Received: 13 March 2019 | Revised: 4 July 2019 | Accepted: 10 July 2019
DOI: 10.1002/we.2402

WILEY

RESEARCH ARTICLE

Offshore wind turbine fault alarm prediction

Alexios Koltsidopoulos Papatzimos[1,2] | Philipp R. Thies[3] | Tariq Dawood[2]

[1]Industrial Doctoral Centre for Offshore Renewable Energy (IDCORE), The University of Edinburgh, Edinburgh, UK
[2]EDF Energy R&D UK Centre, Interchange, London, UK
[3]College of Engineering, Mathematics and Physical Sciences, Renewable Energy Group, University of Exeter, Exeter, UK

Correspondence
Alexios Koltsidopoulos Papatzimos, Industrial Doctoral Centre for Offshore Renewable Energy (IDCORE), The University of Edinburgh, Edinburgh EH9 3JL, UK.
Email: a.koltsidopoulos-papatzimos@ed.ac.uk

Funding information
Energy Technology Institute and the RCUK Energy Programme, Grant/Award Number: EP/J500847/1; EDF Energy

Abstract

Offshore wind operations and maintenance (O&M) costs could reach up to one third of the overall project costs. In order to accelerate the deployment of offshore wind farms, costs need to come down. A key contributor to the O&M costs is the component failures and the downtime caused by them. Thus, an understanding is needed on the root cause of these failures. Previous research has indicated the relationship between wind turbine failures and environmental conditions. These studies are using work-order data from onshore and offshore assets. A limitation of using work orders is that the time of the failure is not known and consequently, the exact environmental conditions cannot be identified. However, if turbine alarms are used to make this correlation, more accurate results can be derived. This paper quantifies this relationship and proposes a novel tool for predicting wind turbine fault alarms for a range of subassemblies, using wind speed statistics. A large variation of the failures between the different subassemblies against the wind speed are shown. The tool uses 5 years of operational data from an offshore wind farm to create a data-driven predictive model. It is tested under low and high wind conditions, showing very promising results of more than 86% accuracy on seven different scenarios. This study is of interest to wind farm operators seeking to utilize the operational data of their assets to predict future faults, which will allow them to better plan their maintenance activities and have a more efficient spare part management system.

KEYWORDS

alarms, failure, maintenance, reliability, turbulence intensity, wind speed

1 | INTRODUCTION

Offshore wind operations and maintenance (O&M) costs could reach up to 30% of the total project cost.[1] These costs are relatively high due to accessibility issues and the need of dedicated vessels and personnel for the turbines' repair. Moreover, weather limitations can delay or abort the operations, which can result in longer waiting times for the turbine to be repaired, causing downtime and loss of energy production. Thus, it is important to have robust planning, maintenance, and monitoring strategies, which can detect faults in advance to critical failures and notify the maintenance managers.

A way to better understand how failures are occurring is to closely investigate the environmental conditions. Previous studies have explored the effect of wind speed and turbulence intensity on turbine failures, by using wind speed measurements and work-order information.[2-5] Although some correlations have been made between the two parameters, it is not always possible to know the exact time a failure has happened and when it was logged in the work orders. This is due to the fact that work orders are logging maintenance actions when completed, which could be several hours or days after the alarm has been triggered, due to unavailability of technicians or spare parts or due to bad weather.

Turbine alarms can be a good indicator to detect when a failure has occurred. Historical work-order information can also help to create failure rate statistics[4] but does not always indicate the exact time that a failure has occurred, only the time a failure has been detected and repaired or a component has been replaced. The alarms provide valuable information to identify potential failure locations and root causes, which could be

The peer review history for this article is available at https://publons.com/publon/10.1002/we.2402

especially efficient as they require the least storage and analysis compared with other sensor signals.[6] Alarms can be triggered when component signals exceed defined threshold limits.[6] These include instances when a wind turbine running state has changed or because of a component malfunction or due to a design defect. Turbine alarms cannot be directly correlated to a failure, as alarms can occur by a system malfunction, corresponding to false alarms. This paper will explore whether careful processing of the alarms can reveal issues with the assets, as they show the exact time stamp of the failure event. This data type can then be easily correlated with any other sensor or environmental information. If faults are predicted in advance, maintenance and repair operations can be combined or planned in advance, which will inform the spare part requirements and ensure that the necessary technical personnel is available. Although previous studies have presented frameworks using alarms to process and better interpret wind turbine failures,[7,8] no previous study has attempted to use turbine alarms and wind speed measurements in order to predict upcoming faults.

