Modelling tower fatigue loads of a wind turbine using data mining techniques on SCADA data

Artur Movsessian[1], Marcel Schedat[2], Torsten Faber[2]

[1]Institute for Infrastructure and Environment, School of Engineering, University of Edinburgh, United Kingdom
[2]Wind Energy Technology Institute, University of Applied Sciences Flensburg, Flensburg, 24943, Germany

Correspondence to: Marcel Schedat (marcel.schedat@hs-flensburg.de)

Abstract. The rapid development of the wind industry in recent decades and the establishment of this technology as a mature and cost-competitive alternative have stressed the need for sophisticated maintenance and monitoring methods. Structural health monitoring has risen as a diagnosis strategy to detect damage or failures in wind turbine structures with the help of measuring sensors. The amount of data recorded by the structural health monitoring system can potentially be used to obtain knowledge about the condition and remaining lifetime of wind turbines. Machine learning techniques provide the opportunity to extract this information, thereby improving the reliability and cost-effectiveness of the wind industry as well. This paper demonstrates modeling damage equivalent loads of the fore-aft bending moments of a wind turbine tower with the advantage of using the neighborhood component analysis as a feature selection technique in comparison to common dimension reduction/feature selection techniques such as correlation analysis, stepwise regression or principal component analysis. For this study a one-year measuring period of data was gathered, pre-processed, and filtered by different operational modes, namely stand still, full load, and partial load. Finally, a sensitivity analysis was performed in the partial load model to determine the required length of the data collection campaign that guarantees the most precise results. The results indicate that applying neighborhood component analysis yields more conservative models regarding the number of features and equally accurate outcomes than traditional feature selection techniques.

1   Introduction

Wind power is becoming the electricity-generation technology with the lowest costs in several areas of the world (REN21, 2018). Given the fierce competition in the industry due to the large deployment of wind turbines (WTs), it is important to continuously look for alternatives to make this technology more cost-effective. The possibility of monitoring with sensors or sensor systems has enabled gathering and supervising data about an object's condition to detect for example failures. Particularly, structural health monitoring (SHM) at WTs allows to monitor the structural behaviour and stresses of structures such as blades, towers, and foundations.

SHM systems could be used to verify structural safety and determine the remaining useful lifetime (RUL) of WTs (Schedat et al., 2016). Moreover, information gathered through SHM during the lifetime of WTs can be potentially used to identify

structural weaknesses and feed this information back to the manufacturers, ultimately improving the design of new turbines (Ziegler et al., 2018). Another potential benefit of SHM is a decrease in maintenance costs. Typically, operation and maintenance costs (including both fixed and variable costs) represent approximately 20-25 % of the total levelized cost of electricity (LCOE) (IRENA, 2015). SHM could reduce this share by allowing the implementation and set in place of more

5   efficient maintenance practices such as predictive maintenance while enabling a better spare-parts inventory management. Consequently, downtime is reduced, and production is increased.

Currently, the assessment and evaluation of the structural condition of WTs without a load measurement system can be challenging. Particularly, the estimation of fatigue loads can be frequently sophisticated due to a lack of information (Melsheimer et al., 2015; Schedat and Faber, 2017). Therefore, exploring the ways to mine data from SHM systems and extract

10   valuable information becomes an interesting and briefly visited field of research.

The reconstruction or estimation of loads using statistics from SCADA data were already presented and tested in the mid-2000s. Cosack and Kühn (Cosack and Kühn, 2006) developed a stepwise regression model for estimating the rotor thrust. Despite the good results (i.e. deviations between the calculated and the estimated loads ranged from 5.4% - 7.3% in the worst case), the presented model was too complex and time-consuming with further restrictions. In a new development of the model,

15   an estimation method for the corresponding target values (damage-equivalent loads and the load magnitude distributions) used neural networks (Cosack, 2010; Cosack and Kühn, 2007).

While machine learning techniques are widely applied in industries such as the automotive, information technology and communication, the wind industry is starting to explore the suitability of these promising methods for its purposes. Although data-driven attempts have been made to estimate the loads acting on the turbine using available information from the SCADA

20   system, there is no consensus yet on the type of a relationship existent between these data and actual load measurements. In the last years, the focus on this topic increased. This section aims to review available scientific literature regarding modeling loads with existing SCADA data.

Ziegler et al. (Ziegler et al., 2018) have recently performed a literature review and have assessed the development of the lifetime-extension market of onshore WTs. The alternative of extending the lifetime of a WT, as opposed to repowering or

25   decommissioning, is appealing given the potential increase of returns on investments (ROIs), however, not much research has been done on this matter. The authors contributed, then, by comparing updated load simulation and inspections for lifetime extension assessment in Germany, Spain, Denmark, and the United Kingdom. Particularly, for lifetime extension to be a feasible alternative, the structural integrity of the turbine should not compromise the level of safety. In this regard, the survey performed by the authors determined that, apart from the use of SCADA systems, no short-term load measurements or

30   monitoring are carried out in the countries surveyed (a few exceptions were identified in the UK, where load reassessment is performed). They found that most interviewees focus on practical assessments for cost reasons. Nevertheless, these practical inspections are no guarantee that the safety level can be maintained during the lifetime extension. The authors concluded that new O&M strategies and data-processing methodologies are necessary for lifetime extension purposes. Moreover, data-driven approaches may contribute to cost reduction of lifetime extension assessments.

In line with the findings of Ziegler et al. (Ziegler et al., 2018), other authors have worked on the aforementioned data-driven approaches. Noppe et al. (Noppe et al., 2018), for example, reconstructed the thrust loads history of a WT based on both simulated and measured SCADA data. The data gathered corresponded to operational 1 s and 10 min data for 2.5 months. Moreover, the data is segregated into different operational modes. The selection of explanatory variables that the authors

5   performed was based on a Pearson correlation analysis. The thrust loads were modeled using neural networks. The model has the following input features: wind speed, blade pitch angle, rotor speed, and generated power. The results of this paper showed that the constructed model was able to estimate thrust loads with a relative error that does not exceed 15 %. The authors also concluded that the use of simulated data yielded slightly better results and that adjustments in the hyperparameters of the neural networks had no significant impact on the estimated thrust loads.

10   Relatedly, Vera-Tudela and Kühn (Vera-Tudela and Kühn, 2014) focused on the selection of variables to be used for fatigue load monitoring and attempted to define an optimum set of explanatory variables for that purpose. The authors identified 117 potential variables (13 statistics of 9 SCADA signals) used in related scientific literature. Among them, the mean of generator speed, electrical power, and pitch angle have been the most commonly used. The authors decided to apply several feature selection methods to six sets of variables. The methods chosen included Spearman coefficients, stepwise regression, cross-

15   correlation, hierarchical clustering, and principal components. To evaluate the outcomes of the feature selection methods a feedforward neural network was employed. The authors concluded that principal components yielded the best set of variables, however the resulting set lost expertise knowledge about the relation between the variables. In this sense, a combination of ranking the variables by their corresponding Spearman coefficients resulted in a fair compromise between the number of features required to monitor the damage equivalent load for blade out of plane bending moment and the available expertise

20   knowledge.

