

Feasibility of machine learning algorithms for classifying damaged offshore jacket structures using SCADA data

D Cevalco¹, J Tautz-Weinert², U Smolka², A Kolios¹

¹ Naval Architecture, Ocean and Marine Engineering Department, University of Strathclyde, G11XQ Glasgow, United Kingdom

² Ramboll Deutschland GmbH, 20095 Hamburg, Germany

Corresponding author: debora.cevasco@strath.ac.uk

Abstract. The best practise for structural damage detection currently relies on the installation of structural health monitoring systems for the collection of dedicated high frequency measurements. Switching to the employment of the wind turbine's SCADA (Supervisory Control and Data Acquisition) signals and their commonly recorded low frequency statistics can lead to a reduction in the number of ad-hoc monitoring sensors and quantity of data required. In this paper, aero-hydro-servo-elastic simulations for a model of a turbine are used to assess its loads and any changes in the dynamics under healthy state and a damaged configuration case study. To prove the feasibility of the damage detection through low-resolution data, the statistics of the typically recorded signals from the SCADA and the structural monitoring systems are fed into a database for training and testing of classification algorithms. The ability of the machine learning models to generalise the classification for both stochasticity and uncertainties in the environmental conditions are tested. Decision tree-based classifiers showed the capability to capture the damage for the majority of the operating conditions considered. Though the setup of the traditional SCADA sensors had to be supplemented with an additional structural health monitoring sensor, the detection of the damage has been shown feasible by referring to low-frequency statistics only.

1. Introduction

Several strategies have been investigated and exploited towards reducing the levelized cost of energy (LCOE) of offshore wind farms. So far, the focus was on a 'race to the bottom' for development costs, pushing for the optimization of design parameters and installations strategies. During the last decade, the LCOE reduction has been targeted through reduction of the operation and maintenance costs [1], which are expected to reach 30% of the asset's life time cost for the next generation of offshore farms [2]. Although structural damage is not very likely, its late detection can lead to critical consequences which will result to high cost of mitigation actions [3]. On the other side, technical assessments and knowledge of the status of the turbine's support structure are necessary to prove that operating assets can maintain the required safety levels during lifetime extensions [4].

Current practice for structural damage detection strongly relies on inspections, i.e. practical assessments on-site, which are associated with significant costs and risks due to the offshore environment, especially for structural failures below sea-water level [5]. With respect to data-driven techniques, research and applications of vibration-based structural damage detection can be found in literature [6], [7]. The methods proposed mostly identify damage using either natural frequencies or mode shapes and their derivatives such as the displacement modal curvature [8]. Natural frequencies



have low sensitivity to structural damage (e.g. member loss, scour, corrosion), so they can usually be used to reflect damage to a moderate degree; in contrast, mode shapes have greater sensitivity to damage [9]. However, the assessment of a mode shape requires high-resolution sensors on the support structure and tower of the wind turbine. Additionally, issues related to the application of traditional operational modal analysis (OMA) to wind turbine dynamics [7], [10] impose difficulties in utilizing mode shapes to represent damage in practical applications. Alternatively, data-driven approaches for anomaly detection can be based on machine learning algorithms and techniques applied to signals from the operational turbine [5]. These have been extensively investigated and applied for the condition monitoring of the drivetrain mechanical and electrical components [11]–[13], where data associated to failure events are more commonly collected. Concerning the structural health monitoring of offshore wind support structures, in [14] the approach for the detection of local damage is supported by experimental data for the training and testing of the algorithm.

2. Scope of the analysis and detection approaches

The aim of this work is to propose an alternative to the current practice for the identification of a damage on an offshore wind turbine jacket support structure. The focus of the analysis is on the feasibility of the detection of a structural damage by capturing abnormal variations of the ten-minute statistics of SCADA signals. As opposed to the on-demand structural health monitoring systems, installed on no more than 10% of the turbines across a farm, SCADA system signals are continuously collected throughout the service life of all assets. Therefore, the proposed approach could be applied to all units across a farm. Furthermore, the requirement for low-frequency statistics of the SCADA data are less of a burden from a data storage and handling point of view.

A summary of the approaches for damage detection based on high-frequency sensor signals (current practice) and low-resolution data (SCADA statistics) is given in Table 1. Two frameworks for implementing machine learning approaches are possible, depending on whether the damage equivalent load (DEL) at different locations and/or components varies in presence of a damage in the structure or not. The accuracy of the fatigue loads estimated from direct monitoring is affected by missing data and noise in the acquired signals [15]. To avoid relying on the installation of strain gauges and dealing with data management issues, several authors proposed and validated the use of ten-minute statistics of available operational SCADA signals for the prediction of structural fatigue loads from trained neural network models [16]–[19]. If a DEL is shown to be a good indicator of an anomaly in the system (see [20], for scour), SCADA data could be employed for quantifying it and its deviation. In contrast, anomalies in the SCADA data and in other measurable signals should be investigated to identify the indicator(s) of the damage if the DEL is shown not to be one. In terms of the machine learning approaches for anomalies in SCADA and other measurable signals, this work is addressing supervised learning only, i.e. classification.

Table 1. Summary of the possible approaches for structural damage detection: current practice and possible approaches based on SCADA data

Damage Indicator(s)	Sensor	Resolution	Detection approach
<i>Natural frequencies and/or mode shapes</i>	Accelerometers	≥ 20 Hz	<i>Vibration-based</i>
<i>Fatigue loads (DEL)</i>	• Strain gauge (direct monitoring)	≥ 20 Hz	<i>Machine learning</i> Monitoring of DEL via regression and/or anomaly detection approach
	• SCADA (indirect monitoring)	10-min	
<i>Anomaly in SCADA data</i>	SCADA		<i>Machine learning</i>
<i>Anomaly in other measurable signals</i>	• Strain gauges	10-min	(1) Classification approach for identification of the damage indicator(s)
	• Accelerometer		(2) Monitoring of quantity via regression and/or anomaly detection approach
	• Inclinometer, etc.		

3. Methodology

Figure 1 shows the flowchart of the approach. The methodology developed in this paper is based on the following three main steps:

- The dynamics of the structure are analysed in response to environmental and operational conditions (EOC) and as a consequence of the structural failure. Semi-coupled simulations of the finite element (FE) model of the support structure, in healthy and damaged condition, and the model of the flexible structure of the turbine tower are run for a representative set of EOC. The time histories of loading conditions and relative structural responses are then post-processed into ten-minute statistics, similarly to SCADA data.
- The data are collected into a database, which is sequentially accessed to construct the datasets for the training and testing of machine learning algorithms. Each dataset for varying environmental loadings of the structure has a 50:50 ratio of healthy and damaged conditions. The features, i.e. the set of independent variables used for the prediction, of each dataset are then standardised before using in the models. The dataset in analysis is then divided into subsets, named “Tr” for the tuning/training subsets and “Te” for testing subsets.
- The tuning, training and testing methods are applied in the same manner for all datasets considered and all algorithms investigated. The detection capability of logistic regression, support vector machine, k-nearest neighbour, random forest, and Gaussian naïve Bayes based classifiers is examined. Each algorithm’s best performance is obtained through an iterative process. If a classifier performs in an unsatisfactory way on the test set for stochastic variation of the loadings, feature and dimension reduction techniques are applied (see Figure 1, in green). The most promising algorithms are then tested on the subsets for variations of the environmental parameters (see Figure 1, with “dashed” lines in the block for “synthetic data generation” and “data processing”). If insufficient detection performance is exhibited at this stage, the subset for the training of the classifier is extended to include additional features and/or the data from the misclassified cases. Tuning, fitting and testing are then repeated (see Figure 1, in blue). Eventually, a recommendation of the best training set and algorithm for the damage detection task is given.

