Received: 26 October 2019 | Revised: 18 June 2020 | Accepted: 7 August 2020

DOI: 10.1002/we.2567

RESEARCH ARTICLE

WILEY

Combining SCADA and vibration data into a single anomaly detection model to predict wind turbine component failure

Alan Turnbull | James Carroll | Alasdair McDonald

Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK

Correspondence
Alan Turnbull, Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G1 1RD, UK.
Email: a.turnbull@strath.ac.uk

Present address
Alan Turnbull, Electronic and Electrical Engineering, Technology and Innovation Centre, 99 George Street, Glasgow, G1 1RD, UK.

Funding information
Engineering and Physical Sciences Research Council, Grant/Award Number: EP/L016680/1

Abstract

Reducing downtime through predictive or condition-based maintenance is a promising strategy to help reduce costs associated with wind farm operation and maintenance. To help effectively monitor wind turbine condition, operators now rely on multiply sources of data to make informed operational decisions which can minimise downtime, increasing availability and profitability of any given site. Two of such approaches are SCADA temperature and vibration monitoring, which are typically performed in isolation and compared over time for both fault diagnostics and reliability analysis. Presenting two separate case studies, this paper describes a methodology to bring multiple data sources together to diagnose faults by using a single-class support vector machine classifier to assess normal behaviour model error, with results showing that anomalies can be detected more consistently when compared to more standard approaches of analysing each data source in isolation.

KEYWORDS

anomaly detection, condition monitoring, failure, machine learning, SCADA, vibration, wind turbine

1 | INTRODUCTION

Operation and maintenance (O&M) can make up a significant proportion of total lifetime costs associated with any wind farm, with up to 30% reported for some offshore developments.[1] For this reason, it is becoming increasingly important for wind farm owners and operators to optimise their assets and extract as much value as possible in order to reduce the levelised cost of energy (LCoE). Reducing downtime through predictive or condition-based maintenance is a promising strategy which is emerging as a real possibility of realising these goals, which is made possible through increased monitoring and gathering of operational data,[2] as well as advances in condition monitoring (CM) systems.[3] With WT generator failure rates in the region of 0.076–0.14 depending on the study, and average downtime per failure second only to the gearbox, faults associated with these components have drawn interest from both wind energy researchers and practitioners over the last decade as a focal point in developing methodologies and approaches to facilitate predictive maintenance strategies.[4-7]

Abbreviations: CMS, condition monitoring system; DFIG, doubly fed induction generator; LCoE, levelised cost of energy; NN, neural network; NBM, normal behaviour model; O&M, operations and maintenance; OEM, original equipment manufacturer; RMSE, root mean square error; SVM, support vector machine; WT, wind turbine.

One approach is through the use of CM systems, with those in operation and development looking at vibration, oil debris, acoustic emission and electrical signatures, all of which have potential to detect and predict component failure through a variety of signal processing and analysis techniques.[3,8] However, with SCADA systems currently fitted to every turbine installed in the last 10–15 years, it is also imperative to extract as much information from these data as possible, especially in cases in which CM systems are not readily available. In most cases, SCADA data includes but is not limited to measurements of power output, wind speed, shaft speeds and temperatures throughout the drivetrain.

As Tautz-Weinert and Watson[9] describe with a comprehensive review, together with more recent studies,[10-14] it has been shown that the multiple approaches and machine learning models can be used to detect faults and predict failure for several components and failure modes throughout WT drivetrains. Most literature utilises SCADA data to develop a normal behaviour model (NBM) to predict temperatures, comparing model predictions with in field measurements to track error and detect anomalies. An example of this would be found in Zhang and Wang[15] and Dinmohammad et al.,[16] which much like the first case study presented in this paper, uses artificial neural networks to detect gearbox faults based on oil temperature monitoring. However, how best to track NBM error is not yet fully appreciated, with literature to date typically suggesting to compare the root mean square error (RMSE) over a chosen time period to the RMSE of the training period. Simple thresholds can be set based on training RMSE to determine anomaly rates leading up to failure,[9,17,18] although other metrics have been suggested such as using Mahalanobis distance,[19] or health degree based on probability.[20]

Demonstrated through the first case study, this paper presents a methodology to robustly track and analyse these errors through time by applying a single-class support vector machine (SVM) classifier to create a more complex decision boundary based on multiple error statistics over a given time period, which is a proven technique used in different scenarios to detect a wide variety of faults.[21,22]

For the second case study, the same methodology is adapted to utilise both SCADA and vibration data to develop a combined anomaly detection model to diagnose a generator bearing fault. With regard to detecting generator faults through vibration, techniques typically are based around Fourier analysis, Hilbert transform and order analysis techniques, which can all be used to identify and extract features from the signal based on rotational frequencies.[23] Zhang et al. introduced the idea of using a variety of techniques to detect bearing faults based on vibration analysis using a direct-drive machine.[24] Using time-domain analysis and the probability density of acceleration along with frequency domain and cepstrum analysis, a fault could be detected in the bearing; however, the process was unable to determine specific cause or location. In another example by Chen et al., a wavelet transform was used to successfully identify a generator bearing fault in both lab conditions and on an operational WT.[25] However, one of the key difficulties of combining SCADA and vibration sources is differing sample rates, in which offering a potential solution is the key motivation behind this paper. In this study, SCADA data were sampled at a constant frequency of 1 kHz and mean values stored at 10-min intervals, while the vibration data were measured with a frequency of approximately 25 kHz for approximately 10 s with one sample stored weekly. Each source has the potential to detect generator bearing failure: in SCADA through bearing temperature and in vibration through rotational and ball passing frequencies. This brings about the second challenge of combining models, while temperature can be extracted and modelled directly from the SCADA data, additional feature extraction is required from the raw vibration signal to gain any meaningful insight into component health. To achieve a combined model, an SVM classifier is again used to create a decision boundary by considering multiple error metrics over a given time period. If a time period is chosen in line with the data source with the lowest resolution, multiple error metrics from different NBMs can therefore be combined into a single model, navigating the issues highlighted above. In this case, generator drive-end bearing temperature and vibration are combined into a model considering weekly error metrics.

Primarily, this paper provides a novel approach to combine multiple normal behaviour models that use different diagnostic indicators from different data sources to detect the same fault. This is presented and discussed through the second case study, in which temperature- and vibration-based models are combined into a single anomaly detection model that can detect deviations in both fault indicators. However, as demonstrated with the first case study, in cases where only a single fault indicator from a single data source is available or required, this approach can still add value to combine and analyse multiple error metrics using a single anomaly detection classifier. This approach could be used and built upon for CM purposes to reduce uncertainty and assess component fault severity across multiple fault indicators.

Section 2 describes the high-level methodology undertaken in the research study before setting out the failure mode and data available in each case study described above. In Section 3, the first case study is presented, outlining both the NBM and classifier used to detect anomalies using SCADA data, before presenting results and comparing these to standard approaches discussed earlier in this introduction. Section 4 then presents the second case study and describes how more data sources can be added to the model by looking at both vibration and SCADA analysis. The two case studies are then compared and key advantages and disadvantages of the methodology discussed in Section 5 conclusion.

2 | METHODOLOGY AND CASE STUDIES

2.1 | Overall methodology

Although the general approach is the same for each case study, depending on how many data sources are available in each case, the number of required NBMs will change accordingly. Sections 2.2 and 2.3 will present each methodology, highlighting the similarities and differences when accounting for different sources of data.