Smolka and Cheng (Smolka and Cheng, 2013) examined the amount and type of data necessary to determine a fatigue estimator for the operational lifetime of a WT. The inputs for the neural network are selected by means of a correlation analysis applied to standard data statistics of available SCADA signals such as electrical power, generator speed, pitch angle, among others. The authors concluded that the minimum training data sample size required is approximately half a month worth of

25   measurements.

Seifert et al. (Seifert et al., 2017), acknowledging the complexity and cost of handling extra measurements, assessed the minimum needed size of a training sample to predict fatigue loads using 10 min statistics of SCADA signals and neural networks. In a sense, Seifert et al.'s work is an extension or continuation of Vera-Tudela and Kühn's (Vera-Tudela and Kühn, 2014) and Smolka et al.'s (Smolka and Cheng, 2013). Seifert et al. (Seifert et al., 2017) tested different sample sizes, like a k-

30   fold cross-validation, varying between one day (i.e., 144 records) and four months (i.e., 4032 records) of measurements. They determined that a sample of 2016 records of 10 min statistics are sufficient to predict flap wise blade root bending moments of a WT independent of seasonal effects.

Artificial neural networks can only perform as good as the information provided to them, thus the features used to train them are key to obtain high accuracy in the results with a parsimonious model. Scarce research has been done regarding feature

selection for modelling tower fatigue loads. The available literature has focused on techniques such as correlation analysis, principal component analysis (PCA) and stepwise regression to select the best subset of information. This paper aims to assess the use of Neighbourhood Component Analysis (NCA) as a feature selection technique to extract relevant information from SCADA data in order to train artificial neural networks and model fatigue loads.

5   The paper is organized as follows: section 2 outlines the applied methodology in this study, section 3 summarizes the results, and, finally, section 4 presents the conclusions derived from the obtained results.

2   Data and Methodology

2.1   Wind turbine and SCADA data

This paper seeks to model tower fatigue loads of a real WT located in Schleswig-Holstein, North Germany. The turbine is used
10   by the Wind Energy Technology Institute at the Flensburg University of Applied Sciences for research purposes. For this study, the readings from the SCADA and strain gauge sensors in the previously mentioned turbine were recorded over one year and collected in 10 min files. The sensors used to extract features for the model are described in Table 1 and were selected based on the literature review and consultations with an application engineer.

Table 1 – Description of SCADA sensors selected

Feature name	Description	Unit of measurement	Frequency [Hz]
Explanatory variables			
Omega	Rotational speed at the rotor	rpm	20
acc_x	Acceleration fore-aft (x-direction)	mm s^{-2}	20
acc_y	Acceleration side-side (y-direction)	mm s^{-2}	20
v_wind	Wind speed	m s^{-2}	20
v_dir	Relative wind direction	degree	10
omega_gen	Rotational speed at the generator	rpm	20
air_density	Air density	kg m^{-3}	20
Pitch	pitch angle	degree	20
ACpow	Active power output	kW	20
Dependent variables			
Bieg1_060_240	Gauge sensor located at 60 & 240 degree inside the tower bottom	kNm	50
Bieg2_150_330	Gauge sensor located at 150 & 330 degree inside the tower bottom	kNm	50

The strain gauges measurements at the turbine were transformed into a resultant fore-aft tower bending moment, which was later used to calculate the short-term damage equivalent load (DEL) for every 10 min time series. This transformation was performed by means of a rainflow counting algorithm and later, the resulting load spectrum was further reduced to a constant

5   load range. After a number of equivalent cycles, this load range results in the same equivalent accumulated damage as the spectrum of loads previously calculated through the rainflow counting algorithm. The short term DELs were calculated following Equation (1):

$$S_0 = \left[\frac{\sum_i n_i * S_i^m}{n_{eq}} \right]^{\frac{1}{m}} \tag{1}$$

where n_{eq} is the equivalent number of cycles, S_i the different load ranges, n_i the corresponding cycle numbers and m is given

10   by the slope of the Stress-Cycle (S-N) curve of the material used for the tower (DNV/Risø, 2002). In this case, it was assumed that $m = 3$. The DELs were then used as the dependent variable of the model.

2.2   Methods

The main methodology used for the development of this paper is graphically described in Figure 1. First, the sensors which provide relevant information to model resultant fore-aft tower bending moments were selected (see Table 1). In the next step,

15   the resulting record was analysed for missing data (e.g. zero values) and outliers. There were periods where the turbine was out of service or measurement failures with no registered data. Subsequently, affected records have been removed. To determine the relationship between the dependent and explanatory variables described previously, each of the 10 min files was summarized by estimating the following descriptive statistics for every explanatory variable: i) minimum value, ii) maximum value, iii) arithmetic mean, iv) range, v) mode, vi) standard deviation, vii) variance.
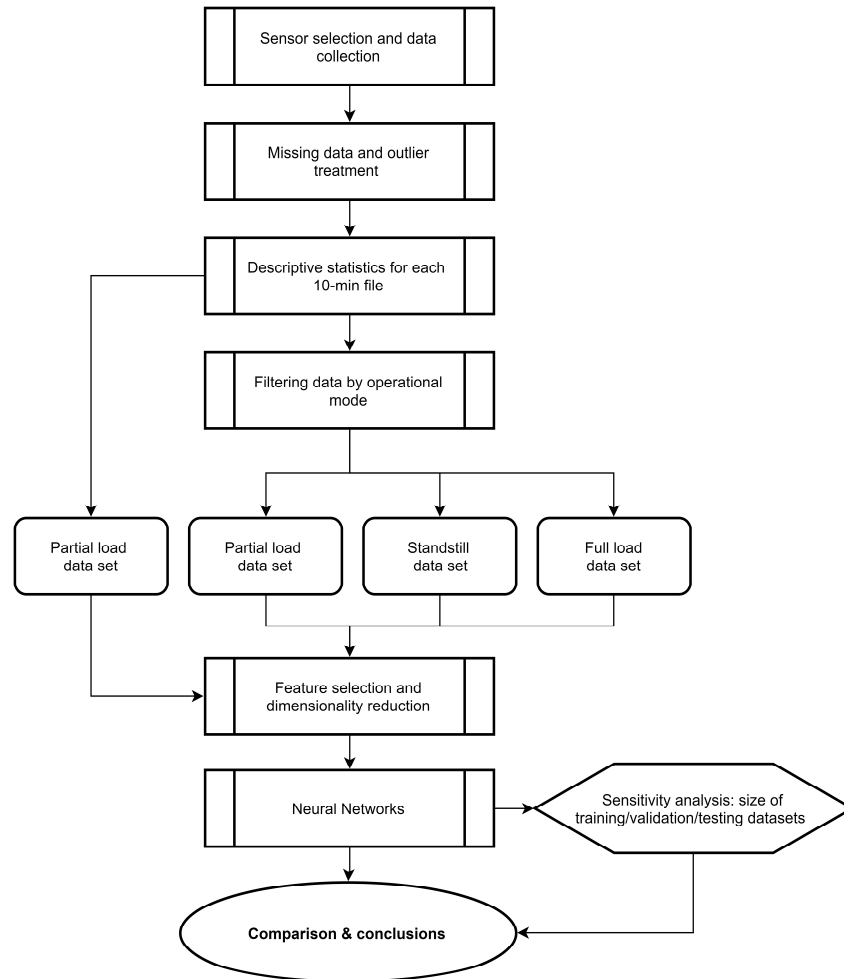
Figure 1: Main methodology steps