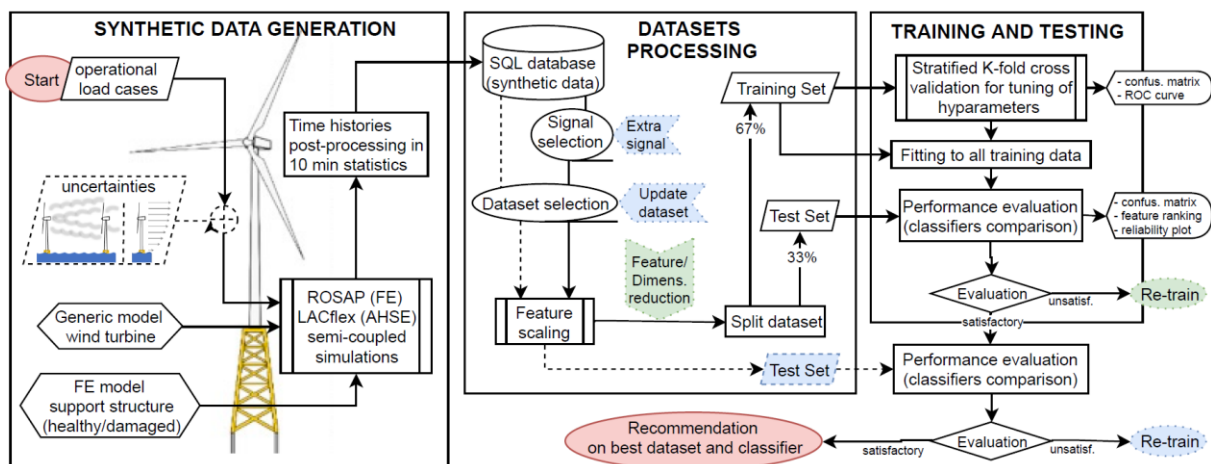


Figure 1. Workflow for the data generation, the datasets processing and the algorithms training and testing. In “dotted lines” are the iterative processes for the re-training of an algorithm.

4. Synthetic data generation

It is very unlikely to have access to real data of damaged states, as such event rarely happens for the subsystem under consideration, and such data would be unlikely to be shared for research purposes due to confidentiality reasons. For the creation and collection of the synthetic data, Ramboll’s in-house software ROSAP (Ramboll Offshore Structural Analysis Programs) [21] and LACflex aero-servo-

elastic code (based on Flex5) [22] are used in a semi-coupled approach for a model of a turbine installed on jacket foundations (cp. Figure 1). The ROSAP packages are used for the definition of the as-designed jacket geometries, material properties, and environmental loads of the jacket structure. The full model of the substructure is reduced into Craig-Bampton superelements [23], approximating its dynamics by including only a limited number of deformation modes. The LACFlex aero-servo-elastic software allows an accurate modelling of the tower and the rotor-nacelle assembly, together with the control strategies in response to the stochastic turbulent wind loads acting on the turbine. The superelement files derived in ROSAP are then coupled with the wind turbine model in LACFlex for each combination of wind and wave loads, and integrated simulations are carried out.

4.1. Damage implementation

As case study, the full loss of a cross-member of the jacket structure is simulated. The stiffness of a brace close to the seabed which connects diagonally two of the legs of the jacket is reduced to a value close to zero. This failure location is selected because it is associated to a high deviation of the global natural frequencies. It is worth mentioning that this approach, which is based on simulated data, carries some uncertainties. On one side, the coupling of the foundation dynamics via the superelement innerly hold some small discrepancies with the substructure dynamics (until 10 Hz), with respect to the full model [23]. On the other side, the global damping of the structure is assumed to be the same as the one defined at the design phase, neglecting thus the possible effect of the structural failure. These model uncertainties are here acknowledged but are not tackled in this paper. Nonetheless, the analysis is harmonized, by extending the first assumption to the healthy model as well. The second assumption is judged acceptable, since the aim of the investigation is to demonstrate only the detection feasibility.

4.2. Simulations setup

Focusing on detection during normal operational conditions (power production) of the turbine, the fatigue limit state load case is used for the setup of the EOC, as specified in DLC 1.2 [24]. A set of representative load combinations is then investigated. Considering the geometry of the jacket and the implemented damage, only 4 wind directions and 12 wave directions are simulated, for 6 values of average wind speed at the hub height (3 below and 3 above rated conditions). Nine realizations of the wind and wave time histories are processed for each loading combination, to guarantee the capability of the detection algorithms to distinguish the response due to load stochasticity from one of the damaged status. Therefore, a total of 2,592 simulations per turbine status are performed.

To account for the uncertainty associated to the real operational conditions of the turbine, the healthy and damaged structural responses are derived for changes in the wind farm flow conditions as illustrated in Figure 2. The wind shear exponent (WS) – which is potentially correlated to multiple factors [25] – is varied from the design base specification to a minimum of 0.08 (WS_L) and a maximum of 0.3 (WS_U), changing the distribution of the normal wind profile. The 90th percentile of the effective turbulence intensity (TI) is used as reference, a value considered representative for the fatigue design calculations [26]. Based on the experience from a similar farm, an upper (TI_U) bound curve is defined to represent the missing extreme cases. Similarly, a lower (TI_L) bound curve is drawn corresponding to the 10th percentile of the effective turbulence intensity. These are then implemented in a Mann turbulence model to represent the fluctuating wind field.

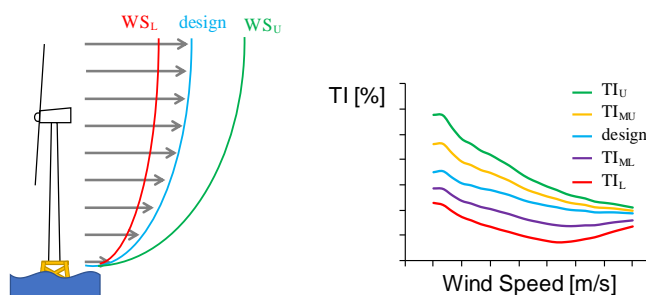


Figure 2. Visualisation of the uncertainties considered (upper and lower bounds) in the wind farm flow condition of the fatigue design simulations: wind shear (left), and TI (right).

