Wind Energy

WIND ENERGY
Wind Energ. (2017)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/we.2102

RESEARCH ARTICLE

An artificial neural network-based condition monitoring method for wind turbines, with application to the monitoring of the gearbox

P. Bangalore[1], S. Letzgus[2], D. Karlsson[1] and M. Patriksson[3]

[1] Department of Energy and Enviroment, Chalmers University of Technology, Gothenburg, Sweden
[2] Institute of Aircraft Design, University of Stuttgart, Stuttgart, Germany
[3] Chalmers University of Technology and University of Gothenburg, Department of Mathematical Sciences, Gothenburg, Sweden

ABSTRACT

Major failures in wind turbines are expensive to repair and cause loss of revenue due to long downtime. Condition-based maintenance, which provides a possibility to reduce maintenance cost, has been made possible because of the successful application of various condition monitoring systems in wind turbines. New methods to improve the condition monitoring system are continuously being developed. Monitoring based on data stored in the supervisory control and data acquisition (SCADA) system in wind turbines has received attention recently. Artificial neural networks (ANNs) have proved to be a powerful tool for SCADA-based condition monitoring applications. This paper first gives an overview of the most important publications that discuss the application of ANN for condition monitoring in wind turbines. The knowledge from these publications is utilized and developed further with a focus on two areas: the data preprocessing and the data post-processing. Methods for filtering of data are presented, which ensure that the ANN models are trained on the data representing the true normal operating conditions of the wind turbine. A method to overcome the errors from the ANN models due to discontinuity in SCADA data is presented. Furthermore, a method utilizing the Mahalanobis distance is presented, which improves the anomaly detection by considering the correlation between ANN model errors and the operating condition. Finally, the proposed method is applied to case studies with failures in wind turbine gearboxes. The results of the application illustrate the advantages and limitations of the proposed method. Copyright © 2017 John Wiley & Sons, Ltd.

KEYWORDS

artificial neural network (ANN); condition monitoring system (CMS); Mahalanobis distance; wind energy; supervisory control and data acquisition (SCADA)

Correspondence

P. Bangalore, Division of Electric Power Engineering, Chalmers University of Technology, Horsalsvagen 11, Gothenburg 41296, Sweden.
E-mail: pramod.bangalore@chalmers.se

Received 27 June 2016; Revised 27 October 2016; Accepted 11 February 2017

1. INTRODUCTION

Wind power has developed to utility scale in many power systems around the globe. The future energy outlook presented by International Energy Agency (IEA)[1] illustrates that the number of turbines is expected to grow further in the future. Furthermore, the wind turbines are erected in remote locations with favorable wind conditions, e.g. offshore, which makes access to wind turbines difficult. This growth in wind energy, and especially its installation at offshore locations, has led to an increased focus on the availability and reliability of wind turbines. The possibility of application of various condition monitoring systems (CMSs) has made condition-based maintenance an attractive strategy to reduce the cost of maintenance and improve availability.

Condition monitoring based on vibration signals has proven to be successful in the monitoring of the wind turbine gearboxes; a few examples for vibration-based monitoring systems are presented in Lu *et al.*[2] However, a study conducted by the National Renewable Energy Laboratory found that the average detection accuracy of the existing vibration monitoring systems is only about 50%.[3] In such a situation, it is desirable to have a complementary CMS that can increase the detection accuracy without requirements of additional measurement sensors. In this regard, condition monitoring using data stored in the supervisory control and data acquisition (SCADA) system presents a potential option. Researchers have published different methods and approaches for using SCADA data for condition monitoring; see previous papers.[4–17] Mathematical modeling methods like artificial neural networks (ANNs) have been frequently utilized for the analysis of SCADA data. The ANN-based methods do not require an in-depth knowledge about the component being monitored. Moreover, the method can be applied to a number of components in the wind turbine given the availability of suitable SCADA data. In addition to being general, the ANN method is easily scalable for application on a large set of wind turbines. These advantages of the ANN-based condition monitoring method have made it a topic of interest for monitoring critical components in the wind turbine. The most prominent studies[4–7] which utilize ANNs for wind turbine condition monitoring using SCADA data are discussed here.

One of the earliest developments that utilized ANN normal behavior models for condition monitoring for wind turbine applications was presented by Garcia *et al*. A software tool called Intelligent System for Predictive Maintenance was presented in Garcia *et al*,[4] and it is divided into six modules responsible for normal behavior modeling, anomaly detection, health condition assessment, failure diagnosis, preventive maintenance scheduling, and maintenance effectiveness assessment. The normal behavior module utilizes a multilayer auto-regressive ANN model for predicting a parameter value based on the selected input parameters. The ANN model output is compared with the measured value in real time, and a difference outside confidence bands, defined by the normal behavior model, is termed as an anomaly. The case study, presented in Garcia *et al.*,[4] shows that the system is able to detect a failure approximately 26 h in advance, which might be sufficient to avoid a catastrophic failure, but not for an effective condition-based maintenance optimization. A similar auto-regressive ANN model was utilized in Zaher *et al*.,[5] where the system was applied to case studies with real data from wind turbines with failures in the gearbox bearings. The case study results showed the possibility of detecting the faults as early as 6 months before the eventual replacement is required. The anomaly detection was achieved by observing an increase in the frequency of the errors between the predicted and measured output parameter. However, such a method for detection of anomaly can be difficult with a large number of wind turbines, and it is desirable to have a threshold value, which can be utilized to automatically generate alarms from the CMS. The feed-forward multilayer ANN normal behavior models with different number of neurons in the hidden layer were investigated in Kusiak and Verma,[6] and the best performing model was selected for predicting faults in the generator bearings of the wind turbine. The case study with 10-s SCADA data illustrated that the method is able to predict faults about 1.5 h before the eventual failure. The detection of an anomaly close to the actual failure does not allow any kind of maintenance planning.

A comprehensive CMS for wind turbine applications utilizing adaptive neuro-fuzzy inference system (ANFIS) modeling approach was presented in Schlechtingen *et al*.,[7] and case studies of application of the system were presented in Schlechtingen and Santos.[18] The adaptive neuro-fuzzy inference system models utilize the concepts of ANN and fuzzy logic to provide a combined model that is capable of detecting anomalies, as well as providing a root cause for such anomalies based on simple If-Then rules. Furthermore, an automatic anomaly detection method was presented, which utilized the standard deviation value of the distribution of errors obtained from the test data set that is used during the ANN model training, to establish the threshold values for anomaly detection. The case studies presented in[18] illustrated that the system is able to detect potential failure months in advance, providing enough opportunity for the operator to make informed decisions. The anomaly detection was based on errors calculated as a difference between the modeled and measured output parameter value. However, as demonstrated in Section 5, utilizing only the numerical values of error could, in some cases, lead to a situation where the anomalies are not detected.

The methods discussed in this paper intend to utilize the valuable knowledge generated by the aforementioned publications and build on this knowledge with an aim to improve the confidence in the SCADA-based condition monitoring process. The main contributions of the paper are focused in two areas: the data preprocessing and the data post-processing. The data preprocessing is implemented in order to ensure that the ANN normal behavior models are trained only with data that represent the true normal operating conditions for the wind turbine. Three data filters are proposed, which enable the removal of abnormal data from the training data set and improve the ANN model performance and generalization. Additionally, the effect of discontinuity of SCADA data on the ANN model output is discussed, and an approach to avoid false alarms under such conditions is presented. The post-processing of data deals with the handling of the output from the ANN models to achieve an early detection of impending failures of the wind turbine components, while avoiding false alarms. In this regard, an approach to overcome the inherent randomness that exists in the ANN model training process is presented. Furthermore, a statistical approach based on the Mahalanobis distance (MHD) measure, proposed in Bangalore and Tjernberg,[19] which provides anomaly detection based on the correlation between the operating condition and the errors, is utilized to improve confidence in the anomaly detection process. In order to illustrate the contribution of this paper, Table I presents a comparative analysis of some important aspects of methods proposed in the publications mentioned earlier and the developments proposed in this paper.

The remainder of the paper is organized as follows: Section 2 presents the outline of the proposed SCADA-based condition monitoring method. Section 3 presents the data preprocessing approach and illustrates the advantages of filtering the data before training and addition of the missing data input parameter to the ANN models. Section 4 presents the

Table I. Summary of the SCADA-based condition monitoring methods utilizing ANN normal behavior models for wind turbines.

	Garcia *et al.*[4]	Zaher *et al.*[5]	Kusiak *et al.*[6]	Schlechtingen *et al.*[7]	This paper
Method used	Multilayer feed-forward auto-regressive ANN	Multilayer feed-forward auto-regressive ANN	Multilayer feed-forward ANN	Adaptive neuro-fuzzy interference system	NARX ANN
Input selection	Domain knowledge	Domain knowledge	Data mining algorithms	Domain knowledge + data mining algorithms	Domain knowledge
Training data filtering	Not mentioned	Manual selection of data	Removal of data-causing error residuals above a control limit during training	Validity and range checks + missing data removal	Validity and range checks + missing data removal and addition of missing data input + cluster data filter
Anomaly detection	Confidence bands based on ANN model training data	Observation of increase in frequency of errors	Observation of errors	Based on a threshold value derived from the distribution of errors from the test data during the ANN model training	Based on a statistical distance measure (Mahalanobis distance) and threshold value based on the correlation between ANN model error and operating point
Maintenance window	26 h	4–6 months	1.5 h	2–3 months	2–3 months

data post-processing methods including the MHD approach for anomaly detection and a method to overcome the inherent randomness in the ANN model training process. Section 5 presents the application of the proposed method to a variety of case studies with data from wind turbines with failures in the gearbox. Finally, Section 6 presents the conclusions.