In this way, the dataset was reduced to 63 features or explanatory variables. Excluding the time where no SCADA data was recorded, the total amount of data results in 36266 observations (corresponds to nearly 6044 hours).

5   Furthermore, for the sensitivity analysis, the data was filtered by operational modes, namely standstill, partial load, and full load. This was done by means of the feature "ACpow" which refers to Active power output. In this sense, standstill corresponds to "ACpow" readings below or equal to 5 kW; partial load to readings higher than 5 and below or equal to 2000 kW; and full load to readings above 2000 kW.

Research by Sharma and Saroha (Sharma and Saroha, 2015) concluded that a reduction of dimensions possibly leads to a better
10   performance of the mining algorithms while maintaining a good accuracy, therefore, it is important to eliminate potential redundant data and select the variables with more predictive power for the model. For this, four different feature selection and dimension reduction techniques were applied to the entire dataset and the datasets resulting from filtering the data by operational mode.

These techniques include Pearson correlation, stepwise regression, NCA, and PCA. Pearson correlation measures the linear correlation between two variables and maps the result to an interval between -1 and 1, where 0 indicates no linear relationship (Boslaugh and Watters, 2008). It can be calculated as per Eq. (2):

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{2}$$

5   where n is the sample size, X_i and Y_i are the observations with index i and $\bar{X}$ represents the arithmetic mean of all the sample. A threshold value of 0.5 was set to define the level of correlation. In this sense, a correlation coefficient between 0 and 0.49 is weak and a correlation coefficient between 0.5 and 0.95 is strong. Correlation coefficients above 0.95 may indicate redundancy in the dataset and lead to lower model accuracy.

Stepwise regression is an iterative method where features are added and removed from a multilinear model based on their
10   statistical significance in the regression (Draper and Smith, 1998). The algorithm begins by constructing an initial model with one feature (forward selection) or all the features (backward selection) and continues adding or removing features by comparing the explanatory power of the larger or smaller models. At each step, the p-value of the corresponding F-statistic is estimated and compared to a threshold p-value to decide which features are included in or excluded from the model. The algorithm repeats this process until the added feature do not improve the model anymore or until all features that do not improve
15   the explanatory power of the model are removed. This method is considered to be locally optimal, yet not globally optimal given that the selection of features included in the initial model is subjective and there is no guarantee that a different initial model will not lead to a better fit.

NCA is a non-parametric classification model used for metric learning and linear dimensionality reduction (Goldberger et al., 2005). It is based on a modelling technique known as k-Nearest Neighbours (k-NN), which is a supervised learning algorithm
20   used for classification or regressions (Han and Kamber, 2006; Parsian, 2015). In its simplest form, the k-NN approach looks for the closest k = 1 observation to the query observation x_q within the training dataset by measuring the distances to the neighbouring data points and selecting the one that satisfies $\min_i$ distance (x_i, x_q). The output is then predicted by applying a function y = h(x). In a multidimensional dataset, the k-NN approach requires to differ between the "relevance" of the explanatory variables for the intended output. For this purpose, different weights can be assigned to the features of the model
25   using the "Scales Euclidian Distance" estimation detailed in Eq. (3):

$$Distance(x_i, x_q) = \sqrt{a_1\big(x_i[1] - x_q[1]\big)^2 + \cdots + a_d\big(x_i[d] - x_q[d]\big)^2} \tag{3}$$

where x_i is a vector of input values, x_q is the query vector, a is the scaling number that defines the relevance of each explanatory value, and d the total number of observations. Other distance metrics can be used, namely Mahalanobis, Manhattan, rank-based, correlation-based, and Hamming (Hazewinkel, 1994).

30   Lastly, principal component analysis (PCA) is a statistical method to reduce the dimensions of a dataset that presumably contains a large number of irrelevant features while retaining the maximum information possible (Vidal et al., 2016). This is done by transforming the original set of multidimensional data into a new set referred to as components by means of

eigenvectors and eigenvalues. A pair of eigenvector and eigenvalue indicate respectively the direction and how much variance is there in the data in that direction. The eigenvector with the highest eigenvalue is the principal component. In this sense, the transformation allows to reduce the dimensions of the dataset to a few components with relatively low loss of information.

In this way, 16 neural networks (NN) were developed corresponding to four datasets (all operational modes, standstill, partial

5   load, and full load) and four feature selection and dimension reduction techniques. Each dataset is divided into training, validation, and testing subsets. 70 % of a dataset is randomly chosen and used by NN for training the model, i.e. this subset is used to adjust the model by means of the mean square error (MSE). This adjustment stops when the MSE does not significantly improve. The validation subset is used as a measure to avoid overfitting the NN and generalizes the transfer function so that the model is applicable to new datasets. The test subset has no effect on training or validation, it is only used to measure the

10 performance of the trained NN.

The NN models used in this paper are trained with the Neural Network Toolbox from MATLAB (MathWorks, 2019). The standard settings consist of a two-layers feed-forward NN with a sigmoid transfer function in the hidden layer and a linear transfer function in the output layer. The number of hidden neurons in the hidden layer are kept, as suggested by MATLAB, at 10 neurons. The Levenberg-Marquardt algorithm is selected as the training algorithm. No changes have been made to the

15 standard configurations suggested by MATLAB.

The results from the 16 models were compared to derive conclusions about the relationship between operational data and tower fatigue loads acting on WTs.

Finally, a sensitivity analysis was developed by varying the sizes of datasets for training, validation, and testing in the partial load model to determine the required length of the data collection campaign necessary to achieve accurate results efficiently.