2.2 | Case study 1—gearbox HSS failure

The first failure mode used for this study is associated with the high-speed shaft (HSS) of the gearbox. The initial point of failure was the HSS; however, on inspection, noticeable damage was also observed to the gears, with the gearbox also requiring to be replaced. Root cause analysis suggested that the failure occurred due to misalignment causing out of plane loads both on the shaft and through the gearbox. For confidentiality reasons, the exact type and model cannot be explicitly stated; however, the WT in question has a doubly fed induction generator (DFIG) and is of between 500 kW and 1 MW at rated power. It has a pitch regulated variable speed control strategy and had been operated onshore for over 15 years prior to failure.

Once the failure date was determined from the O&M log, 10-min averaged SCADA data for 18 months leading up to failure was collected. The NBM model for the first case study consisted of only SCADA data as no CM systems was installed.

The full methodology is described in Figure 1. The first step in the process is to develop the NBM using a neural network (see Section 3.1.1 for model details), which takes a variety of input features to predict gearbox oil temperature. Input features were chosen which could adequately describe WT performance, as well as relevant temperatures throughout the nacelle and components related to the gearbox and generator. Once error had been minimised through the training and validation process (as described in Section 3.1.2), the error between model prediction and measured temperature is then calculated for each time step.

The second stage of the process is to evaluate the error output and determine a threshold which can adequately distinguish between normal and anomalous behaviour. To achieve this, error metrics were first calculated to describe the error distribution over a chosen time period (for this case study, daily metrics were calculated). For the same training data set used for the NBM, a single-class SVM classifier was trained using the error metrics to develop a decision boundary, as described in Section 3.2. While other classifiers could have been chosen, an SVM model was deemed well suited to this specific application, mainly due to the models inherit ability to establish a continuous, complex decision boundary in which a spatial representation can easily be visualised.

Once both models have been trained and validated, they could then be used on new data points to detect anomalies. This is done by using the NBM to first predict temperature based on the same inputs, assessing the error between predicted and measured value, before finally feeding those error metrics into the classifier to determine if the new data point was normal or anomalous. The purpose of presenting the first case study is to prove the methodology with a single data source, showcasing how it can outperform simpler approaches to evaluate NBM error such as using RMSE as single metric. A summary of the data used in this case study can be found in Table 1, which describes both raw data coverage and data available after cleaning.

2.3 | Case study 2—generator bearing failure

The second case study focuses on generator bearing failure, which in this case stems from raised bearing temperatures leading to bearing inner ring growth resulting in the bearing inner ring spinning on the generator rotor shaft at the drive end, with Figure 2 showing damage that can occur

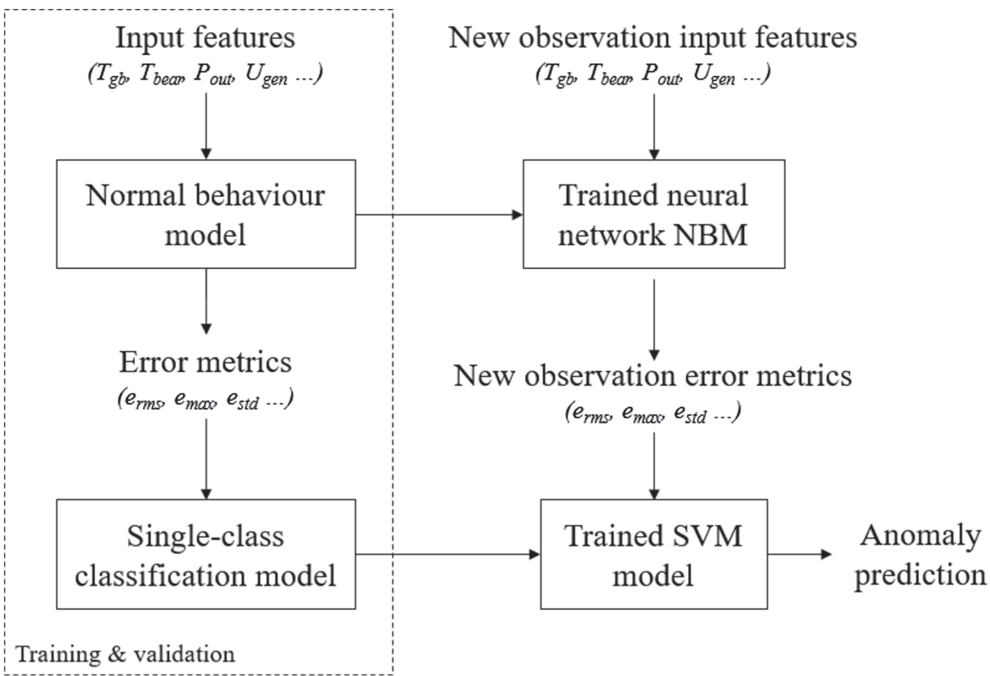


FIGURE 1 Methodology for case study 1

TABLE 1 Case study 1—summary of data

Model phase	Raw data (10 min)	Averaged data (1 h)	Cleaned data
Model development (12 months)	48,096 (92% coverage)	8016	6974
Model development—Training (70%)	-	-	4882
Model development—Validation (15%)	-	-	1046
Model development—Testing (15%)	-	-	1046
Model implementation (6.4 months)	26,634 (94% coverage)	4439	3695

FIGURE 2 Damaged shaft from inner ring of bearing spinning on it (left) and damage to inner ring from spinning on shaft (right) [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Case study 2—summary of SCADA data

Model phase	Raw data (10 min)	Averaged data (1 h)	Cleaned data
Total (12 months)	52,104 (99% coverage)	8785	5275
Model development (6 months)	26,052	4391	2634
Model implementation (6 months)	26,052	4394	2641

TABLE 3 Case study 2—summary of vibration data

Sensor location	No. of samples	Frequency	Sample time
Generator drive end	56 (1 week apart)	Approx. 25 kHz	Approx. 10 s
Generator non-drive end	56 (1 week apart)	Approx. 25 kHz	Approx. 10 s

over time from a fault of this nature. In order to ensure confidentiality, again, the exact power output and wind turbine and bearing type used is not provided; however, it can be stated that it was a DFIG with a rated power of between 2 and 4 MW, again utilising a variable speed, pitch regulated control strategy.

Once the failure was identified in the wind turbine OEM O&M event log, both SCADA data and vibration data were retrieved for the year leading up to failure. To ensure dates collected were correct, SCADA data were checked directly after the failure date to ensure wind turbine downtime occurred as stated in the events log. SCADA data consisted of 10-min mean values taken continuously over the entire year, while vibration data samples were taken 1 week apart at both the drive end and non-drive end of the generator. Each vibration sample consisted of approximately 10 s of data taken with a sampling frequency of approximately 25 kHz. A summary of SCADA and vibration data available for the second case study can be found in Tables 2 and 3, respectively.

The methodology for second case study, which can be found in Figure 3, largely remains the same, however will build on the first case study described in Section 2.2 by encompassing two data sources. NBM 1, which is based on SCADA analysis, predicts generator bearing temperature relying on multiple operational parameters and temperatures as inputs using a random forest regressive model (see Section 4.1 for details) and compares these predictions with observed temperature measurements. NMB 2, which is based on vibration analysis, predicts a summation of fault frequency amplitudes and relies on a variety of spectral features and operating conditions as inputs. Error is evaluated for both models and metrics describing the error distribution are calculated for the same weekly time resolution in each case. These error metrics are then used as inputs features to train a single-class SVM model for anomaly detection.