This paper presents a tool for turbine alarm prediction, using wind speed statistics. The novelty of this work lies in the fact that no such tool has been identified by the authors in the public domain. Such a tool can be used for either short- or long-term planning of offshore wind farm resources, allowing effective and efficient troubleshooting, availability of spare parts, reduced downtime, and optimized offshore wind strategies. The unique element of this paper is that it provides a decision-making tool methodology that can be replicated by offshore wind farm operators allowing them to better plan their activities and reduce operational costs, by using readily available data from the turbines and the met mast.

The paper is structured in four parts. Section 2 presents the methodology for the collection, integration, and postprocessing of the wind turbine data to create the input parameters, as well as the tool itself. Section 3 shows the results of using the tool with 5 years of operational data and testing it under high and low wind conditions. Section 4 discusses the results and the tool's benefits for wind farm operators, as well as its limitation. Finally, Section 5 concludes and proposes future work on this topic.

2 | METHODOLOGY

2.1 | Wind data integration

In order to plan and conduct effective O&M for offshore wind turbines, the wind sector is increasingly facing big data challenges. The data currently generated by multimegawatt wind turbines represent the big data's four Vs: volume, variety, velocity, and veracity. They include multiple sensors, generating live data, as well as different reports with operational information. These typically consist of more than 20 different points of information or data sources at a modern offshore wind farm. Trying to derive useful conclusions from the data is increasingly challenging with the current operation and maintenance workflows. This work suggests an effective methodology based on big data management and processing principles following three steps:

- Extraction, transformation, and loading (ETL) process. This ETL process requires to automatically extract the raw data or some preselected metadata (such as a small description of the work orders or the fault investigated) from all the different data sources. The data then need to be quality checked and filtered if required. In the case of different time stamps, these would need to be aligned in order to speed up the integration process. The next step is to load the data in the warehouse or database.
- Data integration. This paper proposes a relational database (RDB) to integrate the data that have been extracted and organized in a structured format. The data can then be integrated using their time stamps as a primary key.
- Data visualization is the visualization of the data, using a relevant software, in order to better analyze them, to create the required indicators, and to support any data mining needed.
- Data analysis and interpretation.

A detailed description of the steps followed for this study is shown below.

2.2 | Alarms tool

In order to analyze the operational data for offshore wind farms, a methodology to integrate the data from different sources has been developed. This is achieved through an RDB that consolidates all the available data sets and allows convenient analysis and visualization. RDBs are digital databases that are based on relational models of data. Relational models offer a structured way to manage data that allows the users to create a declarative method for specifying data and queries. Most of the RDBs use a structured query language (SQL) for querying and maintaining the database.

There are typically two types of data: structured and unstructured ones. In the offshore wind case, structured data are received from sensors, as well as any metadata generated from inspection and maintenance reports, such as work orders or short inspection report conclusions. Unstructured data are typically generated through reports and any related pictures and videos. A framework for integrating offshore wind data has been suggested by Nguyen et al,[9] and a case study has been presented by Koltsidopoulos Papatzimos et al.[10]

A flowchart of the data handling process is shown in Figure 1, which will be further explained in the following sections.

2.2.1 | Data sources

The data sources used in this paper include 5 years of operational wind farm data for 27 Siemens 2.3-MW turbines located 1.5-km north of Redcar off the Teesside coast, in the North Sea, England. These include met mast, supervisory control and data acquisition (SCADA), alarms,