20 **3   Results and discussion**

This section describes the sensors identified by different methods as potential predictors of tower fatigue loads of the WT. Additionally, it assesses the outcomes of the models used to predict the fatigue loads. Finally, it presents the results of the sensitivity analysis performed that sought to determine the length of the data collection campaign required to appropriately predict the fatigue loads.

25 **3.1   Feature selection and dimension reduction**

Before building a model to predict the desired output, it is important to define which variables could act as predictors. The feature selection methods described in section 2.2 were applied to four different datasets: i) a complete one-year dataset, ii) a full-load dataset, iii) a partial load dataset, and iv) a standstill dataset. The results of the feature selection methods are described below.

3.1.1 Complete dataset: one-year data

A Pearson correlation analysis was applied to the pre-selected features for predicting the DELs of the fore-aft bending moment of the tower finding that only 25 of the 63 features are correlated and used as independent variables in the model. On the one side, the results indicate that most of the descriptive statistics of the acceleration meter in both directions (i.e. x and y-axis) are

5   the most highly correlated features with the DELs. As a matter of fact, the standard deviation of the acceleration in the x-direction presents the highest correlation with a coefficient of 0.97, depicting an almost linear relationship between this feature and the dependent variable. The maximum occurring windspeed, and both the average and maximum power output are also highly correlated with the DELs. These results suggest that these features fluctuate together with the dependent variable, and they could potentially be used to build a model that can estimate the DELs of the fore-aft bending moment of the tower without

10   installing strain gauges sensors.

On the other side, the results show that air density and relative wind direction, along with all their corresponding descriptive statistics, have a very low correlation with the DELs and should be, therefore, disregarded in the model based on this feature-selection technique. The mean wind direction, in particular, has a correlation coefficient close to zero, indicating an insignificant linear relationship with the DELs. Since the DELs are calculated for the thrust loads, the highest observed value

15   for this variable occurs when the turbine is directly facing the wind and the lowest when the wind is not affecting the thrust load directly.

Before using the features with the strongest correlation in a model, it is necessary to check for collinearity, i.e., correlation between independent variables. By observing the correlation between the explanatory variables, it can be determined which variables are highly correlated with each other and, therefore, should be excluded from the model to avoid collinearity issues.

20   From this analysis, it was determined that rotational speed at the rotor should be excluded from the model and only rotational speed at the generator should be included given that these two features are a factor away from each other and, thus, may add bias to the model due to redundancy. Many of the remaining variables are also highly correlated with each other, nevertheless they add potentially valuable information to the model. An alternative would be the use of a method such as PCA which could contribute to avoid multicollinearity by transforming the data while maintaining the information contained in it.

25   After removing the rotational speed at the rotor and eliminating all variables with a correlation below 0.5, there are 21 features remaining that could be included in the model. This data was transformed as explained in section 2.2 estimating the variance explained by each of the first components as seen in Figure 2. It can be observed that 99 % of the information contained in the features is now stored in the first nine components. The remaining 13 components explain less than 1 % of the cumulative variance. A model could be built using the first nine components and the results should be almost as accurate as using the 21

30   features selected after the correlation analysis. The biggest disadvantage with this method is that given the transformation of the data, it is no longer possible to interpret it. The results, nevertheless, remain interpretable and are free of the influence of multicollinearity.
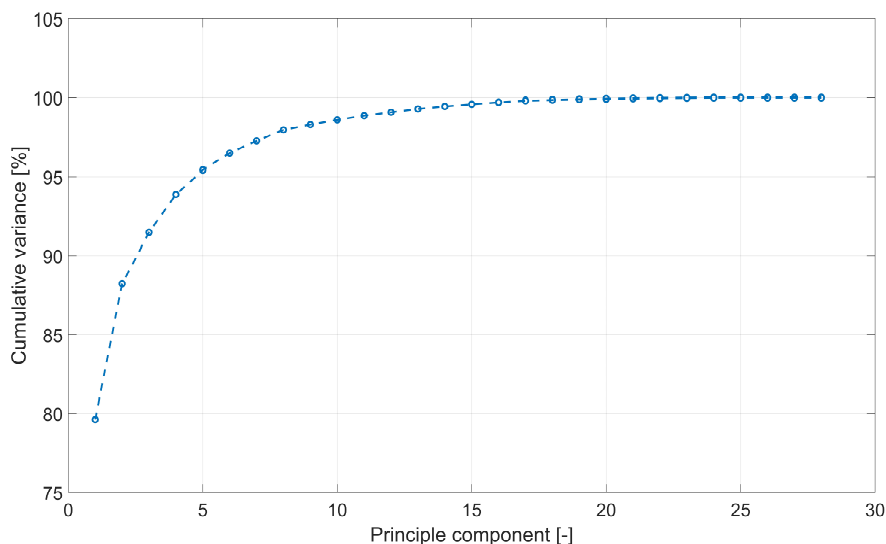
Figure 2: Cumulative variance explained by principal components

Alternatively, an interactive stepwise regression was built using the pre-selected 63 features. Different combinations of features were tested to identify those that should not be included in the model given that they do not contribute to the predictive power

5   or they result in an increase in the error of the model. The features with a p-value above 0.1 should be omitted from the model. Unlike the Pearson correlation analysis, this method avoids multicollinearity among the features. The results suggest excluding a total of 26 variables from the regression model. Among these can be found the minimum, maximum, mean, and range of the rotational speed at the generator, most descriptive statistics of air density, with the exception of the standard deviation, and range, mode, and standard deviation of the acceleration in y-direction.

10   It is important to highlight that the variable that represents the range of the acceleration sensor in the x-direction was identified as statistically insignificant despite its high correlation with DELs. As mentioned earlier, the possible models explored with the stepwise regression are limited. The algorithm builds different models from the 63 features depending on the order in which these features are added to (in the case of forward selection) or removed from (in the case of backward elimination) the models. In this sense, the range of the sensor "acc_x" and the variance of "acc_x", which are correlated with a factor of 0.90, could be

15   considered mutually exclusive. The decision as to which of these variables to include in the model would depend solely on which variable is added or removed first in the stepwise regression. In this case, the algorithm suggests excluding the range of "acc_x", a highly correlated feature, based on the search for the local minimum instead of evaluating all combinations. Ultimately, this method identified 37 features as statistically significant and, thus, these should be included in the model.

The last feature selection method, NCA, was applied as well to the initial 63 features dataset. Ten features were identified by

20   this method as relevant for the prediction of DELs, a significantly smaller number than those selected by applying the correlation analysis and stepwise regression.

To summarize, mean values and standard deviations are the descriptive statistics that can best describe the data according to the three feature selection methods applied. Features selected by all three methods include windspeed, acceleration, and power output.