5. Pre-analysis and datasets processing

5.1. Effect of environmental parameter variations

As outlined in [5], any structural health monitoring method employed in the detection task must be able to distinguish between signal variations related to EOC, as opposed to the ones corresponding to a structural anomaly. The damage implemented was observed to lead to significant changes of the second modes and their natural frequencies with respect to the healthy status. It is then reasonable to expect that this has an impact on the loads at the tower base, building the interface between the turbine and the jacket structure. However, variations of the environmental conditions could affect the global response in a similar manner.

Therefore, a pre-analysis for a reduced number of simulations is performed to investigate the influence of wind flow parameters (upper- and lower-bound values) on the structural response. The deviation of the loads at the tower interface, with respect to the design load case, is presented in Figure 3, for a below rated load combination. The box plots of the time histories of forces and moments in the fore-aft (y) and the side-side (x) direction are reported for the nine stochastic variations of wind and wave loadings. It can be noticed that the wind shear exponent does not seem to significantly affect any of the loads. On the contrary, a high level of turbulence intensity is associated to higher load ranges and standard deviations compared to the design base scenario. Opposite behavior is then observed for the low-turbulence level. Consequently, datasets for variations of the TI need to be fed to the machine learning models during its training, as opposed to variation of the wind shear that can be fully captured in the structural dynamics of the design load case.

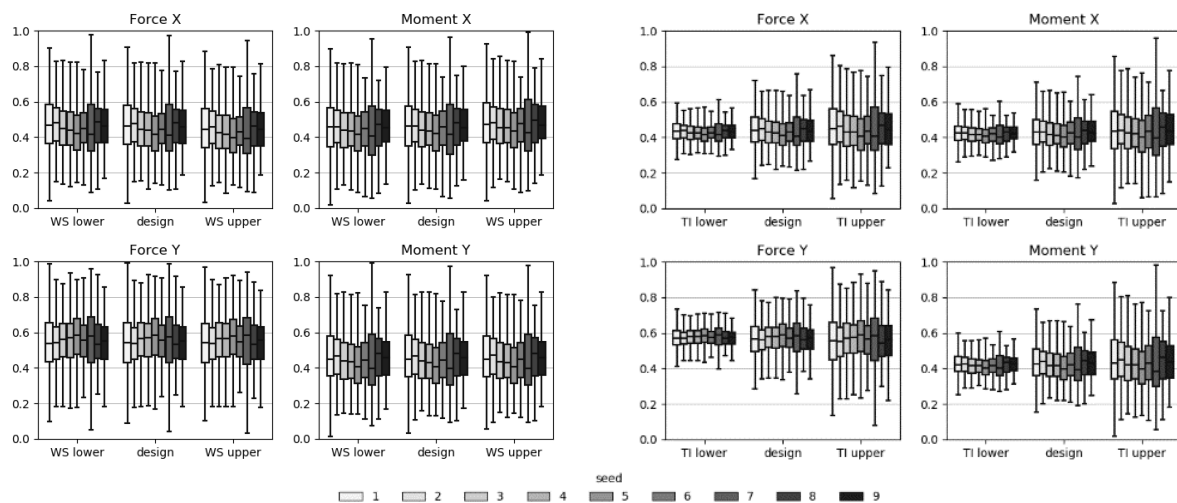


Figure 3. Healthy status turbine box plots of the interface load for lower- and upper-bound EOC parameters (wind shear exponent – left – and TI – right) for a selected below rated load combination.

5.2. Detectability and signals deviation

The 50 Hz time histories output from the aero-hydro-servo-elastic simulations are post processed into ten-minute minimum (min), maximum (max), average (mean), and standard deviation (std). At first, all the measurable signals are collected into the database of operational conditions, potentially being meaningful indicators of the structural failure.

The potential predictors of the damage are then investigated by quantifying the deviation of the statistics, given the scatter for the stochastic variation of the loading, for the healthy and damaged status data. A visual representation of this assessment is given in Figure 4, where trends in the mean of the tower top acceleration and the DEL of tower base bending moments are presented. With respect to the DEL of the bending moments, only the MyF0 related to the side-side motion (Figure 4) deviates slightly from the healthy status for some wind-wave misalignment angles. The DEL of the MxF0 in the fore-aft motion of the tower base (Figure 4), as well as the DELs of the moments at the tower top, at the foundation and at the blade roots remained mainly unaffected. On the contrary, it was observed that the

tower top acceleration (A_{xTT} and A_{yTT} in Figure 4), the power output range (for the extreme turbulence case), the tower bottom acceleration and rotations and some of the forces and moments in the drivetrain (mainly at the main bearing) record discrepancies from the healthy operating conditions, throughout the load combinations considered (not all shown for brevity).

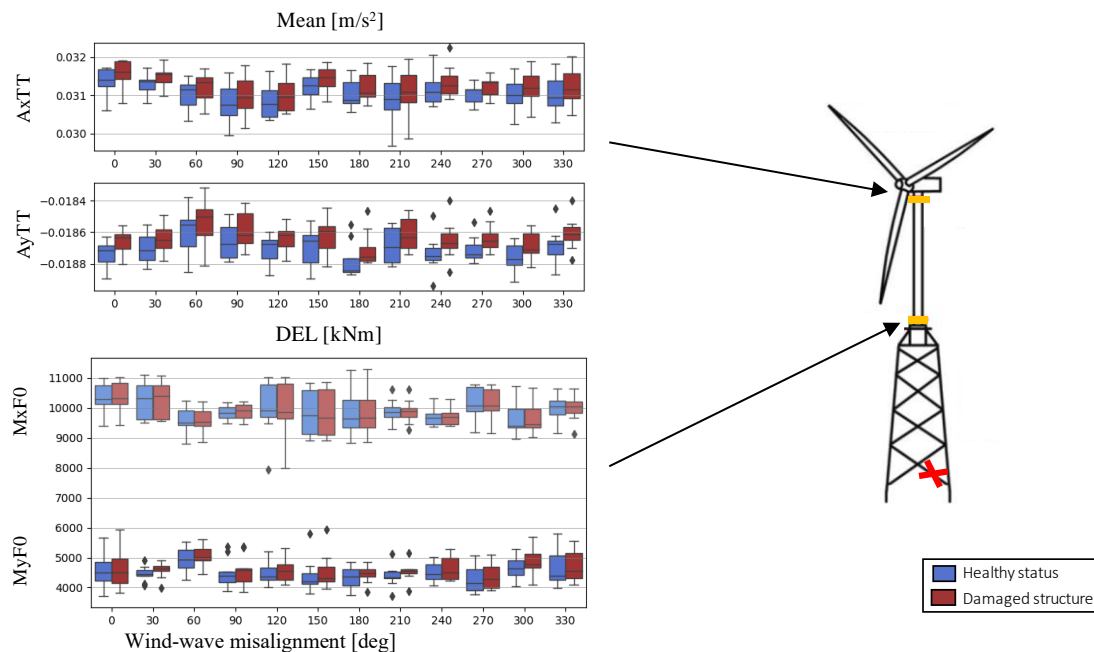


Figure 4. Box plots of the ten-minute mean tower top acceleration ($A_{\bullet TT}$) and interface bending moments DEL ($M_{\bullet F0}$) against wind-wave misalignments for an above rated case load combination.

