availability of SCADA data allows models trained on it to be employed for farm-wide fatigue life estimation[23]. However, in the context of offshore wind where wave-induced loading plays a dominant role, a SCADA-only approach leads to poor results for all structural fatigue loads, as SCADA typically lacks the information regarding waves and/or structural dynamics[24]. This has meant that, in order to accurately capture the complex dynamics of offshore wind turbines, the coupling of acceleration data with SCADA has proven crucial for fatigue-predicting models[24,25,26]. [26] in particular, looked at comparing different sensor setups and their performance for fatigue estimation. Based on varying quality of SCADA and acceleration data acquisition systems, it showed how the inclusion of tower accelerations leads to a sizeable improvement in the ability to estimate fatigue rates. As both tower accelerations and fatigue cycles follow the same equations of motion[27], the higher performance attained by including acceleration data is expected. While high quality accelerometers still cost more than strain gauges, this is readily compensated through the much reduced installation costs. Moreover, the cost of the hardware has been dropping for some time, resulting in the technology gaining traction in recent years[28,29]. Moreover, due to accelerometers' reliability[25], costs and workplace risks are further reduced compared to strain gauges, as these do not require periodic *in-situ* maintenance.

The greater employment of non-intrusive SHM sensing and the resulting increased data flow has coincided with the advent of machine learning (ML) algorithms and other data-driven techniques which capitalize on big data. Some examples of the use of data-driven techniques on SHM wind turbine data include evidence demonstrating the dependability of data-driven fatigue estimators throughout the entire operational lifespan of the turbine[30], assessing the effectiveness of SCADA-based artificial neural network (ANN) models in estimating fatigue load for blade flap- and edgewise bending moments under varying flow conditions[31,32], exploring the applicability of ANNs in predicting blade root flapwise fatigue loads and their relationship to turbine failures[33], evaluate the influence so-called EOPs (Environmental and Operational Parameters) have on the features of the vibration response of the wind turbine blades by employing Gaussian process regression time-series modelling[34], applying conditional variational auto-encoders to infer the probability distribution of accumulated fatigue on the root cross-section of a simulated wind turbine blade, enabling the generation of long-term probabilistic degradation predictions using historical SCADA data[35,36] or of graph neural networks[37], estimating the tower fore-aft bending moments through onshore wind turbines' long-term SCADA[38], utilizing Gaussian processes for damage detection[39] and leveraging SCADA and acceleration data to forecast fatigue loads at the tower-transition piece on a ten-minute basis, providing insights into feature selection, evaluation of different sensor configurations, and tentative implementation across the entire wind farm[26].

The above-mentioned research has pushed the envelope and successfully added to the body of work that advocates for greater instrumentation of farms and better use of the available data for lifetime assessment. However, if this research is to effectively pass from a proof-of-concept phase to industry-wide acceptance, two fundamental issues have to be addressed: (i) the vast majority of the research still concerns itself with ten-minute level predictive capacity of fatigue loads/damage, however, fatigue life assessment is not performed based on a ten-minute basis. Rather, it is necessary to assess the total accumulation of the fatigue damages over the project life and quantifying residual life[40]. And (ii), due to data availability (or better said, the lack thereof), specifically accelerometer installation, no **farm-wide** employment of the methodologies developed for fatigue load estimation on the substructure on one or a couple of turbines has been presented for a real-world case involving a full farm.

(i) has been addressed in[5] (an extension of[26]), where there is a rescaling and accumulation of fatigue loads and damages undertook for two real-world turbines over two years. Thereafter, the necessity to accurately predict fatigue loads on a long-term level led to the introduction of a physics-guided machine learning approach (Φ-ML) in[41]. In Φ-ML the models are taught the governing physical laws which provide *'informative priors'* in the form of strong theoretical constraints and inductive biases on top of the observational ones[42] and can thus be considered hybrids or *'grey-box'* models, as they are data-driven models that include some underlying physical knowledge of the problem at hand, halfway in the spectrum between *'white-box models'* (physics-based) and *'black-box models'* (data-driven)[43].

As for (ii), this is the goal of the current contribution, *i.e.*, present the results for a long-term fatigue load estimation based on SCADA and dedicated, high-quality accelerometers on the scale of a farm, using real-world data. Within the context of a farm-wide employment, we attempt to address the several questions such a employment would pose in the setting of a *continuous monitoring* strategy. Firstly, we discuss at length the problematic nature of missing data and, keeping in line with the intended conservatism of this work, argue for the modelling of missing data points as idling cases, providing a guideline for a long-term employment. Secondly, we juxtapose two different strategies for data usage when considering a population of structures, namely a fleet-leader strategy and a population-based approach. Thirdly, we frame the discussion of the error analysis as being consistent with a continuous monitoring philosophy, and what this means for model re-training. Finally, we present the validation, cross-validation and extrapolation of the trained models, long-term error, farm-wide results and confirm the physicality of the results by focusing on the behaviour of an outlying turbine before drawing conclusions.

1.3 | Article outline and main contributions

With previous work having focused on arguing for the necessity of farm-wide reliable accelerometer data[26] and presenting long-term fatigue accumulation[41], the present contribution addresses the actual farm-wide implementation by presenting results for a real-world farm where all turbines are instrumented with dedicated accelerometers (along with the customary SCADA statistics). Specifically, we discuss at length two distinct approaches to data usage for model training: fleet-leader or population-based. Furthermore, we add a section on data imputation, its motivation and employment and frame the discussion sections within a continuous monitoring strategy.

In this contribution, section 1 introduces this work by framing in subsection 1.1 the motivation behind it and in subsection 1.2 providing a comprehensive literature review of data-driven methods for fatigue monitoring in offshore wind structures. In section 2 the methodology is discussed by firstly introducing the instrumentation setup and its use for fatigue estimation (subsection 2.1). Thereafter, the philosophies behind data use for model training (fleet-leader and population-bases; subsection 2.2), the model training (subsection 2.3) and data imputation (subsection 2.4) are discussed. section 3 discusses the results by dividing these into three main parts: subsection 3.1 compares the fleet-leader and population-based models, subsection 3.3 focuses on the model errors and delves deeper into high loading cases (subsubsection 3.3.1) and finally, subsection 3.4 present farm-wide results and analyses outliers subsubsection 3.4.1. At last, section 4 draws some conclusions on the present work and discusses future work.

2 | METHODOLOGY

2.1 | Instrumentation for fatigue estimation

In the current contribution, data from an ongoing measurement campaign on a real-world offshore wind farm located in the North Sea was utilized. The monitored turbines can be described as sitting on top of *XL* monopiles, so-called because, due to progresses in manufacturing technologies, these extra-large monopiles exceed 8 m diameter (capable of supporting 7-10 MW turbines)[44]. With up to 80% of offshore wind turbine substructures being monopiles and these getting larger and larger[45], we can take the farm under study as being fairly representative of the current offshore wind farm paradigm.

More specifically, SCADA data and accelerations from a nacelle-installed and high quality, tri-axial, microelectromechanical systems accelerometers (MEMS) were collected for all locations within the farm, Figure 1. It has been previously shown that the inclusion of accelerations is essential to capture all complex dynamics of the turbine[26] and, by doing so, the inclusion of wave data is rendered unnecessary.

Additionally, for three locations (WT11, WT14, WT18) – representative of the different seabed-depth clusters present in the farm –, axial strain gauges (6) installed along the inner circumference at the tower-transition piece interface level allowed to get the fore-aft (FA) and side-to-side (SS) bending moments (M_{tn} and M_{tl}, respectively). All data was collected from third-parties, specifically the operator (SCADA) and a specialized monitoring company (monitoring data). Throughout

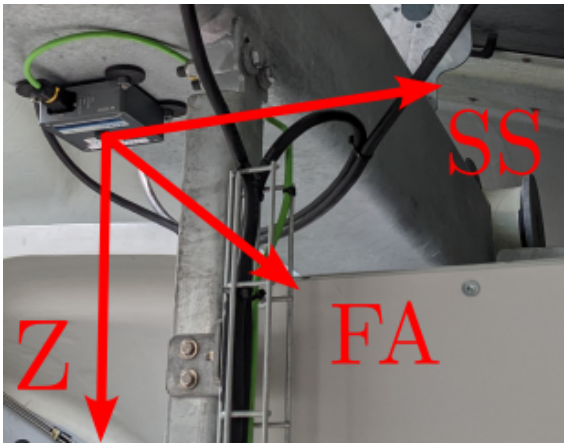


Figure 1 Nacelle-installed tri-axial MEMS accelerometer with measurement directions.

the campaign, extensive calibration and quality checks have been performed. Some periods had no collected data due to power outage; these will be addressed in a ensuing section. For each of the sensors, the signals will have a different sampling frequency. This has meant that all have been postprocessed into the lowest common denominator: ten-minute time instances (a common practice in industry).[46] addressed the shortcomings that a mean ten-minute SCADA approach might entail. However, the inclusion in this contribution of additional statistics such as minima, maxima and standard deviation serves to rectify some of the shortcomings by increasing the valuable information stored from SCADA signals. The full dataset description can be found in Table 1.