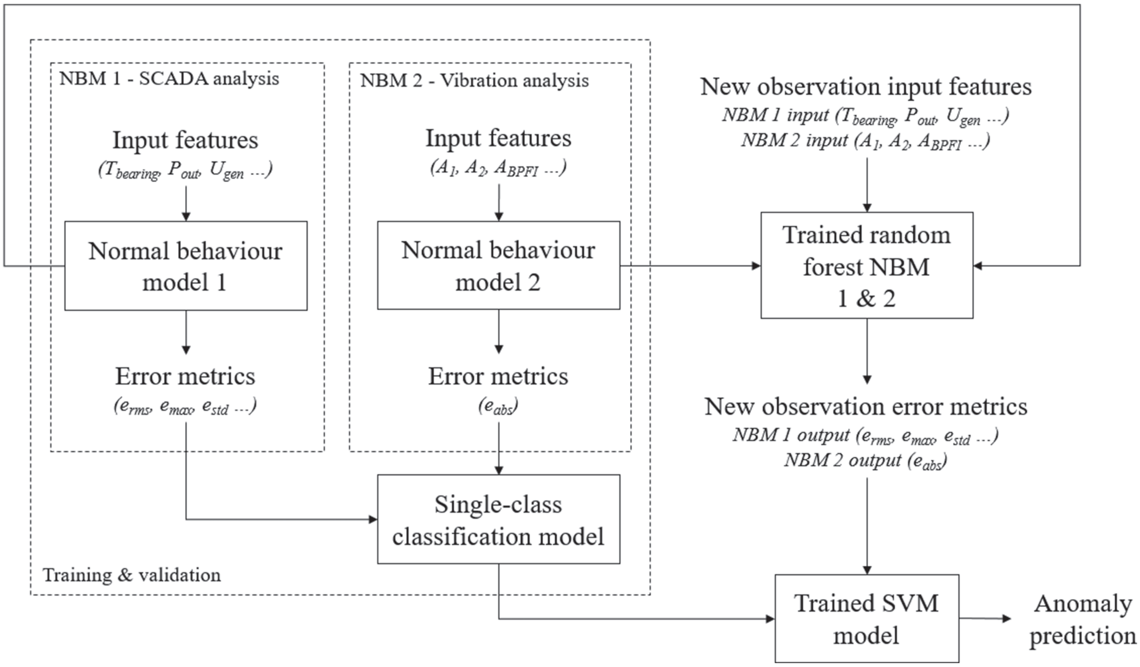


FIGURE 3 Methodology for case study 2

TABLE 4 Case study 1—model features

Feature No.	Feature	Description	Layer
1	P_{out}	Power output	Input layer
2	U_{wind}	Wind speed	Input layer
3	U_{hss}	High-speed shaft speed	Input layer
4	T_{nac}	Nacelle temperature	Input layer
5	T_{gen}	Generator phase temperature	Input layer
6	T_{bear}	Bearing temperature	Input layer
7	T_{gb}	Gearbox oil temperature	Output layer

3 | CASE STUDY 1

3.1 | Normal behaviour model

3.1.1 | Model description

A two-layer feed-forward neural network was utilised for the NBM which had an input layer consisting of six WT operational parameters to predict a single output (Feature No. 7), as described in Table 4. Model input features include parameters to describe the operating conditions (Features No. 1–3) and temperature distribution throughout the nacelle (Features No. 4–6). As previously stated, the model output parameter used for anomaly detection was the gearbox oil temperature. Additional model parameters for the neural network were chosen to reflect the number of input and output features required[26] which in this case utilised single hidden layer with seven neurons. This provided a balance between the accuracy of prediction and computational time for training, and although this was not a restricting factor due to the relatively small data set, there was no need to go beyond seven neurons for this application due to the accuracy achieved. Figure 4 shows a diagram of a two-layer feed-forward neural network used for this case study.

Of the 18 months of SCADA data that was gathered prior to failure, the initial 12 months (which we know had no serious faults through analysing the O&M logs) was used for training and validating the NBM, with the final 6 months used to test and track error leading up to failure. From the initial data set consisting of continuous 10-min mean values, the data were cleaned to remove any periods of curtailment or downtime due to scheduled maintenance. The hourly averages were then calculated to remove some of the lower resolution fluctuations which can decrease model

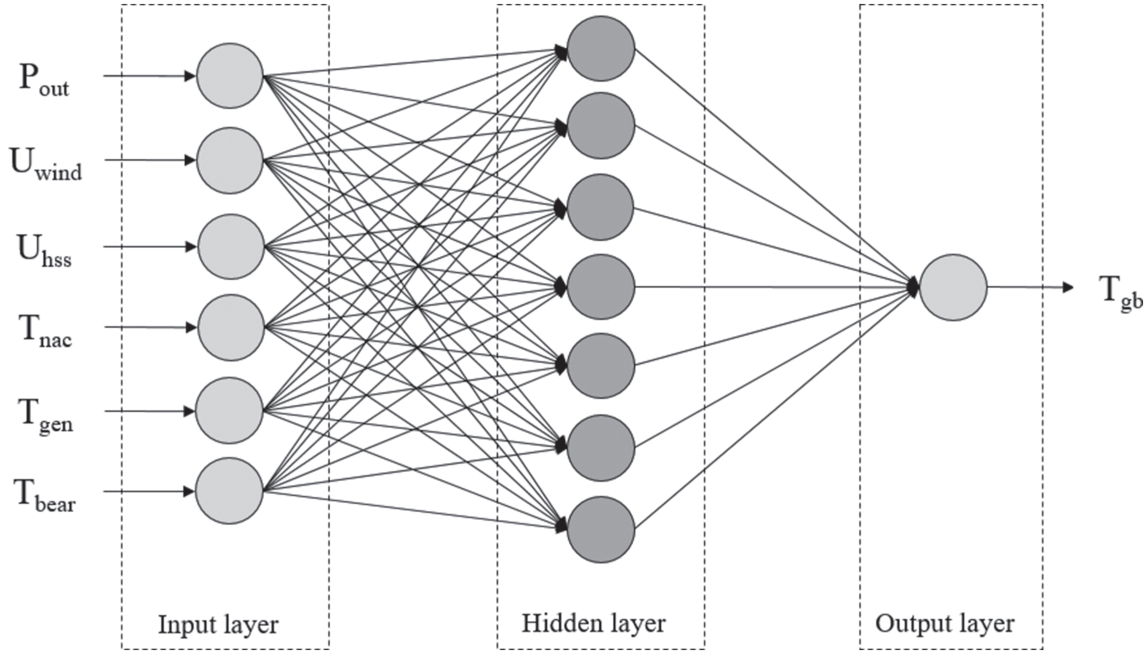


FIGURE 4 Schematic of two-layer feed-forward neural network

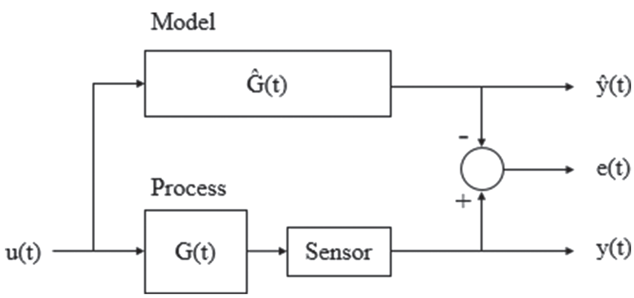


FIGURE 5 Model-based monitoring based on NBM[9,27]

accuracy, allowing the model to be more tailored towards longer time behavioural trends. Once this process had been completed, this left a total of 6974 samples for training and validating the model, with an additional 3695 samples left for testing and tracking the error prior to failure.

