ORIGINALARBEITEN/ORIGINALS

On confidence interval-based anomaly detection approach for temperature predictions of wind turbine drivetrains to assist in lifetime extension assessment

Kelly Tartt[1,3] · Abbas Mehrad Kazemi-Amiri[2] · Amir R. Nejad[3] · James Carroll[2]

Abstract

With the number of wind turbines being installed increasing, due to the commitment of a large number of countries investing more in renewable energy, an informative method to determine when a drivetrain is coming to the end of it's life can be extremely useful. This paper investigates the uncertainty of an output of a methodology used for life extension evaluation of a generator bearing in the drivetrain. A method has been developed to determine when the non-drive end generator bearing is coming to the end of it's life, based upon temperature data extracted from seven years of 10-minute averaged SCADA data. Data from Kelmarsh wind farm was used, which consists of six onshore 2.05 MW Senvion MM92 wind turbines. A number of parameters from the SCADA data are used as the inputs for the model, in order to predict the component temperature and then in turn determine a threshold value, in which if the component's temperature passes, indicates that it is reaching the end of it's life. Due to the consequences that can occur if a component fails, such as loss of power, cost of repair etc. it is extremely important for the model to be as accurate as possible by taking into account any error or uncertainty. Other than the uncertainty of the measurements recorded in the SCADA data, which may be due to noise and/or sensor failure, the other major source of uncertainty comes from the predictive machine learning model that has been developed. Therefore, the model uncertainty is evaluated by a sensitivity analysis, where the input parameters are changed to see how much the output changes. The contribution of this work has investigated the error propagated in the component's remaining life, that have originated from the uncertainty of the machine learning model, as well as the model input parameters/data. The results show that the error arising from the machine learning model and the input data, should fall within a certain range in order to obtain the level of accuracy of the methodology.

✉ Kelly Tartt
kelly.tartt@strath.ac.uk, kelly.tartt@ntnu.no

Abbas Mehrad Kazemi-Amiri
abbas.kazemi-amiri@strath.ac.uk

Amir R. Nejad
amir.nejad@ntnu.no

James Carroll
j.carroll@strath.ac.uk

1   Wind and Marine Energy Systems and Structure CDT, University of Strathclyde, Richmond Street, Glasgow, G1 1XQ, United Kingdom

2   Wind Energy and Control Centre, University of Strathclyde, Richmond Street, Glasgow, G1 1XQ, United Kingdom

3   Department of Marine Technology, Norwegian University of Science and Technology (NTNU), Jonsvannsveien 82, NO-7050 Trondheim, Norway

1 Introduction

Due to the various agreements and policies set worldwide, such as the commitment to limit the global temperature rise to 1.5°C and achieve net zero emissions by 2050, discussions at COP28, the United Nations Climate Change Conference which addresses actions to meet these targets [1], emphasized that the amount of global electricity produced from renewable energy sources needs to increase at a much faster rate. According to the International Energy Agency (IEA) [2], by 2028 42% of global electricity will come from renewable energy sources, with solar and wind accounting for 25%. The Global Wind Energy Council's (GWEC) Global Wind Report 2024 [3], states that 117 GW of wind energy capacity was installed in 2023 and in order to reach the targets, this annual growth needs to triple to 320 GW by the end of the decade. Therefore, 2.75 TW of wind energy capacity needs to be installed by 2030, with the aim of reaching the Net Zero targets.

As mentioned above, 42% of global electricity will come from renewable energy sources by 2028, with over 12% coming from wind, therefore it is extremely important for these wind turbines to operate efficiently and be reliable. Hence, a method to determine the condition of drivetrain components and predict when and/or if it may fail, is important to minimise downtime and in turn maintain reliability and efficiency.

The most common indication of component failure is an increase in vibration and temperature. All turbines record various temperature measurements within the nacelle as part of the Supervisory Control and Data Acquisition (SCADA), whereas vibration measurements are collected by additional sensors as part of a condition monitoring system, which not all turbines have. Therefore, a method to determine remaining useful life (RUL) using temperature would be extremely useful and cost effective.

Previous work has been carried out to develop an informative tool, which can be used for life extension evaluation of wind turbine drivetrain components [4]. To summarise, a method/model has been developed which uses SCADA data together with an existing, straightforward model, to spot any temperature trends or changes, of components within the drivetrain, which may indicate that a component is reaching the end of it's life. SCADA data from the Kelmarsh wind farm is used, which is an onshore wind farm located in Northamptonshire, U.K. and consists of six 2.05 MW Senvion MM92 wind turbines. The 10-minute averaged SCADA data was recorded over seven years, from 2016 to the end of 2022, for all six wind turbines. Three input SCADA parameters: power, rotor speed and nacelle temperature, along with the component's temperature, are used along with a regression tree ensemble model to predict the temperature of the non-drive end generator bearings. These rear generator bearings have been selected because there are two recorded failures of these components, in two separate wind turbines and a proactive replacement in a third. The difference between the predicted and actual temperature values is calculated, followed by determining the cumulative average. This cumulative average is used to predict a threshold value. If the cumulative average for a component crosses this threshold, it indicates that the component may be nearing the end of its operational life.

Model validation is important and this can be achieved by using measurement chains, to ensure the processes are reproducible and thorough. A typical measurement chain can consist of: data collection, preprocessing, feature extraction, model prediction, error analysis, feedback. Reliable models related to wind turbines are vital due to their importance in supplying energy and their costs. An inaccurate prediction could have serious consequences, such as shutting down a wind turbine and replacing a component which is still functional, or a component failing unexpectedly forcing unforeseen downtime. Hence, due to the importance of predicting an accurate threshold value, it is crucial to determine any errors and/or uncertainties within the model and incorporate these.

Model uncertainties can be classed as epistemic or subjective uncertainties, along with measurement and statistical uncertainties [5]. These uncertainties are due to lack of knowledge. Therefore, more information, data and better models can all reduce these type of uncertainties.

Errors in the model refers to the difference between the actual and predicted values, whereas model uncertainty is the ratio between the actual and predicted values. Errors in the model may be caused by factors such as: model bias, variance, noisy data and model limitations. Model uncertainty may be caused by factors such as: insufficient information, complex models, data variability and inherently ambiguous data patterns. Therefore, model errors refer to specific discrepancies between the actual and predicted values, whereas model uncertainty refers to confidence or lack of within the model and recognises that any discrepancies may arise from fundamental issues.

This paper looks at both the model errors and uncertainties. The main sources of errors/uncertainties in this case may come from the machine learning model and the measurements recorded in the SCADA data, which are used for the model's input parameters. Errors in the measurements may be caused by sensor failures, malfunction and/or noise. A method to predict and incorporate the errors has then been proposed, in order to try and obtain a more reliable and accurate predicted value i.e. component temperature and in turn threshold value.

This paper is part of research work to develop and implement a method to determine lifetime extension of wind turbine drivetrains [4, 6, 7].

The novelty of this paper includes the probabilistic-statistical approach, to develop metrics for identification of drivetrain components approaching their end of life. Through applying Monte Carlo simulation on the temperature estimates of the utilised machine learning method, the modelling error is captured. This is to enable quantifying the propagated error in the later stages, due to the machine learning model's inaccuracy and to enhance the robustness of the deterministic approach, presented in [4]. Followed by proposing a confidence interval-based method, the temperature anomalies over the unhealthy stage of the components can be detected. This contribution establishes a method to determine confidence bands for the component temperature, so that any temperature value that exceeds these bands require further investigation, as it may be a sign that the component is reaching the end of it's life, or it may just be noise or a sensor error.

The rest of this paper is structured as follows: Section 2 details the existing literature to determine model errors and

uncertainties. Section 3 describes the methods used, Sect. 4 discusses the results, Sect. 5 sums up the conclusions and Sect. 6 explains any future work.

2 Literature review

Predicting and incorporating uncertainty or error is extremely important when predicting an outcome/output because it improves reliability and accuracy, so a large amount of research has previously been carried out in this area. In this case the output is the predicted component temperatures.

[8] describes four types of uncertainty including: inherent, measurement, statistical and model. They define model uncertainty as "uncertainty due to imperfections and idealizations made in physical model formulations for load and resistance, as well as in the choices of probability distribution types for the representation of uncertainties".

With regards to wind energy/wind turbines, models are used for a variety of reasons. Examples include: predicting wind direction for yaw control [9], modelling of wind speeds to forecast power production [10], predicting wind speed [11], modelling component failures based upon weather conditions [12] and identifying anomalies [13].

Many types of models are used and [14] found that random forest machine learning models, provide the greatest accuracy. They also conclude that the accuracy increased when additional variables were included.

Model uncertainties have been investigated by [15–17] and [18] to name a few. They define model uncertainty as the actual or real value of a variable divided by the predicted value. [8] goes on to explain that bias exists when the mean value does not equal one. They explain that when no detailed information is available, either lognormal or normal distributions are used. Load variables are described using lognormal distributions, whereas normal distribution is typically used for resistance variables. [18] confirm that lognormal distribution is preferred to Gaussian distribution. [17] explain that by dividing the standard deviation by the mean, the coefficient of variation can be determined, which is a "convenient measure of the relative error that the model uncertainty represents". A multiplicative model is then typically used by researchers, including [16], to account for combined uncertainties but [17] suggests considering additive errors.

[19] investigated three different approaches to determine uncertainty in machine learning predictions. They included the Gaussian process and the quantile approach, which used both the absolute difference and square of the difference, where the difference is defined as the difference between the predicted and observed values. They conclude that the quartile approach was the easiest approach but the Gaussian process gave a good estimate. They also found that the best approach was achieved by using the absolute difference between the predicted and expected values.

An error compensation prediction was proposed by [20], based upon an extreme learning machine. This extreme learning machine was used to predict the difference between the predicted and actual value, which is known as the error. This predicted error is obtained from a time lag correction model and is then added to the "raw wind speed prediction result" to determine the "final wind speed prediction result".