	Sensor	Sampling frequency	Variable	Units
Input	SCADA	1 Hz	Rotational speed Yaw angle Pitch angle Power Wind speed Wind direction	rpm deg deg kW m.s^{-1} deg
	Target statistics (10-min): mean, minimum, maximum, standard deviation.			
	Accelerometer	12.5 Hz	FA acceleration SS acceleration Z acceleration	g g g
	Target statistics (10-min): mean, minimum, maximum, root mean square.			
Target	Strain gauges	30 Hz	Normal bending moment (M_{tn}) Lateral bending moment (M_{tl})	MNm MNm
	Target statistics (10-min): damage equivalent loads (**DEL$_{tn}$** and **DEL$_{tl}$**).			

Table 1 Description of datasets from the measurement campaign. Note the conflicting sampling frequencies. Each data-type is processed into 10-minute target statistics.

In subsection 1.2, it was mentioned how strain gauge installation across the entirety of a farm was cost-prohibitive. However, any data-based model utilizing SCADA and accelerations as input will still need accurate reference points (obtained exclusively through strain gauge instrumentation) at the training locations in order to extrapolate across the entire farm. Through the ten-minute average yaw angle from the SCADA the strain gauges enable the calculation of the bending moments M_{tn} and M_{tl}. Then, by employing a rainflow counting algorithm[47,48], holding the linear damage accumulation hypothesis as true (Palmgren-Miner's rule)[49] and through the employment of the Wöhler exponent (the negative inverse slope of the SN curve)[50], a damage-sensitive feature, the damage equivalent loads (DEL), can be calculated for any ten minute window[51,52]. DELs are widely adopted among load engineers to quantify fatigue rates (over *e.g.* damage) and are furthermore used in design documents, which opens the door to comparisons. Therefore, given the Wöhler exponent, m, the number of cycles, n_j, of a given stress range, $\Delta\sigma_j$ and the tower-transition piece outer and inner radii, r_o and r_i, respectively at strain gauge locations we can calculate the DEL, as given by Equation 1[53]. For this contribution, and following the design documentation, the value of 5 was used for m and $N_{eq} = 10^7$, a predefined number of cycles.

$$DEL = \frac{1}{N_{eq}} \cdot \left(\sum_j n_j \cdot \left(\frac{\Delta\sigma_j \cdot \frac{\pi}{4} \cdot (r_o^4 - r_i^4)}{r_i} \right)^m \right)^{1/m} \quad (1)$$

The DELs are calculated for both the FA (DEL_{tn}) and SS (DEL_{tl}) directions. However, as discussed in[41], accuracy on DEL estimation at a ten-minute level is not a sufficient condition to determine the success of the model: one must also be able to accumulate DELs on a longer time-frame.

Thus, by Palmgren-Miner's rule, we can further combine n equivalent load ranges that have been derived for the same reference cycle number and Wöhler exponent through the m-root of the weighted summation of the m-power DEL instance[54], as seen in Equation 2. Here DEL_{LT} represents the long-term DEL re-scaling of j ten-minute DEL instances into a ten-minute DEL representative of the long-term period in question. The performance over this metric will be a particular concern of ours, although not the sole contributor for our evaluation of model performance. In this equation every j ten-minute time-instance DEL represents a damage load with identical occurrence probability of $1/n$[55].

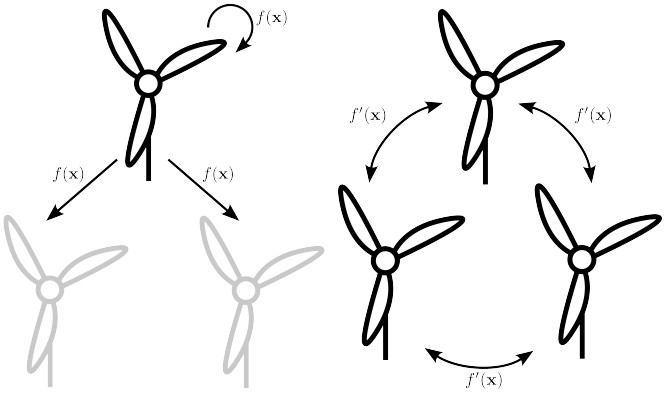
$$DEL_{LT} = \left(\frac{1}{n} \sum_{j=1}^{n} DEL_j^m \right)^{1/m} \quad (2)$$

In this contribution the accurate prediction of accumulated long-term FA and SS DELs is targeted. These two directions represent the two primary loading directions, respectively in the wind direction and crosswind. However, these DELs should not be understood as the accumulated fatigue damage on a physical location on the asset. As the turbine yaws to follow the wind directions FA and SS loads get distributed along the entire circumference. In order to get the full cross-section fatigue profile, further decomposition of the FA and SS components into specific heading is required[10]. While outside of the current scope, part

of future work is to use the found FA and SS DELs and quantify localized fatigue loads given the seen environmental conditions.

2.2 | Fleet-leaders & population-based

The availability of strain measurements (and therefore, DELs) at three locations opens the door at two different methodological philosophies to train the models: fleet-leader models and population-based models (see Figure 2).



(a) Fleet-leader. **(b)** Population-based.

Figure 2 Different training philosophies' diagrams. (a), fleet-leader concept, wherein a neural network model (latent function f for inputs **x**) is trained for a single turbine (fleet-leader) and applied to the remainder turbines. (b), population-based concept, where the latent function f' is trained based on data from all instrumented turbines (population).

In the first, a single instrumented wind turbine is used for the training of a model, cross-validated on other fleet-leaders and which then extrapolates farm-wide[23]. This hypothesis is not that far-fetched, as the population of structures within a wind farm is homogeneous (nominally-identical machines) and the inclusion of accelerations in the training dataset is believed (and empirically confirmed[26]) to sufficiently cover the variability within the farm for most operational cases. As for the second philosophy, it hinges on data availability on a population level. Here, we follow the grammar and discussion of[56], where models of a given feature space attempt to capture the *form* of the object of interest, *i.e.*, the *'essential'* nature of the object (the baseline behaviour, shared across the population) and the variations (deviations from this ideal essential nature) found across any real-world population, in a latent space shared by the population. In the specific case of this contribution, the implementation of a fleet-leader philosophy would lead to a latent function, f, trained on a strain gauge-instrumented turbine. As for a population-based approach, it is only possible to have DEL measurements for three turbines, thus not a full population. However, one can argue that, as each of the instrumented turbines represents a design cluster (dependent on the seabed depth, which ought to be the only variation of note between turbines), a latent function, f', of this sample of the population would be sufficient to accurately represent the form and variation of the farm through a latent space shared by all turbines, enforcing the model to recognize the impact of different structural dynamics in the overall DEL.

2.3 | Model training

Given the instrumentation setup described in subsection 2.1 and subsection 2.2, we can then describe the procedures followed for training the final DEL-predicting model. Firstly, both strategies of subsection 2.2 (population-based and fleet-leader) are opposed to one another; that is to say, the data used for training will either come from each the fleet-leaders (individually) or from pooling one third of the available data of each fleet-leader (population-based approach). So, for both approaches, the same amount of data is pooled and compared for model training. The data available for training comprises of three full months (December 2022 – February 2023).

Prior to model training, the neural network model architecture must be defined. This is achieved two-fold: (i), by reducing the number of input features through a recursive feature elimination algorithm with a random forest regressor and built-in cross-validated selection of the best number of features (RFECV[57]; lower threshold of 15). By removing redundant input variables, computational-, memory- and time costs are reduced[58]. Moreover, it has been shown that this may improve the overall performance of neural network models, as non-informative variables can add uncertainty to the predictions and reduce the overall effectiveness of the model[59]. We can then consider the application of RFECV as a good practice. And (ii), the network architecture is achieved by performing Bayesian hyper-parameter optimization through the employment of Gaussian processes[60] with a Matérn kernel[61] using keras-tuner[62]. Thus, for a differentiable loss function $\mathcal{L}(x)$, we have $x^* = \arg\min_{x \in \mathcal{X}} \mathcal{L}(x)$, with $x*$, the set of hyper-parameters that yields the lowest value for $x \in \mathcal{X}$, trainable hyper-parameters. In this case, the hyper-parameters $x \in \mathcal{X}$ were $h \in \mathcal{H} = \{1, \ldots, 5\}$ hidden layers, $n \in \mathcal{N} = \{32, 64, 96, \ldots, 512\}$ neurons, $a \in \mathcal{A} = \{\text{ReLU}, \text{GELU}, \text{SELU}\}$ activation function types[63], $d \in \mathcal{D} = \{0, 0.1, 0.2, 0.3\}$ dropout rate and $o \in \mathcal{O} = \{1 \cdot e^{-2}, 1 \cdot e^{-3}, 1 \cdot e^{-4}\}$ learning rate of the optimizer (Adam[64]). The monitored loss function was the Minkowski logarithmic error (MLE) introduced in[65]. This function attempts to guide the neural network learning by including physical knowledge specific to the problem at hand, in a so-called physics-guided machine learning approach ($\mathbf{\Phi}-$ML). It is described by Equation 3, with $m = 5$, and $\mathbf{Y}$, $\hat{\mathbf{Y}}$, the measured and predicted vectors. It is based on the L^p norm and the logarithm function, and attempts to prioritise long-term DEL performance (DEL_{LT}) and conservatism (fatigue load over-prediction), whilst maintaining ten-minute level prediction accuracy.