5.3. Data subsets for training and testing

A summary of the datasets and subsets considered is given in Table 2, whereas Table 3 details the considered sensors. A detection through standard SCADA signals (sensor setup S0) is preferred and attempted at first. Initially, the investigation of the feasibility of the status classification is performed on the dataset of the design load combinations (dataset D0). Then the data derived for the different TI levels are added to this base scenario, expecting significant variations in the loadings and response of the structure (datasets D1, D2 and D3). Each of the training subsets (Tr#) consists of 67% of the data of the current set, by randomly selecting six out of the nine realizations for each load case. The remaining set of three realizations per load case, consisting of the 33% of the data in the set, is collected in the subset Te33 and used for testing. Additionally, the mid -upper and -lower TI curves (Figure 2) are derived, and the statistics associated to these loadings are collected into the Te3 and Te4 test sets, over that the one for upper and lower bound values (in the test sets Te1 and Te2).

Table 2. Datasets and subsets for training and testing

Data set	Acronym	Loading conditions	No. simulations
Datasets (D)	D0	design	5,184
	D1	design + TI _U	10,368
	D2	design + TI _L	10,368
	D3	design + TI _U + TI _L	15,552
Training subsets (Tr)	Tr#	-	67% D#
	Te33	-	33% D#
Testing subsets (Te)	Te1	TI _U	5,184
	Te2	TI _L	5,184
	Te3	TI _{MU} *	5,184
	Te4	TI _{ML} *	5,184

* mid-upper and -lower TI curves selected as shown in Figure 2

To achieve satisfactory classification results, also the sensor setup is redefined during the training/testing iterative steps (cp. Table 3). Additionally, an inclinometer and strain gauges are assumed to be installed at the tower base for the indirect and direct measurement of strain. The advantages from the selection of these sensors' signals and location is: 1) in avoiding installing monitoring devices below the water level and, 2) to maintain the analysis as independent as possible from measurement from the drivetrain, which are highly related to the specific control strategies and optimization.

Table 3. Sets of sensors and signals considered for the detection purpose

Sensor type	Measurement	Signal acronym	Unit	Sensor set up			
				S0	S1	S2	S3
SCADA	Nacelle direction	YawPos	[deg]	x	x	x	x
	Wind direction	WDir	[deg]	x	x	x	x
	Yaw angle (misalign. error)	YawErr	[deg]	x	x	x	x
	Wind speed	Whub	[m/s]	x	x	x	x
	Power	Pow	[kW]	x	x	x	x
	Rotor speed	RotSpd	[rpm]	x	x	x	x
	(Collective) Pitch angle	PiPos1	[deg]	x	x	x	x
	2D Tower top acceleration	AxTT, AyTT	[m/s ²]	x	x	x	
Inclinometer	2D Rotation at interface	UrxF, UryF	[deg]		x	x	x
Strain Gauge	2D Bending moment at interface	MxF0, MyF0	[kNm]			x	

6. Training and testing of the detection algorithms

6.1. Training and testing methods

Not having a clear indication on which of the signals (in Table 3) could stand as the best indicator(s) for a damage, the different sensor sets are used for the supervised classification to identify the selection of the best predictors based on the classifiers' performance. The algorithms are implemented with the Python machine learning package (skikit) [27]. The parameter tuning approach and the training methods applied are briefly described in the following. Subsequently, the key performance indicators for the assessment of the detection quality are introduced.

6.1.1. Grid search cross-validation and training. A grid search is applied for the identification of the best hyperparameters for each model. The combination giving the overall best performance on the folds of the training set is selected (cross-validation). These folds, i.e. subsets, are selected by applying a stratified k-fold approach [28], which divides the training set in homogeneous splits of healthy and damaged data samples. Each of the models tuned with the optimal set of hyperparameters, is then fitted to the full set of training data.

6.1.2. Performance estimation. The key performance indicators of the models' detection capability are derived from the confusion matrix (Table 4). These are the accuracy (*acc*), true damage (positive) rate - here referred to as true detection rate (*TDR*) -, and false damage (positive) rate - here referred to as false detection rate (*FDR*) -, as defined in equations from (1) to (3) respectively. While the *acc* gives an overall indication on the goodness of the classifier, the *TDR* refers to the models' ability to detect the damage. A satisfactory *TDR* is assumed for values above 70%. The *FDR*, instead, gives an indication of the percentage of false alarms raised by the classifier. A satisfactory *FDR* is considered if below 30%. Receiver operating characteristic (ROC) curves [29] and feature ranking plots are employed during the tuning and training phases to track the performance and validity of the prediction. Additionally, when performing classification, it is advised to investigate the probability corresponding to the predicted category. This probability gives a measure of the confidence on the prediction and is presented in the so-called reliability curve [30]

Table 4. Confusion matrix terminology

		Predicted	
		Healthy (0 or Negative)	Damaged (1 or Positive)
Actual	Healthy (0 or Negative)	True Healthy (<i>TH</i>)	False Damaged (<i>FD</i>)
	Damaged (1 or Positive)	False Healthy (<i>FH</i>)	True Damaged (<i>TD</i>)

$$acc = \frac{TD+TH}{Total\ population} \quad (1)$$

$$TDR = \frac{TD}{FH+TD} \quad (2)$$

$$FDR = \frac{FD}{TH+FD} \quad (3)$$

6.2. Results

The results of the training and testing of the classification algorithms are reported in Table 5. Unsatisfactory performance at the tuning and training stage are seen by the following models:

- **Gaussian naïve Bayes (NB).** Based on the strong (naïve) independence assumptions between the features, it is here implemented by selecting the first 9 principal components (eigenvalues) corresponding already to more than the 90% of all the variance (as explained in [31]). Yet, unacceptable performance is given, probably because of violation of the normal distribution assumption for the numerical predictors.
- **K-Nearest Neighbour (KNN).** Despite of the broad and small stepped range of K given during the cross-validation tuning and fitting, this algorithm, given its implementation in [27], fails in finding boundaries separating the two classes.

On the contrary, results worthy of further investigation are obtained by:

- **Logistic Regression (LR).** Logistic regression is a linear method that models a binary dependent variable, where the predictions are transformed using the logistic (sigmoid) function. As for linear regression, the model can overfit if there are multiple correlated inputs [32]. Here, it does not seem to happen, despite the high dependency of some of the features.
- **Support Vector Machine (SVM).** The SVM approach aims to find a line, surface or hypersurface for the separation of the classes. When applied to the data, it fails in finding a linear hyperplane for a correct classification. On the other hand, by projecting the data into a higher-dimensional space defined by polynomials (poly) and Gaussian (radial) basis functions (rbf), the models manage to capture the nonlinearity of the classification problem yet requiring a higher computational time.
- **Random Forest (RF).** This method fits a number of decision tree classifiers on various subsamples of the learning datasets, providing as output the average among the single trees' predictions. This 'trick', together with the limitation in the number of features for the trees and their depth, aims to control the over-fitting.