3.1.2 Data filtered by operational modes

5   The dataset was divided by operational modes resulting in 21.6 % of the data corresponding to standstill, 7.8 % to full load, and 70.6 % to partial load. Each dataset contains 56 features and the corresponding DEL. Unlike with the entire dataset, the rotational speed at the rotor was removed due to redundancy with the rotational speed at the generator.

The first feature selection method used in these new datasets is the again the Pearson correlation analysis. The results show that most of the descriptive statistics for wind speed and acceleration are highly correlated with the DELs in all operation

10   modes. The first differences appear in the generator speed. As expected, the generator speed is not relevant during stand still since the rotor is not moving or is only idling. The mean wind speed is not as relevant in full load as it is in partial load. During full load, the rotational speed is around a specified number and must be kept as stable as possible. Therefore, the mean rotational speed does not change significantly during full load. During partial load, the mean rotational speed is within a higher range, therefore it has a higher correlation with the corresponding DEL.

15   During full load, the standard deviation and variance of the rotational speed are highly correlated with the DELs. The standard deviation explains how the values differ from the mean, thus conclusions about the dynamics of the turbine can be derived based on these spreads. For example, a large deviation in the rotational speed around the mean during full load has a significant effect of the tower movement. The features of the pitch angle are correlated with a factor greater than 0.5 with the output. This correlation is only significant during full load. The pitch angle is held at the most efficient lift-to-drag ratio during the partial

20   load and, therefore, not many variations can be observed during standstill and partial load. During full load, the turbine pitches constantly to keep the rotational speed nearly constant.

The second features selection method applied is a stepwise regression. The results are not consistent with the correlation analysis. Air density and wind direction had no correlation with the DELs, however, they were chosen by the stepwise regression during standstill and partial load as potential predictors. Also, the pitch was chosen as a significant variable during

25   standstill, even though the turbine is not pitching. In general, the modeler needs to be careful when interpreting the results from a stepwise regression as described in section 2.2.

Lastly, NCA was applied to the three datasets. Examining the results, one significant difference to the correlation analysis is that the wind direction was identified as significant during standstill and partial load by the NCA, whereas the correlation analysis showed no correlation of these features with the output during any operational mode. Furthermore, the range of the

30   pitch angle was identified as relevant during partial load, which was not the case in the correlation analysis. Mean acceleration in the x-direction and the standard deviation were the only two features identified as significant by the NCA in all three operational modes.

3.2 Modeling fatigue loads

Once the features with predictive power have been identified for the different datasets, NN are built to evaluate the predictions. The outcomes of these models are described hereafter.

3.2.1 One-year data

5 The NN stopped its training at epoch (iteration) 30 with the best performance at a mean squared error of 152758 kNm (see Figure 3). The performance plot shows how the mean square error drops rapidly as the NN learns. Six more iterations were made to confirm that the value does not change significantly.
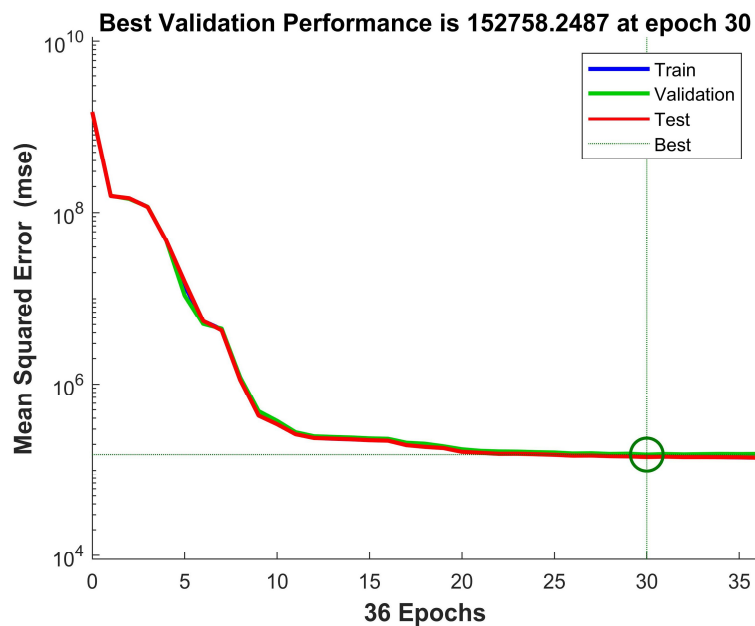
Figure 3 – Neural Networks performance validation

10 From the error histogram in Figure 4, it can be observed that most errors lay in the interval from -1000 to 1000. Approximately 23567 of the data points (65 % of the total) have an error between -500 and 500. Data points close to -3500 kNm and 4500 kNm differ significantly from the mean and occur less frequently, which could indicate the presence of outliers in the data used for modeling. These potential outliers can also be observed in the regression plot in Figure 5, where some values are clearly distanced from the regression line. The regression plot shows the calculated DELs from the strain gauges measurements

15 with respect to the trained DELs obtained from the NN. The correlation "R" is significantly high for training, validation, and test data. The R on the testing data is 0.99486 and higher in the cases of training and validation data, which indicates a nearly perfect positive linear correspondence. All values need to lay on the 45-degree line to have a perfect fit. This can only happen if the network output equals the target values.

The calculation for mean error in percent following Eq. (4):

$$Error\ [\%] = \frac{Target_{value} - Modeled_{value}}{Target_{value}} * 100 \qquad (4)$$

Applying the Eq. (4) and dividing by the number of observations, a mean error of 9.45 % and a maximum error of 526 % can be calculated. The maximum error indicates a possible outlier in the dataset that confirms the errors observed in the histogram.
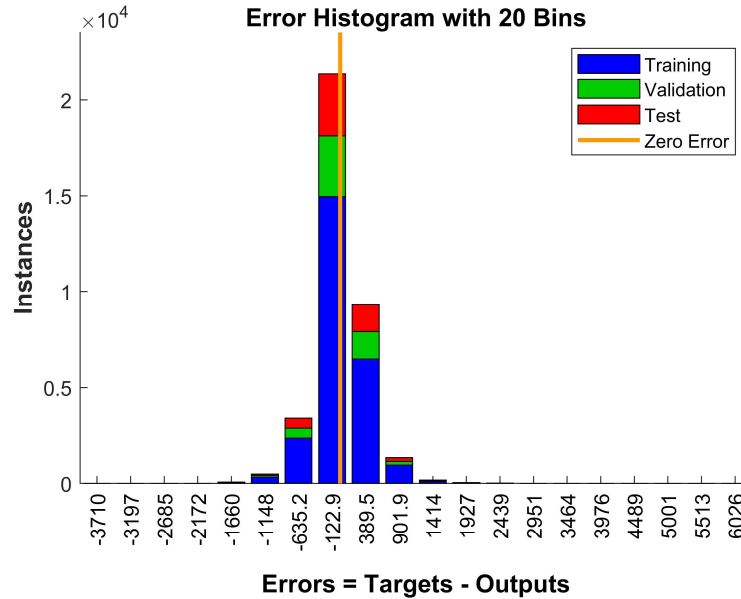
Figure 4 – Neural Networks error histogram