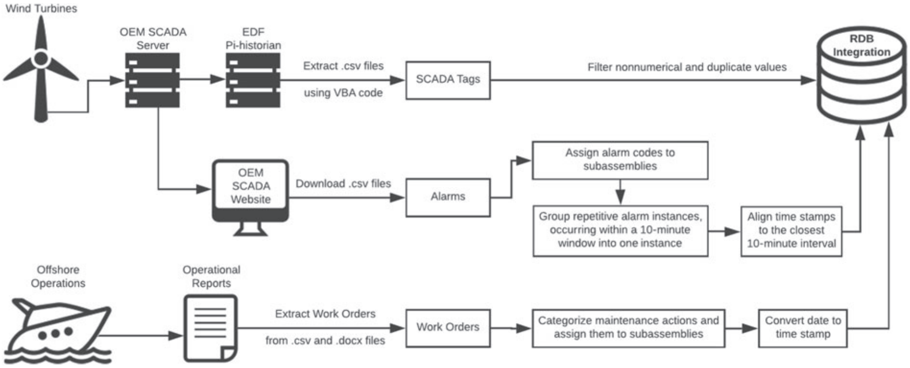


FIGURE 1 Data retrieval and integration

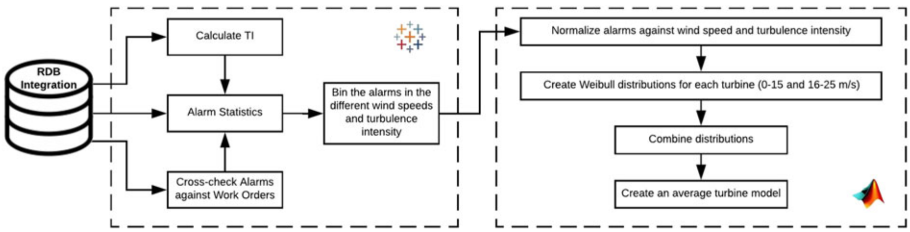


FIGURE 2 Data analysis [Colour figure can be viewed at wileyonlinelibrary.com]

and maintenance operation description from the maintenance logs. Only the met mast data are presented for a shorter period of 2 years from September 2015 to September 2017; this does not have any influence on the design and training of the tool, as the met mast data are only used for validating the model.

2.2.2 | Data transformation

The performed data handling will be typical of the data processing needs of offshore wind farms and is thus presented in some more detail here: First, data handling step was to manipulate the data and convert them into the required format, preparing it for later integration. This process is formed on the following steps:

- Extraction of the information included in the maintenance logs and of the relevant SCADA and met mast information.
- Filtering of all the data. Once all the required information was collected and extracted, it was assessed for consistency. The met mast and SCADA data were filtered (using Bash commands) in order to remove duplicate values and to avoid any nonnumerical values from sensor readings caused by connection issues. For nonnumerical instances, the entire time stamp was discarded for all the data sources.
- Grouping of turbine alarms. For reoccurring alarms within half an hour from the previous one, the alarms were considered as a single instance.
- Identification of critical alarms. The alarms that have led to a further investigation, either by sending personnel offshore to inspect or to repair the potential failure or by shutting the turbine down and investigating the issue remotely, are classified as critical alarms.
- Alignment of work order, alarm, and SCADA sensor time stamps. The alarm time stamps were rounded to the closest 10 minutes because of their complex time stamps, in order to ease the integration with the SCADA data. The work orders were then assigned to the alarm time stamp indicating that the turbine is shut down for maintenance. Every time that the turbine is in local control for a maintenance activity, a work order is logged.
- Organization of the data sources. The data were then prepared to be imported into the database. Furthermore, the O&M data were organized by categorizing the actions to failures and adding the relevant components and subassemblies. The alarms had to be filtered in order to group any repetition on the same turbine and to separate any remote/local/manual stops or upgrades of the turbine's software with any alarm that indicates failure or degradation of components. The taxonomy used followed the one from Carroll et al[4] for consistency.

2.2.3 | Data integration

The RDB structure has been chosen so that it is flexible, scalable, and adaptive to any additional sensors or data sources that will need to be added in the future or when the DB will be updated. The initial tables were created, using the column that contains the date and time (date/time/time stamp) information as a primary key. This guarantees that all the time values inserted will be unique, if not an error will be returned. In that case, the data will need to be filtered again or checked. Using the time stamp as a primary key saves computational time, as less data will be processed, and could save up to 20% storage space per table.

2.2.4 | Data postprocessing

The analysis of the data was performed in two stages. Initially, the data were visualized from the data base using Tableau, which helped to visually interpret the data and create some quick data aggregating, and then, the data were transferred to MATLAB for the statistical analysis and the implementation of the tool. A visual representation of the process is shown in Figure 2.

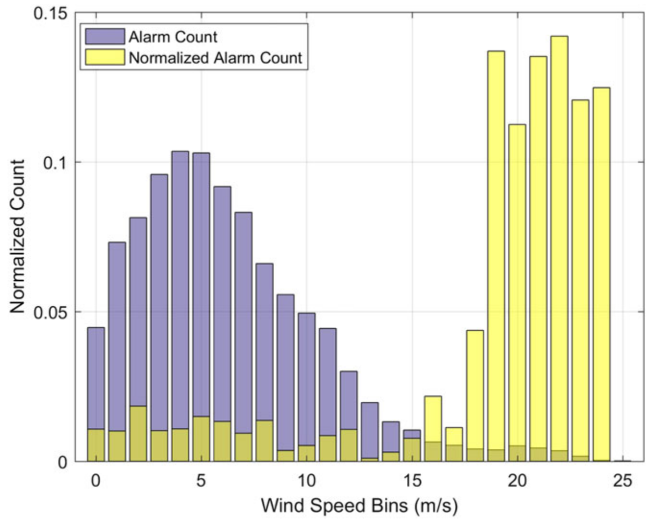


FIGURE 3 Alarm count against wind speed for normalized alarm count data against wind speed and nonnormalized ones for wind turbine number 13 [Colour figure can be viewed at wileyonlinelibrary.com]