Confidence intervals can be used to quantify uncertainty from data-driven models. [21] proposed two approaches with regards to confidence intervals, pointwise and simultaneous, to "measure the uncertainty associated with an SVM-based power curve model". They conclude that pointwise confidence intervals gave the most accurate results, when measuring the uncertainty of the power curve.

Monte Carlo simulation and methods have been used by a number of researchers to account for uncertainty and errors in models. [22] proposed an "uncertainty quantification approach" to determine remaining useful life predictions, which consists of kernel density estimation and Monte Carlo dropout. They conclude a high accuracy based upon this proposed method. A method of combining both a sensitivity analysis and Monte Carlo method was proposed by [23], in order to try and determine how uncertainties may effect the financial risks of wind projects. [24] looked at modelling the effect of both aerodynamic and structural uncertainties in wind turbine blades. They used the Monte Carlo method to characterise the uncertainties in the material properties in a structural model of the blade. [25] used Monte Carlo simulation to obtain the "annual energy production and its uncertainty for the wind farm".

3 Method

A deterministic approach to identify drivetrain components reaching the end of their operational life is documented in [4]. In [4], the main aim was to use a machine learning model to predict component temperatures, which are then used to calculate the temperature differences between actual and predicted temperatures. The actual component temperature, T_{ad_j} and predicted component temperature, T_{pd_j}, for each day, j, are calculated from the instantaneous values of actual and predicted temperature, T_{a_i} and T_{p_i}, respectively, as shown in Eqs. 1 and 2, where n refers to the number

of ten minute blocks per day (i.e. 6 blocks in one hour, 24 hours a day, therefore n = 6 × 24 = 144).

$$T_{ad_j} = \frac{\sum_{i=x}^{x+(n-1)} T_{a_i}}{n} \tag{1}$$

$$T_{pd_j} = \frac{\sum_{i=x}^{x+(n-1)} T_{p_i}}{n} \tag{2}$$

The daily temperature difference, δT_{d_j}, is then calculated as shown in Eq. 3.

$$\delta T_{d_j} = T_{ad_j} - T_{pd_j} \tag{3}$$

From discrepancies between "healthy" and "unhealthy" years/data, metrics including: temperature difference as described above, cumulative sum and cumulative average are obtained. "Healthy" years/data is assumed to be either the first year of operation, or the year after a major repair has occured, where there are no recorded failures. Whereas "unhealthy" data is assumed to be all subsequent years.

For further information on the machine learning model, definitions and metrics, please refer to [4].

Different methods are followed according to whether the error within the model is investigated or the model uncertainty determined.

3.1 Model uncertainty

The model uncertainty can be used in stochastic analysis, for example for reliability analysis [16, 26], to represent a probabilistic model of the uncertainty associated with the data-driven model.

The first step in determining model uncertainty specifically uses the first year of SCADA data, which in this case is all the data from 2016 because this data is assumed to be "healthy" data. The data is randomly split into 70% training data and 30% test data and applied to the regression tree ensemble model, which has been developed previously. The model is trained and tested on different data in order to get an unbiased view of the model. The input parameters are: power, rotor speed, nacelle temperature and rear generator bearing temperature for the training dataset and just the first three input parameters for the test dataset. The output from the model is the predicted rear generator bearing temperature. Due to the data split being randomly generated, this process is carried out a total of ten times, in order to get a range of predicted rear generator bearing temperatures for the test dataset. The predicted temperature values for all ten runs are recorded and stored, both the predicted temperatures for each 10-minute data value, as well as the daily average values.

Once the first step has been completed, the next step is to calculate model uncertainty, χ, which as mentioned in Sect. 2, is the ratio of the actual rear generator bearing temperature, X_{true}, to predicted rear generator bearing temperature, $X_{\text{Predicted}}$ [16], as shown below.

$$\chi = \frac{X_{\text{true}}}{X_{\text{Predicted}}} \tag{4}$$

The model uncertainty is calculated for each 10-minute averaged data value and for the daily averaged value. The uncertainty value is then modelled using the lognormal distribution to determine the mean and standard deviation. Lognormal distribution is typically used because it only allows the use of positive values, thus shortening the "random variable range" [27].

The cumulative distribution function (CDF) of a lognormal distribution can be defined as:

$$F_x(x) = \phi\left(\frac{\ln x - a}{c}\right) \tag{5}$$

Where $\phi()$ is the standard Gaussian CDF, ln is natural logarithm, a is the mean and c is the standard deviation of $\ln x$.

The next step is to work out the 95% confidence interval, which is a range of values in which there is a 95% certainty, that the predictions are expected to fall within this upper and lower limit.

The above process is then repeated for different dataset splits, i.e. 60% training/40% testing and 80% training/20% testing, to determine if this will improve or affect the predicted values in any way.

3.2 Errors in model

The error in the model is the difference between the actual and predicted rear generator bearing temperatures, which were determined in the previous section. Once the error has been calculated the Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) can then be determined, as these are all valuable performance indicators for regression models.

The Mean Absolute Error (MAE) calculates the average absolute difference or error between the actual and predicted values [28], so obviously the smaller the value the better.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |X_{\text{true}} - X_{\text{Predicted}}| \tag{6}$$

The Mean Squared Error (MSE) squares the average absolute difference or error between the actual and predicted values.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (X_{\text{true}} - X_{\text{Predicted}})^2 \qquad (7)$$

The Root Mean Squared Error (RMSE) calculates the root of the mean of the differences between the actual and predicted values squared [29].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_{\text{true}} - X_{\text{Predicted}})^2} \qquad (8)$$

The Mean Absolute Percentage Error (MAPE) calculates the average magnitude of the difference or error between the actual and predicted values. It is the percentage version of MAE [30].

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{X_{\text{true}} - X_{\text{Predicted}}}{X_{\text{true}}} \right| \times 100 \qquad (9)$$

These error metrics are calculated in order to try and observe how the model is performing, for example to see if there is any bias.

3.3 Monte Carlo confidence interval-based anomaly detection approach and metrics

The next step is to use the Monte Carlo method to include an 'error' value in any future predictions. This error value is determined from the values of mean and standard deviation, extracted from the probability distribution of the calculated error values, in the "healthy" data and explained in further details later in this Section. The following process is run for each turbine separately.

The data preparation process that has been described previously and which is shown in Fig. 21 is carried out on all data from year 2016. For all data from years 2017 to 2022, the same process is carried out except for the last stage.

The next step is to set up the Monte Carlo simulation and this is detailed in a flowchart shown in Fig. 1. The process is ran for each turbine separately. The number of simulations is set to one thousand (1000). The "healthy" data, which in this case is the 2016 data is randomly split into 70% training data and 30% test data and the model ran. The differences in actual and predicted component temperatures are then calculated and these errors are plotted using a Matlab tool to fit the probability distribution to these errors. From this normal probability distribution, the mean and standard deviation are determined, which are then used
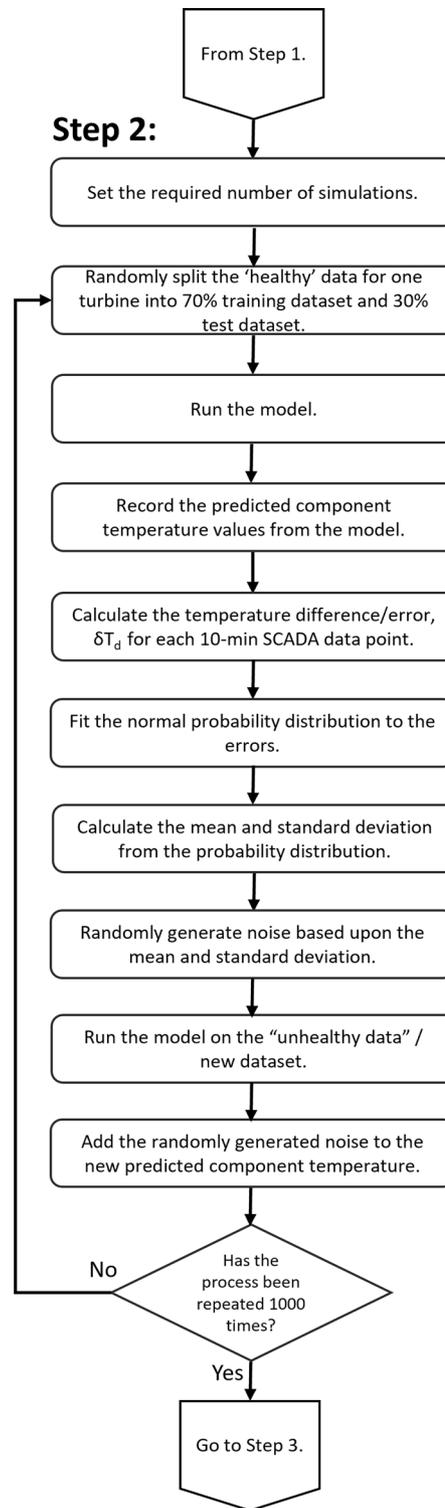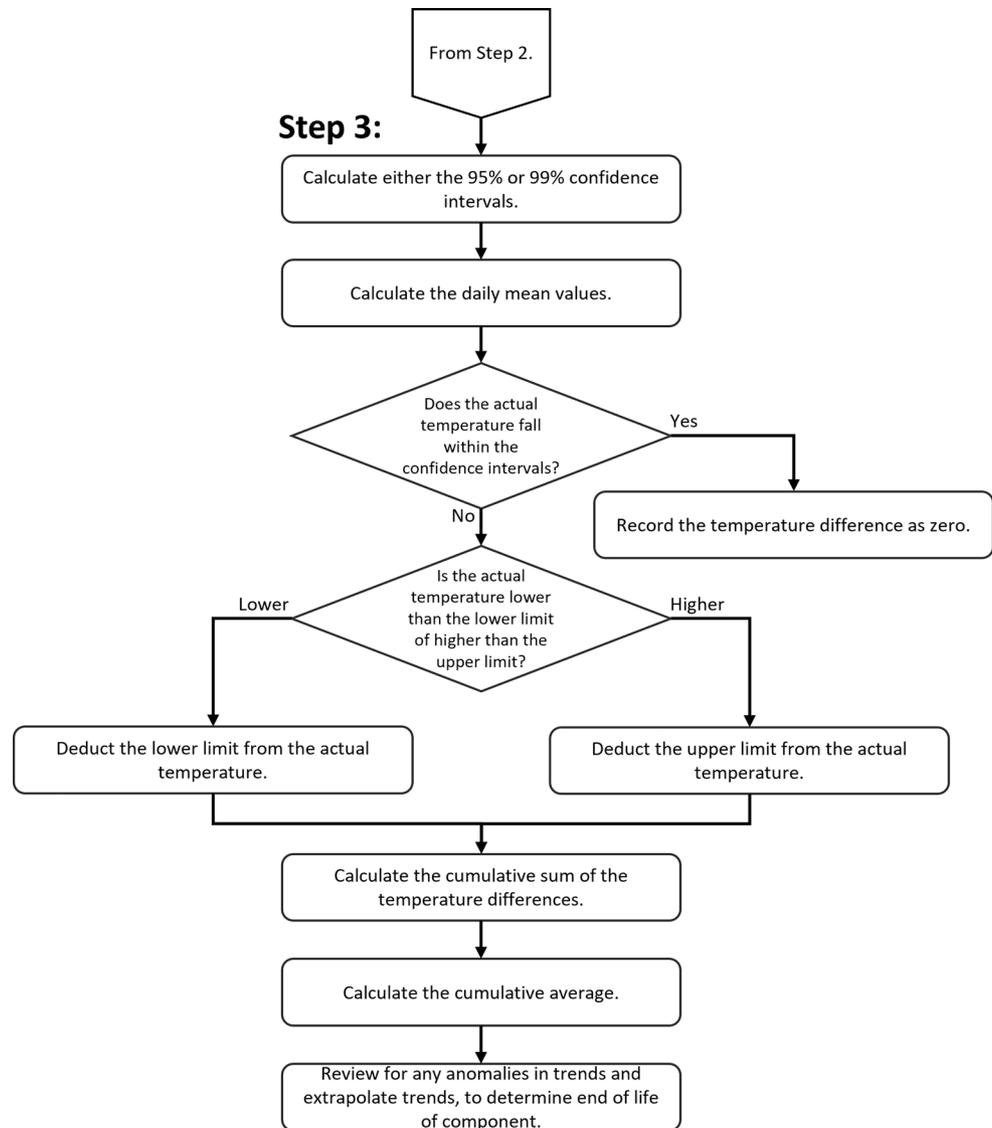