5   Table 2 summarizes the results of neural networks built from the different predictor sets.

Table 2 - Summary of results from Neural Networks for the complete one-year dataset and data subsets

N°	Subset	No. of features	R			Iterations (epoch)	Max Error [%]	Mean abs. Error [%]	Mean abs. Error [kNm]
			Training	Validation	Test				
1	Correlation	31	0.99584	0.99554	0.99536	30	526	9.45	237
2	Correlation & PCA	9	0.99437	0.99379	0.99405	83	444	10.85	276
3	Stepwise	38	0.99632	0.99569	0.99556	28	1134	10.38	224
4	NCA	13	0.99581	0.99593	0.99568	49	980	10.49	228
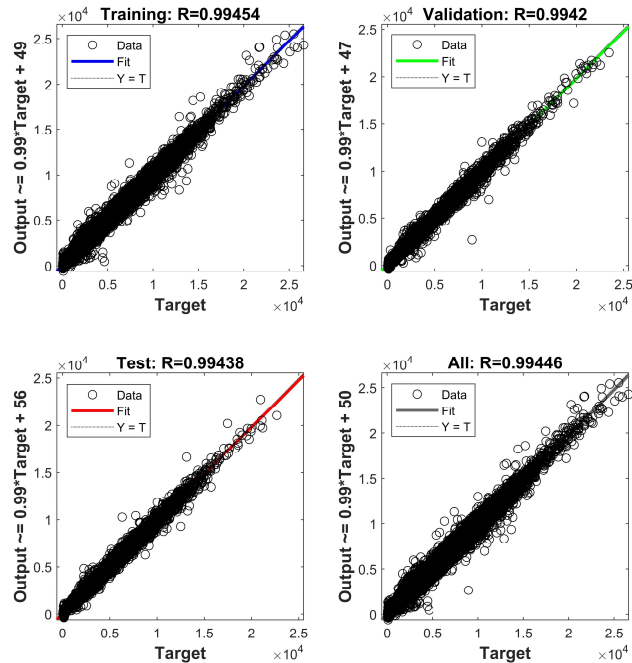
Figure 5 – Linear regressions between the Neural Networks prediction and the DELs

Comparing the different subsets trained with the same settings in the NN, it can observe that the correlation "R" value is high in all cases. More significant differences can be observed in the maximum and mean error. It is presumed that outliers lead to

5 a 1000 % maximum error in the case of the stepwise regression subset and maximum errors above 400 % in the remaining models. The mean absolute error of approximately 10 % shows that most values do not present high errors such as 1000 %, therefore, it can be concluded that the general performance of the model is acceptable and can be improved through outlier treatment and data smoothening.

3.2.2 Data filtered by operational modes

10 This section explores the performance of the NN when built with data subsets from different operational modes. It can observe that the mean error in percent is significantly high in standstill compared to other operational modes. Nevertheless, the mean absolute error in kNm is the lowest. The high maximum error observed previously when using the complete one-year dataset (i.e., when using "all operational modes") for the different training sets could be explained by the poor predictive power of the data from the standstill mode. When the NN was built using filtered data for partial and full load, the errors of the predictions

15 decreased significantly. Thus, the data from the standstill mode adds noise and reduces the accuracy of the model.

Figure 6 shows that DELs estimations during standstill were significantly lower compared to those during full load (DELs in Figure 6 (b) are around three times lower than during full load Figure 6 (a)). Therefore, an error of over 20 % in standstill mode is equivalent to a 7 % error in full load. This, nevertheless, does not justify the 20 % error seen during standstill mode.

The NN performs worse when using data from the standstill mode than when using data from the other operational modes. Presumably, the small variations observed in the readings from the sensors during standstill do not provide sufficient information to predict the DEL. This is consistent with the R values of the model during standstill mode. These values are the lowest among the different operational modes. A more detailed look at this case would be necessary to derive valuable insight.

(a)

(b)
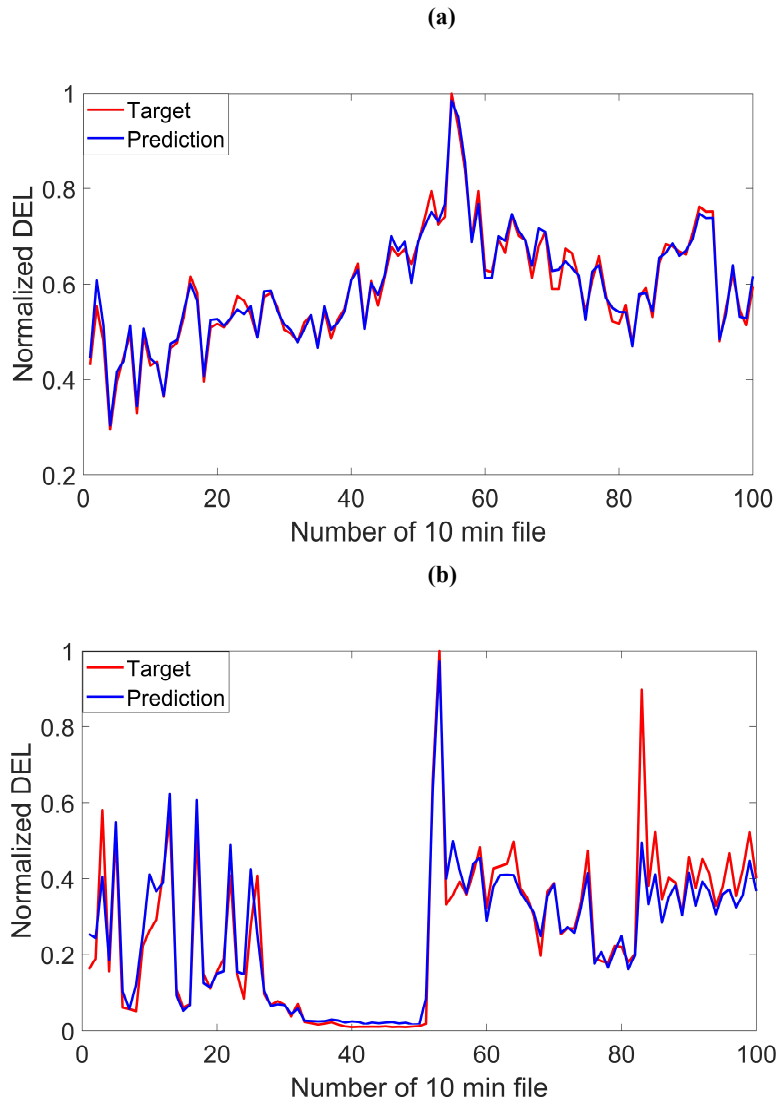
Figure 6 – Comparison of target vs. prediction in full load (a) and standstill (b) mode

5   Moreover, Table 3 shows that those models constructed using smaller sets of features derived from the application of methods such as PCA or NCA have approximately the same predictive power as those models constructed using larger sets of features derived from applying methods such as stepwise regression or correlation analysis. This can be observed from the comparison of the measures of goodness-of-fit (i.e. R^2 values) among the models. It is important to highlight that this observation is valid

for all operating modes. Thus, the application of feature-selection and dimension-reduction methods can be considered a good practice.