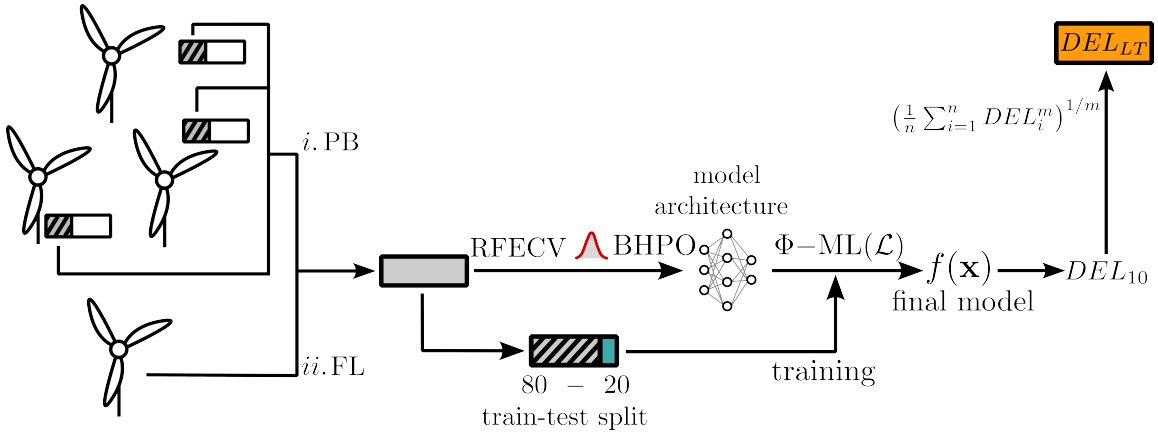


Figure 3 Overview of methodology employed, using either $i.$ a population-based approach (PB), where we randomly pick a third of each turbine's data, or $ii.$ a fleet-leader approach (FL). The model architecture is attained by performing dimensionality reduction through a recursive feature elimination algorithm with cross-validation (RFECV) and through Bayesian hyper-parameter optimization (BHPO). The final model (latent function, f) employs physics-guided machine learning (Φ-ML) through its loss function ($\mathcal{L}$, the Minkowski logarithmic error) and estimates ten-minute DELs (DEL_{10}), further re-scaled long-term (DEL_{LT}).

$$\mathcal{L}(\mathbf{Y}, \hat{\mathbf{Y}}) = \left(\sum_{i=0}^{n} |\log(y_i + 1) - \log(\hat{y}_i + 1)|^m \right)^{1/m} \quad (3)$$

Thus, given the MLE loss function, the neural network architecture is trained by employing a 80-20% train-test split on the available data. The final model is then capable of predicting DELs on a ten-minute basis given SCADA and MEMS accelerations. These DELs can later, with Equation 2, be re-scaled into a DEL representative of any given timeframe. The full methodology is summarized in Figure 3.

2.4 | Data imputation

The methodology ascribed in subsection 3.1 produces a model, $f(\boldsymbol{x})$, capable of predicting tower-transition piece interface DELs given the availability of SCADA and acceleration data. However, if this data is not available, the model is not capable of estimating any value. The alternative would be either to ignore missing data and reduce their analysis to intervals of complete time-series, under the assumption these are representative [66], or to linearly interpolate the mean DELs from the available data.

Meanwhile, the most likely cause of the lack of SCADA and/or acceleration data is a non-operational turbine.. As discussed in the introduction, for the 7+MW generation of turbines [67], non-operating conditions are often result in greater fatigue loading due to the waves exciting unimpeded the structure (lack of aerodynamic damping). Therefore, if these missing timestamps are not included, or imputed from operational data, significant contributors to fatigue damage will be ignored in the global analysis of the structural health and subsequent underestimation of consumed fatigue life.

As explained with the use of the MLE loss function, the present work attempts to retain a certain degree of conservatism (a *'safety-factor'*). Figure 4.

Hence, we devise a missing data imputation strategy that works under the conservative assumption that when data is not available, then the turbine is experiencing loads as under a idling period of operation. Our data imputation approach hinges on the use of wind, wave and tidal data from an external source. In this case the public Flemish maritime database (Meetnet Vlaamse Banken [68]), with a measuring station (measuring pile and buoy) in the North Sea at the Westhinder site. Its original data sampling frequency is of 30 minutes, but resampled using linear interpolation to the target ten minutes. The data set description is presented in Table 2.

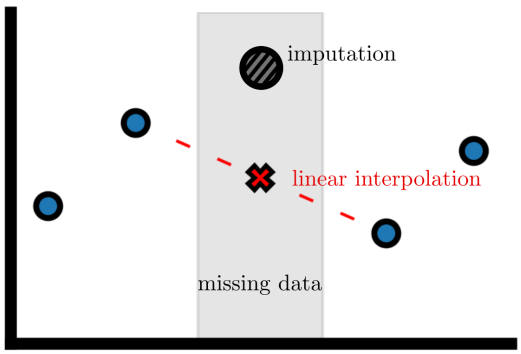


Figure 4 Diverging missing data handling strategies: linear interpolation between known points or data imputation (based on the assumption of idling operation).

One could also have relied on wind-farm data itself, so long as not coming from the turbine with missing data, however the strategy adopted

allows for greater flexibility, as weather data is collected very reliably.

The public database data is used to train an *'imputation'* model, $f^I(\mathbf{x})$,

Sampling frequency	Variable	Units
5.6×10^{-4} Hz	Wave height	cm
	10% highest waves	%
	Average wave period	s
	Height waves with period > 10 s	cm
	High frequent wave direction	deg
	Low frequent wave direction	deg
	Sea water temperature	°C
	Max 3s wind gust (at 10m height)	$m.s^{-1}$
	Average wind direction	deg
	Average wind speed	$m.s^{-1}$
	Tide TAW	cm
	Air pressure	hPa
	Air temperature	°C
Target statistics (10-min): damage equivalent loads ($\mathbf{DEL_{tn}}$ and $\mathbf{DEL_{tl}}$) for idling turbines.		

Table 2 Measuring pile and buoy data set description for the imputation model. Sampling frequency of 30 minutes.

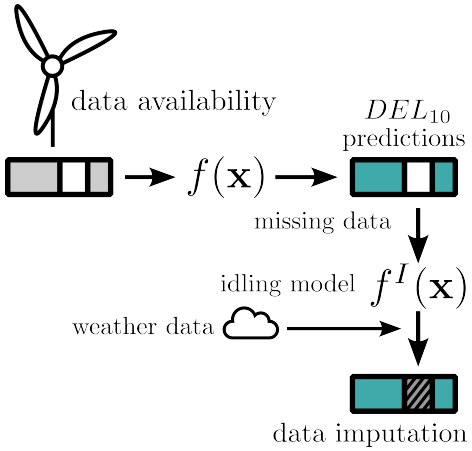


Figure 5 Data imputation methodology: depending on the data availability, ten-minute DEL predictions are performed. Missing data points are estimated by an imputation model $f^I(\mathbf{x})$, trained to predict idling DELs using XGBoost and based on weather data (see Table 2), taken from a public database.

which is trained on a dataset of idling loads during the training period (we define idling as RPM < 2 & pitch > 77 or Power < 0;see also Table 4), pooling data from all three turbines to increase the dataset size. During idling, loads are predominantly wave- and tidal-driven, and therefore, the use of the public database can provide an accurate and flexible assessment of weather conditions. The model trained is an extreme gradient tree boosting algorithm (XGBoost[69]). For our main DEL model, the use of neural networks enhanced by physical-guided learning is a powerful tool, as the largeness and complexity of the dataset is best captured by the non-linear smoothing that the neural network activation functions allow. However, for a smaller dataset consisting of a single, well-defined operational case, the use of the fast-trained XGBoost algorithm is more appropriate, as it has been seen that for easy to moderate tasks with low amounts of data it may over-perform regular neural networks[70]. We can visualize the data imputation methodology in Figure 5.

3 | RESULTS AND DISCUSSION

3.1 | Model comparison and selection

Following the discussion in subsection 2.2 and the methodology introduced in subsection 2.3, four different models were compared: three models using the individual fleet-leader wind turbines data (WT11, WT14, WT18) and a population-based model combining data from the fleet-leaders. Five runs were used to train each model, with the only change being the random seed value used for a 80-20 train-test split to ensure variability. As discussed above, the focal point for model performance evaluation is the error on the long-term ten-minute DELs, based on Equation 2. In Figure 6a we present the long-term fore-aft DEL error ($\delta_{LT} = 100 \times ((\mathbf{DEL_{real}} - \mathbf{DEL_{pred}})/\mathbf{DEL_{real}})$) for each of the instrumented turbines (and the average error) based on the training dataset (fleet-leader/population). The data used for computing DEL_{LT} comprised of the full three training months for each turbine.