Fig. 1 Flow Chart Showing the Monte Carlo Method used to Determine Predicted Values

Fig. 2 Flow Chart Showing the Post-Processing Process used to Determine Prediction Levels

to randomly generate an 'error'. The model is then ran on the new dataset, which consists of the future years data, to predict the component's temperature, in which the randomly generated 'error' is added too. Once the simulation has ran one thousand times, both the 95% and 99% confidence intervals are determined from these simulated predictions and the post-processing process is carried out, which is detailed in a flowchart in Fig. 2. The daily values are then calculated, along with the temperature difference of any actual temperatures which fall outside of the confidence intervals, the cumulative sum of these temperature differences and cumulative average.

The process mentioned above is then repeated for a "healthy" data split of 80%/20% and 60%/40%, in order to observe if more or less training data affects the accuracy of the predicted component temperature. It is also repeated using different amounts of future data, i.e. 1 year–6 years,

in order to see if the amount of data available affects the reliability and accuracy of the output.

4 Results and discussions

4.1 Model uncertainty

The lognormal probability distribution graphs of all turbines are very similar. Turbines 3 and 4 are shown in Figs. 3 and 4. Turbine 3 is an example of a turbine which did not have any reported failures, whereas Turbine 4 did record a failure in the rear generator bearing.

Tables 1 and 2 show the average mean and standard deviation for each 10-minute SCADA data point, which will be referred to as 'all data' going forth and the daily averages, respectively, of the model uncertainty. They show the mean
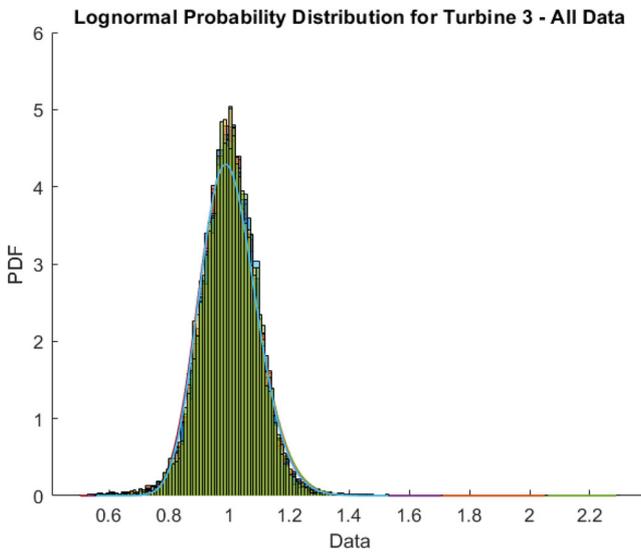
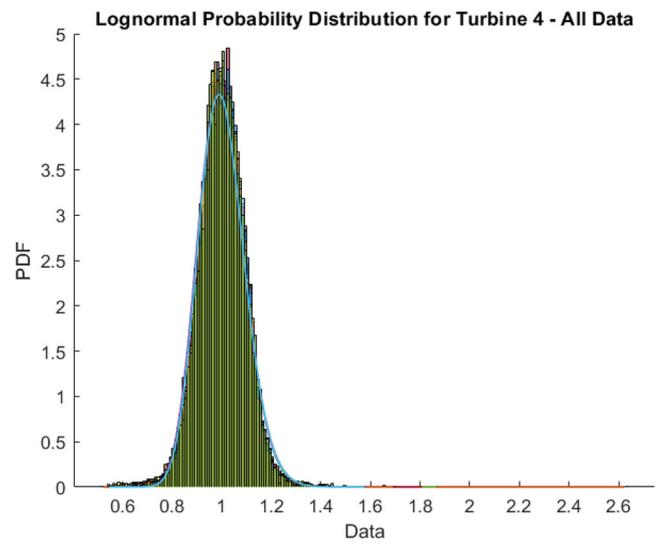**Fig. 3** Probability Distribution for Turbine 3

Fig. 4 Probability Distribution for Turbine 4

Table 1 Average Values for All Turbines Using Model Uncertainty

Turbine	Mean	Standard Deviation
1	1.0011	0.0918
2	1.0010	0.0899
3	1.0012	0.0951
4	1.0012	0.0933
5	1.0012	0.0936
6	1.0018	0.0997

Table 2 Average Values for All Turbines Using Model Uncertainty (Daily Values)

Turbine	Mean	Standard Deviation
1	1.0010	0.0358
2	1.0005	0.0334
3	1.0017	0.0369
4	1.0010	0.0359
5	1.0014	0.0343
6	1.0017	0.0381

Table 3 Average Error Metrics for All Turbines

Turbine	MAE	MSE	RMSE	MAPE
1	1.8321	5.5723	2.3605	7.0493
2	1.8376	5.4687	2.3385	7.0438
3	1.9485	6.2626	2.5025	7.3174
4	1.9355	6.0592	2.4615	7.2834
5	1.9341	5.9778	2.4449	7.3240
6	1.9411	6.3569	2.5211	7.5029

Table 4 Average Error Metrics for All Turbines (Daily Values)

Turbine	MAE	MSE	RMSE	MAPE
1	0.7735	0.9453	0.9720	2.9141
2	0.7204	0.8252	0.9082	2.7076
3	0.8035	1.0944	1.0460	2.9460
4	0.7950	0.9927	0.9960	2.9230
5	0.7436	0.9000	0.9483	2.7284
6	0.7982	1.0413	1.0202	2.9999

to be around 1 and standard deviation to be around 0.09 for all data values and 0.03 for the daily average values. All the mean values are slightly above 1, which indicates that the model is pretty accurate, although a little conservative. The standard deviation of 0.09 shows minimum variability, demonstrating the model's consistency. Converting the 10-minute averages to daily averages, reduces the standard deviation, thus decreasing variability because it smooths out any short-term fluctuations.

Tables 7 and 8 show the mean and standard deviation of the model uncertainty when using different training/test data splits. Table 7 shows that the standard deviation values are slightly lower when using an 80%/20% split, whereas this is the opposite in Table 8, where the average daily

values are used. Although in general, changing the percentage split of training/test data, does not change the values significantly.

Therefore, in terms of model uncertainty, the results indicate that the model demonstrates a reasonably high level of accuracy.

4.2 Errors in model

The average error metrics for all data and the average daily values for the "healthy" (2016) data are shown in Tables 3 and 4.