Data visualization

The data were initially visualized using Tableau. An initial sanity check was performed in order to make sure that the work orders and the alarms match. Once this was completed, the alarms were binned against the different wind speed and turbulence intensity (TI) values for the different turbines. TI was calculated by using Equation (1), where U is the wind speed and σU is the standard deviation of the wind speed for a 10-minute average data sampling period.

$$TI = \frac{\sigma U}{U} \tag{1}$$

Reliability analysis

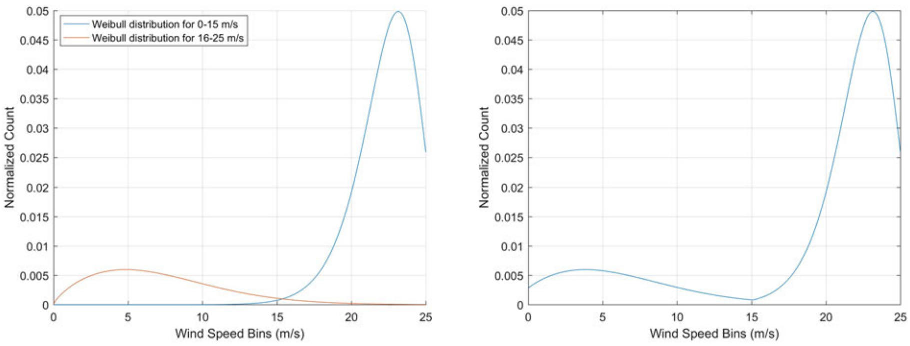
The binned data were then transferred to MATLAB. The data were normalized against wind speed and TI to account for the environmental conditions; ie, the alarms were normalized regarding the given occurrence of each bin. An example of normalized and nonnormalized data is shown in Figure 3 for wind turbine 13. The example shown in Figure 3 is an extreme case, in order to emphasize this relationship. Two different Weibull distributions were generated for the data from 0 to 15 and 16 to 25 m/s. The range of the distributions was decided due to the shape of the data shown in Figure 3, and it could vary depending on the analysis. This essentially creates two different probabilistic models for lower and for higher wind speeds. After that, the distributions were combined into a single one, and they were averaged for the whole wind farm, as shown in Figure 4. This allows the generation of a generic model for the farm. The data shown are normalized against the frequency of the wind speed, meaning that, for example, for every 10-minute average 23-m/s recordings, there is a 5% chance that there will be an alarm. The normalization of the alarms against the wind speed was done by dividing the alarm count by the number of wind speed instances (ie, wind speed occurrence) in every bin. This seeks to evaluate the alarms. The combination of the distributions was done by neglecting the tails and only taking into account the more dominant part of the distribution, as shown in Figure 4B. A chi-square goodness-of-fit test was used for the two Weibul distributions, and the P values calculated were 0.1625 for the 0- to 16-m/s distribution and 0.1108 for the 17- to 25-m/s distribution. For the TI, a Rayleigh distribution has been used, as shown in Figure 5, with a goodness of fit of 0.1259. The generic steps for the distribution generation are shown in Algorithm 1.

Algorithm 1 Generic distribution generator

```
1: Input: binned failure rates for individual subassemblies and generic turbine one
2: for all failure rate inputs do
3:     create distributions
4:     repeat
5:      test distribution goodness of fit
6:     until distribution satisfies fit requirements
7: end for
       return Weibull and Rayleigh distributions for the wind farm and subassemblies
```

2.2.5 | Tool

The tool, Figure 6, reads as inputs time series of 10-minute average wind speed and standard deviation data. The TI for the farm is then calculated, Equation 1, by averaging all the individual TI values. The tool then reads the failure alarm data for the different subassemblies. In this case, this includes 5 years of historical data collected from Teesside offshore wind farm. Then, the data are analyzed as described in Section 2.2.0. The tool can then generate an estimate of the alarms for the different subassemblies and an overall one for the wind farm.