By testing the LR, SVM and RF classifiers on the test set for stochastic variations of the environmental conditions, it is observed that, for below rated conditions, their predictions are generally satisfactory, distinguishing successfully the normal operating conditions from the damage status (cp. Table 5). In contrast, substandard performance is obtained for above rated conditions. This could be explained by the higher fluctuation of the tower top acceleration in above rated operating conditions.

In a next step, the algorithms are tested on the subsets of data corresponding to the response of the structure to variation of the TI, according to the curves in Figure 2 (test subsets Te1 to Te4). It is evident from Table 5 that none of the models can perform such generalization. Consequently, a re-training iteration is carried out by updating the training subset by either increasing the amount of data samples considered (varying the D datasets) or changing the amount and/or type of sensors employed (varying the S sensor set up per dataset). For brevity, the results are following reported only for the RF classifier in Table 6. The results of LR and SVM are given in the Appendix.

Table 5. On the left, the performance of the best classifiers for the dataset D0 (trained on Tr0), sensor set up S0. On the right, the colour legend (targeted performances in green).

Classifier	Tr0			Te33			Te1	Te2	Te3	Te4	
	acc	TDR	FDR	acc	TDR	FDR	acc	acc	acc	acc	
Below rated	LR	69%	67%	28%	70%	67%	27%	50%	50%	52%	52%
	SVM (poly)	91%	92%	9%	71%	72%	30%	50%	50%	53%	54%
	RF	100%	100%	0%	86%	86%	14%	55%	68%	66%	72%
Above rated	LR	61%	60%	39%	59%	59%	41%	50%	50%	52%	50%
	SVM (rbf)	89%	86%	9%	64%	66%	37%	50%	50%	52%	50%
	RF	100%	100%	0%	69%	66%	28%	56%	56%	60%	59%

	acc/TDR	FDR
	below 60	above 40
	[60;75)	(30;40]
	[75;90)	(10;30]
	[90;100]	[0;10]

6.2.1. *Varying training datasets.* The algorithms are at first re-trained by adding the data samples corresponding to the varying turbulence levels to the dataset of the design load case (datasets D1, D2 and D3). Acceptable results are achieved for RF classifier only, while LR and SVM do not exceed 60% accuracy. The targeted accuracy and detection rate are generally achieved for prediction on the test set Te33 and the mid-low TI (test set Te4) conditions, when adding either the low-TI or both the extreme-TI data samples to the training set (Tr2 and Tr3, respectively). However, this achievement is associated to a high false alarm rate (up to 45% for above rated conditions).

Table 6. RF classifier performance for the datasets D1, D2 and D3, in combination with the sensor S0, S1, S2 and S3.

Dataset	Sensor	Te33			Te1			Te2			Te3			Te4			
		acc	TDR	FDR	acc	TDR	FDR	acc	TDR	FDR	acc	TDR	FDR	acc	TDR	FDR	
Below rated	D1	S0	85%	86%	14%	-	-	-	63%	50%	40%	69%	62%	26%	72%	67%	36%
	D2	S0	91%	96%	4%	57%	73%	41%	-	-	-	68%	69%	8%	80%	90%	38%
	D3	S0	88%	96%	3%	-	-	-	-	-	-	73%	67%	37%	82%	73%	20%
	D0	S1	96%	96%	4%	66%	73%	41%	80%	69%	8%	76%	90%	38%	84%	71%	3%
	D0	S2	96%	96%	3%	68%	73%	38%	81%	71%	9%	78%	92%	36%	85%	72%	1%
	D0	S3	95%	94%	3%	82%	76%	13%	86%	77%	5%	90%	89%	9%	91%	84%	1%
D2	S3	97%	96%	2%	82%	85%	21%	-	-	-	91%	92%	11%	96%	93%	2%	
Above rated	D1	S0	85%	67%	32%	-	-	-	63%	34%	20%	69%	75%	40%	72%	29%	12%
	D2	S0	91%	73%	19%	57%	94%	84%	-	-	-	68%	92%	68%	80%	79%	38%
	D3	S0	88%	70%	25%	-	-	-	-	-	-	73%	70%	35%	82%	84%	45%
	D0	S1	96%	90%	9%	66%	99%	51%	80%	41%	0%	76%	100%	33%	84%	44%	0%
	D0	S2	96%	91%	8%	68%	98%	51%	81%	34%	0%	78%	99%	26%	85%	39%	0%
	D0	S3	95%	91%	10%	82%	96%	36%	86%	42%	3%	90%	98%	25%	91%	60%	0%
D2	S3	97%	92%	6%	82%	97%	29%	-	-	-	91%	98%	23%	96%	88%	5%	

6.2.2. *Varying sensor sets.* By adding the ten-minute statistics of the time signals from the tower base inclinometer to the SCADA data (sensor set up S1), significant improvements are achieved by RF. Due to the yet high amount of false alarms for TI level above the 90th percentile curve (test sets Te1 and Te2), the classifier is re-trained by adding the statistics of tower base strain gauge signals (sensor set up S2), including the DELs as additional features. While only a slight improvement is recorded again for RF (at below rated condition mainly), it is interesting to observe that LR and linear SVM accomplish exceptionally good results (cp. Table and Table in the Appendix). This is identified as symptom of the instability of the models due to the addition of collinearity into the analysis. The rotation and bending moment signals are, indeed, strongly correlated. The effect of this phenomenon is somewhat reduced in the RF classification thanks to random selection of a reduced number of features at each node. In line with this logic, and by recognizing a generally high correlation of the acceleration signals with the power, the shaft rotational speed and the wind speed (R values above 0.9), a re-training based on the use of the inclinometer statistics instead of the tower top accelerometer is attempted (sensor setup S3). The

results confirm the hypothesis, showing improved performance for the RF, and unsatisfactory results for the LR and SVM classifiers. Finally, by extending the training subset with data from the lowest TI level (Tr2), it is observed that the targeted performance is achieved on all the test sets (Te33, Te1, Te3 and Te4) for the RF classifier, at both below and above rated conditions. It is worth noting that also the SVM classifier achieves generally satisfactory performance for this dataset-sensors combination (D2-S3). In particular, acceptable detection rate and low number of false alarms are achieved on the test sets Te33 and Te4, at below rated conditions, by employing a rbf kernel transformation (cp. Table in the Appendix).

7. Discussion

The confidence of the classifiers in their prediction is discussed based on the reliability curve in Figure 5. The predicted probabilities for the damaged class are divided into bins - along the x-axis. The number of predicted damaged events are then counted for each bin and normalized on the y axis (observed relative frequency). A well calibrated binary classifier should classify the samples such that, for instance, among the samples to which is associated a probability of 0.9, approximately 90% of the cases are classified as damaged. Therefore, the more reliable a forecast is, the closer the points will appear along the main diagonal ("perfectly calibrated"). For points of the curve below the diagonal, the model has over-forecasted the probability, while above the diagonal the probability forecasted is too small.