All the error metrics: MAE, MSE, RMSE and MAPE are reduced when each 10-min data value is transformed
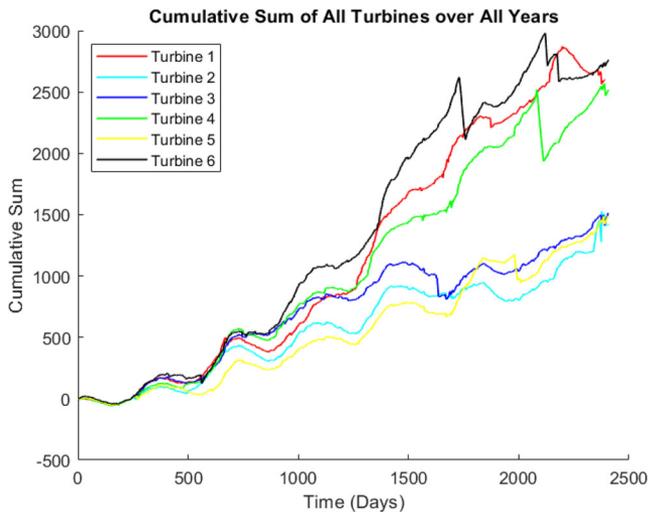
Fig. 5 Cumulative Sum of the Temperature Differences for All Turbines Over All Years

Fig. 6 Cumulative Average of the Temperature Difference

into the daily average. This transformation can be advantageous because it smooths out any short-term fluctuations, resulting in smaller error metric values, which can indicate a better correlation between the actual and predicted values, thus signifying a model with higher accuracy. These error metrics determine the error between actual and predicted values, so lower values demonstrate improved performance of the model. Therefore, with regards to model evaluation, minimizing these errors is a crucial goal.

A sensitivity analysis is carried out by adjusting the training/test data split. Tables 9 and 10 show the error metrics for the "healthy" data only i.e. 2016 data, for the 70%/30%, 60%/40% and 80%/20% data splits, for all data and the average daily values, respectively. Table 10 shows that MAE and MSE increases as the training dataset increases from 60% to 80%, but RMSE and MAPE reduce in value as the training dataset increases. Whereas for all the data (Table 9), all the error metrics, i.e. MAE, MSE, RMSE and MAPE, reduce in value as the size of the training dataset increases. Again, varying the training/test data split does not alter the results significantly.

4.3 Monte Carlo confidence interval-based anomaly detection approach and metrics

The original graphs of the cumulative sum of the temperature differences and the calculated cumulative average, prior to investigating the model errors and uncertainty, taken from previous work, are shown in Figs. 5 and 6.

This work has been previously documented [4] but to briefly summarise:

Figure 5 shows the cumulative sum of the temperature differences between the actual and predicted rear generator bearing temperatures, for all turbines, across all years.
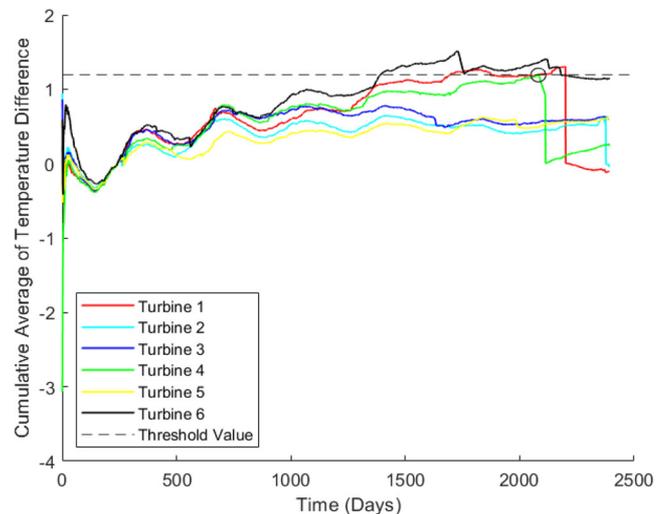
The graph shows that the turbines appear to split into two groups around day 1000, which corresponds to the first quarter of 2019. One group consists of Turbines 1, 4 and 6 and the other consists of Turbines 2, 3 and 5, with the former group having a much higher cumulative sum. The data logs recorded generator NDE bearing failures in Turbines 2 and 4, with the proactive replacement of the same component in Turbine 1, which all occurred within the year 2022. Therefore, this explains the higher cumulative sum values in Turbines 1 and 4. With regards to the higher values in Turbine 6, there was around 600 hours of preventative maintenance carried out in the first quarter of 2021 and forced outage in early 2022, with both events reducing the temperature. Turbine 2 doesn't appear to follow the same failure pattern as Turbine 4 but appears to fail shortly after sharp, sudden temperature rises, as opposed to degradation over time.

Figure 6 displays the cumulative average, which has been calculated in order to see if there is an obvious threshold value, in which if the turbine crosses then it is reaching it's end of life. This has been calculated by dividing the cumulative sum by the day. It was decided to use the cumulative average for finding this threshold value because the cumulative average is quite a straightforward process, which is also effective in detecting any trends. The graph shows that just prior to the time Turbine 4 failed, the threshold value was 1.2. The threshold value was captured at this point because prior to that point the cumulative average had been increasing but it was only once it reached that point, that the rear generator bearing failed, resulting in turbine downtime. Turbine 1 also passed this value around the same time and went on to have the same component proactively replaced about four months later. Turbine 6 crosses this threshold value twice and underwent scheduled maintenance, although un-
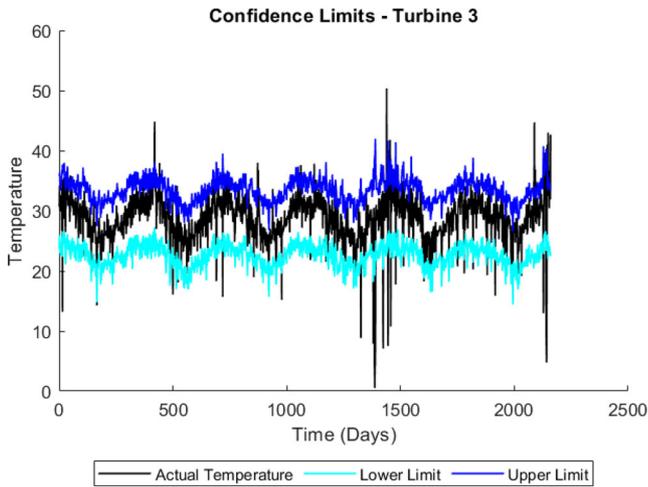
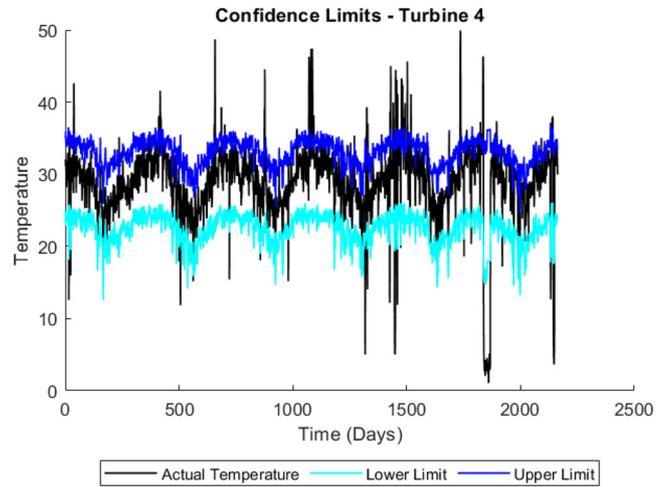**Fig. 7** Turbine 3—Actual Temperature and Confidence Limits (95%)

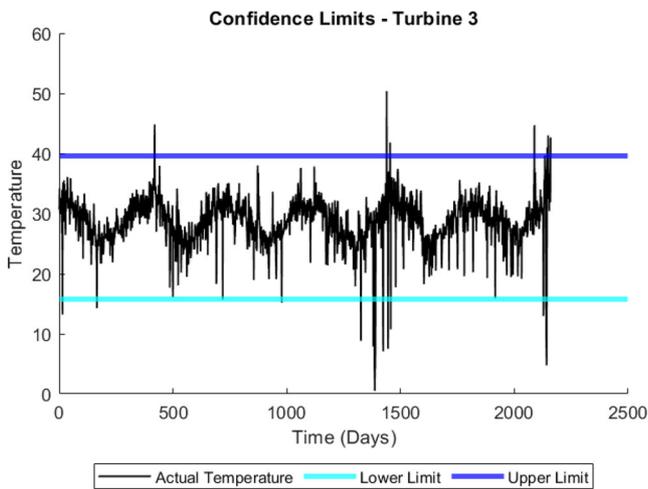**Fig. 9** Turbine 4—Actual Temperature and Confidence Limits (95%)

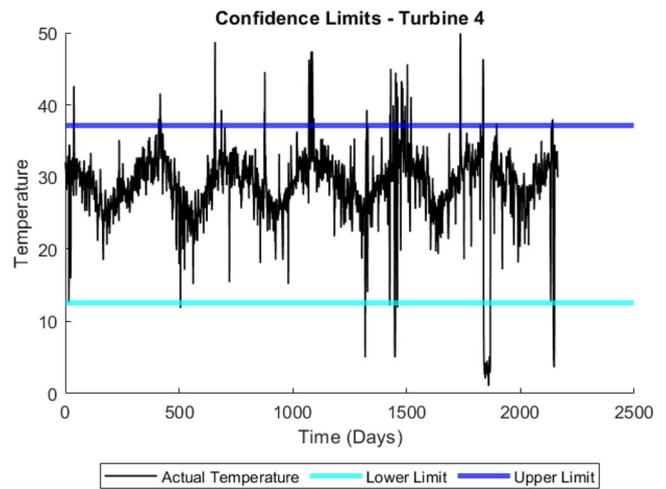**Fig. 8** Turbine 3—Actual Temperature and Confidence Bands (95%)

Fig. 10 Turbine 4—Actual Temperature and Confidence Bands (95%)

fortunately we do not have anymore details. Therefore, 1.2 was deemed a suitable threshold value, due to three turbines crossing this point and either failing or undergoing maintenance, even though the time between passing this point and failing varied for each turbine, so didn't really assist with determining life extension, which was the ultimate aim of the research.

As described in Sect. 3, the proposed Monte Carlo process/method is originally carried out on all turbines, using a training/test data split of 70%/30%. The 95% and 99% prediction or confidence intervals of the predicted component temperatures are then calculated. Figures 7 and 9 show the actual rear generator bearing temperatures, along with the 95% confidence upper and lower limits for the daily values, for Turbines 3 and 4. Figures 8 and 10 show the corresponding confidence bands. It can be seen from the graphs that some actual component temperatures exceed either the upper or lower limits, this may be due to excessive noise, a temporary sensor issue or component failure,

especially in the case of Turbine 4, where there are many more values exceeding the upper limit.