(A) 2 Weibull distributions (B) Combined Weibull distributions

FIGURE 4 A,B, Weibull distributions for a generic farm model [Colour figure can be viewed at wileyonlinelibrary.com]

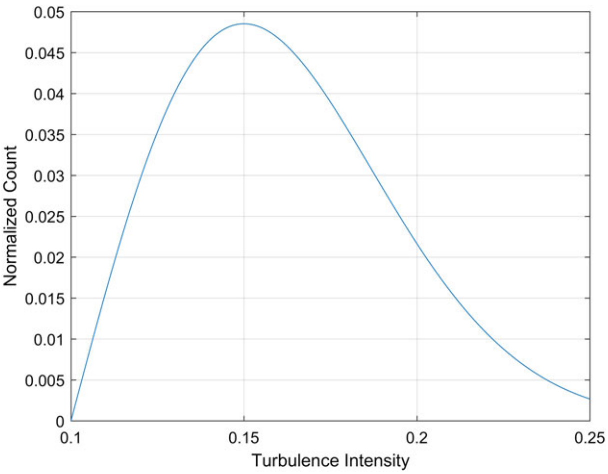


FIGURE 5 Rayleigh distribution for turbulence intensity [Colour figure can be viewed at wileyonlinelibrary.com]

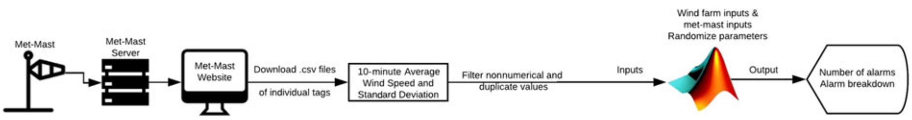


FIGURE 6 Model inputs [Colour figure can be viewed at wileyonlinelibrary.com]

2.2.6 | Testing

Once the 10-minute average wind speed and standard deviation data have been correlated with the fault alarms, a prediction can be made based on the provided predicted wind speeds. Once the data from all the turbine sensors have been aggregated and the average turbine model has been built, the tool can be tested. In order to test the capabilities of the tool, the inputs from the met mast were used, as shown in Figure 6. This gives values for a single data source, making it easier and less complicated to test, as just one source of information is needed in order to test the tool. A generic step-by-step pseudocode is presented in Algorithm 2.

Algorithm 2 Generic tool algorithm

```
Input: met-mast 10-minute average and standard deviation timeseries
Calculate TI
Input: Weibull and Rayleigh distributions for the wind farm and subassemblies
for individual subaasemblies do
    initialize probability for subassembly
    for met-mast parameters do
        select met-mast parameters
        select equivalent parameters from Weibull and Rayleigh distributions
        assign probabilities for wind speed and TI
        average the two probabilities
        add to previous probability
    end for
    Multiply final probability with the number of turbines
end for
    return wind farm and subassembly number of alarms expected
```

TABLE 1 Tool outputs for the different failure alarms as a percentage of the total turbine alarms

Alarms	% of Failure Alarms	% of Critical Failure Alarms
Converter	28.90	2.75
Electrical	8.15	1.85
Gearbox	2.85	1.66
Generator	26.15	4.02
Inverter	2.36	1.53
Pitch	3.89	1.00
Sensors	2.45	1.50
Transformer	0.89	0.28
Yaw	8.95	0.85

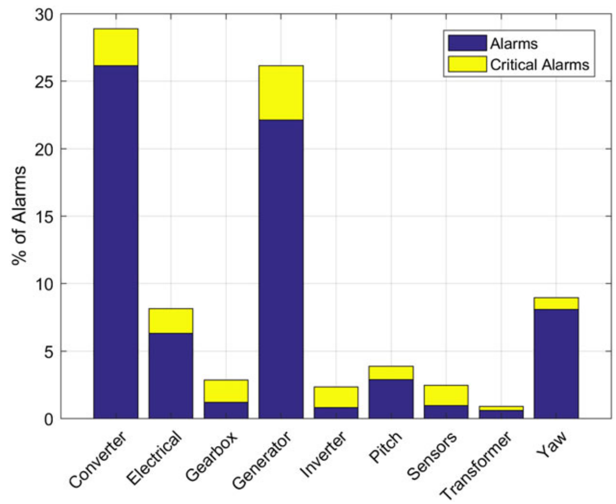


FIGURE 7 Tool outputs for the different failure alarms as a percentage of the total turbine alarms [Colour figure can be viewed at wileyonlinelibrary.com]

The tool was then tested in two different scenarios:

- Short-term planning. This use of the tool would be for daily operational decision making. It is mainly dependent on the accuracy of the available forecasted data, and it cannot exceed 10 days.
- Long-term planning. This use of the tool would be useful to owners and operators planning a long-term strategy for their farm, and given the available wind data, it could provide estimates for a longer period of time, informing future planning and spare part management strategies.