In this figure, notable differences can be observed among the models trained on different datasets. Firstly, when examining the fleet-leader models (WT11, WT14, WT18), it is evident that some outperform others. For instance, models trained on WT11 exhibit errors centered around zero with a spread of approximately $\pm 10\%$, while WT18 shows a considerably higher average error for all turbines, centered at $+10 \pm 5\%$. WT14 falls somewhere in between these two extremes. These results suggest that WT11 experiences greater variability in loading during its operation, enabling it to better account for the cases encountered at other locations. Additionally, the models trained using a population-based approach, which combines a third of the data from each turbine, demonstrate the best performance. They exhibit errors centered around zero, with a smaller spread between different runs compared to WT11, as well as the lowest average δ_{LT}. This outcome is expected since the population-based approach incorporates data from all three instrumented turbines, providing a broader coverage of the operational conditions faced by the turbines. However, it is important to note that relying solely on the average δ_{LT} may not provide a comprehensive assessment of performance. It was also seen that the ten-minute level performance, in the form of the coefficient of determination (R^2), was also seen to be stronger for the population-based models. These had the best R^2 (above 0.9; rather positive), followed by WT11. This seems to indicate that a population-based approach enables a more accurate and consistent performance both on long-term and ten-minute level. As for Figure 6a, we plot the side-to-side δ_{LT} as for the fore-aft.

Once again, we can reiterate the main findings in general terms. WT11 consistently produces superior models compared to the other fleet-leader turbines. However, the distinction between WT11 and the

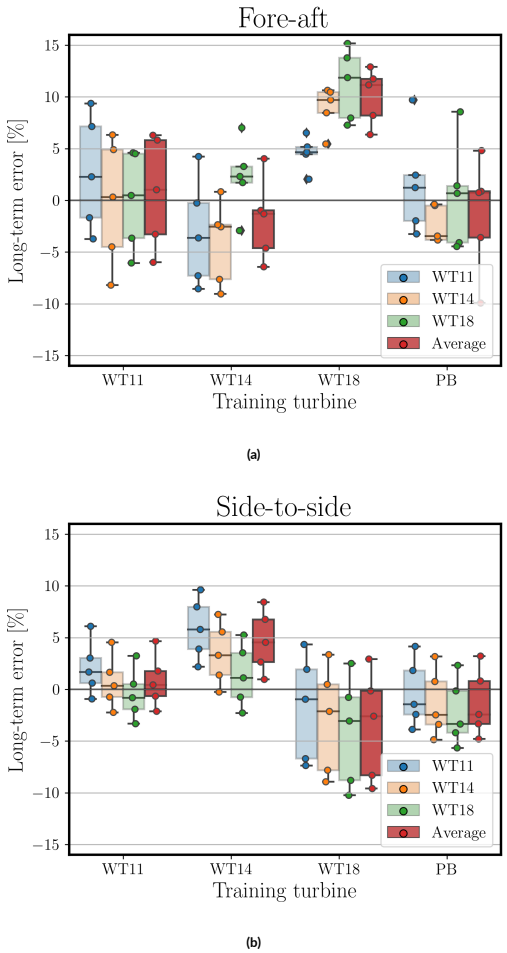


Figure 6 Box plot of long-term DEL relative error (δ_{LT}, [%]) over three months based on the training turbine models (five runs) for (a) fore-aft and (b) side-to-side.

population-based models is not as definitive in this case. Both types of models demonstrate relatively good performance, although WT11 exhibits a slight positive error bias (i.e., tending to under-predict), while the population-based (PB) models show the opposite trend (which was intended by our Φ-ML guided learning strategy). Both types of models have a spread of approximately $\pm 5\%$. There are two possible explanations for the similarity between these models. Firstly, the dynamics encountered by the instrumented turbines are already quite similar as the considered farm is quite homogeneous. Secondly, the population-based approach, which samples one third of the data from each turbine, may result in the loss of some relevant information. However, in general, it can be concluded that the population-based approach better captures the overall characteristics of the farm in terms of fore-aft and side-to-side variations. Nevertheless, it is important to note that there is a clear trade-off with the population-based strategy, as it does not allow for cross-validation in the strictest sense with a previously unseen turbine. Although it is not possible to cross-validate the model for an unseen turbine, PB models can be tested using the remaining two-thirds of unseen data from the three instrumented turbines. In this study, the decision is made to proceed with the best-performing PB model, as it is believed to provide better generalization across the entire farm.

3.2 | Imputation model

In subsection 2.4 a data imputation strategy was discussed at length — its reason of being and necessity, as well as its implementation. As previously mentioned, we assume that the missing data points are best modelled conservatively as under idling. Therefore, we must first train and validate a model capable of predicting idling DEL in the current section. Its employment for missing data points (which, naturally, can not be further validated) will be presented in subsection 3.4. In our data set we use almost 3000 data points labelled as idling from the three fleet-leaders — due to the rarity of such cases, we will not make a by-turbine distinction of the loads both during training and validation of the model thus merging the data of all fleet-leaders into a PB *'idling'* data set. We perform a 80—20 train—test split on the idling data set and train an XGBoost with 1000 estimators and a maximum depth of 7. We can see a timeseries of WT11 of the imputation model in Figure 7, where the predicted fatigue loads are juxtaposed to the actual values for idling cases over a background of all loads.

As we can see in this figure, the predicted loads closely follow the true

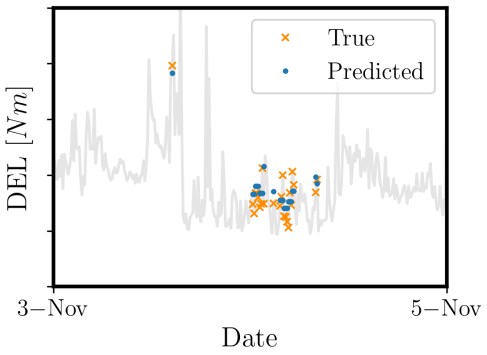


Figure 7 Timeseries of imputation model for WT11 — real idling loads (orange cross) and imputed load (blue) on a backdrop of all fore-aft fatigue loads (grey).

values. This can be furthermore seen in the actual error over the entire idling dataset by observing the FA and SS errors in the box plots in Figure 8.

In this figure we can see how the errors are distributed around zero $\pm 30\%$. In Table 3 additional error metrics are provide.

Here we can also see that the mean error $\bar{\delta}$ and long-term error δ_{Lt} are close to zero. If we take a look at $\bar{\delta}$, we can see that it is negative (overall over-prediction), which can also be observed in the slight negative skew

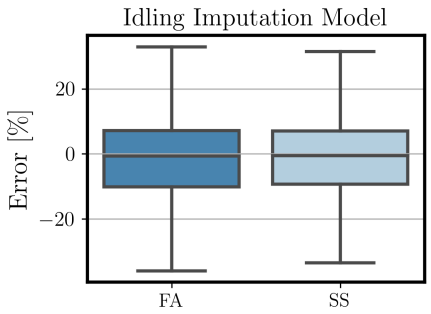


Figure 8 Boxplots of idling imputation model errors for FA and SS.

Direction	FA	SS
mean error, $\bar{\delta}$ [%]	-2.35%	-1.97%
long term error, δ_{LT} [%]	1.47%	1.77%
R^2	0.94	0.92

Table 3 Error Metrics of the imputation model for FA and SS.

of the box plots. The overall errors, specially δ_{Lt}, are very positive, with $R^2 > 0.9$ further confirming the soundness of this model. Similarly to FA, the SS long term error is kept below 2%.

3.3 | Error analysis

Having selected the best-performing PB model, we can proceed by analysing its long-term performance. To this end, we combine all DELs of a given month by re-scaling them with Equation 2, giving a ten-minute DEL representative of the month. Then, we multiply this representative DEL by the number of time instances on the month, and we are left with the DEL accumulated for that month (here, as we are comparing errors, we abstain from including imputed timestamps for which there is no measured value). This procedure is extended to subsequent months by again re-scaling DELs to represent the period in question (two months, three, . . .) and multiplying it by the number of ten-minute instances. In our analysis we use 9 months of data (October 2022—June 2023) of a turbine instrumented with strain gauges (WT11). We can observe these results in Figure 9.