The number of daily average data values, that fall outside the upper and lower limits, are shown in Table 5, for both 95% and 99% confidence intervals. It shows that for the 95% confidence interval, between 5% to 11% of the total dataset values fall outside the limits, whereas between 2%

Table 5 Number of Data Points that fall outside the Lower and Upper Limits for both 95% and 99% Confidence Intervals

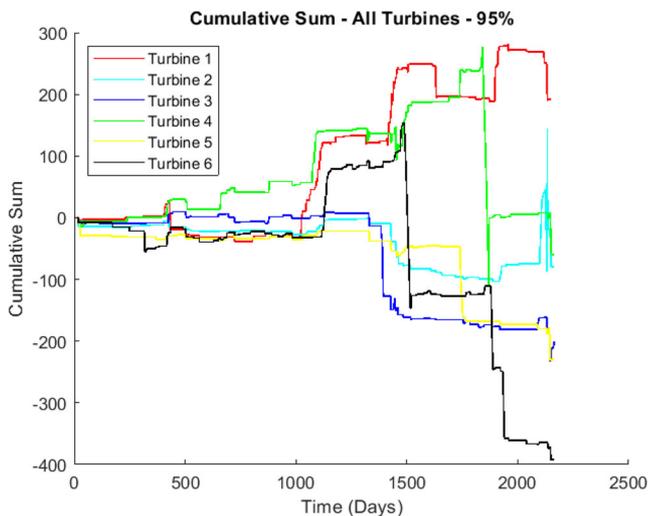| Turbine	No. of Values Outside Limits	
	95%	99%
1	171 (8%)	108 (5%)
2	123 (6%)	73 (3%)
3	108 (5%)	58 (3%)
4	202 (9%)	131 (6%)
5	119 (6%)	52 (2%)
6	236 (11%)	132 (6%)

Fig. 11 Cumulative Sum of the Temperature Differences for All Turbines Over All Years Using the New Process (95% Prediction Level)
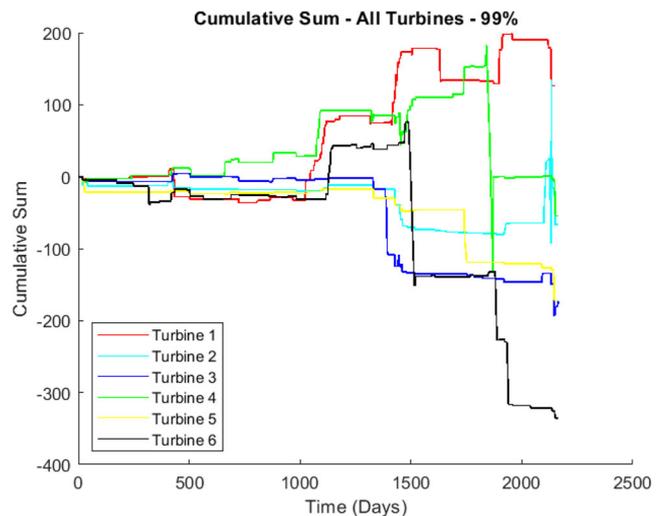

Fig. 13 Cumulative Sum of the Temperature Differences for All Turbines Over All Years Using the New Process (99% Prediction Level)
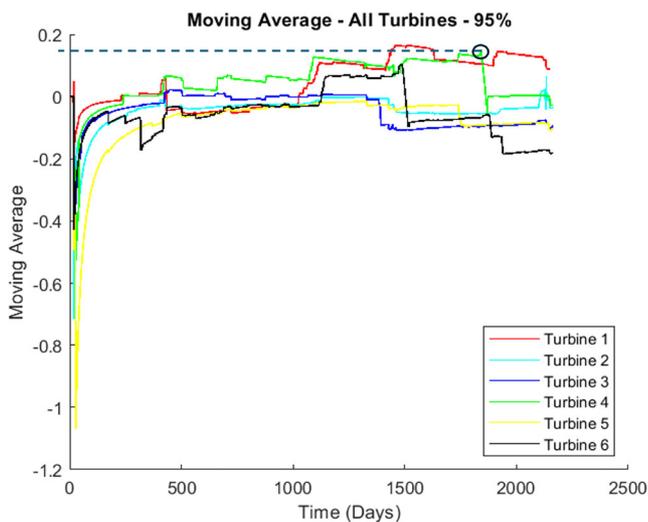

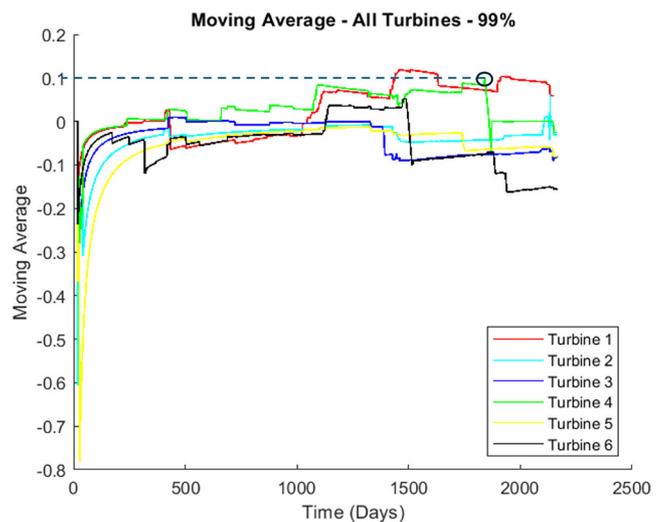**Fig. 12** Cumulative Average (95% Prediction Level)

Fig. 14 Cumulative Average (99% Prediction Level)

to 6% fall outside the limits when using the 99% confidence interval.

The new graphs of the cumulative sum of the temperature differences and the cumulative average for all turbines are displayed in Figs. 11–14. The graphs appear different to the original because only actual temperatures that do not fall within the prediction/confidence limits for each daily average value, are recorded as having a temperature difference. Any values that fall within the limits, are recorded as having zero temperature difference, which means that the cumulative sum can stay at the same value for a period of time, until an actual temperature outside the limits is recorded. This explains why there are more horizontal and vertical lines on the new graphs.

From these graphs, it can be seen that by using the same technique that was used to determine the threshold value in Fig. 6, that the threshold value is approximately 0.15 for 95% confidence interval and 0.1 for 99% confidence interval, as opposed to 1.2 in the original results. They also show that only Turbines 1 and 4 cross the threshold value, which matches the information recorded in the data logs. The data logs recorded generator NDE bearing failures in Turbines 2 and 4, as well as the proactive replacement of the component in Turbine 1. Turbine 6 does not cross the threshold value in this case, even though it gets close, this may be due to the fact that the issues it had were not related to the rear generator bearing, or that maintenance was carried out before it failed.

Some of the error metrics, specifically the mean absolute error (MAE) and mean squared error (MSE), are calculated for the results obtained via the new process and compared with the results from the original method. The results are
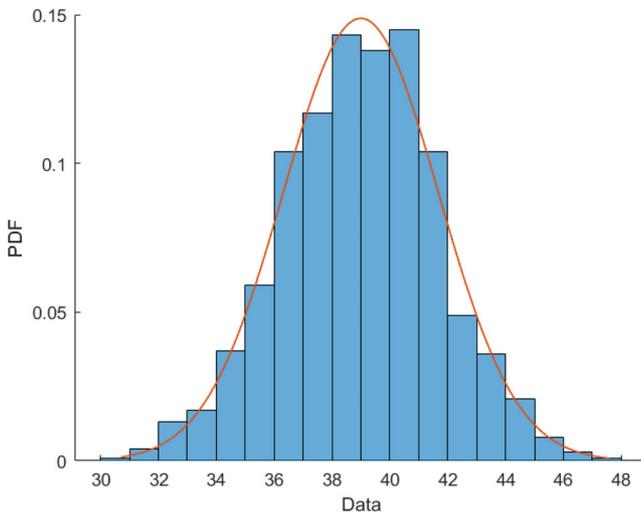
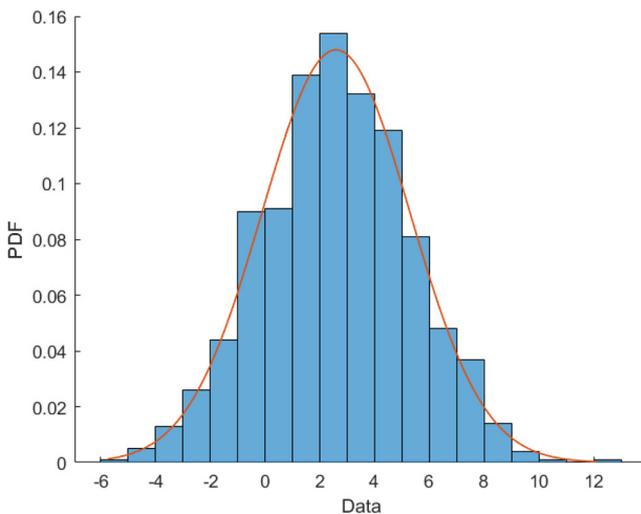**Fig. 15** Turbine 4—Probability Distribution Graph Two Days Prior to Failure

Fig. 17 Cumulative Sum of the Temperature Differences for Turbine 3 Over All Years: Comparison Using Training/Test Split of 60%/40% vs. 70%/30% vs. 80%/20% (99% Prediction Level)

Fig. 16 Turbine 3—Probability Distribution Graph

Fig. 18 Cumulative Average for Turbine 3 Over All Years: Comparison Using Training/Test Split of 60%/40% vs. 70%/30% vs. 80%/20% (99% Prediction Level)

displayed in Table 11. It shows that the proposed Monte Carlo method reduces both the MAE and MSE, in all turbines, which is advantageous.