The tool was also tested for high and low wind conditions. The separation of the high and low wind conditions was chosen due to the data availability, as there were not long enough periods of time for wind speeds well over 10 m/s.

3 | RESULTS

3.1 | Findings

For a 2-year period, the predicted alarm distribution is shown in Table 1 and Figure 7. The overall accuracy achieved was 99.3% for all the alarms. Over that period, the tool overpredicted the number of alarms by 1.5%. This value is the average of the total number of alarms predicted, the alarms from the different subassemblies, and the number of critical alarms. The actual number of alarms is not shown due to confidentiality reasons. Instead, the percentage of the total number of alarms is shown. The rest of the alarms not shown on the table are the ones that are either automatically reset and resolved or ones not indicating failures but changes of state of the turbine, such as "remote stop," or other warnings, such as "high/low wind speed."

Figure 8 also shows a cumulative distribution function for the different fault alarms during the 5-year period. In general, a wide variation is shown among the different subassemblies. It is shown that yaw and gearbox systems are more affected by the higher wind speeds, whereas sensor, pitch, and electrical systems experience more issues during the lower wind speeds and thus during higher turbulence intensity values. Generators, converters, transformers, and inverters have a more uniform failure distribution. A more detailed analysis of these findings can be found at another study of Koltsidopoulos Papatzimos et al.[11]

3.2 | Comparison

The tool was then tested at the following cases:

- High wind condition—considered as a period of time with average wind speeds higher than 10 m/s, shown in Figure 9A, 9C, and 9E.

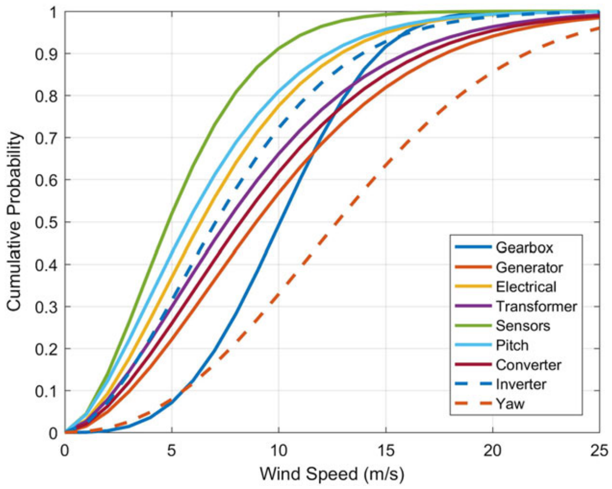
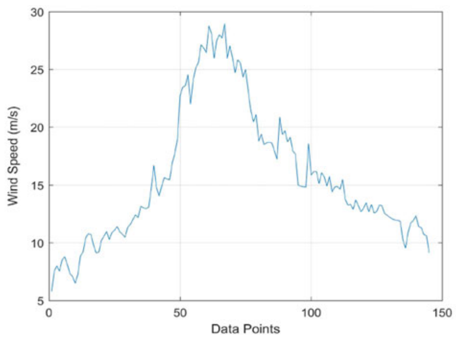
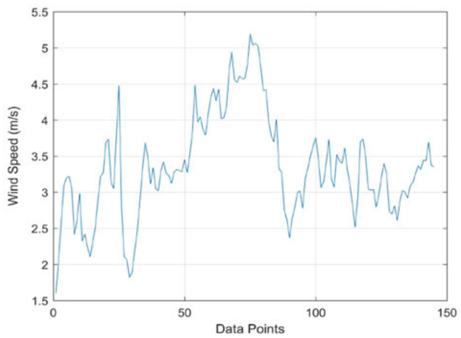


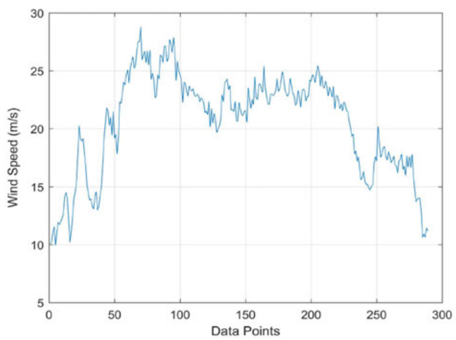
FIGURE 8 Cumulative distribution function for the different subassemblies (gearbox, generator, electrical systems, transformer, sensors, pitch system, converter, inverter, and yaw) against wind speed [Colour figure can be viewed at wileyonlinelibrary.com]