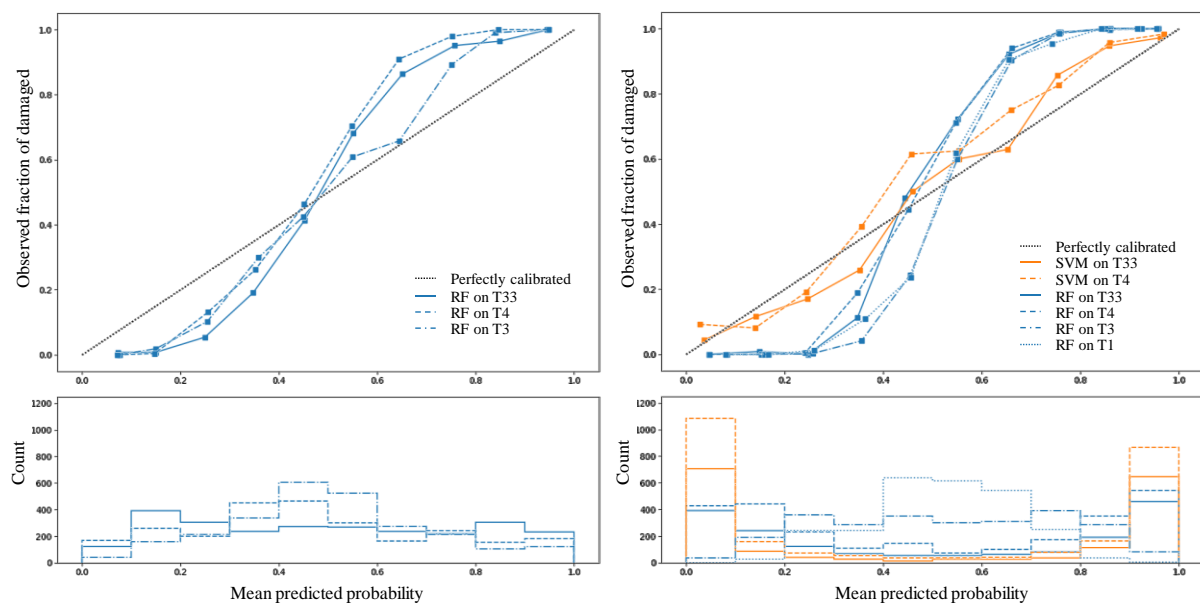


Figure 5. Reliability plots of the predictions from SVM (orange) and RF (blue) classifiers when trained and tested on the dataset-sensor combination D3-S0 (on the left) and D2-S3 (on the right), for below rated conditions.

The reliability curves of the RF classifiers present their typical sigmoid shape (Figure 5) [33]. This means that the algorithm is overconfident on small predicted probability and underconfident for big predicted probabilities. This behavior is common for RF, because the average predictions from the base-level trees can have high variance due to feature sub-setting. A slightly better confidence of the RF prediction is given when trained on D3-S0 and tested on medium-high TI levels (Te3). Applying specific calibration techniques seems necessary for the predictions associated to high-TI levels, where a higher under-confidence of the RF models is shown by the histogram peaks moving further away from 0 and 1 (Figure 5, right). Therefore, although the derivation of the optimal algorithm is not in the scope of this paper, a re-calibration activity is suggested for further study, potentially improving the RF classifiers' performance [33].

When trained/tested on the D2-S3 combination, it is shown in Figure 5 that the RF-based algorithms are able to extend their prediction to all operating conditions, even for significant variation of the TI levels. Although the SVM models do not exhibit as good performance as the ones of the RF classifiers, the rbf-SVM classifier in the combination D2-S3 for below rated conditions is very confident in its predictions (Figure 5, right).

8. Conclusions and future work

The analysis carried out highlights the feasibility of an approach for the indirect monitoring of a structural failure through the low-frequency statistic of the operational data from the offshore wind turbine. Supervised algorithms are trained on ten-minute SCADA data derived from simulating the structural response and power performance associated to the healthy-status jacket structure, and to a disconnection of one of the jacket brace members. The overfitting of the algorithms is controlled by applying a cross validation approach and by extensively testing their performance on subsets for variation of the stochastic representation of wind and wave time histories and turbulence intensity levels.

It is observed that, although the tower top accelerometer can give indications on the presence of the structural damage, its signals are highly affected by variations in the environmental conditions, making the classification activity of the algorithms harder. Acceptable performance in the accuracy and detection rate of tree-based classifiers are obtained mainly for below rated conditions. However, the dataset for the algorithm training must be extended by additional data samples for correctly classifying the structural integrity status through a wide range of turbulence intensity levels. Furthermore, the high number of false alarms recorded can be reduced either by prior information on the turbulence intensity level (installing a specific sensor for this purpose) or better, by replacing the information from the tower top accelerometer with this obtained of an inclinometer positioned at the tower base. In this way, significant improvements are achieved in the detection skills of the random forest classifier at all operating conditions, and of the support vector classifier for the below rated case. This suggests the use of this sensor setup for further analysis of this type of damage.

It should be noted that the classification models developed in this paper depend on either the availability of data associated to the damaged structure, or their simulation through a true digital twin model of the structure [34]. Furthermore, a broad set of isolated ten-minute SCADA data is required for the off-line training and testing of the algorithms, to ensure satisfactory detection performances among the several operating and environmental conditions. The practical value of this approach for the detection of the status of an operational system can be achieved by extending analysis to multiple damaged conditions, either by increasing the number of labels classified, or the number of classifiers. Therefore, future work needs to prove the applicability of this detection method by addressing the following:

- understand the limits of an evaluation based on simulated data and establish the requirements for the integration of real data from the operating turbine.
- investigate the capabilities of unsupervised (or semi-supervised) approaches for the creation of normal behaviour models of the turbine.
- handle the detection of multiple failure modes, failure locations, and their potential simultaneous occurrence.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 745625 (ROME0) ("Romeo Project" 2018). The dissemination of results herein reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains. Furthermore, this work was supported by grant EP/L016303/1 for the University of Strathclyde, Centre for Doctoral Training in Renewable Energy Marine Structures - REMS (<http://www.rems-cdt.ac.uk/>) from the UK Engineering and Physical Sciences Research Council (EPSRC). Finally, the authors would like to thank the reviewers for their thoughtful comments and efforts towards improving the manuscript.