It can also be seen from the table that the MSE values for Turbines 2 and 4 are higher than the others in the new Monte Carlo method, probably due to the fact that failures occurred in both these turbines.

The probability distribution related to the temperature differences for Turbine 4 just prior to failure and Turbine 3 are shown in Figs. 15 and 16.

Figure 15 shows a range of temperature difference values from 30 to 48 for Turbine 4, compared to a range of values
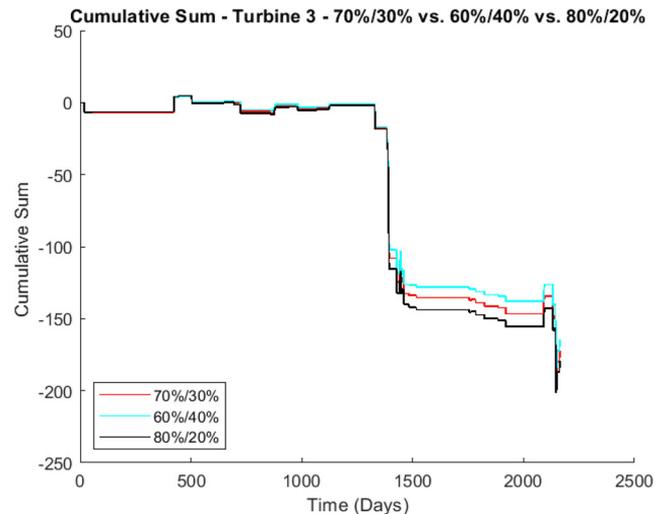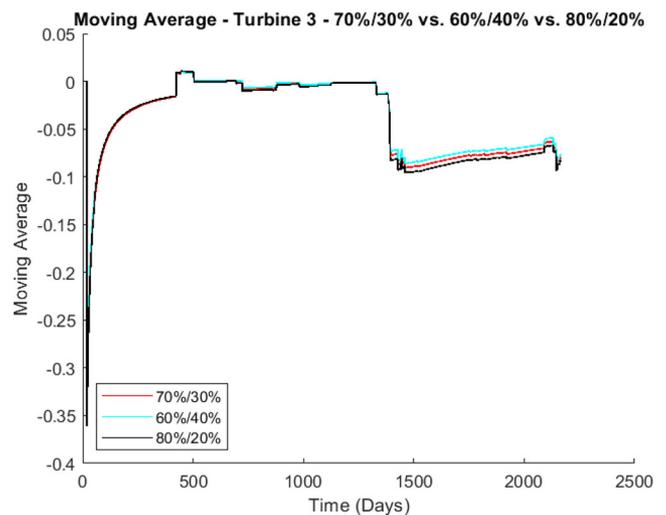
from -8 to 12 for Turbine 3, which is a noticeable difference between a turbine that failed and one that did not.

A sensitivity analysis is performed, which involves repeating the process using a different data split for the "healthy" data, a split of both 80%/20% and 60%/40% is applied and the results are shown in Figs. 17–20. The tabular results from the sensitivity analysis are shown in Table 6.

The error metrics for the complete process, i.e. the error calculated from applying the proposed Monte Carlo method to years 2017–2022, using the different "healthy" dataset splits are calculated and are shown in Table 6. The table shows that both the MAE and MSE increase as the train-
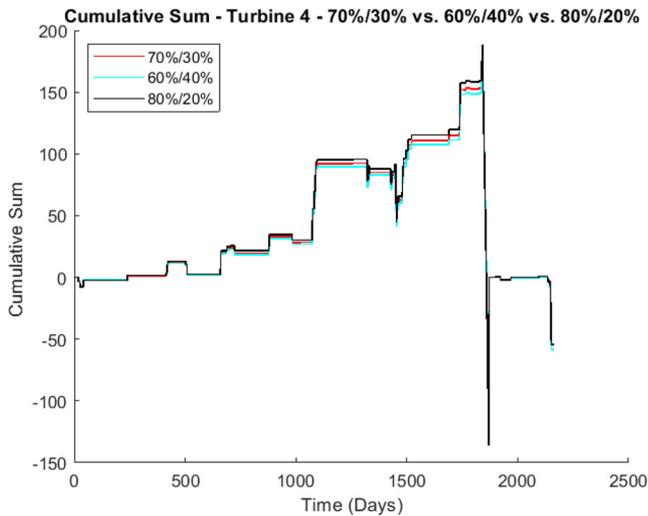
Fig. 19 Cumulative Sum of the Temperature Differences for Turbine 4 Over All Years: Comparison Using Training/Test Split of 60%/40% vs. 70%/30% vs. 80%/20% (99% Prediction Level)
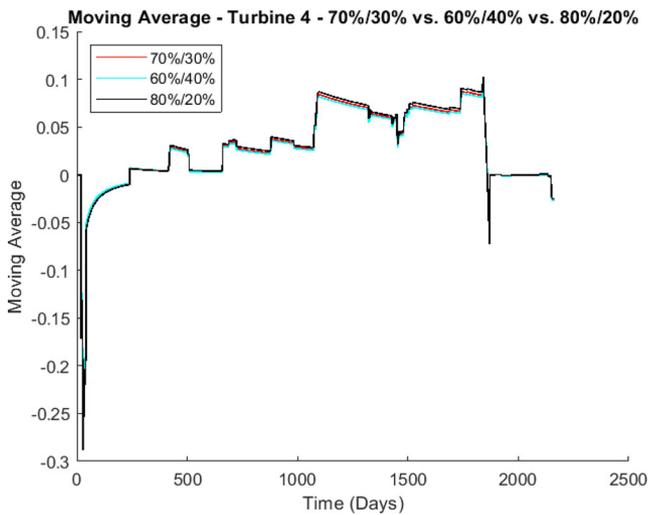

Fig. 20 Cumulative Average for Turbine 4 Over All Years: Comparison Using Training/Test Split of 60%/40% vs. 70%/30% vs. 80%/20% (99% Prediction Level)

ing dataset increases from 60% to 80% and testing dataset reduces from 40% to 20%.

With regards to the amount of data required to obtain accurate results, the process is repeated and the model applied to different durations, i.e. 1-6 years, to Turbine 4. The results for years 1, 3 and 5 are shown in Figs. 22–27. These results show that no matter how much data is available, this process can still be applied in order to see any anomalies or trends.

Table 6 Error Metrics for Different Data Splits—Daily Values for All Years

Turbine	Error Metric	60%/40%	70%/30%	80%/20%
1	MAE	0.36	0.37	0.38
1	MSE	3.50	3.59	3.67
2	MAE	0.36	0.37	0.38
2	MSE	15.32	15.43	15.63
3	MAE	0.21	0.22	0.23
3	MSE	2.24	2.36	2.40
4	MAE	0.49	0.50	0.52
4	MSE	5.08	5.17	5.36
5	MAE	0.17	0.17	0.19
5	MSE	1.27	1.34	1.48
6	MAE	0.47	0.49	0.49
6	MSE	4.47	4.68	4.76

5 Conclusion

The aim of this research was to determine the uncertainty and errors of a regression tree ensemble model, which was used to predict the temperature of a non-drive end generator bearing in multiple wind turbines, in order to predict if it was coming to the end of it's life. Determining errors and uncertainties in a model, especially ones concerned with wind turbines are extremely important due to the severe implications of inaccurate and unreliable models, such as sudden component failure causing unexpected downtime. Therefore, not only is it important to try and predict when a component is coming to the end of it's life, it is also important that the predictions are as accurate as possible.

For this research, data had been obtained from an onshore wind farm, which consists of six wind turbines and their SCADA data recorded from years 2016 to 2022.

The first step was to determine model uncertainty, which is defined as the ratio of actual to predicted temperature. Once the model uncertainty was calculated for the first year of the turbine operating, it was displayed as a lognormal probability distribution graph, in which the mean and standard deviation were calculated. Obviously the ideal mean value is 1 and in this case the mean was determined to be just above 1, ranging from 1.001 to 1.0018 for all turbines.

The next step was to calculate the error metrics, such as the mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE) and mean absolute percentage error (MAPE). These were calculated in order to get an idea as to how the model performs.

The final step was to utilise Monte Carlo simulation. The difference between the actual and predicted temperatures were calculated for "healthy" data. These errors were then displayed as a normal probability distribution and the mean and standard deviation calculated. These values were then used to randomly generate an error, which was added

to future temperature predictions. Both the 95% and 99% confidence intervals were then determined, prior to calculating the temperature difference for any point that falls outside the upper and lower limits. From here, the cumulative sum of the temperature differences were calculated along with the cumulative average, which was used to determine a new threshold value. The process was repeated for each turbine. The inputs were also varied by using different "healthy" training/test dataset splits i.e. 70% /30%, 60%/40%, 80%/20%, to see if this affected the predicted values, as well on different periods of time i.e. 1-6 years.

The results showed that by using the proposed Monte Carlo method along with the confidence intervals, the error metrics reduced, which is promising. The reduction in MAE was between 78% and 90% across the turbines and between 33% and 82% for MSE, although the MSE values for Turbines 2 and 4 stayed higher than the others, due to the fact they had component failures. The revised cumulative average graph, showed Turbines 1 and 4 crossing the new threshold value, which coincides with the failure/issue data log. As mentioned previously, Turbine 2 appears to follow a different failure pattern to Turbine 4, in the sense failure occurred shortly after a sudden, sharp temperature rise, instead of failure due to operating at higher temperatures over a long period of time (Turbine 4).

The sensitivity analyses concerning adjusting the training/test data split did not seriously affect the results, there was an increase from the 60%/40% results to 80%/20%, which ranged from 2% to 16% (MSE) and 4% to 12% (MAE). With regards to the use of 95% or 99% confidence intervals, obviously the 99% confidence interval includes more data points and with regards to calculating a threshold value, there was a difference in values of around 33% i.e. 0.1 as opposed to 0.15.