The NBM works by empirically modelling the gearbox oil temperature based on the input parameters described in Table 1. The process is summarised in Figure 5, where $u(t)$ are the input variables at timestep t, $\hat{G}(t)$ represents the data-driven NBM to predict target variable, $\hat{y}(t)$ while $G(t)$ constitutes the process of obtaining the measured target variable $y(t)$ through the required in field sensor. Finally, $e(t)$ represents the error between the predicted and measured value.

3.1.2 | Training and validation

Out of the 3695 samples in the 12-month training data set, 70% was used for training and 15% for validation, with the remaining 15% used to test the model independently as described in Table 1. From experience, this provides a good balance between the number of samples required for training, validating and testing the model. During the training phase, the training and validation samples are chosen at random and then fed into the neural network, which is adjusted in line with the error between predicted and known values of the target variable. It is then validated with the validation samples, and mean square error is calculated for the new data points. This process is repeated until the mean square error no longer increases for the validation data set, indicating that the neural network generalises well and is no longer over-fitting to the training set. Once this has been achieved, the model is then independently tested with the testing data set (the 15% of data leftover after randomly selecting training and validation data) to ensure generalisation.

To get a sense of the error distribution for this particular model, a histogram can be observed in Figure 6, showing the model error over the 12-month training period, broken down into the phases described above. The mean squared error over the training, validation and testing phases were 2.364, 2.254 and 3.075, respectively, with correlation coefficient (r) values of 8.585, 8.592 and 8.316. With R values deviating approximately 3.2% between data sets, this model generalises relatively well, which sets a good foundation when trying to detect anomalies leading up to failure.

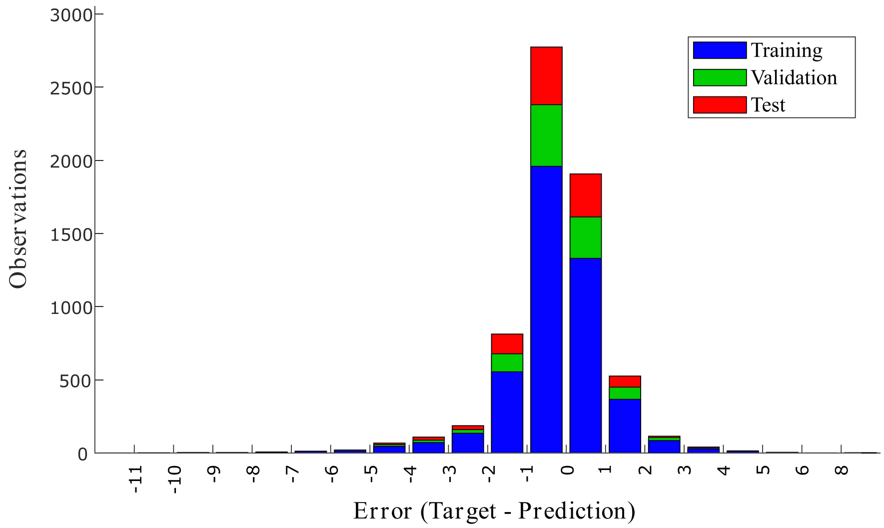


FIGURE 6 Distribution of error from 12-month training period [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 5 SVM error model feature

Feature No.	Feature	Description	Input	Case study
1	e_{rms}	RMSE	SCADA	1 & 2
2	e_{min}	Min error	SCADA	1 & 2
3	e_{max}	Max error	SCADA	1 & 2
4	e_{std}	Standard deviation of error distribution	SCADA	1 & 2
5	$e_{kurtosis}$	Kurtosis of error distribution	SCADA	1 & 2
6	e_{abs}	Absolute error	Vibration	2

3.2 | Single-class SVM model

Once a baseline expected error has been established from the training period, there are several ways in which to then compare residuals moving forward. Typically, this is achieved by comparing the daily or weekly RMSE with the RMSE of the training period to give an indication of whether the temperature (or other chosen parameter) is acting as predicted. This approach however does come with limitations, stemming mainly from the fact that only one parameter is used to describe an error which, over any particular time period is multifaceted and has a unique error distribution associated with it. More than one parameter can be looked at in isolation, or indeed the entire distribution could be described and tracked in time; however, this introduces a different problem: how to robustly set thresholds that will indicate the fault. Using a single-class SVM aims to address these limitations, first of all by considering multiple parameters that can effectively describe the distribution of error over a chosen time period and second, to set more complex boundaries that can more precisely describe the threshold to indicate a fault.

A single-class SVM model was developed to evaluate the error distribution each day based on the NBM output. For each daily error distribution, the parameters stated in Table 5 were calculated and used as inputs to the SVM model.

To begin with, let us consider a single-class SVM model with only two features as an example. Using RMSE and maximum error, the model can be trained with the cleaned 12-month data set, giving a total of 290 samples. Figure 7 shows each observation in the trained model, with support vectors (which influence the decision boundaries) circled in red. On the right-hand side of Figure 7, there is a scale that corresponds to the contour line boundaries indicating an anomaly. In the model shown, any score below zero would constitute an outlier or anomaly, which lies outside the decision boundary. The more negative a score, the further away from the central cluster the observation is. This process was then repeated with all parameters which describe the daily error distribution as shown in Table 5. Although these models could not be visualised in the same manner, the scoring system remained the same and could be used to detect outliers in new observations leading up to failure. The model was trained to recognise 1% of data as anomalies in the training period; therefore, a similar percentile would be expected moving forward if no fault was present in the system.

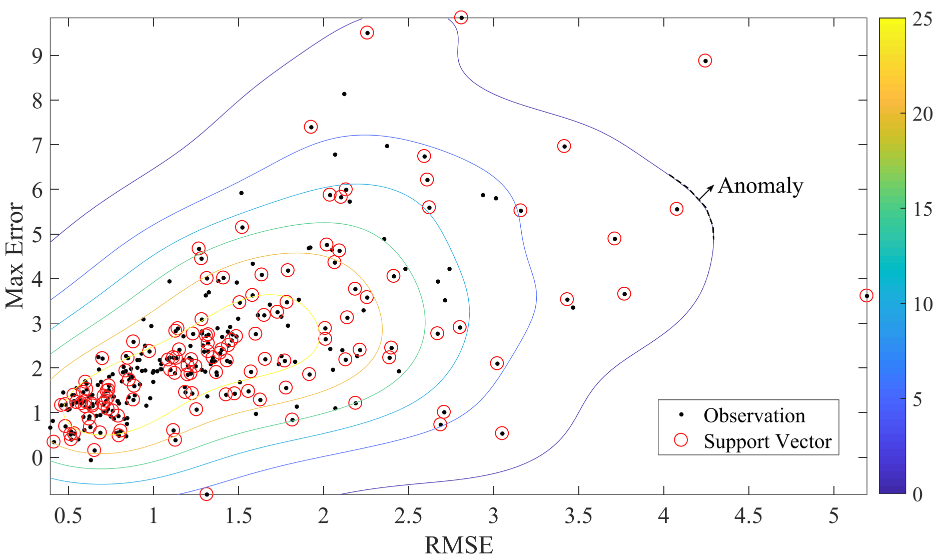


FIGURE 7 Trained SVM model with only two features: (1) RMSE and (2) max error [Colour figure can be viewed at wileyonlinelibrary.com]