Appendix

The results of the logistic regression (LR) and the support vector machine (SVM) -based algorithms are here presented. For brevity, the performance on the test sets are reported in terms of accuracy (*acc*) only. However, the confusion matrix (*CM*) of each classification test is given to allow deriving the other performance indicators (see Table 4). Overfitting is identified when a strong collinearity is introduced among the classification predictors, thus by implementing both the inclinometer and strain gauge measurements (sensor setup S2). Indicators of this phenomenon are: i) the improvement of the accuracy for both LR and SVM classifiers (cp. Table A and B), ii) the switch to a linear separation function for the SVM classifier, while more complex kernel transformations are employed for the all the other sensor combinations. Finally, the performance of the SVM-based classifiers on the combination D2-S3 are reported in Table C. At below rated conditions, the SVM model implementing a rfb kernel scores satisfactory accuracy (above 90%), detection rate (of about 90%) and alarm rate (below 10%) on the Te33 and Te4 tests set.

Table A. LR classifiers performance on the dataset D0, sensor set up S0, S1, S2 and S3

Sensor	Tr0		Te33		Te1		Te2		Te3		Te4		
	CM	acc	CM	acc	CM	acc	CM	acc	CM	acc	CM	acc	
Below rated	S0	[[622, 242], [288, 576]]	69%	[[314, 118], [143, 289]]	70%	[[8, 1288], [2, 1294]]	50%	[[1296, 0], [1293, 3]]	50%	[[54, 1242], [15, 1281]]	52%	[[1267, 29], [1206, 90]]	52%
	S1	[[630, 234], [245, 619]]	72%	[[308, 124], [136, 296]]	70%	[[12, 1284], [6, 1290]]	50%	[[1294, 2], [1290, 6]]	50%	[[87, 1209], [16, 1280]]	53%	[[1247, 49], [1184, 112]]	52%
	S2	[[864, 0], [0, 864]]	100%	[[432, 0], [0, 432]]	100%	[[1273, 23], [2, 1294]]	99%	[[1255, 41], [122, 1174]]	94%	[[1294, 2], [0, 1296]]	100%	[[1293, 3], [18, 1278]]	99%
	S3	[[538, 326], [332, 532]]	62%	[[266, 166], [175, 257]]	61%	[[327, 969], [163, 1133]]	56%	[[1269, 27], [1192, 104]]	53%	[[435, 861], [227, 1069]]	58%	[[1144, 152], [902, 394]]	59%
Above rated	S0	[[529, 335], [342, 522]]	61%	[[256, 176], [175, 257]]	59%	[[7, 1289], [5, 1291]]	50%	[[1296, 0], [1296, 0]]	50%	[[107, 1189], [48, 1248]]	52%	[[1296, 0], [1296, 0]]	50%
	S1	[[549, 315], [310, 554]]	64%	[[261, 171], [138, 294]]	64%	[[539, 757], [276, 1020]]	60%	[[1295, 1], [1294, 2]]	50%	[[622, 674], [325, 971]]	61%	[[1261, 35], [1120, 176]]	55%
	S2	[[864, 0], [0, 864]]	100%	[[431, 1], [0, 432]]	100%	[[1277, 19], [0, 1296]]	99%	[[1296, 0], [882, 414]]	66%	[[1292, 4], [0, 1296]]	100%	[[1295, 1], [184, 1112]]	93%
	S3	[[538, 326], [334, 530]]	62%	[[262, 170], [146, 286]]	63%	[[18, 1278], [10, 1286]]	50%	[[1296, 0], [1296, 0]]	50%	[[90, 1206], [33, 1263]]	52%	[[1296, 0], [1296, 0]]	50%

Table B. SVM classifiers performance on the dataset D0, sensor set up S0, S1, S2 and S3

Kernel	Sensor	Tr0		Te33		Te1		Te2		Te3		Te4	
		CM	acc	CM	acc	CM	acc	CM	acc	CM	acc	CM	acc
Below rated	poly (4) S0	[[787, 77], [73, 791]]	91%	[[303, 129], [121, 311]]	71%	[[916, 380], [922, 374]]	50%	[[358, 938], [346, 950]]	50%	[[707, 589], [632, 664]]	53%	[[683, 613], [574, 722]]	54%
	poly (2) S1	[[815, 49], [67, 797]]	93%	[[378, 54], [84, 348]]	84%	[[1191, 105], [1140, 156]]	52%	[[1285, 11], [1074, 222]]	58%	[[1107, 189], [995, 301]]	54%	[[1169, 127], [574, 722]]	73%
	linear S2	[[864, 0], [0, 864]]	100%	[[431, 1], [1, 431]]	100%	[[1266, 30], [3, 1293]]	99%	[[1294, 2], [122, 1174]]	95%	[[1287, 9], [1, 1295]]	100%	[[1296, 0], [25, 1271]]	99%
	rbf S3	[[860, 4], [6, 858]]	99%	[[374, 58], [49, 383]]	88%	[[0, 1296], [0, 1296]]	50%	[[0, 1296], [0, 1296]]	50%	[[130, 1166], [84, 1212]]	52%	[[503, 793], [181, 1115]]	62%
Above rated	rbf S0	[[789, 75], [123, 741]]	89%	[[272, 160], [149, 283]]	64%	[[0, 1296], [0, 1296]]	50%	[[0, 1296], [0, 1296]]	50%	[[119, 1177], [61, 1235]]	52%	[[0, 1296], [0, 1296]]	50%
	poly (2) S1	[[740, 124], [126, 738]]	86%	[[328, 104], [117, 315]]	74%	[[1287, 9], [1271, 25]]	51%	[[1296, 0], [1296, 0]]	50%	[[1200, 96], [1030, 266]]	57%	[[1293, 3], [1293, 3]]	50%
	linear S2	[[862, 2], [0, 864]]	100%	[[431, 1], [0, 432]]	100%	[[1294, 2], [2, 1294]]	100%	[[1228, 68], [177, 1119]]	91%	[[1294, 2], [0, 1296]]	100%	[[1285, 11], [16, 1280]]	99%
	rbf S3	[[864, 0], [0, 864]]	100%	[[252, 180], [126, 306]]	65%	[[0, 1296], [0, 1296]]	50%	[[0, 1296], [0, 1296]]	50%	[[76, 1220], [33, 1263]]	52%	[[0, 1296], [0, 1296]]	50%

Table C. SVM classifiers performance on the dataset D2, sensor set up S3

	Te33			Te1			Te2			Te3			Te4		
	acc	TDR	FDR	acc	TDR	FDR	acc	TDR	FDR	acc	TDR	FDR	acc	TDR	FDR
Below rated	94%	93%	6%	53%	42%	36%	-	-	-	64%	65%	37%	91%	87%	4%
Above rated	78%	77%	22%	53%	86%	79%	-	-	-	60%	86%	67%	52%	43%	39%