2. THE SCADA-BASED CMS METHOD USING ANN NORMAL BEHAVIOR MODELS

The ANNs attempt to emulate the data processing capabilities of a human brain. The similarity between ANN and the function of the brain is the process of knowledge acquisition through experience or learning and the retention of this knowledge with the interneuron connections, known as synaptic weights.[20] The synaptic weights in the ANN model are decided based on the training of the ANN models with a representative data set. The training can be performed using standard optimization algorithms such as the steepest descent algorithm. However, in this paper, the Levenberg–Marquardt algorithm (LMA)[21,22] is utilized. It has the combined advantage of the convergent steepest descent algorithm and Newton's method, which usually is fast near an optimum. The LMA gives better performance in terms of accuracy for neural networks with less than 100 neurons.[23] Hence, as the number of neurons required for modeling is less than 100, in this paper the LMA has been used for training of ANN models.

The schematic representation of the ANN-based CMS method is shown in Figure 1. The condition monitoring method can be divided into two blocks. The block on the left in Figure 1 presents a one-time process, during which the ANN model is trained to emulate normal operating conditions in the monitored component. The block on the right presents the continuous application process of anomaly detection and condition monitoring.

The process of building ANN-based normal behavior models can be divided into three sub-tasks:

1. Specification of an ANN configuration,
2. Input/output parameter selection,
3. Data preprocessing and post-processing.

Together, they represent an iterative development process.

2.1. Specification of ANN configuration

The configuration of the ANN model should be determined based on the knowledge about the data and the source, and as suggested in Tarassenko,[24] it is an experimental process where different configurations should be tested in order to realize the most suitable one. The application of multilayer auto-regressive neural networks in Garcial *et al.*[4] and Zaher *et al.*[5]

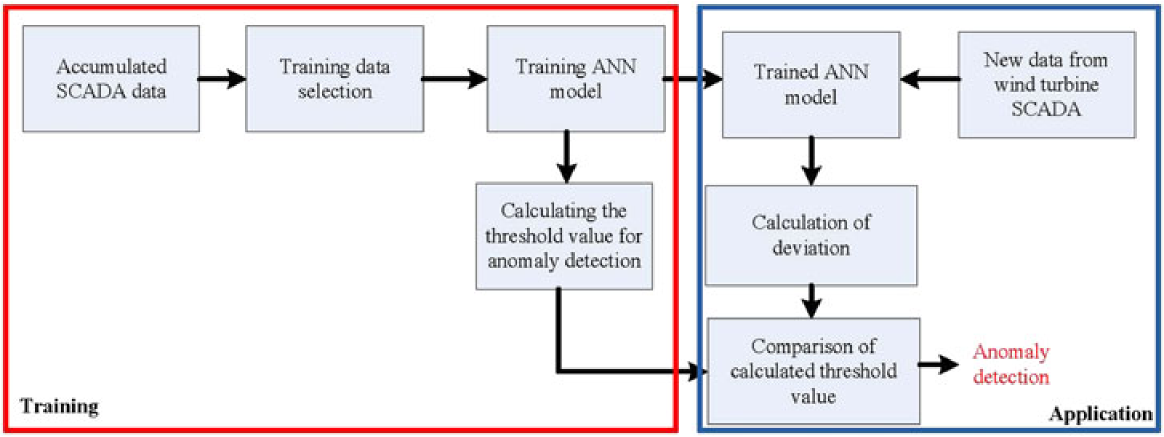


Figure 1. The proposed ANN-based condition monitoring method using SCADA data. [Colour figure can be viewed at wileyonlinelibrary.com]

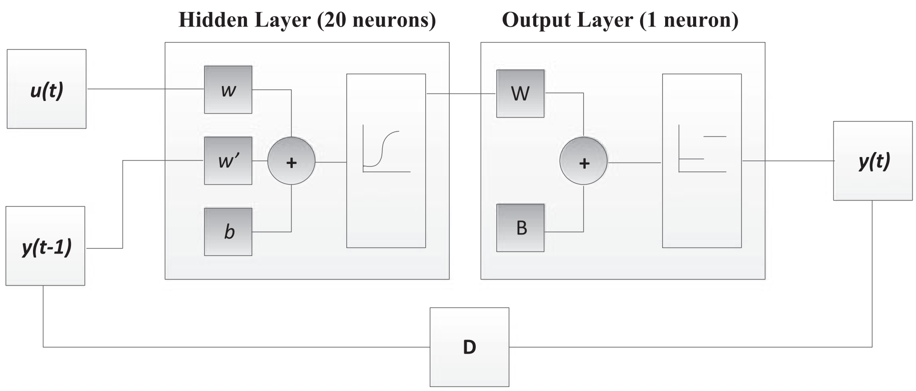


Figure 2. A schematic representation of the NARX ANN model configuration.

has shown the suitability of this configuration for wind turbine applications. The nonlinear auto-regressive with exogenous input (NARX) ANN configuration is a variation of the auto-regressive ANN model, and the difference between these two configurations can be found in the manner the delayed output values are used in the model. In the auto-regressive ANN configuration presented in Garcial *et al.*[4] and Zaher *et al.*,[5] the delayed output parameter value $y(t-1)$ is extracted from the SCADA system, while in the NARX model, the regressive input is the value estimated by the model itself as shown in Figure 2. Unlike the auto-regressive ANN configuration, the NARX configuration provides a possibility to isolate the influence of an anomaly in the component being monitored on the output of the ANN model. Furthermore, an analysis performed in Karlsson,[25] where three configurations—the feed-forward ANN with one hidden layer, the feed-forward ANN with two hidden layers, and the NARX model with one hidden layer—were tested with application to data from different wind turbines, also showed that the NARX ANN configuration is the best among the investigated options. These results are also in-line with the observations made in Salmasi *et al.*[26] and Sundermeyer *et al.*[27] Hence, in this paper, the NARX ANN configuration with one hidden layer with 20 neurons is selected for creating the normal behavior models.

2.2. Input parameter selection

The selection of input parameters for modeling a particular output parameter can be performed using various data mining approaches; for example, three data mining algorithms were utilized in Kusiak and Verma[6] to establish a cause and effect relationship between different measurements available in the wind turbine SCADA system. However, such an approach may lead to a selection of a large number of input parameters, and domain knowledge has to be applied to keep the number of parameters to a reasonable value as suggested in Schlechtingen *et al.*[7] Moreover, the selection of input parameters based on domain knowledge has been demonstrated to be successful in Garcial *et al.*[4] and Zaher *et al.*[5] Hence, in this paper, an understanding of the physics and the domain knowledge are considered to be the best method to decide suitable input parameters for modeling. The case studies presented in Section 5 discuss the application of the proposed method for monitoring of wind turbine gearboxes. Hence, the discussion presented here is limited to the selection of input parameters for ANN models that are used for gearbox monitoring.

The gearbox bearing and lubrication oil temperature values are important from a condition monitoring perspective, as the most common failure modes in the gearbox will, potentially, manifest themselves into a deviation in these measurements. Hence, normal behavior models for the gearbox bearing and lubrication oil temperatures are utilized to achieve condition monitoring of the gearbox.

The gearbox bearing and lubrication oil temperatures are directly connected to the nacelle and ambient temperature values, and there exists a state of thermal equilibrium between these temperatures under normal operating conditions.[17] The ANN normal behavior model can be used to emulate this thermal equilibrium condition, and any disturbance in the equilibrium may then indicate an anomalous operation in the gearbox. Consequently, the ambient and nacelle temperature measurements are utilized as input for the ANN normal behavior models. Furthermore, the temperatures inside the nacelle are directly related to the power being produced by the wind turbine, as the electrical and mechanical losses are proportional to the power produced; this concept was also exploited in Feng *et al.*[16] for the monitoring of wind turbine gearboxes. Hence, power produced from the wind turbine will also be included as an input to the ANN normal behavior models. The wind turbines utilized for case studies in this paper have a mechanical pump, mounted on the low-speed side of the gearbox, in the lubrication oil system. Hence, the flow of lubrication oil will depend on the rotational speed of the main shaft, and consequently, the rotational speed is also included as an input for the ANN normal behavior models.