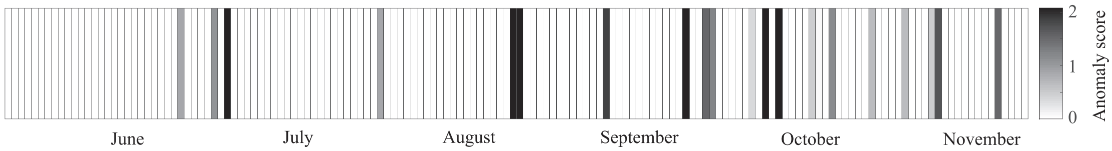


FIGURE 8 Heatmap of anomalies leading up to failure

TABLE 6 Anomaly rate using single-class SVM approach

Date	Number of anomalies	Percentage of anomalies	Turbine active days per month (days of month in data)
May	0	0%	22 (22)
June	2	8%	25 (30)
July	1	3%	30 (31)
August	4	13%	31 (31)
September	7	24%	29 (30)
October	6	25%	30 (31)
November	1	14%	7 (7)

3.3 | Anomaly detection results

Using the models described in Sections 3.1 and 3.2, the error over the 6 months leading up to failure was evaluated. For every 1-h time step, the error output was determined and the daily statistics shown in Table 5 calculated. If features fed into the single-class SVM model were considered part of the same class (of normal behaviour), it resulted in a positive score and was assigned an anomaly score of zero. If the model output score was below zero, this was considered an outlier. In order to create a heatmap, this score was then inverted to give a positive anomaly score that provides an indication of fault severity or distance from the determined frontier that defined the class.

Figure 8 shows a heatmap of the daily anomaly score leading up to failure. A total of 154 days are shown, which accounts for all 24-h periods in each month while the turbine was operating under normal operating conditions. If the WT was curtailed, shutdown due to planned maintenance activities or the wind was below cut in speed or above cut out speed for an entire 24 h, data were removed during the cleaning process described in Section 3.1 If during the 24-h period the WT partially experienced normal operating conditions, these data were still used to calculate the daily error metrics.

Table 6 gives a breakdown of anomalies detected per month along with associated percentages taking into consideration the number of active days. In general, results show that anomalies increase towards failure, with a maximum of 8% per month in the 4–6 months before failure, increasing to 25% in the 2 months directly before failure. In terms of detection time, the first meaningful and consistent anomalies are observed in

TABLE 7 Anomaly rate using standard RMSE approach

Date	Number of anomalies (thresholds 1, 2, 3)	Percentage of anomalies (thresholds 1, 2, 3)
May	8, 1, 0	36%, 4.5%, 0%
June	11, 1, 0	44%, 4%, 0%
July	9, 0, 0	30%, 0%, 0%
August	11, 2, 0	35.5%, 6.5%, 0%
September	14, 6, 2	48%, 20.7%, 6.9%
October	23, 2, 0	76.6%, 6%, 0%
November	4, 1, 0	57%, 14.3%, 0%

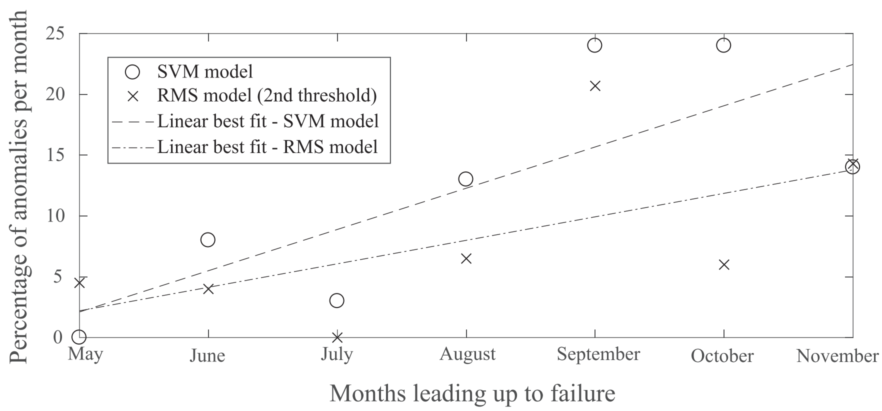


FIGURE 9 Comparison of SVM classification approach to RMSE

August, giving approximately 3 months lead time. A slight decrease in anomalies are reported in November just before failure; however, this could be down to the significantly less active days (1 week) in the month to assess and detect anomalous behaviour before failure occurred.

3.4 | Comparison of results with other methods

It is important to compare results obtained using this methodology with standard approaches observed in literature.[9,16,18] To do this, the RMSE and standard deviation was first calculated of the entire training period, which could then be compared to the daily RMSE for the 6 months prior to failure. Three different thresholds were used to compare the daily RMSE with the training period; the first threshold was simply the RMSE, the second was 1 standard deviation above the RMSE, while the third threshold was 2 standard deviations above the RMSE. This gave error thresholds of 1.57, 3.15 and 4.72, respectively, whereby any new observation with a daily RMSE of above these limits was deemed an anomaly. Results using all three thresholds individually are shown in Table 7.

The first threshold shows a clear increase in anomalies leading up to failure; however, with such large proportion of errors across the entire 6 months, in practice, this would lead to a large amount of false alarms. The second threshold actually performs relatively well, with an average of less than 5% of anomalies detected over the first 2 months. Moving closer to failure also results in a clear increase in anomaly rate using this threshold; however, it does not perform as well as the SVM model presented in Section 3.3. This is best observed through Figure 9, which shows a direct comparison of results. By plotting a simple linear regression of each set of results, it becomes obvious that the SVM model not only detects on average more anomalies per month but is also more consistent over the final 2 months. It is also important to note that both models have a very low anomaly detection rate in May, with the SVM model actually detecting less anomalies than the threshold based on the RMSE and standard deviation, giving confidence that increase in anomaly rate actually relates to the fault. Using the third threshold, anomalies were only detected for a short spell 2 months prior to failure, indicating that the limit was set too high to track progression, however does offer an insight into the largest changes in behaviour, albeit for a short period of time.

3.5 | Conclusions and discussion

Results have demonstrated that single-class SVM models are a successful method to evaluate errors output by a NBM, which in this case used a two-layer feed-forward neural network. Using a SVM model acts as an improvement to existing techniques by first of all allowing multiple