In this figure we can observe the progression of the accumulated damage equivalent loads, slowly building up from October through June (including). We can observe that, for both FA (Figure 9a) and SS (Figure 9b), there is a steady increase on the fatigue loads registered at the structure, with the predicted accumulated DEL closely following the true values both at hindcasting (before the training period) and forecasting (after the training period). We can also observe how for FA the long-term error (in green) is kept below $\pm 5\%$, hovering around

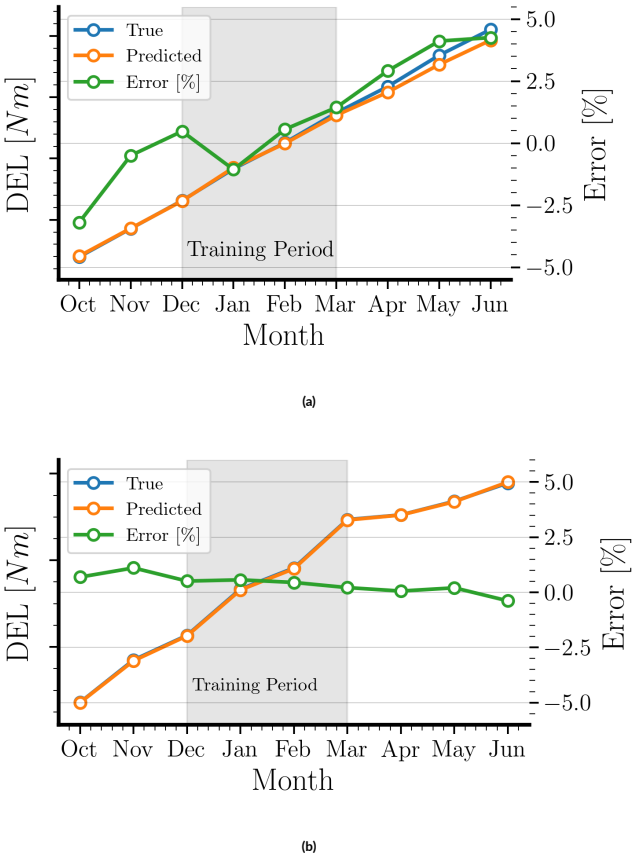


Figure 9 Monthly accumulation of true (blue) and predicted (orange) DEL from October—June for (a) fore-aft and (b) side-to-side on WT11. The training period in highlighted in grey. Error on monthly accumulated DEL in green.

0. However, this error appears to be progressing steadily with each month. Literature suggests that a minimum of 9—12 months is required for models to accurately capture the seasonality of fatigue loads[51]. Here, we hypothesize that, because the model has been trained during winter, is error grows over the summer months. In SS the error hovers around 0% very closely. Both results are similarly confirmed for the remaining strain gauge instrumented turbines (WT14 and WT18).

In this work we focused, given the hitherto collected data, on showing the model in hind- and forecasting scenarios, limiting the training period to just three months. In an actual real-world implementation, this time-window has naturally to be extended to at least 9–12 months.

The current approach should also be understood in the context of *continuous monitoring*: with data being continuously collected, it is essential to continuously monitor the progression of the long-term errors at the fleet leaders. In case of a growing error (e.g. beyond 5%) one can decide to retrain the model at any stage of the project.

In some cases model re-training might be even required due to external reasons, e.g. changes in the controller strategy, previously unseen

environmental conditions or changes in the turbine e.g. a rotor replacement or scour development. These will invariably change the underlying dynamics affecting fatigue damage in the substructure and render previously trained models unsuitable. The first two reasons for dynamics change have the possibility of becoming rather prevalent as, (i) changes in controller action programming have become more common, as for (ii), with environmental conditions changing progressively (*e.g.* drawn-out variations on wind and wave speed and direction), the structure's physics may radically change. Keeping track of the validity of the models as time progresses is thus an intrinsic part of the proposed methodology.

We can perform a further test on the model (WT11; FA) by analysing the performance of the predictive model *vs.* wind speed. For this, we plot the real and predicted representative re-scaled DELs for each wind speed bin (with a step of 2 $m.s^{-1}$), along with the absolute difference between both (over-predictions in green, under-predictions in red).

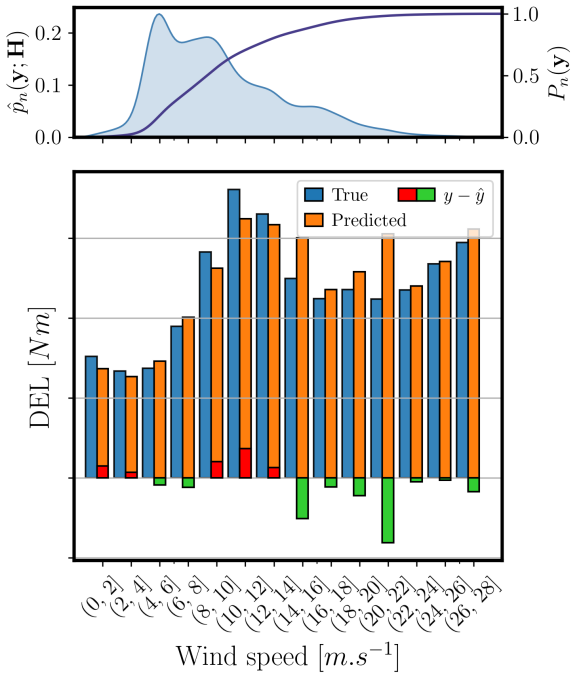


Figure 10 Above, probability density function or PDF ($\hat{p}_n(\boldsymbol{y};\boldsymbol{H})$; blue distribution) of wind speed and cumulative density function or CDF ($P_n(y) = \int_{-\infty}^{x} \hat{p}_n(\boldsymbol{y};\boldsymbol{H})dy$, purple curve). Below, real (y) and predicted ($\hat{y}$) representative re-scaled DELs for each wind speed bin (with a step of 2 $m.s^{-1}$), along with the absolute difference between both ($y - \hat{y}$; over-predictions in green, under-predictions in red).

In Figure 10 we can see a familiar curve: the steady increase of fatigue loads with the wind speed until the rated power (wind-driven fatigue), after which there is a sudden drop and subsequent slow increase in fatigue loads. As it can be seen, for the most frequent wind speeds the loads (and consequently, the absolute error) are rather small. We can also see that, for the majority of wind speeds (see the PDF), the model is over-predicting (green differential). However, for some wind speeds, namely between 8 and 14 $m.s^{-1}$, the model's under-predicting (positive error, red differential).

3.3.1 | High loads

The findings of the previous section pose us the question: why is it that the model is under-predicting (also globally, as seen by the positive accumulated error on Figure 9a) when we explicitly included the logarithm in Equation 3 to force over-predictions? In order to discern the reason, as well as see the impact small, frequent loads and big, rarer loads have on the global DEL, we can take a look at the DEL accumulated based on the magnitude of the loads (*i.e.*, for every successive DEL quantile, $\hat{q}_i$, we progressively accumulate through Equation 2, such that $DEL = \int \hat{q}_i \, d\hat{q}$) in Figure 11.

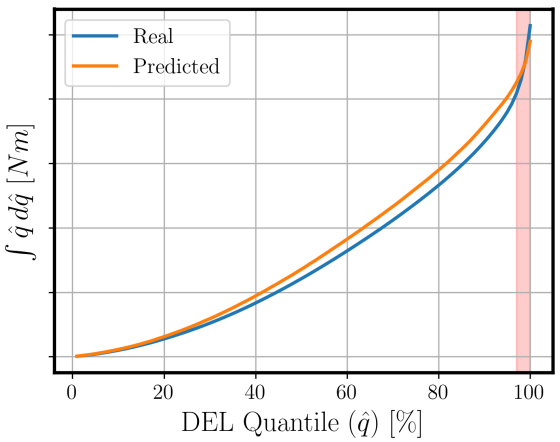


Figure 11 Accumulated FA DEL (WT11) per fatigue load magnitude quantile $\hat{q}$, such that $DEL = \int \hat{q}_i \, d\hat{q}$). Full data set is used. *N.b.* the inversion on the curves for the highest loads, identified in red.

Here we can see how the predicted curve (orange) is generally above the real DEL curve (blue), which means the model is over-predicting as intended. *E.g.* for the 80% smallest fatigue loads (or, excluding the 20 % highest loads) the predicted re-scalled DEL is greater than the real. However, we can observe that for the highest loads this trend is inverted, and the model under-predicts (highest 3%). The observed errors on a ten-minute level on the top 1% DELs for all strain gauge instrumented turbines produced a severe under-prediction of around +25%. This mean that, because of severely under-predicting the highest loads, has a big impact on long-term DELs given the effect of the Wöhler exponent in Equation 2. These higher loads almost invariably are linked with so-called events (*i.e.* rotor-stops). Although the model is — correctly —

inferring higher damages for events, these cases remain very difficult to accurately capture by any model because, (i), they are infrequent, meaning that there is not a lot of data points available for training; and (ii), DELs are estimated on a ten-minute basis, but *e.g.* a rotor-stop is a single instance in that ten-minute window. This means that the position of the event (if the rotor-stop happens early or late on the ten-minutes) becomes relevant as having a rotor-stop followed by idling (which also introduces higher fatigue loads) in a ten-minute window is not the same thing as have normal operating conditions followed by an event in a single ten-minute window. Because we are dealing with ten-minute data, it is very difficult (if not impossible) to retrieve the position of the event within the time window. The alternative would consist in full timeseries reconstruction but this has been seen to produce higher errors further along the line[71]. One possible way to circumvent this issue with ten-minute signals would be to increase the statistical information of the signal, by including metrics sensitive to peak position, such as skewness and kurtosis (spectral moments). However, it must be reiterated that the global accumulated DEL errors are well within $\pm 5\%$.