In order for the anomaly detection to be improved using the ANN models, parameters that are highly correlated with the output parameter should not be utilized as an input. Such a situation could lead the ANN model to correctly predict an abnormal operating condition. For example, it was found that utilizing the gearbox lubrication oil temperature as an

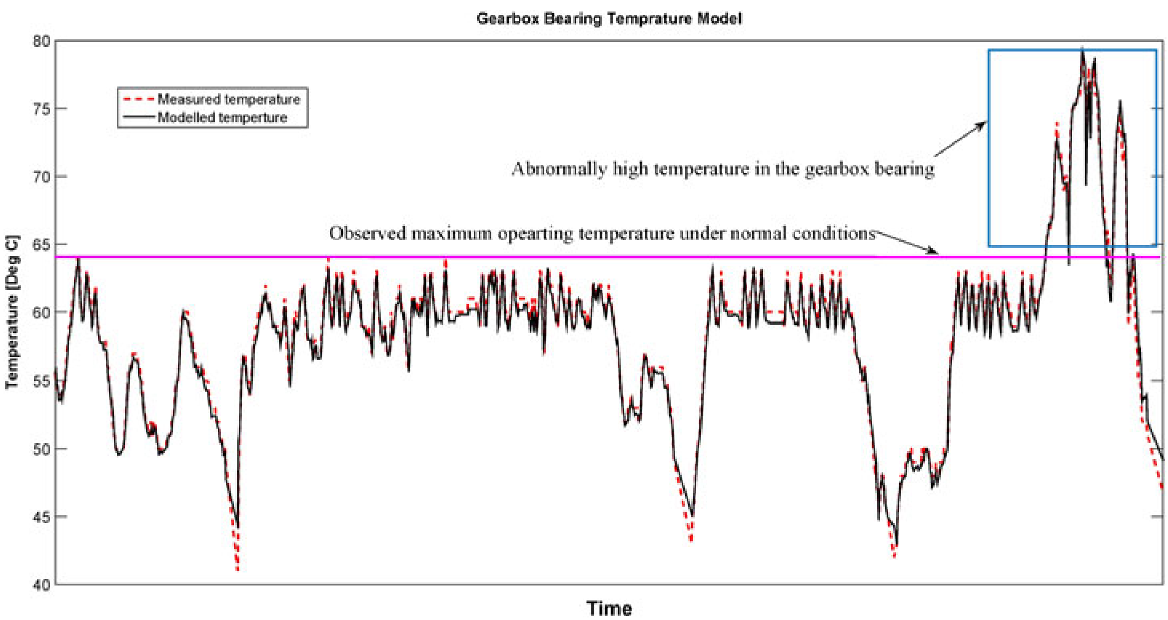


Figure 3. A demonstration of ANN model correctly estimating an abnormal operating condition. [Colour figure can be viewed at wileyonlinelibrary.com]

Table II. Input and output parameters.

Output parameters	Input parameters
Gearbox bearing temperature [°C]	Power production [kW] Rotor RPM
Gearbox lubrication oil temperature [°C]	Nacelle temperature [°C] Ambient temperature [°C] Missing data input [–]

input for modeling the gearbox bearing temperature could lead the ANN model to correctly estimate abnormally high temperatures in the gearbox bearings, as shown in Figure 3. Consequently, the model might fail to detect an anomaly.

The input and output parameters for the ANN normal behavior models created for condition monitoring of the wind turbine gearbox are presented in Table II.

3. DATA PREPROCESSING

The ANN models learn the input/output mapping based, solely, on the data provided during the training stage. Hence, it is important that the training data are free from errors. In the real world, however, there seldom exists a perfect data set, and often, SCADA data are found to be discontinuous and to contain inconsistencies. These inconsistencies lead to inaccuracies in the ANN models and hence need to be dealt with in an appropriate manner. Three types of filters are presented here, which can be applied to remove data points that might reduce the performance of the ANN models.

3.1. General filter

Malfunctions in the SCADA communication system, sensors or signal processing errors, and standstill during maintenance and repair actions lead to missing or faulty data points. The following three simple rules are utilized for filtering the missing and garbage data:

1. Filter out all data vectors where one of the input or output parameter values is missing.
2. Filter out all data vectors that correspond to a situation when the wind turbine is not producing any power.
3. Filter out all data vectors where one or more parameters have a value higher than a predefined threshold. In this paper, the threshold values are decided based on the manufacturer specifications. For example, all measurements with a bearing temperature greater than 90°C are filtered out.

3.2. Cluster filter

Wind turbines are subjected to highly variable operating conditions and have a nonlinear operating characteristic, and hence, it is difficult to detect outliers by setting simple threshold values. Moreover, during power curtailment conditions, the wind turbine operation cannot be considered as normal as they are producing less than normal power. The cluster filter is used to remove data outliers and data corresponding to curtailment conditions from the training data set.

Clustering presents a suitable solution to the problem of the classification of wind turbine SCADA data and has been demonstrated to be successful in Kusiak and Verma.[28] The approach presented in Kusiak and Verma[28] illustrated the application of clustering to the wind power curve for detection of faults. This approach has been extended here, and the clustering method is applied to a multidimensional data set that consists of all the input and output parameters of the ANN model. The algorithm for cluster filter is described in Algorithm 1.

Algorithm 1 The algorithm for cluster filter

1: Decide the maximum number N of clusters
2: Use the clustering method to assign a cluster number $n \in \{1, \dots, N\}$ to each input data vector D_i, $i \in \{1, \dots, \text{length(Dataset)}\}$, in the training data set
3: Find the centroid $C_n, n \in \{1, \dots, N\}$, for each cluster
4: Calculate the Mahalanobis distance MHD_i, $i \in \{1, \dots, \text{length(Dataset)}\}$, of each data vector D_i^n from its cluster center C_n
5: Estimate the probability distribution for the Mahalanobis distances in the vector MHD
6: Eliminate those data vectors, whose probability of occurrence is lower than a threshold value

The training data set is divided into N clusters utilizing Ward's minimum variance algorithm, described in Ward.[29] Consequently, the clusters are decided to minimize the inner square distance, calculated over Euclidean space, between cluster centers. The number of clusters N is decided based on the understanding of the behavior of the wind turbine, such that each cluster represents a different operating condition. The criterion utilized for deciding the number of clusters is presented in Table III, based on which the training data set is divided into 12 clusters.

In the next step, the MHD, discussed in Section 4, is calculated for each data vector in the training data set from its cluster center. Figure 4 presents the distribution of the MHD values of the training data set for a case study. It can be observed that the MHD values can be estimated by a loglogistic probability distribution function.

Table III. The criteria for deciding the number of clusters.

Operating parameter	Interval 1	Interval 2	Interval 3	Interval 4
Wind speed [m/s]	$0 \rightarrow 5$	$5 \rightarrow 8$	$8 \rightarrow 11$	$11 \rightarrow 25$
Ambient temperature [°C]	$-30 \rightarrow -3$	$-3 \rightarrow 5$	$5 \rightarrow 30$	–

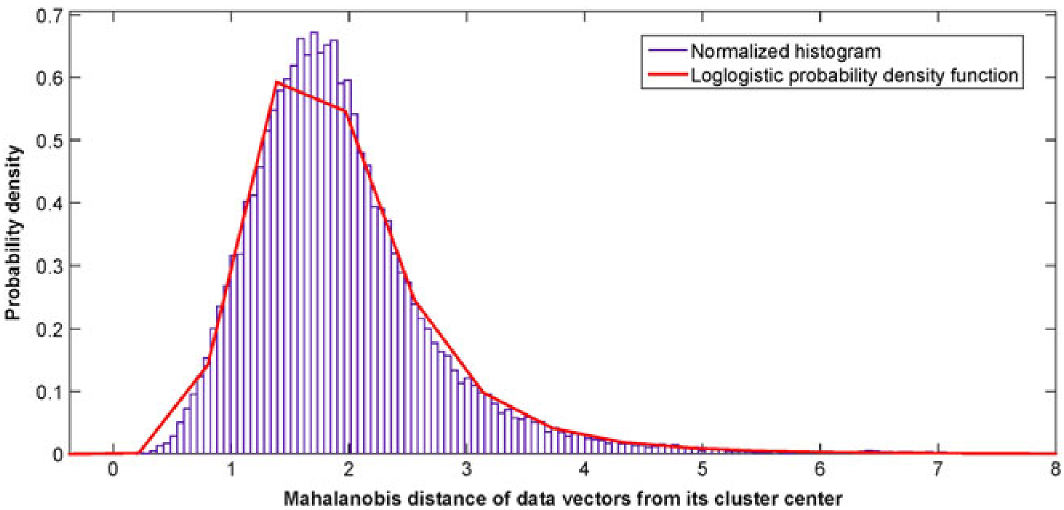


Figure 4. The histogram and probability density function fit for Mahalanobis distance values of data vectors from its cluster center. [Colour figure can be viewed at wileyonlinelibrary.com]

Finally, a probability threshold of 2.5% is chosen, and data vectors with a lower probability of occurrence are filtered out. The probability threshold is selected based on the knowledge that power curtailment, even though possible, is not a common practice for the wind turbines considered for case studies, and hence, a low threshold value is acceptable as a small amount of data might be affected because of power curtailment. The cluster filter is used only on the training data set and is not utilized during the application stage. Hence, during the application stage, power curtailment could lead to false alarms. Consequently, it is suggested that the condition monitoring process be blocked during power curtailment. The modern wind turbine SCADA systems include signals that indicate if a power curtailment has been initiated, and such signals can be utilized to block the condition monitoring.

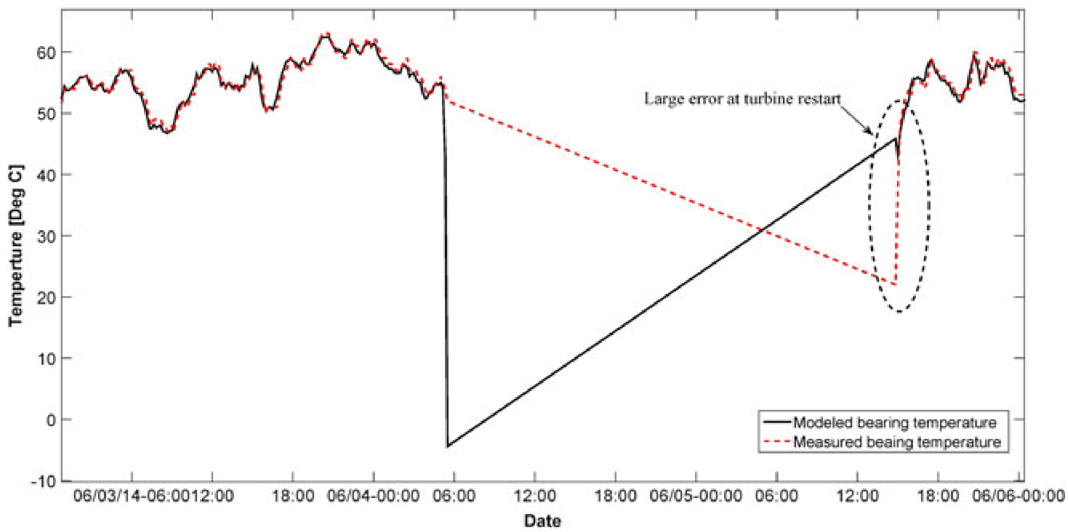
3.3. Missing data filter

In the ANN models used in this paper, the feedback loop for the NARX models is adjusted such that the model considers the value of the output at time instant $t-1$ to estimate the output at time instant t. Hence, it is important that continuous data are available during the training and application stage when using NARX models. However, because of communication issues, on occasion, continuous data might not be available and such data can cause false alarms during the condition monitoring stage. One such case of large error between estimated and actual values due to missing data is presented in Figure 5(a).