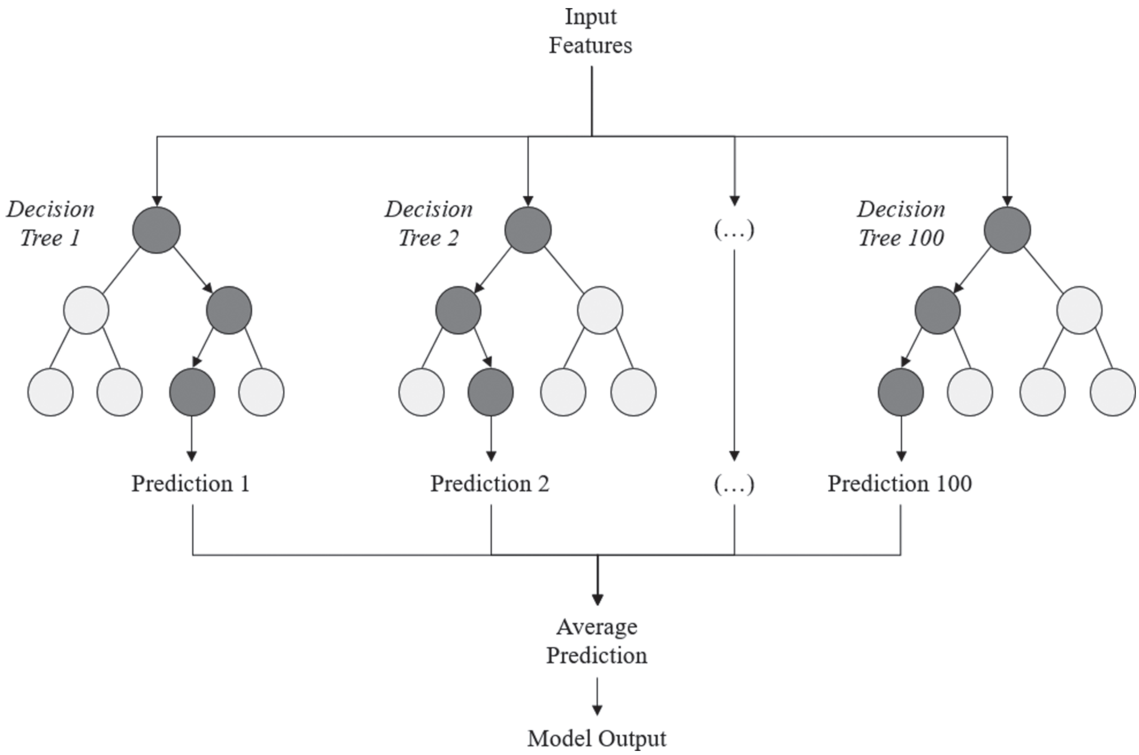


FIGURE 10 Schematic of random forest regressive model—input features and output dependant on model

parameters to be used which can effectively describe the error distribution over a chosen time period. Second, it allows for a more complex decision boundary to be formed which can detect and distinguish anomalous behaviour from normal behaviour. For this first case study, results showed that anomalies in gearbox oil temperature can be detected up to 3 months before the HSS failure. It also shows a larger and more consistent increase when compared to simple thresholds based on RMSE alone. The decrease in anomaly rate observed in November could be attributed to the fact there was only 7 days, therefore less data in which to make the calculation.

4 | CASE STUDY 2

4.1 | Normal behaviour model

For the second case study, both SCADA and vibration are used; however, for the NBM, independent analysis was performed for each data source following the same methodology as used previously. Again, Figure 5 illustrates the process of using these types of models to detect faults, with the residual error, $e(t)$, this time between either measured temperature or vibration, $y(t)$, and model prediction $\hat{y}(t)$ used as an indicator for a potential fault.

Although a neural network was used in the first case study, in reality, there are many supervised regressive machine learning models that could be chosen to represent normal behaviour of the target variable. To highlight this fact, a random forest algorithm was selected for the second case study. Random forests are ensemble learning models that build multiple decision trees and merge them together to get a more accurate and stable prediction. Using ensemble methods allows for overall better performance and reduces the risk of over-fitting, which can be an issue when utilising solitary decision trees. Figure 10 shows a diagram of this type of random forest regressive model in relation to model input features and prediction output.

4.1.1 | SCADA NBM description

The SCADA channels available for this analysis are stated in Table 8. Like the previous study, data describes the operating conditions and power output at each timestamp, as well as a range of temperatures throughout the nacelle at selected locations, this time to give a representation of overall generator performance and condition at each instance. Feature importance was determined through a combination of an iterative model based approach and expert knowledge, which will be discussed in greater detail throughout Section 4 as the NBM is described.

TABLE 8 Case study 2—SCADA model features

Feature No.	Feature	Description	Model
1	P_{out}	Average power output	Predictor
2	U_{rotor}	Average rotor speed	Predictor
3	U_{gen}	Average generator speed	Predictor
4	U_{wind}	Average wind speed	Predictor
5	$T_{ambient}$	Average ambient temperature	Predictor
6	$T_{nacelle}$	Average nacelle temperature	Predictor
7	$T_{slipring}$	Average generator slip ring temperature	Predictor
8	T_{phase1}	Average generator phase 1 temperature	Predictor
9	T_{phase2}	Average generator phase 2 temperature	Predictor
10	T_{phase3}	Average generator phase 3 temperature	Predictor
11	$T_{bearing}$	Average bearing temperature over	Response

TABLE 9 Case study 2—vibration model features

Feature No.	Feature	Description	Description
1	A_1	Amplitude of peak at order number 1	Predictor
2	A_2	Amplitude of peak at order number 2	Predictor
3	A_3	Amplitude of peak at order number 3	Predictor
4	P_{out}	Generator power output at time of vibration sample	Predictor
5	$A_{diagnostic}$	Sum of amplitude peaks at fault frequencies	Response

As with the first case study, data cleaning is a crucial component in producing any NBM which can detect anomalies effectively. For this case study, data were cleaned to remove times in which the wind turbine was operating in conditions out with the normal torque speed operating strategy. Primarily, this was any sustained periods of shutdown or power derating, which could occur for a number of reasons including maintenance work or grid constraints.

4.1.2 | Vibration NBM description

Unlike the SCADA data, in which features can be taken directly, features must first of all be identified and extracted from the vibration signal. The features that are extracted depend on the failure mode that is trying to be detected, with different failures having different markers, or fault identifiers, within the signal. The raw time domain signal is not sufficient to extract these identifiers; therefore, frequency domain analysis is required, where the time domain signal is converted into the frequency domain using a fast Fourier transform (FFT).

To add additional complexity, the generator shaft speed varies over the 10-s sample, meaning that the signal is not stationary (as FFT analysis requires). Order analysis was therefore performed, which is a resampling technique aimed to mitigate the effects of this on the FFT spectrum. More information about this technique and its applications can be found in previous studies,[23,28,29] but it essentially takes a signal with varying rotational speed and constant sample rate and resamples with a constant rotational speed and varying sample rate. It then uses a short-time Fourier transform to compute the order spectrum, where order number 1 is a reference speed, in this case, the average shaft rotational velocity of the generator. Table 9 highlights the input and output parameters used for the vibration based NBM. Input predictors consisted of the power output at the time the vibration sample was taken, along with vibration amplitudes at shaft frequency and two harmonics. The output response, or diagnostic, was the sum of vibration amplitudes at known fault frequencies, which in this case was the ball passing frequency of the inner race (BPFI) together with two harmonics and associated side bands at shaft frequency. All amplitudes were identified and values extracted in the order domain.

4.1.3 | NBM training and validation

The NBM developed for both SCADA and vibration data were trained, validated and tested with a similar methodology as described in case study 1 (see Section 3.1.2 for details). However, because vibration samples were only taken 1 week apart, there was less data available in this case study to validate and test the model; therefore, fivefold cross validation was used. This means that 80% of data was used for training, with 20% kept for independent validation. This process was then repeated with different data splits and an average error and *r* value used as a performance indicator. The key hyperparameters and performance indicators for each model are shown in Table 10.

4.2 | Single-class SVM model

Features describing the error from both the SCADA and vibration NBMs were used for the classification model. The features used for the second case study are described in Table 5 alongside those used in the first. The key difference in case study 2 being the fact that weekly stats were used instead of the daily features used in the first case study. This is in line with the lowest data resolution of the vibration samples; therefore, only a single feature could be used to train the classifier from the vibration NBM. For this particular case study, sample size was deemed more important than the number of features per sample, which could be increased, for example, if monthly metrics were used instead.

Figure 11 shows an example of the trained SVM classifier for the 6 months of data using the weekly SCADA RMSE and vibration absolute error. This was again extended to include all features described in Table 5 in order to get a more comprehensive understanding of weekly error distribution. It should be noted that although the scale on Figure 11 ranges from 2 to −1.5, zero remained the learned frontier, with 1 data point from the training data in this example outside this threshold.