Therefore, it can be concluded that when trying to determine a threshold value, in which if an actual temperature value passes then the component may be reaching the end of it's life, using the Monte Carlo method to add a random 'error' to the predicted temperature based upon the probability distribution of errors in the "healthy" data, along with using confidence intervals, can be useful to take into account any uncertainties or errors that the model may contain. Thus, creating a slightly more accurate model, which is important when trying to evaluate component life extension. Although, as previous research and literature has stated, using only temperature data to try and determine remaining useful life of a component long in advance, is not such a good indicator but can assist the operator in identifying problematic components.

6 Future work

Future work will include, trying to improve the method further, it may be helpful to investigate whether improving the pre-processing method may help remove any false positives, or values that exceed the limits due to noise or sensor issues. It will also be advantageous to apply the method/process to another wind farm, which has had more failures and also failures of various different components.

7 Appendix: Flowchart showing the data preparation
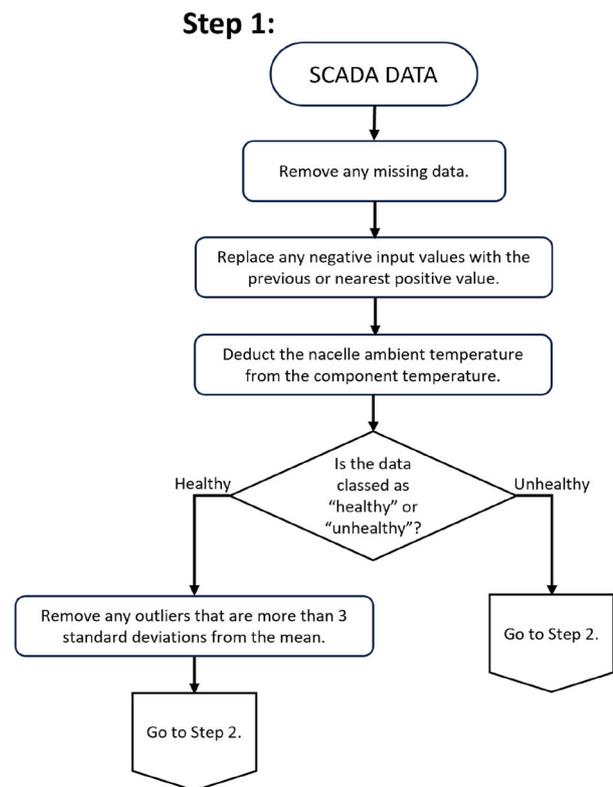
NOTE: "Healthy" data is data from year 2016. "Unhealthy" data refers to data from all other years.

Fig. 21 Flow Chart Showing the Data Preparation

8 Appendix: Tables showing the model uncertainty—mean and standard deviation—for different training/test data splits

Table 7 Average Mean and Standard Deviation Values for All Turbines Using Different Training/Test Data Split

Tur-bine	Training/Test : 60%/40%		Training/Test : 80%/20%	
	Mean	Standard Deviation	Mean	Standard Deviation
1	1.0010	0.0929	1.0011	0.0908
2	1.0017	0.0913	1.0012	0.0893
3	1.0014	0.0960	1.0008	0.0940
4	1.0013	0.0941	1.0005	0.0924
5	1.0014	0.0945	1.0011	0.0926
6	1.0017	0.1004	1.0018	0.0978

Table 8 Average Mean and Standard Deviation Values for All Turbines Using Different Training/Test Data Split (Daily Values)

Tur-bine	Training/Test : 60%/40%		Training/Test : 80%/20%	
	Mean	Standard Deviation	Mean	Standard Deviation
1	1.0010	0.0348	1.0012	0.0364
2	1.0014	0.0330	1.0006	0.0352
3	1.0018	0.0369	1.0011	0.0384
4	1.0009	0.0350	1.0005	0.0373
5	1.0017	0.0333	1.0015	0.0366
6	1.0013	0.0371	1.0015	0.0391

9 Appendix: Tables showing the error metrics for all data and daily values—sensitivity analysis

Table 9 Error Metrics for Different Data Splits for 2016 Data—All Values

Turbine	Error Metric	60%/40%	70%/30%	80%/20%
1	MAE	1.847	1.832	1.817
1	MSE	5.694	5.575	5.460
1	RMSE	2.386	2.361	2.337
1	MAPE	7.126	7.060	7.002
2	MAE	1.854	1.829	1.817
2	MSE	5.560	5.403	5.345
2	RMSE	2.358	2.324	2.312
2	MAPE	7.113	7.022	6.968
3	MAE	1.972	1.946	1.931
3	MSE	6.427	6.282	6.176
3	RMSE	2.535	2.506	2.485
3	MAPE	7.404	7.316	7.254
4	MAE	1.950	1.933	1.920
4	MSE	6.168	6.051	5.948
4	RMSE	2.483	2.460	2.439
4	MAPE	7.340	7.275	7.239
5	MAE	1.951	1.935	1.917
5	MSE	6.107	5.998	5.883
5	RMSE	2.471	2.449	2.425
5	MAPE	7.396	7.339	7.265
6	MAE	1.962	1.949	1.920
6	MSE	6.487	6.423	6.214
6	RMSE	2.547	2.534	2.493
6	MAPE	7.592	7.541	7.414

Table 10 Error Metrics for Different Data Splits for 2016 Data—Daily Values

Turbine	Error Metric	60%/40%	70%/30%	80%/20%
1	MAE	0.747	0.770	0.775
1	MSE	0.885	0.935	0.968
1	RMSE	2.386	2.361	2.337
1	MAPE	7.126	7.060	7.002
2	MAE	0.706	0.715	0.751
2	MSE	0.784	0.808	0.890
2	RMSE	2.358	2.324	2.312
2	MAPE	7.113	7.022	6.968
3	MAE	0.791	0.794	0.829
3	MSE	1.052	1.049	1.140
3	RMSE	2.535	2.506	2.485
3	MAPE	7.404	7.316	7.254
4	MAE	0.773	0.796	0.820
4	MSE	0.950	0.994	1.058
4	RMSE	2.483	2.460	2.439
4	MAPE	7.340	7.275	7.239
5	MAE	0.731	0.748	0.781
5	MSE	0.855	0.897	1.014
5	RMSE	2.471	2.449	2.425
5	MAPE	7.396	7.339	7.265
6	MAE	0.782	0.803	0.816
6	MSE	1.009	1.053	1.106
6	RMSE	2.547	2.534	2.493
6	MAPE	7.592	7.541	7.414

10 Appendix: Table showing a comparison of the error metrics—original vs new process

Table 11 Error Metrics for Original vs. New Process

Turbine	Error Metric	Original Method	New MC Method
1	MSE	12.69	3.59
2	MAE	1.90	0.37
2	MSE	23.19	15.43
3	MAE	1.84	0.22
3	MSE	9.81	2.36
4	MAE	2.30	0.50
4	MSE	15.61	5.17
5	MAE	1.71	0.17
5	MSE	7.46	1.34
6	MAE	2.79	0.49
6	MSE	20.88	4.68

11 Appendix: Figures showing actual temperature and confidence intervals for 1, 3 and 5 years of data for turbine 4
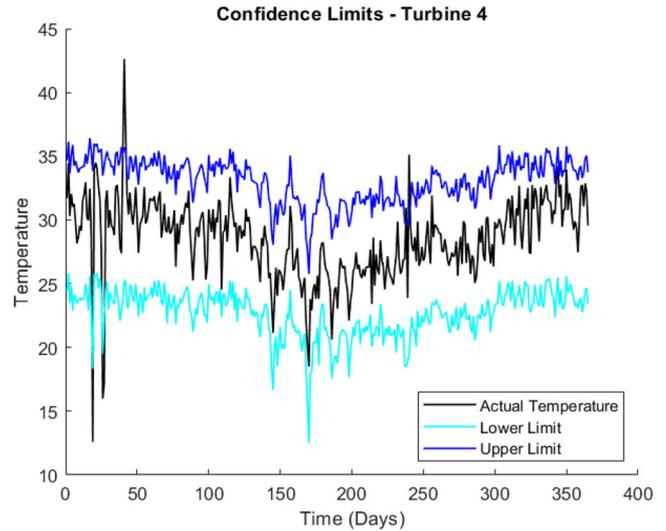
Fig. 22 Actual Temperature and Confidence Intervals for 1 Year Data for Turbine 4
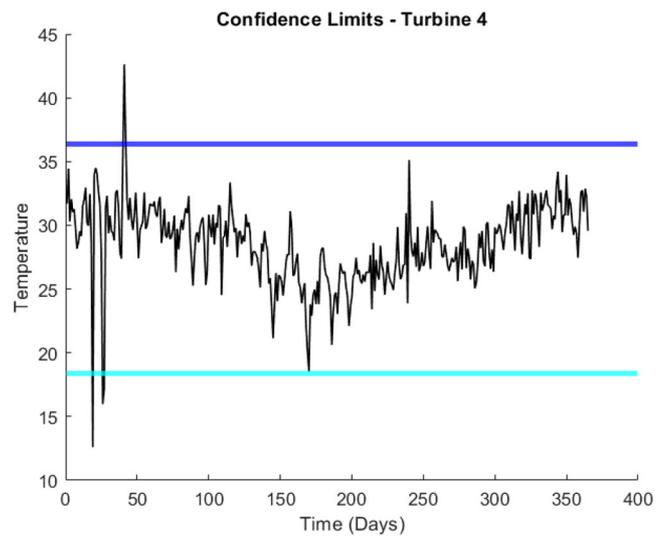
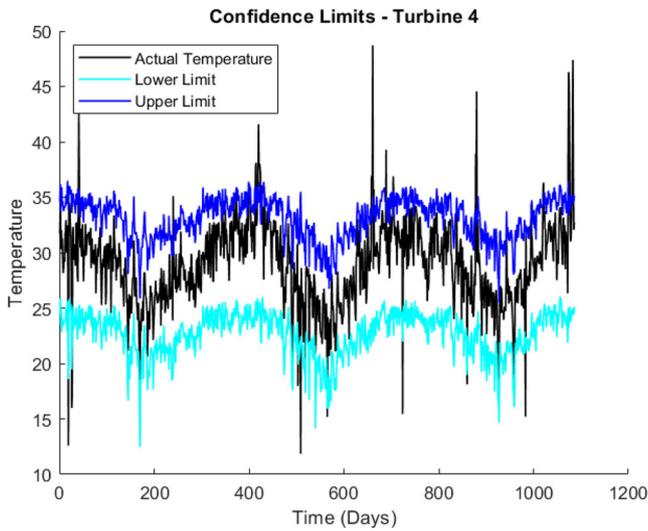**Fig. 23** Actual Temperature and Confidence Bands for 1 Year Data for Turbine 4