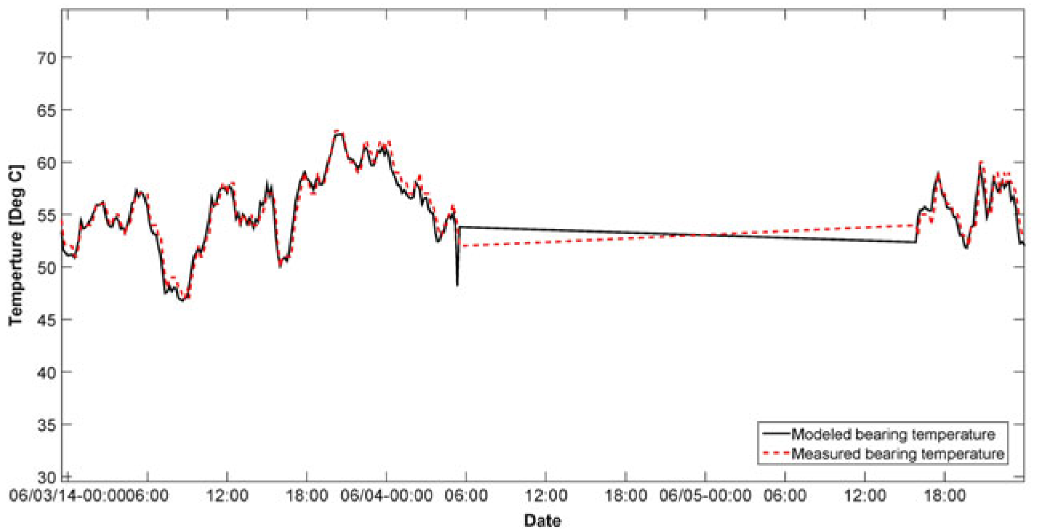
In order for the problem to be rectified, the missing data filter is implemented, which is designed to ensure that at least 1 h of continuous data is available for a parameter vector to be considered during the training and application stages. The data vectors that do not fulfill this criterion are eliminated from the training and application data sets. In addition to the filtering of the missing data, it is important to inform the ANN model that a data vector has been removed and that there is a break in continuity of the data. In order for this information to be transferred to the ANN model, a missing data parameter is defined and included as an input to the model. The missing data parameter is defined as shown as follows:

$$MD_i = \begin{cases} 1, & \text{if Timestamp}_i - \text{Timestamp}_{i-1} > 10\,\text{min}, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where MD_i is a new input parameter that transfers the information of continuity of data to the ANN model. The improvement in the model due to missing data input is presented in Figure 5(b).



(a) Measured and modeled temperature without the missing data input parameter



(b) Measured and modeled temperature with the missing data input parameter

Figure 5. The effect of missing data input. (a) Measured and modeled temperature without the missing data input parameter. (b) Measured and modeled temperature with the missing data input parameter. [Colour figure can be viewed at wileyonlinelibrary.com]

The aforementioned filters eliminate some amount of data from the training and application stages. A pictorial representation of the extent of data removed because of the mentioned filters is presented in Figure 6.

4. DATA POST-PROCESSING

The purpose of creating ANN normal behavior models is to be able to detect anomalies in the operation of the components being modeled. Researchers have proposed different methods for anomaly detection using ANN models. One such approach that utilized the probability distribution of errors during the ANN model training to create threshold values for anomaly detection was proposed in Schlechtingen *et al.*[7] However, a threshold based on the distribution of errors during the ANN model's training might not be sufficient, as the model might be skewed and prone to be inaccurate at some operating points, as shown in Figure 7, where a higher error can be seen for target parameter values around 40. Hence, in order for the dependence of the operating condition on the ANN model output to be considered, an approach using the MHD measure, which was proposed in Bangalore and Tjernberg,[19] is utilized in this paper.

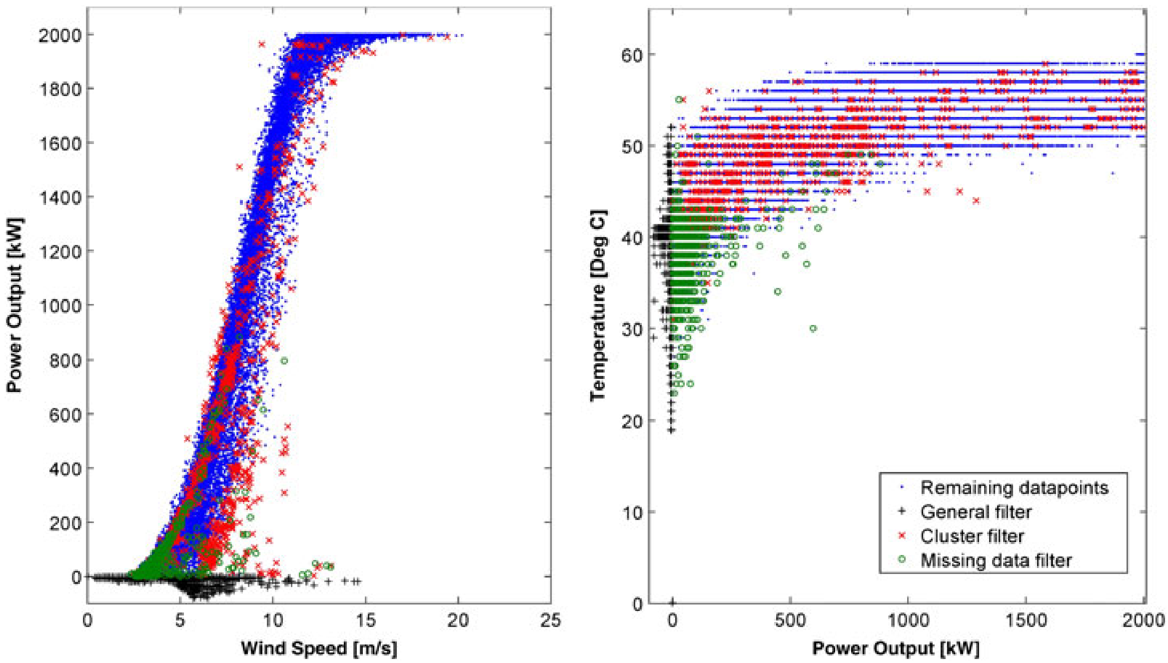


Figure 6. A pictorial representation of data removed from the training data set because of filtering. [Colour figure can be viewed at wileyonlinelibrary.com]

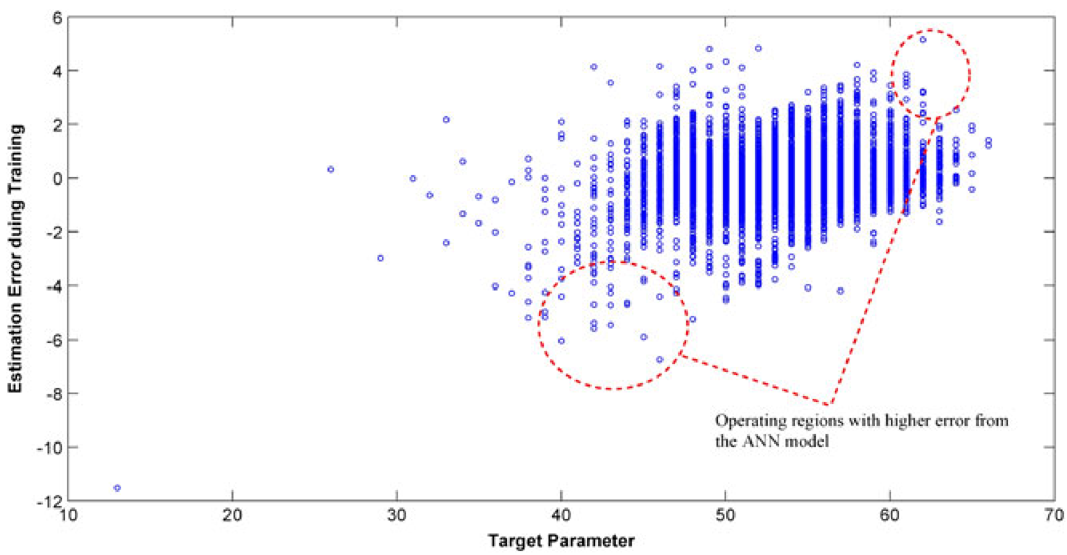


Figure 7. A pictorial representation of modeling errors from the training stage. [Colour figure can be viewed at wileyonlinelibrary.com]