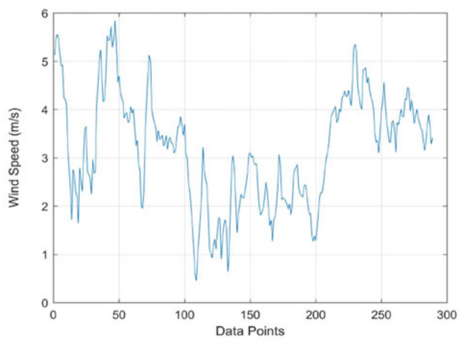
(A) 24-hour high wind condition



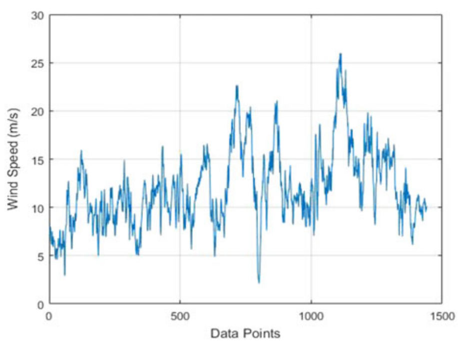
(B) 24-hour low wind condition



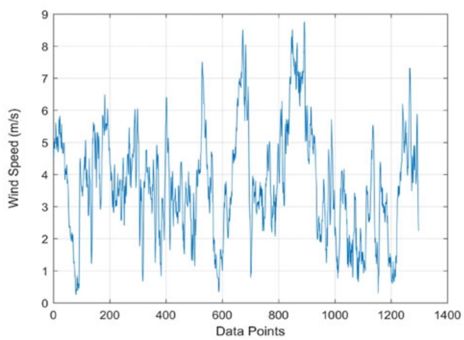
(C) 48-hour high wind condition



(D) 48-hour low wind condition



(E) 10-day high wind condition



(F) 9-day low wind condition

FIGURE 9 A-F, Met mast wind speed inputs for the model [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Tool's overall accuracy

Duration	Low Wind	High Wind
24 h	88.5% (11.5% underestimated)	87.8% (12.2% underestimated)
48 h	92.7% (7.3% overestimated)	90.3% (9.7% underestimated)
9/10 days	86.9% (13.1% overestimated)	93.1% (6.9% underestimated)
2 y	99.3% (0.7% overestimated)	

TABLE 3 Comparison of actual and predicted number of alarms for a 24-hour low wind (LW) and high wind (HW) periods

Subassemblies	LW Actual	LW Predicted	HW Actual	HW Predicted
Converter	1	1.47	1	1.48
Electrical	0	0.54	0	0.21
Gearbox	0	0.03	0	0.15
Generator	1	1.07	2	1.96
Inverter	0	0.13	0	0.09
Pitch	0	0.28	0	0.10
Sensors	0	0.22	0	0.02
Transformer	0	0.07	0	0.04
Yaw	0	0.13	1	1.33

TABLE 4 Comparison of actual and predicted number of alarms for a 48-hour low wind (LW) and high wind (HW) periods

Subassemblies	LW Actual	LW Predicted	HW Actual	HW Predicted
Converter	3	2.79	7	7.10
Electrical	1	1.04	1	0.54
Gearbox	0	0.07	0	0.17
Generator	1	2.02	11	10.52
Inverter	0	0.25	0	0.13
Pitch	0	0.55	0	0.27
Sensors	0	0.44	0	0.09
Transformer	0	0.13	0	0.06
Yaw	1	0.24	6	5.04

TABLE 5 Comparison of actual and predicted number of alarms for a 9/10-day low wind (LW) and high wind (HW) periods

Subassemblies	LW Actual	LW Predicted	HW Actual	HW Predicted
Converter	10	12.84	15	13.99
Electrical	5	4.53	3	2.65
Gearbox	0	0.46	2	1.85
Generator	9	9.62	18	16.79
Inverter	1	1.12	1	0.97
Pitch	2	2.36	1	1.20
Sensors	2	1.84	1	0.42
Transformer	0	0.53	1	0.42
Yaw	2	1.40	9	10.07

- Low wind condition—considered as a period of time with average wind speeds less than 10 m/s, shown in Figure 9B, 9D, and 9F.

All of the cases used are within the same time period used to complete the original distribution, but they only represent a few days, taking into account that the original distribution was created using a 5-year period. Moreover, they are not coming from the same data source, as the Weibull distribution data are from the individual turbine wind anemometers and the ones used for the model's validation are from the met mast.