Table 3 – Summary of results from Neural Networks for different operational modes and data subsets

N°	Subset	No. of features	R Training	R Valida-tion	R Test	Iterations (epoch)	Max Error [%]	Mean abs. Error [%]	Mean abs. Error [kNm]	R^2
				Standstill						
1.1	Correlation	18	0.97029	0.95458	0.96413	17	477	26.54	189	0.935
1.2	Correlation & PCA	7	0.96498	0.9623	0.95612	26	350	21.80	199	0.928
1.3	Stepwise	35	0.98551	0.9809	0.97639	34	1159	22.63	138	0.967
1.4	NCA	16	0.97957	0.98052	0.97744	21	600	24.64	145	0.959
					Partial Load					
2.1	Correlation	28	0.99297	0.99233	0.99201	19	63	6.65	242	0.985
2.2	Correlation & PCA	12	0.99169	0.99046	0.99108	43	84	6.97	256	0.983
2.3	Stepwise	28	0.9932	0.99211	0.99202	18	69	6.57	239	0.986
2.4	NCA	11	0.99277	0.99236	0.99216	38	82	6.59	240	0.985
					Full Load					
3.1	Correlation	28	0.99124	0.98997	0.98899	13	52	2.27	276	0.981
3.2	Correlation & PCA	12	0.99021	0.98793	0.98775	25	55	2.37	290	0.979
3.3	Stepwise	23	0.99169	0.99128	0.98713	16	58	2.24	272	0.982
3.4	NCA	11	0.99208	0.98754	0.98371	26	30	2.24	273	0.980

3.3 Sensitivity analysis

In this section, the results of the sensitivity analysis are presented. The aim is to identify if and how the error in the outcomes of the model varied by reducing the size of the shares of observations used for training and validation of the NN. Since the majority (70.6 %) of the data gathered corresponds to partial load mode, this subset was used for the sensitivity analysis.

The results from section 3.2.2 indicated that the errors did not differ significantly among the different sets of predictors used to predict the DELs during partial load, therefore the set of predictors obtained from the stepwise regression application were used for this sensitivity analysis.

Figure 7 shows that both the absolute error share and the root mean squared error decrease as the share of data used for training and validation increases. Nevertheless, this behavior stabilizes when using approximately 40 % of the data for training and validation. There is, then, a cost-reduction potential. Instead of investing time and effort in collecting large amounts of data over long periods of time, the same results can be obtained with a smaller dataset. In this case, 10241 observations were

sufficient to yield an error that does not decrease significantly when increasing the share of training and validation data in the NN.
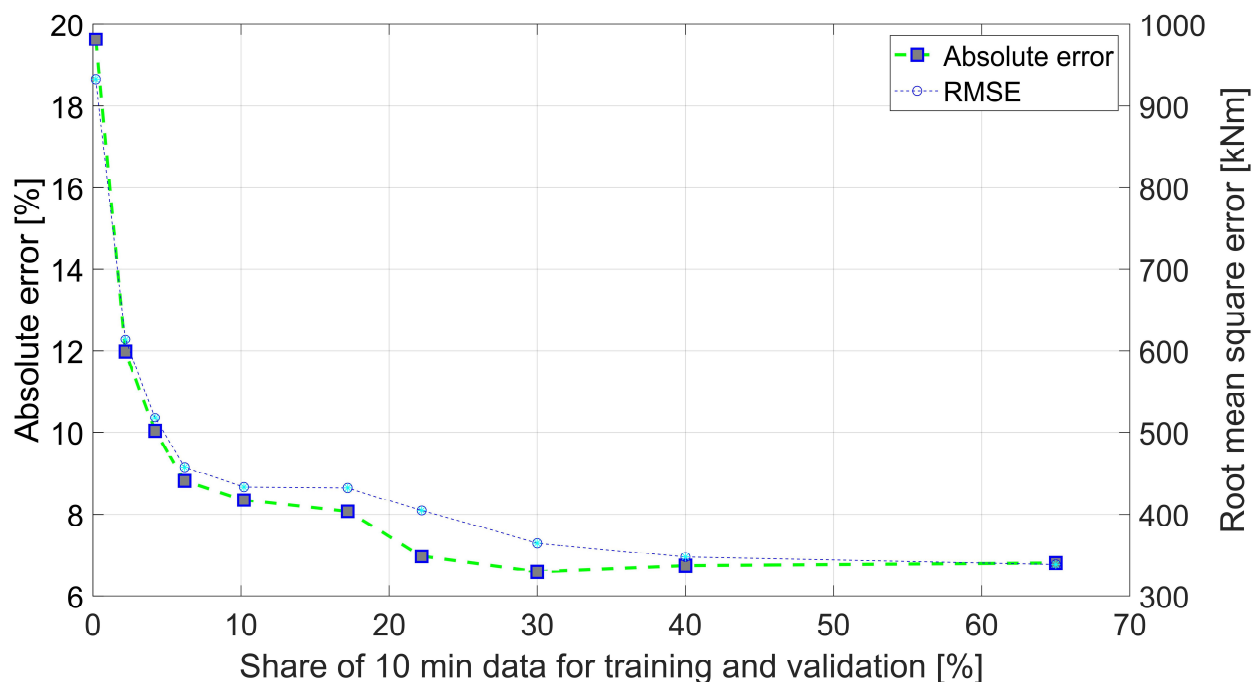
Figure 7 – Evaluation of dataset size for training / validation / testing a six month partial load predictive model

5 **4   Conclusions**

This paper used available SCADA data as well as strain gauges measurements from a research WT to develop a predictive model to estimate the DELs of the fore-aft bending moments of a WT tower. The dataset included a period of 6044 hours useful data. Different feature selection methods and a dimension reduction technique were applied to choose the sensors with the strongest predictive power. The data were then inputted into a feedforward neural network. The methodology and data
10 used reproduces and enhances the approaches of similar studies in the field of SHM.