The MHD is a unit-less distance measurement that has the ability to capture correlations of variables in a process or a system and is defined as shown as follows:

$$MHD_i = \sqrt{(X_i - \mu)\, C^{-1}\, (X_i - \mu)^T},\ i = 1, \ldots, m, \tag{2}$$

where MHD_i is the MHD measure for the ith observation vector $X_i = [X_{i1}, \ldots, X_{im}]$, where m is the total number of parameters. The vector $\mu = [\mu_1, \ldots, \mu_m]$ is the vector of mean values and C is the covariance matrix. The MHD for the training data set is calculated as follows:

$$MHD_{ref_i} = \sqrt{\left(X_{ref_i} - \mu_{ref}\right) C_{ref}^{-1} \left(X_{ref_i} - \mu_{ref}\right)^T}, \tag{3}$$

where $X_{ref_i} = [\text{Training Error}_i, \text{Target Value}_i]$ is the ith vector that consists of the target value and the corresponding error value obtained from the training data set. The covariance matrix C_{ref} and the vector or means μ_{ref} is calculated from the vector X_{ref}. It was found that the distribution of the MHD values in the reference vector X_{ref} can be approximated by a Weibull distribution as shown in Figure 8. Finally, the Weibull distribution fit is utilized for deciding the anomaly detection threshold, such that only the MHD values from the anomaly detection stage with probability of occurrence less than 1%

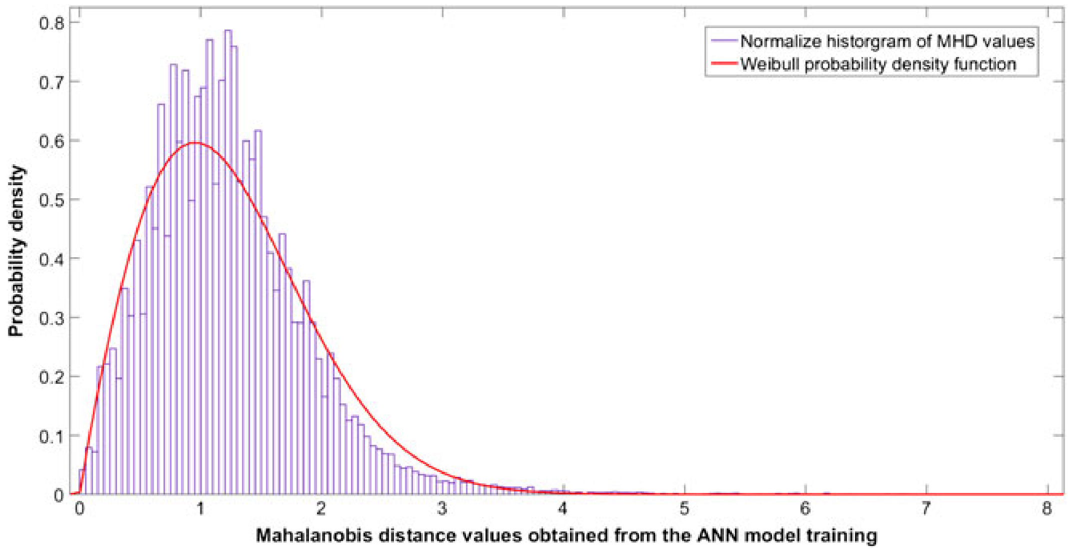


Figure 8. Histogram and Weibull probability density function for MHD values obtained during ANN training. [Colour figure can be viewed at wileyonlinelibrary.com]

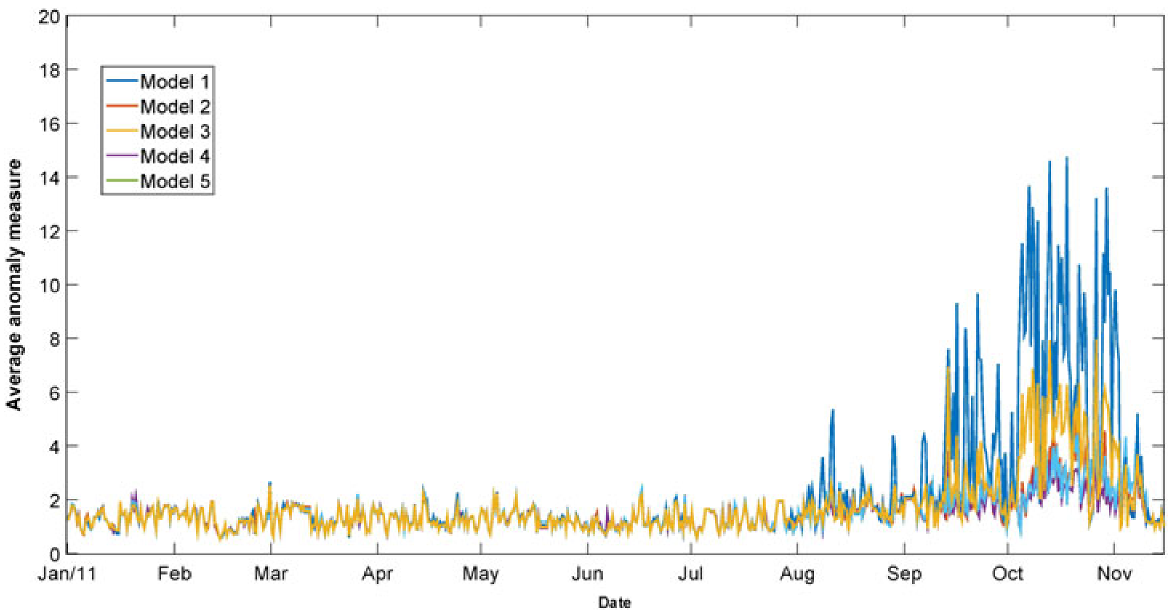


Figure 9. Output of five ANN models trained with same data. [Colour figure can be viewed at wileyonlinelibrary.com]

are tagged as an anomaly. The MHD during the anomaly detection stage is calculated as follows:

$$MHD_{app_i} = \sqrt{\left(X_{app_i} - \mu_{ref}\right) C_{ref}^{-1} \left(X_{app_i} - \mu_{ref}\right)^T}, \tag{4}$$

where $X_{app_i} = [\text{Estimation error}_i, \text{Measured paramter}_i]$ is the vector that consists of the model prediction error and the target value, which is extracted from the SCADA system during the anomaly detection stage.

In addition to the MHD approach for anomaly detection, it is important to consider the inherent randomness in the ANN model training process. The training of ANN models is, in general, a non-convex optimization problem. The training might stop at a local optimum, when the performance parameter is minimized; however, the generalization capability of the model might not be good. During the validation of the ANN-based CMS method to case studies, it was realized that the ANN models trained with the same data behave differently when there is an indication of an anomaly in the monitored component, as can be seen in Figure 9, where the difference in the anomaly measure varies widely among the five models after the month of August. All the five models have the same configuration and have been trained using the same data with randomly initialized weights.

A sensitive model, like model 1 in Figure 9, will be prone to false alarms, and an insensitive model, like model 4, might not detect the anomaly; both situations are undesirable. In order for the effect of this randomness in the modeling process to be avoided, several ANN models are trained, out of which a small subset of the best performing models are selected, and the final output is averaged over all of these.

5. CASE STUDIES

In order for the results of the method proposed in this paper to be compared, the performance of the NARX normal behavior model for generator bearing temperature is compared with the performance of previously published ANN models in other studies.[5–7] Table IV presents two performance measures: the mean absolute error and the root mean squared error, for the generator bearing temperature model. The NARX models have been trained on data from 1 year of turbine operation with no recorded failures in the generator system.

The results illustrate that the NARX models have a better performance compared with the previously mentioned methods. Furthermore, the methods presented for filtering the training data result into an improvement in the performance of the models.

In order for the proposed ANN-based CMS to be validated, the method is applied to data from four onshore wind turbines with a failure in the gearbox; the access to the data was provided by the industry partners in the project. All the wind turbines are rated 2 MW and located in South and Central parts of Sweden. The details of the component failures in each of the considered wind turbines are provided in Table V. The input and output parameters for the ANN model are presented in Table II. An additional input that transfers the information about a break in continuity of SCADA data is included, in-line with the discussion presented in Section 3.3. The NARX ANN model with 20 neurons in the hidden layer, with a delay of one time unit, was selected as the ANN configuration for the modeling. The output from the anomaly detection stage, i.e, the MHD value, has been averaged over the 100 best ANN models selected from a total of 300 trained models. It should be noted that the ANN-based condition monitoring method presented here is capable of detecting an anomaly in the operation of the monitored component but does not provide any information about the failure mode responsible for the anomalous

Table IV. The performance measures for generator bearing temperature model.

Performance measure	Zaher *et al.*[5]	Kusiak *et al.*[6]	Schlechtingen *et al.*[7]	NARX (without filter)	NARX (with filter)
Mean absolute error	–	0.663	–	0.46	0.44
Root mean squared error	1.23 (gearbox bearing model)	–	1.3	0.81	0.77

Table V. Details of component failures in the wind turbines used in the case study.

Turbine ID	Turbine rating [kW]	Failure mode
Turbine A	2000	Spalling in drive end planet carrier bearing of the gearbox
Turbine B	2000	Gearbox stuck and intermediate shaft bearing severely damaged
Turbine C	2000	Gearbox temperature sensor calibration error
Turbine D	2000	Gearbox intermediate shaft gear failure

DOI: 10.1002/we

operation. In-line with current industry practice, it is suggested that an inspection is carried out following an alarm from the ANN-based CMS system to determine the correct failure mode.