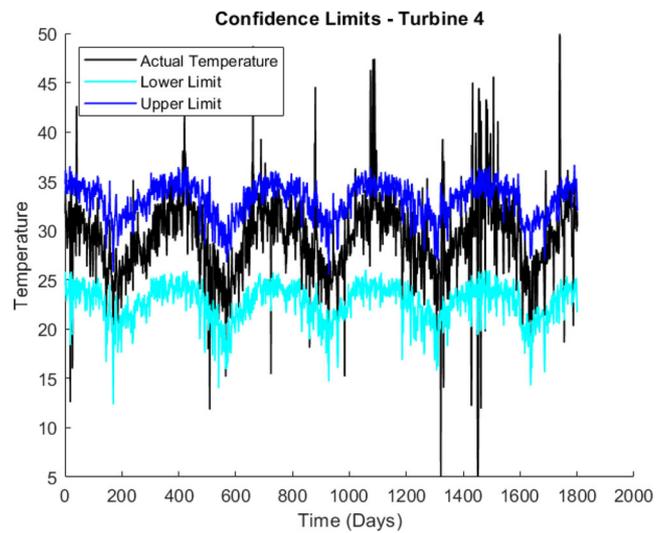**Fig. 24** Actual Temperature and Confidence Intervals for 3 Year Data for Turbine 4

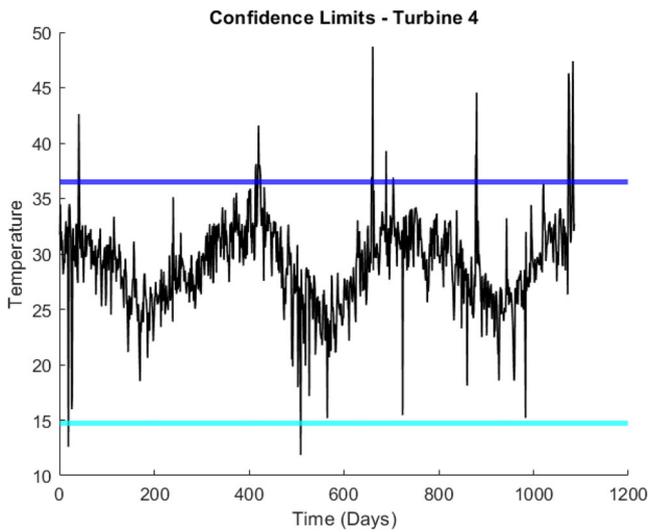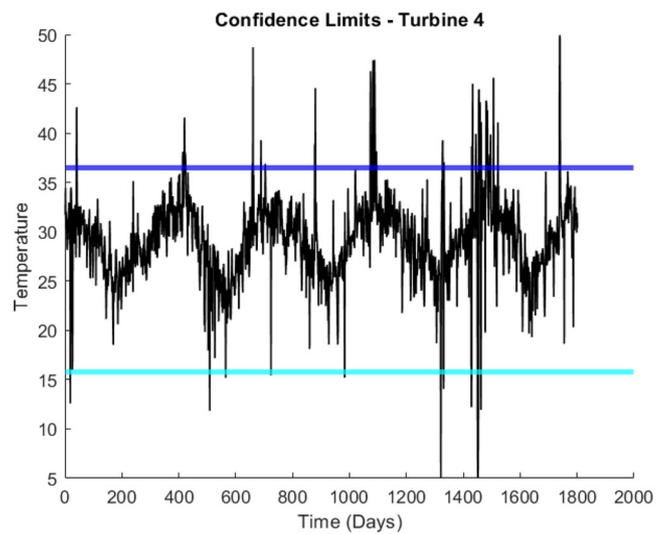**Fig. 26** Actual Temperature and Confidence Intervals for 5 Year Data for Turbine 4

Fig. 25 Actual Temperature and Confidence Bands for 3 Year Data for Turbine 4

Fig. 27 Actual Temperature and Confidence Bands for 5 Year Data for Turbine 4

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Change UNC (2023) About cop 28. https://unfccc.int/process-and-meetings/conferences/un-climate-change-conference-united-arab-emirates-nov/dec-2023/about-cop-28#What-will-be-discussed-at-COP-28

2. Campbell SL, Gear CW (1995) The index of general nonlinear DAES. Numer Math 72(2):173–196

3. Council GWE (2024) Global wind report 2024. https://gwec.net/global-wind-report-2024/

4. Tartt K, Kazemi-Amiri AM, Nejad AR et al (2024) Life extension of wind turbine drivetrains by means of scada data: Case study of generator bearings in an onshore wind farm. Results Eng 24:102,921

5. Sørensen JD, Toft HS (2010) Probabilistic design of wind turbines. Energies 3(2):241–257

6. Tartt K, Nejad AR, Kazemi-Amiri A et al (2021) On lifetime extension of wind turbine drivetrains. In: International Conference on Offshore Mechanics and Arctic Engineering. American Society of Mechanical Engineers, p V009T09A025

7. Tartt K, Kazemi-Amiri A, Nejad A et al (2022) Development of a vulnerability map of wind turbine power converters. In: Journal of Physics: Conference Series. IOP Publishing, p 32052

8. Bai Y, Jin WL (2016) Random variables and uncertainty analysis. Mar Struct Des 12:615–625

9. Ouyang T, Kusiak A, He Y (2017) Predictive model of yaw error in a wind turbine. energy 123:119–130

10. De Giorgi MG, Ficarella A, Tarantino M (2011) Error analysis of short term wind power prediction models. Appl Energy 88(4):1298–1311

11. Knudsen T, Bak T, Soltani M (2011) Prediction models for wind speed at turbine locations in a wind farm. Wind Energy 14(7):877–894

12. Reder M, Melero JJ (2018) Modelling the effects of environmental conditions on wind turbine failures. Wind Energy 21(10):876–891

13. Sun P, Li J, Wang C et al (2016) A generalized model for wind turbine anomaly identification based on scada data. Appl Energy 168:550–567

14. Barboza F, Kimura H, Altman E (2017) Machine learning models and bankruptcy prediction. Expert Syst Appl 83:405–417

15. Jiang Z, Hu W, Dong W et al (2017) Structural reliability analysis of wind turbines: A review. Energies 10(12):2099

16. Nejad AR, Gao Z, Moan T (2014) On long-term fatigue damage and reliability analysis of gears under wind loads in offshore wind turbine drivetrains. Int J Fatigue 61:116–128

17. Tarp-Johansen NJ (2003) Examples of fatigue lifetime and reliability evaluation of larger wind turbine components

18. Dong W, Moan T, Gao Z (2012) Fatigue reliability analysis of the jacket support structure for offshore wind turbine considering the effect of corrosion and inspection. Reliab Eng Syst Saf 106:11–27

19. Tavazza F, DeCost B, Choudhary K (2021) Uncertainty prediction for machine learning models of material properties. Acs Omega 6(48):32,431–32,440

20. Jiao X, Zhang D, Wang X et al (2023) Wind speed prediction based on error compensation. Sensors 23(10):4905

21. Pandit R, Kolios A (2020) Scada data-based support vector machine wind turbine power curve uncertainty estimation and its comparative studies. Appl Sci 10(23):8685

22. Cao L, Zhang H, Meng Z et al (2023) A parallel gru with dual-stage attention mechanism model integrating uncertainty quantification for probabilistic rul prediction of wind turbine bearings. Reliab Eng Syst Saf 235:109,197

23. Afanasyeva S, Saari J, Kalkofen M et al (2016) Technical, economic and uncertainty modelling of a wind farm project. Energy Convers Manag 107:22–33

24. Gonzaga P, Toft H, Worden K et al (2022) Impact of blade structural and aerodynamic uncertainties on wind turbine loads. Wind Energy 25(6):1060–1076

25. Liu R, Peng L, Huang G et al (2023) A monte carlo simulation method for probabilistic evaluation of annual energy production of wind farm considering wind flow model and wake effect. Energy Convers Manag 292:117,355

26. Dong W, Nejad AR, Moan T et al (2020) Structural reliability analysis of contact fatigue design of gears in wind turbine drivetrains. J Loss Prev Process Ind 65:104,115

27. Chakrabarty A, Mannan S, Cagin T (2015) Multiscale modeling for process safety applications. Butterworth-Heinemann

28. Ahmed MW (2023) Understanding mean absolute error (MAE) in regression: A practical guide. https://medium.com/@m.waqar.ahmed/understanding-mean-absolute-error-mae-in-regression-a-practical-guide-26e80ebb97df

29. Hodson TO (2022) Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. Geosci Model Dev Discuss 2022:1–10

30. Roberts A (2023) Mean absolute percentage error (MAPE): What you need to know. https://arize.com/blog-course/mean-absolute-percentage-error-mape-what-you-need-to-know/#:~:text=One%20of%20the%20most%20common,off%20predictions%20are%20on%20average

Supplementary information Not applicable.