3.4 | Farm-wide

As stated throughout this contribution, the goal of this work does not lie with accumulating long-term DELs for a given instrumented turbine, rather to assess fatigue loads across the entire offshore wind farm. As such, after having thoroughly investigated the physicality, feasibility and expected error of the model in subsection 3.3, we concluded that it compared favourably to current industry standards and that its farm-wide application is sensible. Therefore, we can plot in Figure 12 the FA DEL (ten-minutes) representative for each month for which data is available farm-wide (December 2022 — June 2023). All results in this section include the use of imputed data.

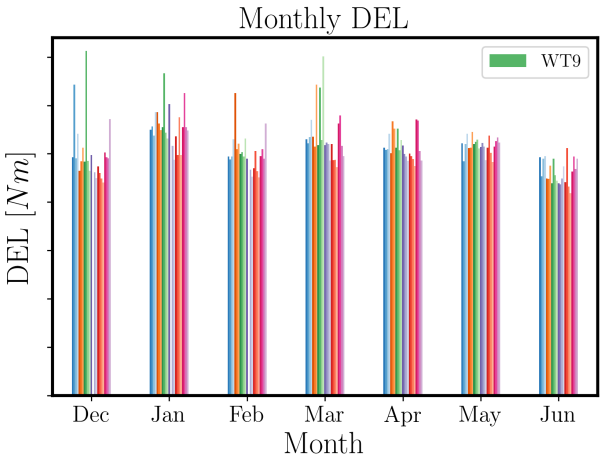


Figure 12 Farm-wide monthly representative FA DEL for December 2022 — June 2023 period.

Figure 12 allows us to effectively compare the fatigue loads experienced by each turbine within the farm over a month. This is precisely the sort of outcome which can enable operators taking informed decisions regarding the structural life management of their assets. Here, we can easily identify outliers (turbines with above average damage) and take the necessary actions: (i) figure out why a specific turbine is experiencing higher damage by delving into operational conditions and SCADA data and (ii) act accordingly and intervene if necessary. This figure underlines how this methodology can successfully be used for identifying *'problematic'* turbines, and react in the shortest amount of time, if deemed necessary. For example, we can see in the months of December and January how WT9 has a higher-than-average DEL.

Naturally, the month-representative DELs can be progressively updated with each new value to a long-term DEL, representative of the entire period. This value can then be multiplied by the number of ten-minute time-instances of said period, providing an accumulated long-term DEL (see Figure 9). However, one does not need to restrict oneself to a monthly accumulation. In Figure 13 we present the daily DEL accumulation for all turbines of the farm.

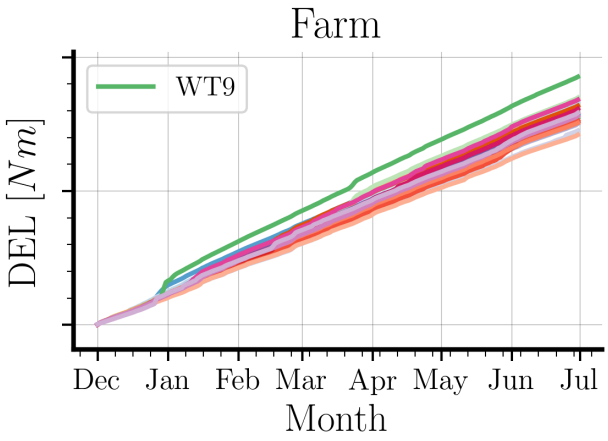


Figure 13 Daily FA DEL accumulation (December 2022 — July 2023) for all turbines. WT9 labeled in green.

This figure can be seen as the culmination of this contribution, as it shows the accumulated damage equivalent loads on a daily basis for all turbines within a real-world farm based on SCADA and reliable acceleration measurements. Interestingly, we can still identify an outlier: WT9. It is visible how in the end of December / beginning of January the accumulated fatigue damage rapidly increases above the rest of the turbines (as seen in Figure 12). This effect is still felt several months after, with WT9 being clearly the turbine with the most accumulated damage at the period in question. We can delve deeper, and identify the causes behind this behaviour.

3.4.1 | Outlier analysis

In order to identify why WT9 sees a greater fatigue damage in December and January we can take a look at how each operational condition influences the monthly representative DEL, as seen in Figure 14.

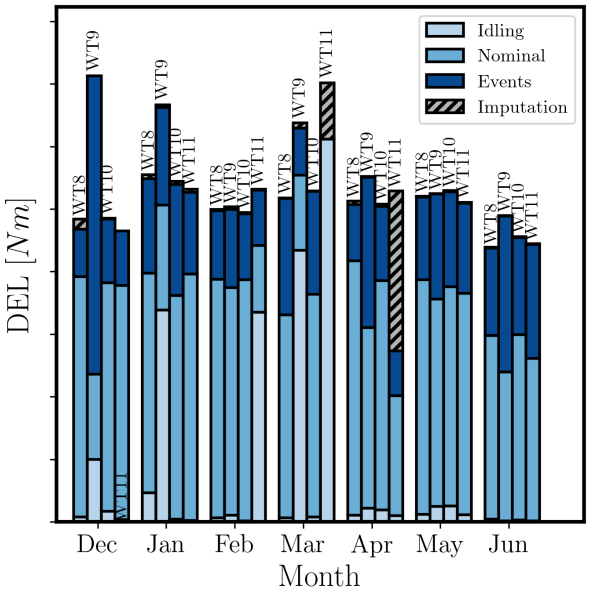


Figure 14 Impact of operational case on monthly DEL. In different shades of blue: idling (including shut-down), nominal and events (also includes other unlabelled data). In grey, imputed DELs (*e.g.*, *same as idling*).

In this figure, we can see the contribution each operational condition (along with the imputed data, that is to say, estimated as idling fatigue loads; grey) has on the monthly DEL. The operational conditions present are idling (light blue; includes shut-downs), nominal operating condition (middle blue) and events (dark blue). The respective case definitions can be consulted in Table 4.

A caveat has to be made regarding the crude labelling of an operational conditions as an *'event'*: the definition of each operational condition is based on SCADA thresholds, and the events category is composed of data-points which fall off the idling and nominal categories. The event label might therefor not catch every single event, or might mislabel nominal data. However, when the events category becomes more sizeable it still means that actual events are occurring. This is what happens for WT9 in December, where the events are the biggest fatigue contributor. It must also be said that this does not mean that WT9 operates the majority of time under events: they represent a minority of cases (below 10%) but, due to their magnitude, they have an oversized impact on lifetime. We can verify that it is indeed events which cause the higher fatigue loads experienced by WT9 in late December by inspecting Figure 15.

Case	Definition
Idling	RPM < 2 & pitch > 77 or Power < 0
Operational	0 < RPM < 10.5 ; 1 < pitch < 66
Events	Cases not included in the other categories

Table 4 Operational cases definition. *'Events'* category consists of all timestamps not included in other case definitions.

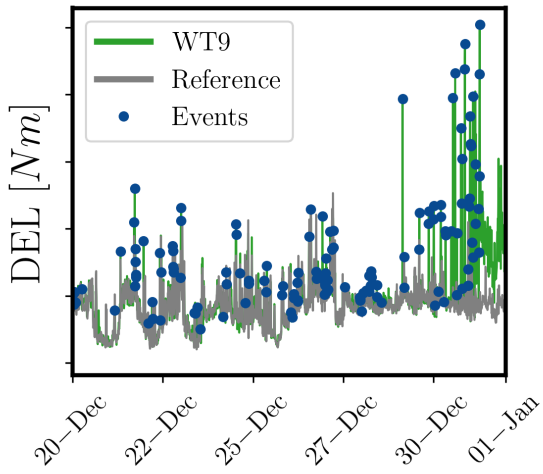


Figure 15 Bottom: estimated FA ten-minute DEL for WT9 (light green) and a reference turbine from the same string (dark green) with timestamps categorized as *'event'* superimposed (purple). On top, difference on pitch angle standard deviation between WT9 and reference (δ_β).

In this figure, we can observe the estimated DELs for WT9 (light green) and a reference turbine from the same string (grey) with WT9's timestamps categorized as *'event'* superimposed (purple). Here, we can see how the fatigue loads for WT9 are much bigger than the reference after the 30th of December, with the peaks corresponding well to the timestamps categorized as events. We can affirm that the hypothesis that WT9 experiences higher fatigue in December due to events is physically coherent, specially when we consider that, for this period at the end of December, WT9 had over 200% more labelled events than the reference turbine. We can also observe that after this period of events, loads for WT9 remain high, indicating that the turbine was idling. This behaviour is carried through to January, where WT9 faces higher idling fatigue loads, see Figure 14. We can confirm this by analysing Figure 16.