The 10-min-average SCADA data are used for monitoring purposes. Hence, in 24 h, there are a maximum 144 measurements. The results for the anomaly detection are presented as an average of 12-h periods, resulting in two MHD values per day. In order for the confidence in the prediction to be increased and in-line with the missing data filtering approach, it is ensured that at least 1 h of data is available for an output from the anomaly detection to be considered. In cases where sufficient data are not available, the previous valid output is copied and an indication of missing data is presented in the output.

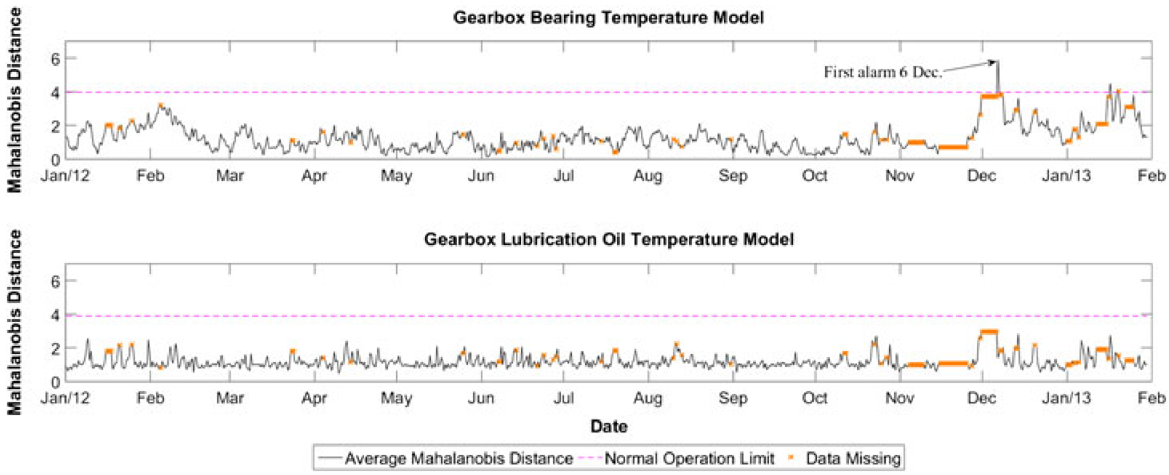


Figure 10. Anomaly detection output for Turbine-A. [Colour figure can be viewed at wileyonlinelibrary.com]

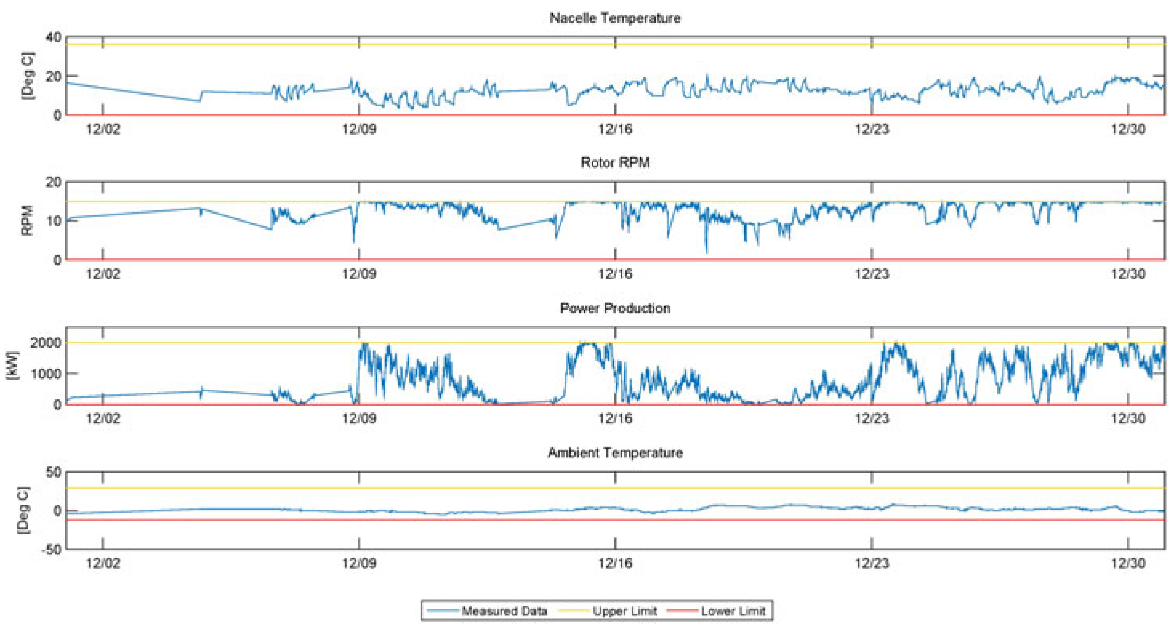


Figure 11. Inputs to the ANN model for Turbine-A. [Colour figure can be viewed at wileyonlinelibrary.com]

5.1. Case study for Turbine-A

The ANN models for Turbine-A were trained on data from the year 2011 and applied for anomaly detection during the year 2012, and the output is presented in Figure 10. The first alarm was seen from the gearbox bearing model on December 6, 2012. From the maintenance records, it was realized that the gearbox was replaced in February 2013. Furthermore, the vibration monitoring system reported an alarm on November 23, 2012, and an inspection was carried out on November 28, which revealed a spalling damage in the monitored bearing. The proposed method was able to detect the failure approximately 3 months in advance and was almost at par with the vibration-based CMS.

The ANN models lack any physical understanding of the system, and hence, it is necessary to ensure that the anomaly detected by the ANN model is not a result of incorrect inputs to the model. The four inputs to the ANN model are presented in Figure 11 for the month of December 2012. The upper limit and the lower limit for the data in Figure 11 correspond to the maximum and minimum values of the data used in the training set. It can be observed that the input parameters for the month of December have been within the limits of the data provided to the ANN model during the training process. Hence, it can be concluded that the anomaly detected using the models is, in fact, a result of the component having an abnormal operation.

In order for the advantage of the MHD approach over the traditional approach of utilizing the error between the ANN model estimated and the actual temperature value to be demonstrated, the plot of the absolute error for case study for

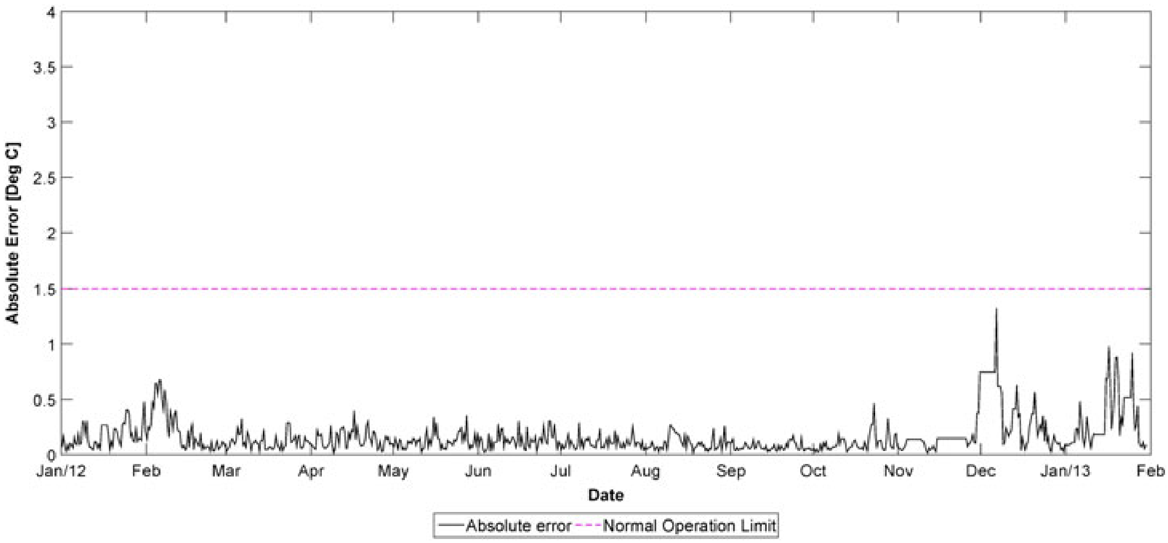


Figure 12. The plot of absolute error values for Turbine-A. [Colour figure can be viewed at wileyonlinelibrary.com]

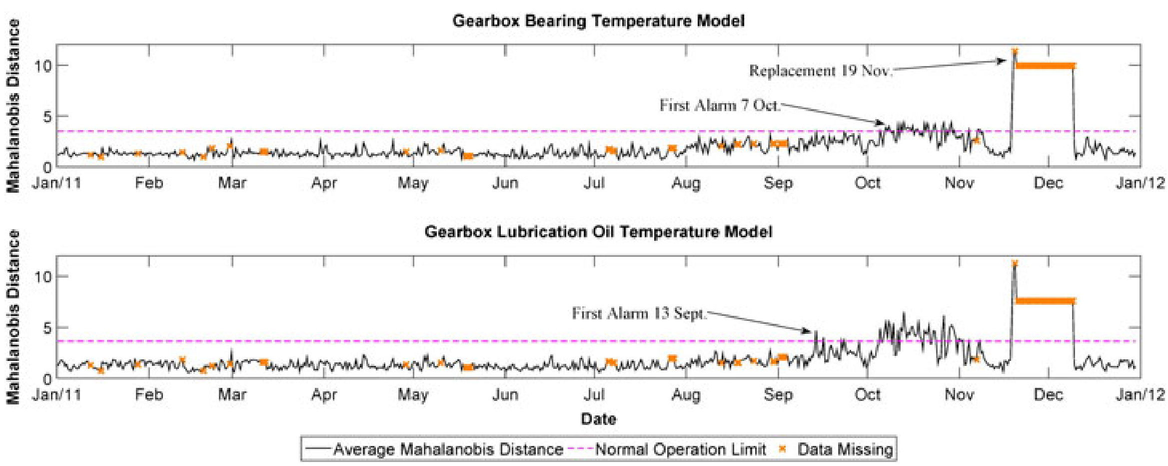


Figure 13. Anomaly detection output for Turbine-B. [Colour figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1002/we

Turbine-A is presented in Figure 12. It can be observed that even with a low threshold value of 1.5°C, the absolute error value does not cross the threshold, and hence, anomaly detection would not have been possible.