4.3 | Anomaly detection results

Results show that anomalies were detected consistently 16 weeks prior to failure, with errors associated with the highest anomaly score appearing between 10 and 12 weeks before failure. As with the previous case study, if the new observation was deemed normal behaviour, it was simply given a score of zero, with anomalies given a higher score the further from the decision boundary they stray. Figure 12 shows the heatmap of anomalies for approximately 10 months leading up to failure.

TABLE 10 NBM training and validation parameters for both SCADA and vibration data

Parameter	SCADA model	Vibration model
No. of training samples	1000	30
Random forest branch No. estimators	100	100
Random forest branch max depth	None	None
Random forest branch max features	4	3
Random forest min samples	3	3
MSE—training and validation	4.93	0.02
R value—training and validation	0.88	0.71

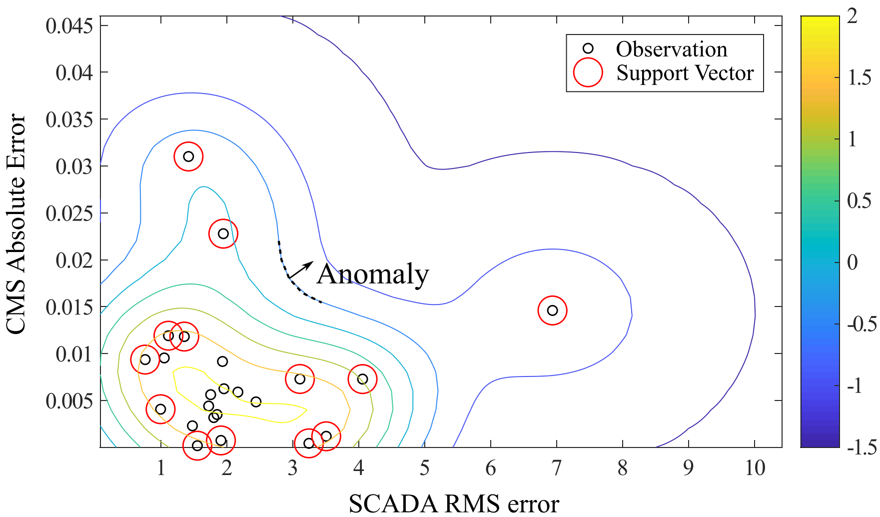


FIGURE 11 Trained SVM model with only two features: (1) SCADA RMSE and (2) vibration absolute error [Colour figure can be viewed at wileyonlinelibrary.com]

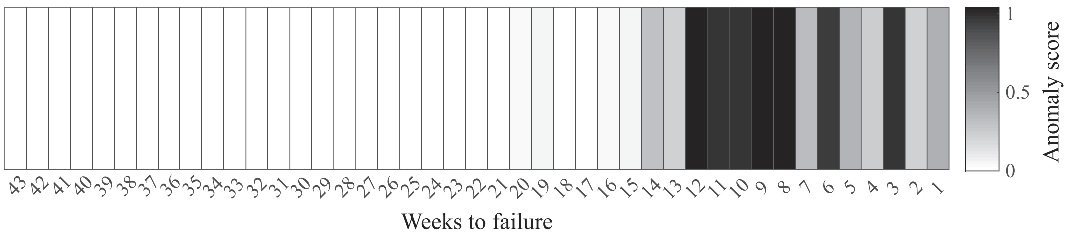
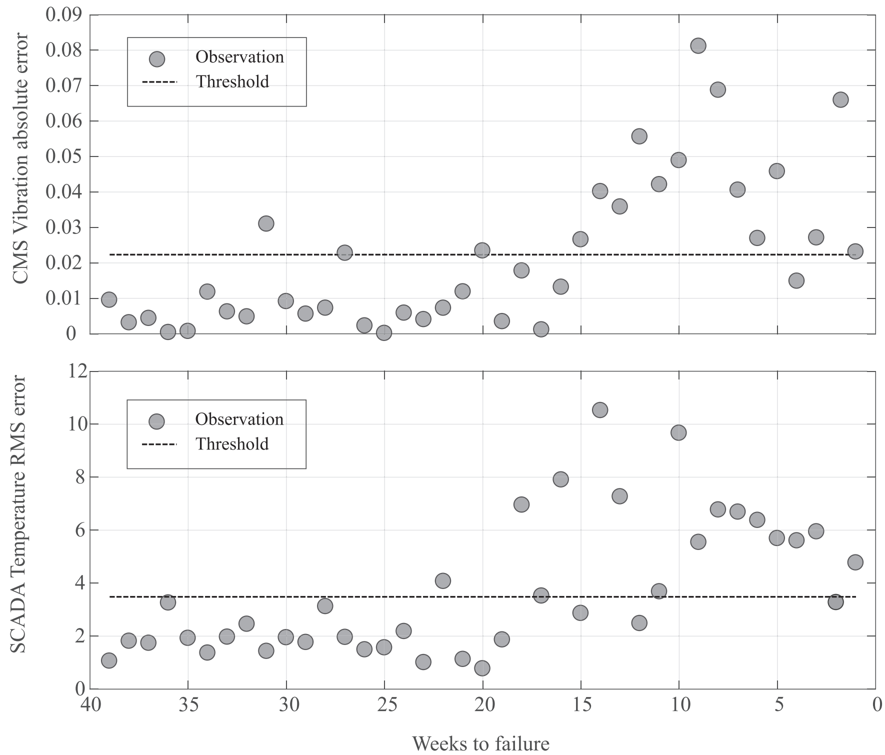


FIGURE 12 Heatmap of anomalies leading up to failure

FIGURE 13 Plot of weekly errors leading up to failure for individual NBMs



4.4 | Comparison with single models

For completeness, the results from Section 4.3 have been compared with each NBM individually using a single threshold based on the RMSE and standard deviation of the training period (see threshold 2 in Section 3.4 for details). Figure 13 shows the weekly error for each model along with the calculated threshold. Results shows an initial step change in vibration levels at approximately 14 weeks to failure, with an increase in temperature actually occurring in the 3 weeks prior to this. The heatmap shown in Figure 12 demonstrates that the SVM model captures changes in both temperature and vibration. This rise is then followed by a slight decrease in average vibration and temperature in the 7 weeks directly before failure, which is again in line with Figure 12.

4.5 | Conclusion and discussion

Results show that an SVM classifier is an effective method in evaluating the error from two different NBMs that can describe bearing failure, each with a different data source and analysis techniques. Figure 11 shows that the decision boundary can be modelled to account for each error and allows for anomalies to be detected for new observations in which the temperature error, vibration error or indeed both deviate from expected levels.

The decrease in vibration and temperature levels which occur approximately 7 weeks directly before failure could be attributed to the type of failure used in this case study. Bearing faults can see increases in temperature or vibration due to pitting (in this case, on the inner race) which can smooth over time due to further wear. This can often then be followed by additional step changes as more pitting or damage to the inner race and bearing assembly occurs, or in this case, failure. As this analysis focuses on detecting the bearing fault at the end of life, the step changes in vibration and temperature are quite apparent. To see any initial changes in vibration due to early fault onset, more vibration samples would have to be taken going further back in time as it likely occurred prior to this data set. These samples were not made available for use in this study.

As highlighted previously, the main benefit for using the SVM classifier is to combine models into a single threshold which in this case produces consistent anomalies in the 14 weeks prior to failure. If looking at temperature and vibration in isolation, this is not the case, with each model dropping below the anomaly threshold at some stage in the final weeks. In terms of early fault detection or failure prediction, this study does not show substantial improvements to actual lead time from first anomaly detection to failure.