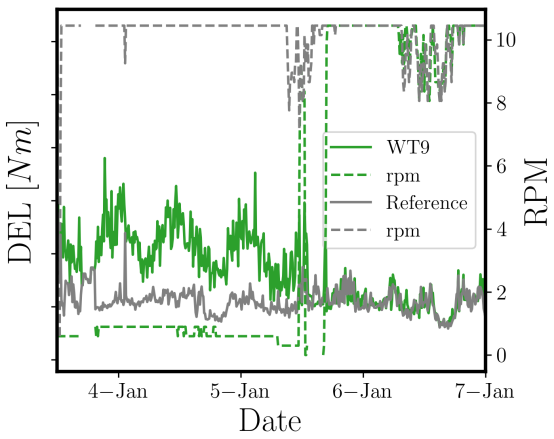


Figure 16 Estimated FA ten-minute DEL for WT9 (light green) and a reference turbine from the same string (dark green) and the rpm corresponding to each turbine (dotted lines).

In Figure 16 we present a zoomed instance of four days were one can see the reason behind the higher loads faced by WT9: in the period running up to January 6th WT9 DELs are higher than the reference turbine and this maps clearly with the RPM. This is, WT9 loads are higher because the rpm is zero (or close to it), thus indicating that the turbine is idling. As discussed in the introduction, for bigger monopiles, idling represents an increased consumption of lifetime as the lack of aerodynamic damping leads to unimpeded wave-excitation. Noteworthy is the apparent periodicity in the estimated FA DELs during idling conditions. The two full cycles in DEL correspond to the variations in tidal level, resulting in higher fatigue loads during at high tide. This is behaviour that is also observed on instrumented turbines and now is correctly predicted by the model.

4 | CONCLUSIONS AND FUTURE WORK

In this contribution we have presented a data-driven methodology for farm-wide tower-transition piece fatigue load estimation. Based on ten-minute SCADA and dedicated MEMS acceleration data, we attempted to ensure model performance beyond ten-minute estimations by physically guiding a neural network, making it focus both on long-term accuracy and conservatism. We specifically discussed what a real-world farm-wide employment of such a methodology means when taking place under continuous monitoring. This relates with the appropriate strategy to deal with missing data points, which we argue should consist in training a model (XGBoost) to estimate missing data as under idling, therefore ensuring an underlying conservatism. Additionally, based on the strain gauge instrumentation availability over the farm (on three turbines representative of the design clusters), we also discussed two competing approaches for data usage in model training: fleet-leader or population-based.

In the results section we began by verifying the better performance of population-based (lowest long-term error) models and we posited that this was due to their greater ability to better model the overall form of the turbines. Thereafter we focused in analysing the errors of the trained model over a long-term period of seven months, with monthly accumulated errors always kept well below $\pm 5\%$. Furthermore, we discussed how this error ought to be framed in relation to a continuous monitoring strategy, which enables model re-training and the minimal amount of data required for capturing the dynamics' seasonality (9−12 months). We additionally focused on the relative worse performance on high load estimation and hypothesized this error is due to the ten-minute resolution which does not enable the localization of an event within the timeseries, inducing errors. One mitigation strategy suggested to address this issue was to also compute spectral moments as a statistic for each ten-minute signal. Finally, we presented farm-wide results on DEL estimation, which allowed to identify outliers, whose behaviour we correlated with their operational conditions.

Lastly, this contribution has successfully present a real-world farm-wide estimation of fatigue loads based on SCADA and acceleration data, providing continuous fatigue monitoring. As discussed throughout this work, this ability has a considerable value on its own, opening the door to cost-effective structural health decisions: better scheduled preventive maintenance, with greater margins for turbines where a sufficient structural reserve is found, and tighter control for turbines which see more damage. Finally, this methodology can also serve as a springboard for further lifetime-based decision-making: which turbines require more or less maintenance, farm-wide curtailment impact quantification and DEL-based anomaly detection. This future use of the current methodology ought to be furthermore coupled with the improvement of the actual methodology, namely through evolving into uncertainty quantification (with a special focus to epistemic uncertainty) and by testing other architectures.

Acknowledgements

The authors would like to acknowledge 24SEA and Parkwind for the data collection and management and the financing provided by VLAIO through the ICON Supersized 4.0 project.

Author contributions

Francisco de N Santos: conceptualization; formal analysis; writing—original draft preparation. **Nymfa Noppe:** Data curation. **Wout Weijtjens:** conceptualization; supervision; writing—review and editing. **Christof Devriendt:** conceptualization; writing—review and editing; funding acquisition.

References

1. Statistics Time Series, International Renewable Energy Agency. https://www.irena.org/Statistics/View-Data-by-Topic/Capacity-and-Generation/Statistics-Time-Series; . Accessed: 2023-05-31.

2. Miedema R. Offshore Wind Energy Operations & Maintenance Analysis. *Hoge school van Amsterdam* 2012.

3. Bhattacharya S. Challenges in design of foundations for offshore wind turbines. *Eng. Technol. Ref* 2014; 1(1): 922.

4. Sparrevik P. Offshore Wind Turbine Foundations State of the Art. In: IOS Press. 2019 (pp. 216–238).

5. N Santos dF, Robbelein K, D'Antuono P, Noppe N, Weijtjens W, Devriendt C. Towards a Fleetwide Data-Driven Lifetime Assessment Methodology of Offshore Wind Support Structures Based on SCADA and SHM Data. In: Rizzo P, Milazzo A., eds. *European Workshop on Structural Health Monitoring*. 253 of *Lecture Notes in Civil Engineering*. Springer International Publishing; 2023; Cham: 123–132.

6. Society for Underwater Technology*PISA: new design methods for offshore wind turbine monopiles*. 142; 2017.

7. Velarde J, Kramhøft C, Sørensen JD, Zorzi G. Fatigue reliability of large monopiles for offshore wind turbines. *International Journal of Fatigue* 2020; 134: 105487.

8. Gengenbach J, Mikkelsen K, Rudinger F, Brommundt M, Gretlund J. Design challenges of XL monopiles. *Proceedings of the EWEA (European Wind Energy Association) Offshore* 2015.

9. Robbelein K, Daems P, Verstraeten T, et al. Effect of curtailment scenarios on the loads and lifetime of offshore wind turbine generator support structures. In: . 2507. IOP Publishing. ; 2023: 012013.

10. Pacheco J, Pimenta F, Pereira S, Cunha Á, Magalhães F. Fatigue assessment of wind turbine towers: Review of processing strategies with illustrative case study. *Energies* 2022; 15(13): 4782.

11. Nielsen JJ, Sørensen JD. On risk-based operation and maintenance of offshore wind turbine components. *Reliability engineering & system safety* 2011(1): 218–229.

12. García D, Tcherniak D. An experimental study on the data-driven structural health monitoring of large wind turbine blades using a single accelerometer and actuator. *Mechanical Systems and Signal Processing* 2019; 127: 102–119.

13. Schedat M, Faber T, Sivanesan A. Structural health monitoring concept to predict the remaining lifetime of the wind turbine structure. In: IEEE. ; 2016: 1–5.

14. Iliopoulos A, Weijtjens W, Van Hemelrijck D, Devriendt C. Fatigue assessment of offshore wind turbines on monopile foundations using multi-band modal expansion. *Wind Energy* 2017; 20(8): 1463–1479.

15. Ziegler L, Smolka U, Cosack N, Muskulus M. Brief communication: Structural monitoring for lifetime extension of offshore wind monopiles: can strain measurements at one level tell us everything?. *Wind Energy Science* 2017; 2(2): 469–476.

16. Worden K, Farrar CR, Manson G, Park G. The fundamental axioms of structural health monitoring. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2007; 463(2082): 1639–1664.

17. Wymore ML, Van Dam JE, Ceylan H, Qiao D. A survey of health monitoring systems for wind turbines. *Renewable and Sustainable Energy Reviews* 2015; 52: 976–990.

18. Ziegler L, Cosack N, Kolios A, Muskulus M. Structural monitoring for lifetime extension of offshore wind monopiles: Verification of strain-based load extrapolation algorithm. *Marine Structures* 2019; 66: 154–163.

19. Hübler C, Rolfes R. Probabilistic temporal extrapolation of fatigue damage of offshore wind turbine substructures based on strain measurements. *Wind Energy Science* 2022; 7(5): 1919–1940.

20. Pacheco J, Pimenta F, Pereira S, Cunha Á, Magalhães F. Experimental evaluation of strategies for wind turbine farm-wide fatigue damage estimation. *Engineering Structures* 2023; 285: 115913.

21. Vera-Tudela L, Kühn M. Evaluation of a wind turbine fatigue load monitoring system based on standard SCADA signals in different wind farm flow conditions. *Dewek 2015* 2015; 49(441).

22. Movsessian A, Schedat M, Faber T. Modelling tower fatigue loads of a wind turbine using data mining techniques on SCADA data. *Wind Energy Science Discussions* 2020: 1–20.