References

- [1] GL Garrad Hassan, “A guide to UK offshore wind operations and maintenance,” 2013.
- [2] T. Rubert, P. Niewczas, and D. Mcmillan, “Life extension for wind turbine structures and foundations,” in *International Conference on Offshore Renewable Energy 2016*, 2019, no. September 2016.
- [3] M. N. Scheu, L. Tremps, U. Smolka, A. Kolios, and F. Brennan, “A systematic Failure Mode Effects and Criticality Analysis for offshore wind turbine systems towards integrated condition based maintenance strategies,” *Ocean Eng.*, vol. 176, no. October 2018, pp. 118–133, 2019.
- [4] L. Ziegler, N. Cosack, A. Kolios, and M. Muskulus, “Structural monitoring for lifetime extension of offshore wind monopiles: Verification of strain-based load extrapolation algorithm,” *Mar. Struct.*, vol. 66, no. March, pp. 154–163, 2019.
- [5] M. Martinez-Luengo, A. Kolios, and L. Wang, “Structural health monitoring of offshore wind turbines: A review through the Statistical Pattern Recognition Paradigm,” *Renew. Sustain. Energy Rev.*, vol. 64, pp. 91–105, 2016.
- [6] W. Weijtjens, T. Verbelen, G. De Sitter, and C. Devriendt, “Foundation structural health monitoring of an offshore wind turbine—a full-scale case study,” *Struct. Heal. Monit.*, vol. 15, no. 4, pp. 389–402, 2016.
- [7] W.-H. Hu, S. Thöns, R. G. Rohrmann, S. Said, and W. Rücker, “Vibration-based structural health monitoring of a wind turbine system. Part I: Resonance phenomenon,” *Eng. Struct.*, vol. 89, pp. 260–272, 2015.
- [8] W. Weijtjens, T. Verbelen, E. Capello, and C. Devriendt, “Vibration based structural health monitoring of the substructures of five offshore wind turbines,” *Procedia Eng.*, vol. 199, pp. 2294–2299, 2017.
- [9] M. Richmond, U. Smolka, and A. Kolios, “Damage Identification Feasibility through Modelling of Global Dynamics in Offshore Wind Jacket Structures,” *Submitt. to Mar. Struct.*, 2019.
- [10] C. Devriendt, P. J. Jordaens, G. De Sitter, and P. Guillaume, “Damping estimation of an offshore wind turbine on a monopile foundation,” in *European Wind Energy Conference and Exhibition 2012, EWEC 2012*, 2012, vol. 2, pp. 1112–1121.
- [11] J. Tautz-Weinert and S. J. Watson, “Using SCADA data for wind turbine condition monitoring - A review,” *IET Renew. Power Gener.*, vol. 11, no. 4, pp. 382–394, 2017.
- [12] A. Stetco *et al.*, “Machine learning methods for wind turbine condition monitoring: A review,” *Renew. Energy*, vol. 133, pp. 620–635, 2019.
- [13] R. K. Pandit, D. Infield, and A. Kolios, “Comparison of advanced non-parametric models for Wind Turbine Power Curves,” *IET Renew. Power Gener.*, pp. 1–11, 2019.
- [14] B. Puruncajas, Y. Vidal, and C. Tutivén, “Vibration-Response-Only Structural Health Monitoring for Offshore Wind Turbine Jacket Foundations via Convolutional Neural Networks,” *Sensors*, no. 3429, p. 20, 2020.
- [15] M. Martinez-luengo, M. Shafiee, and A. Kolios, “Data management for structural integrity assessment of offshore wind turbine support structures : data cleansing and missing data imputation,” *Ocean Eng.*, vol. 173, no. November 2018, pp. 867–883, 2019.
- [16] U. Smolka and P. W. Cheng, “On the Design of Measurement Campaigns for Fatigue Life Monitoring of Offshore Wind Turbines,” *Twenty-third Int. Offshore Polar Eng. Conf.*, vol. 9, pp. 408–413, 2013.
- [17] N. Noppe, A. Iliopoulos, W. Weijtjens, and C. Devriendt, “Full load estimation of an offshore wind turbine based on SCADA and accelerometer data,” *J. Phys. Conf. Ser.*, vol. 753, no. 7, 2016.
- [18] L. Vera-Tudela and M. Kühn, “Evaluation of a wind turbine fatigue load monitoring system based on standard SCADA signals in different wind farm flow conditions,” *Dewek 2015*, vol. 49, no. 441, 2015.
- [19] L. Vera-Tudela and M. Kühn, “Analysing wind turbine fatigue load prediction: The impact of

- wind farm flow conditions,” in *Renewable Energy*, 2017, vol. 107, pp. 352–360.
- [20] J. Weinert, U. Smolka, B. Schümann, and P. W. Cheng, “Detecting Critical Scour Developments at Monopile Foundations Under Operating Conditions,” *Proc. Eur. Wind Energy Assoc. Annu. Event, EWEA 2015*, pp. 135–139, 2015.
- [21] Ramboll, “ROSAP 5.3 - Ramboll Offshore Structural Analysis Program Package.,” 2019.
- [22] Ramboll, “LACFlex aero-servo-elastic tool v1.9.2.0,” 2019.
- [23] M. B. Nielsen, J. F. Jensen, D. Augustyn, and R. R. Pedersen, “Efficient response recovery procedures for detailed design of jacket foundations,” in *Insights and Innovations in Structural Engineering, Mechanics and Computation - Proceedings of the 6th International Conference on Structural Engineering, Mechanics and Computation, SEMC 2016*, 2016, no. January 2018, pp. 2060–2065.
- [24] *BS EN 61400-3:2009 - Wind turbines - Part 3: Design requirements for offshore wind turbines*. 2009.
- [25] A. Pena Diaz *et al.*, “Offshore vertical wind shear: Final report on NORSEWIND’s work task 3.1.,” 2012.
- [26] *IEC 61400-1 Edition 3 Amendment 1*. 2010.
- [27] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *JMLR*, 2011. [Online]. Available: <http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>.
- [28] “Scikit-learn 0.22 Online Guide - 3.1. Cross-validation: evaluating estimator performance - 3.1.2.2.1. Stratified k-fold.”
- [29] “Scikit-learn 0.22 Online Guide - 3.3. Metrics and scoring: quantifying the quality of predictions - 3.3.2.14. Receiver operating characteristic (ROC).” .
- [30] J. Brownlee, “How and When to Use a Calibrated Classification Model with scikit-learn,” *Last Updated on September 25, 2019*. [Online]. Available: <https://machinelearningmastery.com/calibrated-classification-model-in-scikit-learn/>.
- [31] “Scikit-learn 0.22 Online Guide - 2.5. Decomposing signals in components (matrix factorization problems) - 2.5.1. Principal component analysis (PCA).”
- [32] J. Brownlee, “Logistic Regression for Machine Learning,” *Last Updated on August 12, 2019*. .
- [33] A. Niculescu-Mizil and R. Caruana, “Predicting Good Probabilities With Supervised Learning,” in *Appearing in Proceedings of the 22 nd International Conference on Machine Learning*, 2005.
- [34] U. T. Tygesen, M. S. Jepsen, J. Vestermark, N. Døllerup, and A. Pedersen, “The true digital twin concept for fatigue re-assessment of marine structures,” *Proc. Int. Conf. Offshore Mech. Arct. Eng. - OMAE*, vol. 1, no. June, 2018.