5 | CONCLUSION

This paper has presented a methodology to detect anomalies by first of all building a NBM which can adequately describe a particular failure and secondly by assessing the error distribution over time through a SVM classifier. Through case study 1, results have shown that when considering SCADA data alone, anomalies can be detected more consistently using this methodology when compared to only monitoring error through a simple fixed threshold. Furthermore, this modular approach also allows multiple NBMs from different data sources to be modelled and combined in a single SVM classifier, as shown in case study 2, which builds on the first case study by adding a vibration NBM. This allows a single model to track component health and detect deviations in both temperature and vibration. Key advantages of this methodology are summarised below:

- Multiple parameters can be used to describe the error distribution over a chosen time period.
- A more complex decision boundary can be formed to classify new observations as normal or anomalous behaviour.
- A modular approach allows error output from multiple NBMs from different data sources to be combined into a single classifier for monitoring component behaviour.
- Using SVM classifiers allow for anomalies to be given an anomaly score rating, which describes how far from normal conditions each new anomalous observation lies.

Results from both case studies have supported the advantages described above; however, it should be noted that limitations do exist from this study in which future case studies could address, the most critical being the lack of vibration samples to build the NBM in the second case study. Additional vibration samples collected more frequently and further back in time would allow more features and samples to be used while training the NBM and SVM classifier, providing better opportunities to test and validate the model while reducing problems associated with model over-fitting. Further insight into the benefits of building a model using a modular approach could also be achieved with an example with more than two data sources. For example, looking at gearbox failure with SCADA, vibration and oil analysis could add even more value to the model and increase anomaly detection times and rates. That being said, when multiple features are used to build the SVM classifier, it becomes hard to visualise which NBM the anomaly stems from, which if using the second case study as an example, makes it difficult to distinguish if the anomaly is associated with temperature or vibration. This problem would be compounded as additional NBMs are introduced.

ACKNOWLEDGEMENTS

This study is supported by Engineering and Physical Sciences Research Council (project reference number EP/L016680/1).

ORCID

Alan Turnbull https://orcid.org/0000-0002-1404-6371
James Carroll https://orcid.org/0000-0002-1510-1416

REFERENCES

1. Crabtree CJ, Zappala D, Hogg SI. Wind energy: UK experiences and offshore operational challenges. *Proc Ins Mech Eng Part J Power Energy*. 2015;229 (7):727-746.
2. Valpy B, English P. Future renewable energy costs: offshore wind, how technology innovation is anticipated to reduce the cost of energy from european offshore wind farms; 2014.
3. Crabtree C. Survey of commercially available condition monitoring systems for wind turbines. *SupergenWind Energy Technologies*. Durham: Durham University; 2010.
4. Carroll J, McDonald A, McMillan D. Reliability comparison of wind turbines with DFIG and PMG drive trains. *IEEE Trans Energy Conv*. 2015;30(2): 663-670.
5. Hahn B, Durstewitz M, Rohrig K. Reliability of wind turbines—experiences of 15 years with 1,500 wts. *Wind Energy*. 2007:328-330.
6. Pfaffel S, Faulstich S, Rohrig K. Performance and reliability of wind turbines: a review. MPDI energies; 2017.
7. Spinato F, Tavner PJ, van Bussel GJW, Koutoulakos E. Reliability of wind turbine subassemblies. *IET Renew Power Gen*. 2009;3(4):387-401.

8. Salameh JP, Cauet S, Etien E, Sakout A, Rambault L. Gearbox condition monitoring in wind turbines: a review. *Mech Syst Sig Process*. 2018;111: 251-264.
9. Tautz-Weinert J, Watson S. Using scada data for wind turbine condition monitoring—a review. *IET Renew Power Gen*. 2017;11(4):382-394.
10. Koukoura S, McDonald A, Charalambous A, Weiss S, Carroll J, McArthur S. Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques. *Wind Energy*. 2019;22(3):360-375.
11. Wang Y-F, Lin K-C, Chen M-Y, Hsu JH-Y, Hsu J-Y. Wind turbine fault diagnosis and predictive maintenance through statistical process control and machine learning. *IEEE Access*. 2020;8:23427-23439.
12. Elasha F, Shanbr S, Mba D, Li X. Remaining useful life prediction of rolling element bearings using supervised machine learning. *Energies*. 2019;12(14): 2705.
13. Dethlefs N, Chatterjee J. Deep learning with knowledge transfer for explainable anomaly prediction in wind turbines. *Wind Energy*. 2020.
14. Elforjani M. Diagnosis and prognosis of real world wind turbine gears. *Renew Energ*. 2020;147:1676-1693.
15. Zhang ZY, Wang KS. Wind turbine fault detection based on SCADA data analysis using ANN. *Adv Manufact*. 2014;2(1):70-78.
16. Dinmohammadi F, Zhao X, Robu V, et al. Machine learning methods for wind turbine condition monitoring: a review. *Renew Energ*. 2019;133:620-635.
17. Schlechtingen M, Santos I. Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: system description. *Appl Soft Comput*. 2013;13(1):259-270.
18. Hong YS, Cho YM, Ahn SH, Song CK, Hameed Z. Condition monitoring and fault detection of wind turbines and related algorithms: a review. *Renew Sust Energ Rev*. 2009;13(1):1-39.
19. Bangalore P, Tjernberg LB. An artificial neural network approach for early fault detection of gearbox bearings. *IEEE Trans Smart Grid*. 2015;6:980-987.
20. Sun P, Jian L, Caisheng W, Xiao L. A generalized model for wind turbine anomaly identification based on SCADA data. *Appl Energy*. 2016;168: 550-567.
21. Laouti N, Sheibat-Othman N, Othman S. Support vector machines for fault detection in wind turbines. *IFAC Proc Vol (IFAC-PapersOnline)*. 2011;18 (PART 1):7067-7072.
22. Hsu C, Chih-chung C, Chih-Jen L. A practical guide to support vector classification. Department of Computer Science, National Taiwan University; 2003.
23. Mollasalehi E. Data-driven and model-based bearing fault analysis—wind turbine application. *Thesis*. Calgary, AB: University of Calgary; 2017.
24. Zhang DH, Qian LY, Mao BJ, Huang C, Huang B, Si YL. A data-driven design for fault detection of wind turbines using random forests and XGboost. *IEEE Access*. 2018;6:21020-21031.
25. Chen JL, Pan J, Li ZP, Zi YY, Chen XF. Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals. *Renew Energy*. 2016;89:80-92.
26. Muller AC, Guido S. Introduction to machine learning with python. O'Reilly Media; 2017.
27. Garlick W, Dixon R, Watson S. A model-based approach to wind turbine condition monitoring using SCADA data; 2009.
28. Mollasalehi E, Wood D, Sun Q. Indicative fault diagnosis of wind turbine generator bearings using tower sound and vibration. *Energies*. 2017;10(11): 1853.
29. Borghesani P, Pennacchi P, Chatterton S, Ricci R. The velocity synchronous discrete fourier transform for order tracking in the field of rotating machinery. *Mech Syst Sig Proc*. 2014;44(1):118-133. Special Issue on Instantaneous Angular Speed (IAS) Processing and Angular Applications.

How to cite this article: Turnbull A, Carroll J, McDonald A. Combining SCADA and vibration data into a single anomaly detection model to predict wind turbine component failure. *Wind Energy*. 2021;24:197–211. https://doi.org/10.1002/we.2567