23. Noppe N, Hübler C, Devriendt C, Weijtjens W. Validated extrapolation of measured damage within an offshore wind farm using instrumented fleet leaders. In: . 1618. IOP Publishing. ; 2020: 022005.

24. Smolka U, Kaufer D, Cheng P. Are sea state measurements required for fatigue load monitoring of offshore wind turbines?. In: . 555. IOP Publishing. ; 2014: 012095.

25. Noppe N, Iliopoulos A, Weijtjens W, Devriendt C. Full load estimation of an offshore wind turbine based on SCADA and accelerometer data. In: . 753. IOP Publishing. ; 2016: 072025.

26. N Santos dF, Noppe N, Weijtjens W, Devriendt C. Data-driven farm-wide fatigue estimation on jacket-foundation OWTs for multiple SHM setups. *Wind Energy Science* 2022; 7(1): 299–321.

27. Géradin M, Rixen DJ. *Mechanical vibrations: theory and application to structural dynamics*. John Wiley & Sons . 2014.

28. Noppe N, Tatsis K, Chatzi E, Devrient C, Weijtjens W. Fatigue stress estimation of offshore wind turbine using a Kalman filter in combination with accelerometers. In: KU Leuven, Department of Mechanical Engineering. ; 2018: 4693–6701.

29. Venu A, Lüdde M, Omole J. Prediction and Validation of Fatigue loads using Artificial Intelligence on Real World Measurement Data. In: . 1618. IOP Publishing. ; 2020: 022006.

30. Smolka U, Cheng PW, others . On the design of measurement campaigns for fatigue life monitoring of offshore wind turbines. In: International Society of Offshore and Polar Engineers. ; 2013.

31. Vera-Tudela L, Kühn M. On the selection of input variables for a wind turbine load monitoring system. *Procedia Technology* 2014; 15: 726–736.

32. Vera-Tudela L, Kühn M. Analysing wind turbine fatigue load prediction: The impact of wind farm flow conditions. *Renewable Energy* 2017; 107: 352–360.

33. Schröder L. *Towards digital twin technology: Wind farm operation analysis and optimization using model-supported data analytics*. PhD thesis. Danmarks Tekniske Universitet, 2020.

34. Avendaño-Valencia LD, Abdallah I, Chatzi E. Virtual fatigue diagnostics of wake-affected wind turbine via Gaussian Process Regression. *Renewable Energy* 2021; 170: 539–561.

35. Mylonas C, Abdallah I, Chatzi E. Deep unsupervised learning for condition monitoring and prediction of high dimensional data with application on windfarm scada data. In: Springer. 2020 (pp. 189–196).

36. Mylonas C, Abdallah I, Chatzi E. Conditional variational autoencoders for probabilistic wind turbine blade fatigue estimation using Supervisory, Control, and Data Acquisition data. *Wind Energy* 2021; 24(10): 1122–1139.

37. Mylonas C, Chatzi E. Remaining useful life estimation for engineered systems operating under uncertainty with causal GraphNets. *Sensors* 2021; 21(19): 6325.

38. Movsessian A, Schedat M, Faber T. Feature selection techniques for modelling tower fatigue loads of a wind turbine with neural networks. *Wind Energy Science* 2021; 6(2): 539–554.

39. Avendano-Valencia LD, Chatzi EN, Tcherniak D. Gaussian process models for mitigation of operational variability in the structural health monitoring of wind turbines. *Mechanical Systems and Signal Processing* 2020; 142: 106686.

40. Yeter B, Garbatov Y. Structural integrity assessment of fixed support structures for offshore wind turbines: A review. *Ocean Engineering* 2022; 244: 110271.

41. N Santos dF, D'Antuono P, Robbelein K, Noppe N, Weijtjens W, Devriendt C. Long-term fatigue estimation on offshore wind turbines interface loads through loss function physics-guided learning of neural networks. *Renewable Energy* 2023.

42. Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. *Nature Reviews Physics* 2021; 3(6): 422–440.

43. Cross EJ, Gibson S, Jones M, Pitchforth D, Zhang S, Rogers T. Physics-Informed Machine Learning for Structural Health Monitoring. In: Springer. 2022 (pp. 347–367).

44. Igwemezie V, Mehmanparast A, Kolios A. Current trend in offshore wind energy sector and material requirements for fatigue resistance improvement in large wind turbine support structures–A review. *Renewable and Sustainable Energy Reviews* 2019; 101: 181–196.

45. Negro V, López-Gutiérrez JS, Esteban MD, Alberdi P, Imaz M, Serraclara JM. Monopiles in offshore wind: Preliminary estimate of main dimensions. *Ocean Engineering* 2017; 133: 253–261.

46. Pandit R, Astolfi D, Hong J, Infield D, Santos M. SCADA data for wind turbine data-driven condition/performance monitoring: A review on state-of-art, challenges and future trends. *Wind Engineering* 2022: 0309524X221124031.

47. Dirlik T. *Application of computers in fatigue analysis*. PhD thesis. University of Warwick, 1985.

48. Marsh G, Wignall C, Thies PR, et al. Review and application of Rainflow residue processing techniques for accurate fatigue damage estimation. *International Journal of Fatigue* 2016; 82: 757–765.

49. Kauzlarich J. The Palmgren-Miner rule derived. In: . 14. Elsevier. 1989 (pp. 175–179).

50. Ziegler L, Muskulus M. Comparing a fracture mechanics model to the SN-curve approach for jacket-supported offshore wind turbines: Challenges and opportunities for lifetime prediction. In: . 49972. American Society of Mechanical Engineers. ; 2016: V006T09A054.

51. Hübler C, Weijtjens W, Rolfes R, Devriendt C. Reliability analysis of fatigue damage extrapolations of wind turbines using offshore strain measurements. *Journal of Physics: Conference Series* 2018; 1037(3): 032035.

52. Sadeghi N, Robbelein K, D'Antuono P, Noppe N, Weijtjens W, Devriendt C. Fatigue damage calculation of offshore wind turbines' long-term data considering the low-frequency fatigue dynamics. In: . 2265. IOP Publishing. ; 2022: 032063.

53. Hendriks H, Bulder B. Fatigue Equivalent Load Cycle Method. 1995.

54. Cosack N. *Fatigue load monitoring with standard wind turbine signals*. PhD thesis. Universität Stuttgart, 2010.

55. Seidel M, Voormeeren S, Steen v. dJB. State-of-the-art design processes for offshore wind turbine support structures: Practical approaches and pitfalls during different stages in the design process. *Stahlbau* 2016; 85(9): 583–590.

56. Bull L, Gardner P, Gosliga J, et al. Foundations of population-based SHM, Part I: Homogeneous populations and forms. *Mechanical systems and signal processing* 2021; 148: 107141.

57. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Machine learning* 2002; 46: 389–422.

58. Guyon I, Elisseeff A. An introduction to variable and feature selection. *Journal of machine learning research* 2003; 3(Mar): 1157–1182.

59. Kuhn M, Johnson K, others . *Applied predictive modeling*. 26. Springer . 2013.

60. Rasmussen CE, Williams CK, others . *Gaussian processes for machine learning*. 1. Springer . 2006.

61. Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems* 2012; 25.

62. O'Malley T, Bursztein E, Long J, et al. KerasTuner. https://github.com/keras-team/keras-tuner; 2019.

63. Ramachandran P, Zoph B, Le QV. Searching for activation functions. *arXiv preprint arXiv:1710.05941* 2017.

64. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* 2014.

65. N Santos dF, D'Antuono P, Noppe N, Weijtjens W, Devriendt C. Minkowski logarithmic error: A physics-informed neural network approach for wind turbine lifetime assessment. In: ESANN. ; 2022.

66. Martinez-Luengo M, Shafiee M, Kolios A. Data management for structural integrity assessment of offshore wind turbine support structures: data cleansing and missing data imputation. *Ocean Engineering* 2019; 173: 867–883.

67. N Santos dF, Noppe N, Weijtjens W, Devriendt C. Results of fatigue measurement campaign on XL monopiles and early predictive models. In: . 2265. IOP Publishing. ; 2022: 032092.

68. Meetnet Vlaamse Banken. https://meetnetvlaamsebanken.be/; . Accessed: 05-06-2023.

69. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: ; 2016: 785–794.

70. Zou M, Jiang WG, Qin QH, Liu YC, Li ML. Optimized XGBoost Model with Small Dataset for Predicting Relative Density of Ti-6Al-4V Parts Manufactured by Selective Laser Melting. *Materials* 2022; 15(15): 5298.

71. N Santos dF, Noppe N, Weijtjens W, Devriendt C. SCADA-based neural network thrust load model for fatigue assessment: cross validation with in-situ measurements. In: . 1618. IOP Publishing. ; 2020: 022020.

How to cite this article: F. d. N. Santos, N. Noppe, W. Weijtjens and C. Devriendt (2023), Farm-wide fatigue loads estimation: a data-driven approach, *Wind Energy*, *vol*.