5.2. Case study for Turbine-B

The ANN models for Turbine-B were trained on data from the year 2010 and applied for anomaly detection during the year 2011, and the output is presented in Figure 13. The first alarm was seen in the gearbox lubrication oil model on September 13, 2011, and in the gearbox bearing model on October 7, 2011. Unlike the case for Turbine-A, in this case, there was no alarm from the vibration-based CMS. The gearbox was replaced on November 19, 2011, after it got stuck and the wind turbine could not be restarted. The reason for failure was a severe damage in the intermediate shaft bearing. The measured gearbox bearing and lubrication oil temperatures for the month of September, when the anomaly was first detected, are presented in Figure 14. It can be observed that the gearbox bearing and lubrication oil temperature values are within the range of normal operation. This case study demonstrates the capability of the method to detect failures that might be missed by the vibration-based CMS.

5.3. Case study for Turbine-C

The ANN models for Turbine-C were trained on data from the year 2013 and applied for anomaly detection during the year 2014, and the output is presented in Figure 15. The first alarm was observed in the gearbox lubrication oil model on March 8, 2014. The alarms continued to occur for most part of the year. Furthermore, as shown in Figure 16, in the month of June 2014, the values of the gearbox bearing and lubrication oil temperature are outside the limits of the training data provided to the ANN models.

Following a consultation with the owner, it was realized that the temperature measurement sensors were incorrectly calibrated after a routine maintenance in the wind turbine in February 2014. The situation was detected only after a SCADA alarm for a high temperature in the gearbox lubrication oil was received in October 2014. As the vibration-based CMS might not be connected to the temperature sensors, the sensor fault was not detected by the vibration-based CMS. This case study illustrates the capability of the system to indicate errors in the sensor measurements as well.

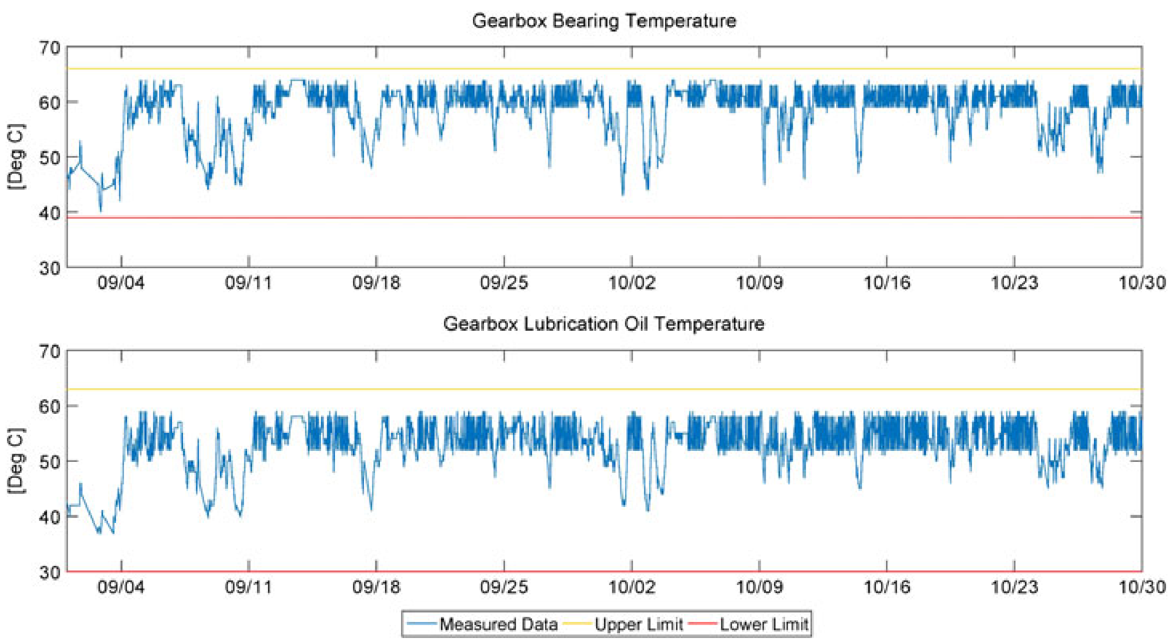


Figure 14. Output parameters recorded in SCADA for Turbine-B. [Colour figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1002/we

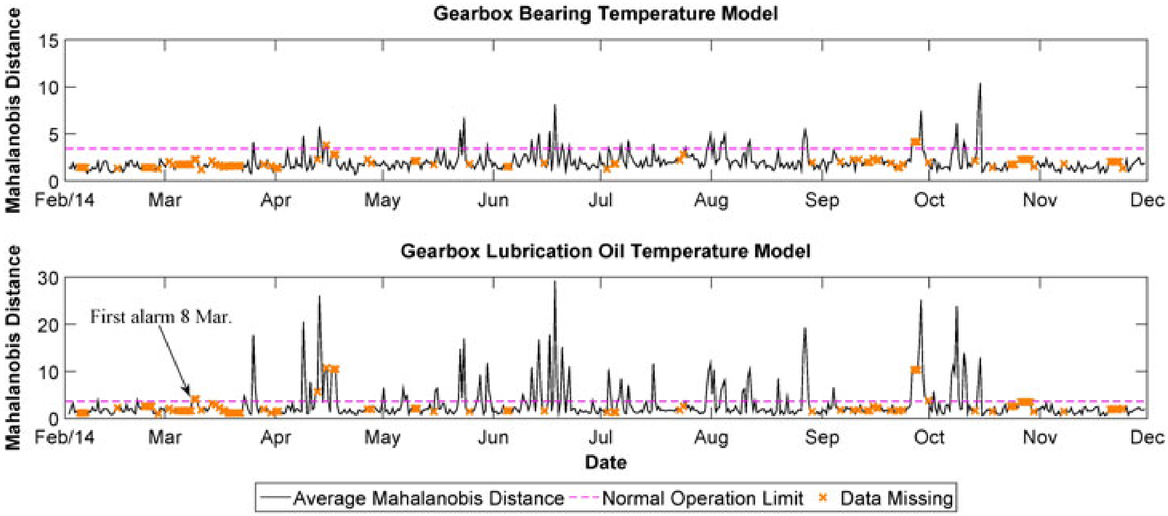


Figure 15. Anomaly detection output for Turbine-C. [Colour figure can be viewed at wileyonlinelibrary.com]

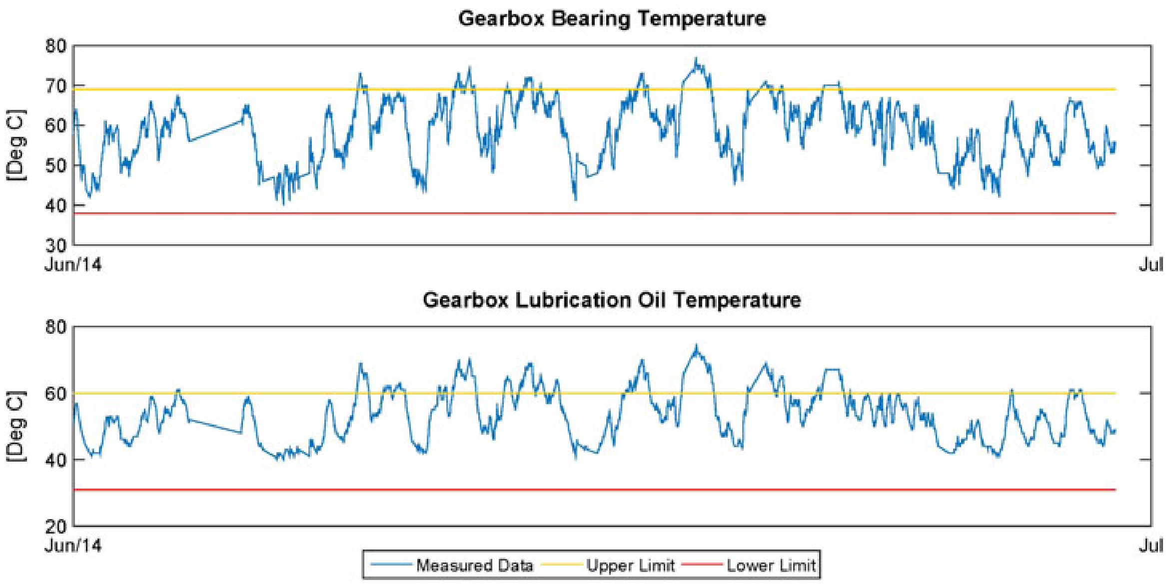


Figure 16. Output parameters recorded in SCADA for Turbine-C. [Colour figure can be viewed at wileyonlinelibrary.com]

5.4. Case study for Turbine-D

The ANN models for Turbine-D were trained on data from the year 2010 and applied for anomaly detection during the year 2011, and the output is presented in Figure 17. The ANN-based condition monitoring was not able to detect any anomaly in the system. The failure in the gearbox for Turbine-D originated in the gears of the intermediate shaft, which was detected after an inspection following an alarm from the vibration-based CMS on January 13, 2012. The gearbox was repaired on site on January 20, 2012. This case study provides an insight into the limits of the proposed method. In general, it can be inferred that the method is able to detect those failure modes of the gearbox that manifest as a change in the behavior of the gearbox bearing and lubrication oil temperatures.