The results indicate that using all data and applying neighborhood component analysis for feature selection yields the most conservative model regarding the number of features and the most accurate outcomes with the lowest mean absolute error. Additionally, dimension reduction techniques such as principal component analysis can contribute to a more parsimonious model reducing the number of features needed, however compromising the interpretability of the inputs given the
15 transformation of the data.

The results were significantly better, i.e. yielded lower mean absolute errors, when the dataset was divided by operational mode. Particularly, the models were significantly more accurate when analyzing the operation of the turbine at full load and partial load. The outcome of the model using signals from when the turbine was standing still was rather inaccurate with mean

absolute errors approximately four times higher than those from the partial load and approximately 10 times higher than those from full load.

Finally, a sensitivity analysis performed in the partial load model determined that a data collection campaign of 1706 hours would suffice to build a neural network that provides outcomes with mean absolute errors equivalent to that of a neural network using data from 3627 hours.

This paper examined data from only one turbine. To be able to generalize the results obtained from this study, the NN need to be trained with information from different turbines and data collected during different operating conditions. By doing this, it will be possible to determine the relationship between SCADA data and fatigue loads with more precision, thereby eliminating the need to install expensive gauge sensors to estimate these loads and contributing to more efficient structural health monitoring methods.

Furthermore, the methodology developed during this study could be further tested by means of an aero-elastic analytical model. Such a model would provide larger datasets to train the NN more precisely without the significant costs that this would imply if done empirically. The results of the NN trained with information from the aero-elastic model can be compared to the results presented in this paper to derive conclusions on the reliability and accuracy of this methodology.

Finally, the results could benefit from exploring alternative machine learning algorithms such as support vector machine and k-Nearest Neighbors.

5   Data availability

The high-frequency measurements from the SCADA and strain gauge sensors are not available due to confidentiality issue.

6   Author contributions

The work was carried out by AM, based on his Master Thesis at the Wind Energy Technology Institute under the supervision of MS and TF. AM preprocessed the data for machine learning purposes, implemented feature selection techniques, modeled fatigue loads with neural networks and ran a sensitivity analysis. MS initiated the issue, ran the data gathering campaign and processed the raw data. All authors were involved in the development of the manuscript.

7   Competing interests

The authors declare that they have no conflict of interest.

8   Acknowledgements

We acknowledge financial support by Land Schleswig-Holstein within the funding program Open Access-Publikationsfonds.

9   References

Boslaugh, S. and Watters, P. A.: Statistics in a nutshell, O'Reilly., 2008.

Cosack, N.: Fatigue Load Monitoring with Standard Wind Turbine Signals, University of Stuttgart. [online] Available from: https://d-nb.info/1009926721/34, 2010.

5   Cosack, N. and Kühn, M.: Ueberwachung von Belastungen an Windenergieanlagen durch Analyse von Standardsignalen.pdf, in AKIDA Tagungsband, 6. Aachener Kolloquium für Instandhaltung, Diagnose und Anlagenüberwachung, 14.-15.11.2006, pp. 277–283., 2006.

Cosack, N. and Kühn, M.: Prognose von Ermüdungslasten an Windenergieanlagen mittels Standardsignalen und neuronaler Netze.pdf, in DMK 2007 - Dresdner Maschinenelemente Kolloquium : 5. und 6. Dezember 2007, pp. 461–476., 2007.

10   DNV/Risø: Guidelines for Design of Wind Turbines - 2nd Edition., 2002.

Draper, N. R. and Smith, H.: Applied regression analysis, Wiley., 1998.

Goldberger, J., Roweis, S., Hinton, G. and Salakhutdinov, R.: Neighbourhood Components Analysis, 2005.

Han, J. and Kamber, M.: Data Mining: Concepts and Techniques., 2006.

Hazewinkel, M.: Encyclopaedia of Mathematics (set), 1st ed., Springer, Netherlands., 1994.

15   IRENA: Renewable Power Generation Costs in 2014, Bonn, Germany., 2015.

MathWorks: Fit Data with a Shallow Neural Network - MATLAB & Simulink - MathWorks, [online] Available from: https://de.mathworks.com/help/deeplearning/gs/fit-data-with-a-neural-network.html (Accessed 5 June 2019), 2019.

Melsheimer, M., Kamieth, R., Liebich, R. and Heilmann, Christoph, Grunwald, A.: No Title, in German Wind Energy Conference (DEWEK) Proceedings - 19th to 20th May 2015 in the Congress Centrum Bremen, pp. 1–4., 2015.

20   Noppe, N., Weijtjens, W. and Devriendt, C.: Modeling of quasi-static thrust load of wind turbines based on 1 s SCADA data, Wind Energy Sci., 3, 139–147, doi:10.5194/wes-3-139-2018, 2018.

Parsian, M.: Data algorithms : recipes for scaling up with Hadoop and Spark, "O'Reilly Media, Inc.", 2015., 2015.

REN21: Renewables 2018 Global Status Report, Paris, France., 2018.

Schedat, M. and Faber, T.: Fatigue load reconstruction on wind turbine structures using structural health monitoring, in Proceedings DEWEK 2017, 13th German Wind Energy Conference, pp. 1–4., 2017.

5   Schedat, M., Faber, T. and Sivanesan, A.: Structural health monitoring concept to predict the remaining lifetime of the wind turbine structure, in Proceedings of the 24th Conference on the Domestic Use of Energy, DUE 2016, pp. 1–5, IEEE., 2016.

Seifert, J., Vera-Tudela, L. and Kühn, M.: Training requirements of a neural network used for fatigue load estimation of offshore wind turbines, Energy Procedia, 137, 315–322, doi:10.1016/j.egypro.2017.10.356, 2017.

Sharma, N. and Saroha, K.: Study of dimension reduction methodologies in data mining, in International Conference on
10 Computing, Communication and Automation, ICCCA 2015, pp. 133–137, IEEE., 2015.

Smolka, U. and Cheng, P. W.: On the Design of Measurement Campaigns for Fatigue Life Monitoring of Offshore Wind Turbines, in The Twenty-third International Offshore and Polar Engineering Conference, vol. 9, pp. 408–413, International Society of Offshore and Polar Engineers, Anchorage, Alaska, USA,., 2013.

Vera-Tudela, L. and Kühn, M.: On the Selection of Input Variables for a Wind Turbine Load Monitoring System, Procedia
15 Technol., 15, 726–736, doi:10.1016/j.protcy.2014.09.045, 2014.

Vidal, R., Ma, Y. and Sastry, S. S.: Principal Component Analysis, pp. 25–62., 2016.

Ziegler, L., Gonzalez, E., Rubert, T., Smolka, U. and Melero, J. J.: Lifetime extension of onshore wind turbines: A review covering Germany, Spain, Denmark, and the UK, Renew. Sustain. Energy Rev., 82, 1261–1271, doi:10.1016/j.rser.2017.09.100, 2018.