DOI: 10.1002/we

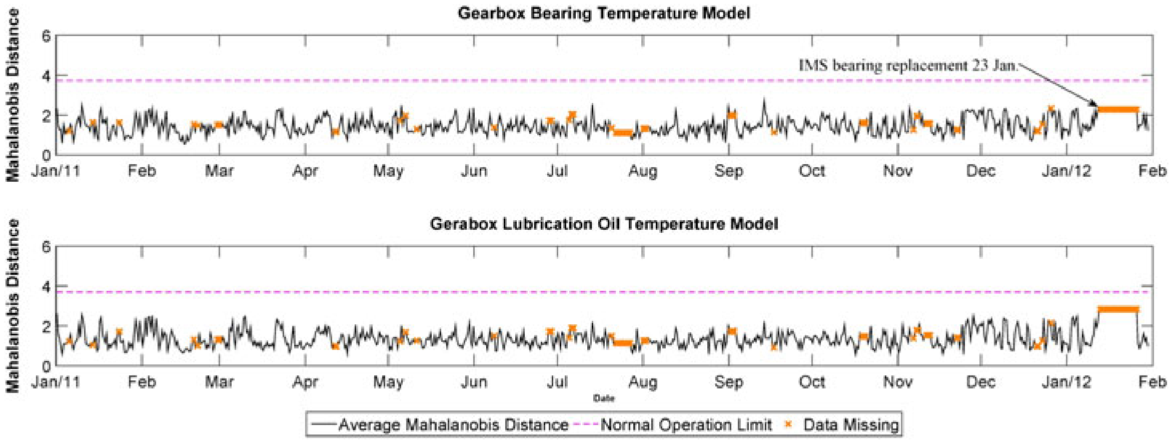


Figure 17. Anomaly detection output for Turbine-D. [Colour figure can be viewed at wileyonlinelibrary.com]

6. CONCLUSIONS

The wind turbine SCADA system collects a range of measurement data for various wind turbine components. An analysis of the recorded SCADA data can provide a better understanding of the current operating condition of critical components in the wind turbine. A literature survey was presented, which illustrated the success of methods using ANN models for the condition monitoring of wind turbine components based on data from the SCADA system.

The knowledge created through the selected publications, mentioned in the paper, was utilized and improved upon with the NARX models by focusing on two areas: the data preprocessing and the data post-processing methods. Three filtering methods were presented, which ensure that the ANN models are trained, strictly, with data that represent the true normal operating characteristics of the wind turbine. Furthermore, an additional input, called the missing data input, which transfers the information about break in continuity of SCADA data, was suggested to be included in the ANN model. The missing data input adds value to the condition monitoring by avoiding false alarms that might occur because of discontinuous SCADA data.

The anomaly detection in the ANN model application stage was improved by utilizing the MHD method, which considers the correlation between the error from the ANN model and the operating condition. The MHD enables an early detection of failures, especially in cases where the monitored temperature values do not deviate significantly from the normal operating conditions. Additionally, the inherent randomness that exists in the ANN model training and its effect on the anomaly detection were highlighted and a method for its mitigation was suggested.

Finally, case studies with application to data from four wind turbines were presented. The case study results illustrated the advantages and limitations of the proposed condition monitoring method. The proposed method can be extended for an application to other components in the wind turbine by creating different normal behavior models from the available SCADA data.

ACKNOWLEDGEMENTS

This project is financed through the Swedish Research Council (Dnr. 621-2014-5138) and the Swedish Wind Power Technology Centre (SWPTC) at Chalmers University of Technology. The purpose of the center is to support Swedish industry with knowledge of design techniques, as well as maintenance, in the field of wind power. The center is funded by the Swedish Energy Agency, Chalmers University of Technology, industry, and academic partners. The authors would also like to specifically acknowledge the support and guidance from Stena Renewable. Finally, the authors would like to thank the reviewers for their valuable comments that improved the quality of the paper.

REFERENCES

1. World energy outlook 2015, Technical Report WEO-2015, International Energy Agency, Paris, 2015. http://www.iea.org/Textbase/npsum/WEO2015SUM.pdf.
2. Lu B, Li Y, Wu X, Yang Z. A review of recent advances in wind turbine condition monitoring and fault diagnosis. *Power Electronics and Machines in Wind Applications, 2009. PEMWA 2009*, IEEE, Lincoln NE; USA, 2009; 1–7.

3. Wind turbine gearbox condition monitoring round robin study – vibration analysis, Technical Report NREL/TP-5000-54530, National Renewable Energy Laboratory, Denver, Colorado, USA, 2012. http://www.nrel.gov/docs/fy12osti/54530.pdf.
4. Garcia MC, Sanz-Bobi MA, del Pico J. SIMAP: intelligent system for predictive maintenance. *Computers in Industry* 2006; **57**(6): 552–568.
5. Zaher A, McArthur SDJ, Infield DG, Patel Y. Online wind turbine fault detection through automated SCADA data analysis. *Wind Energy* 2009; **12**(6): 574–593.
6. Kusiak A, Verma A. Analyzing bearing faults in wind turbines: a data-mining approach. *Renewable Energy* 2012; **48**: 110–116.
7. Schlechtingen M, Santos IF, Achiche S. Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: system description. *Applied Soft Computing* 2013; **13**(1): 259–270.
8. Baydar N, Ball A. Detection of gear deterioration under varying load conditions by using the instantaneous power spectrum. *Mechanical Systems and Signal Processing* 2000; **14**(6): 907–921.
9. Xiang J, Watson S, Liu Y. Smart monitoring of wind turbines using neural networks. *Sustainability in Energy and Buildings - Proceedings of the International Conference in Sustainability in Energy and Buildings, SEB'09*, Brighton, UK, 2009; 1–8.
10. Bennouna O, Héraud N, Camblong H, Rodriguez M, Kahyeh MA. Diagnosis and fault signature analysis of a wind turbine at a variable speed. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 2009; **223**(1): 41–50.
11. Wang X, Makis V, Yang M. A wavelet approach to fault diagnosis of a gearbox under varying load conditions. *Journal of Sound and Vibration* 2010; **329**(9): 1570–1585.
12. Kim K, Parthasarathy G, Uluyol O, Foslien W, Sheng S, Fleming P. Use of SCADA data for failure detection in wind turbines. *ASME 2011 5th International Conference on Energy Sustainability*, ASME, Washington DC, USA, 2011; 2071–2079.
13. Gill S, Stephen B, Galloway S. Wind turbine condition assessment through power curve copula modeling. *IEEE Transactions on Sustainable Energy* 2012; **3**(1): 94–101.
14. Qiu Y, Feng Y, Tavner P, Richardson P, Erdos G, Chen B. Wind turbine SCADA alarm analysis for improving reliability. *Wind Energy* 2012nov; **15**(8): 951–966.
15. Yang W, Court R, Jiang J. Wind turbine condition monitoring by the approach of SCADA data analysis. *Renewable Energy* 2013; **53**: 365–376.
16. Feng Y, Qiu Y, Crabtree CJ, Long H, Tavner PJ. Monitoring wind turbine gearboxes. *Wind Energy* 2013; **16**(5): 728–740.
17. Qiu Y, Feng Y, Sun J, Zhang W, Infield D. Applying thermophysics for wind turbine drivetrain fault diagnosis using scada data. *IET Renewable Power Generation* 2016; **10**(5): 661–668.
18. Schlechtingen M, Santos IF. Wind turbine condition monitoring based on SCADA data using normal behavior models. part 2: Application examples. *Applied Soft Computing Journal* 2014; **14**(PART C): 447–460.
19. Bangalore P, Tjernberg LB. An artificial neural network approach for early fault detection of gearbox bearings. *IEEE Transactions on Smart Grid* 2015; **6**(2): 980–987.
20. Haykin S. *Neural Networks and Learning Machines* Third edn. Pearson International Edition: Upper Saddle River, New Jersey 07458, 2009.
21. Levenberg K. A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Applied Mathematics* 1944; **II**(2): 164–168.
22. Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* 1963; **11**(2): 431–441.
23. Hagan MT, Menhaj MB. Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks / a publication of the IEEE Neural Networks Council* 1994; **5**(6): 989–93.
24. Tarassenko L. 7 - design, training and testing of the prototype. In *Guide to Neural Computing Applications*, Tarassenko L (ed.). Butterworth-Heinemann: New York, 1998; 77 –97.
25. Karlsson D. Wind turbine performance monitoring using artificial neural networks with a multi-dimensional data filtering approach, *Masters' Thesis*, Dept. of Energy and Environment, Chalmers University of Technology, Gothenburg, Sweden, 2015.

26. Salmasi M, Mahdavi-Nasab H, Pourghassem H. Comparison of feed-forward and recurrent neural networks in active cancellation of sound noise. *2011 International Conference on Multimedia and Signal Processing*, Vol. 2, IEEE, Guilin, Guangxi; China, 2011may; 25–29.
27. Sundermeyer M, Oparin I, Gauvain J-L, Freiberg B, Schluter R, Ney H. Comparison of feedforward and recurrent neural network language models. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, Vancouver, BC; Canada, 2013; 8430–8434.
28. Kusiak A, Verma A. Monitoring wind farms with performance curves. *IEEE Transactions on Sustainable Energy* 2013; **4**(1): 192–199.
29. Ward JH. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 1963; **58**(301): 236–244.