In both cases, a short term period (24 and 48 h) and a longer period are tested of 9/10 days, in order to evaluate the capabilities of the tool in shorter- and longer term forecasting. The different inputs from the met mast are shown in Figure 9, and the overall average accuracy results are shown in Table 2. Accuracy for the individual subassemblies has been calculated by using Equation 2. As it can be seen, the overall low wind accuracy is lower compared with the high wind one. For the high wind cases, the accuracy increases with the increase of the period tested, whereas the low wind case is inconsistent.

$$Accuracy = \frac{|Real - Predicted|}{Real} \tag{2}$$

A more detailed output of the model and a comparison with the actual data are shown in Tables 3 to 5. The tables show a comparison between the actual and the predicted values for the different cases considered. For the predicted values, as the results are probabilistic, the decimals are also shown. In reality, those values could be either visualized like that or could be rounded to the closest whole number, as the number of the alarms can only be an integer.

The comparison section of the results is limited to the real and predicted values only, and no comparison with other tools was made, as there is no similar tool identified by the authors in the public domain.

4 | DISCUSSION

The overall analysis of the results indicates that about one fifth to one fourth of the alarms generated by the turbine need some kind of action, which could either be a remote reset or an intervention. Around 15% of the overall fault alarms are critical and will need in situ investigation or replacement of the turbine components. Moreover, the large variation of the failures between the different subassemblies against the wind speed is also highlighted.

The normalized alarms against wind speed in Figure 3 indicate that there are comparatively more alarms at wind speeds higher than 18 m/s, indicating the influence of high wind speeds on the number of alarms. It is worth noting that at wind speeds 13 to 15 m/s, there is a dip in the number of alarms, compared with the ones experienced at lower wind speeds. These are the wind speeds where the turbine is reaching its rated power. It can be assumed that since these are the rated wind speed velocities, the turbine is designed to operate under those conditions and thus, less failure alarms are triggered.

The forecasting capabilities of the tool have shown an overall 99.3% accuracy considering the prediction based on the 2 years of data provided by the met mast. This is a really high accuracy value, which indicates that the results can be potentially used in long-term prediction of failure alarms in strategy and operational expenditure tools. The short-term prediction of 24/48 hours and up to 10 days has shown accuracies ranging from 83.1% to 93.1%, indicating that the tool does not perform that well for short-term forecasting. This is expected as the models are data driven and the failures occurrence is a probabilistic process, based also on the physics of failure, which would result in a higher accuracy when dealing with larger data sets. Moreover, short-term wind forecasting depends on the forecasting capabilities of the different weather models, which can add additional uncertainties into the estimation of future failure alarms. In this study, we have assumed a perfect forecasting, as historical data have been used. On average, the low wind cases produce less accurate results compared with the high wind ones, probably due to the higher TI values at low wind speeds, which can cause the underestimation of the generated failure alarms. The 48-hour window provides the best short-term estimate with an accuracy of over 90% for both high and low wind cases, which means that the tool could be used for for a 48-hour forecasting with a high level of accuracy.

Some limitations of this work are discussed below:

- The training data for the tool include readings from the turbines' anemometers, which can be influenced by the wakes generated by the turbines' rotors. This is an unavoidable uncertainty in the inputs of this study but is characteristic of the type of operational data that a wind farm owner would have typically access to. The anemometers would give point measurements for every turbine, which can then be correlated to the individual failures. For future uses, the use of nacelle lidar systems can be investigated to reduce this uncertainty in the readings, but the benefits of lidar will have to outweigh its cost.
- The tool does not work for individual turbines. It generates generic farm outputs that indicate the overall expected failures for the assets. Although this information might not specify the turbine, it is still very useful for planning of component replacement and repairs.
- The tool is producing accurate results for the first 5 years of the asset, after that some deterioration factors might need to be taken into account to represent more realistic data.
- The tool and methodology presented in this paper are data-driven, so a considerable amount of data are needed in order to generate accurate distributions for the different turbines.
- If the tool is used in a live case study, the level of uncertainty in the short-term applications might be higher, due to any uncertainties related to the weather forecasting.
- The tool can only be used for failures detected by the alarm system. This is not always the case though, as failures either can occur without any warning or can be detected earlier by anomaly detection or trending analysis.[12] As an effect, wind farm operators should not exclusively rely on such a tool but use it to inform their future actions.

The overall tool's performance is very promising, as it shows that with readily available operational SCADA and met-mast data and reasonable computation time, the tool can rerun when needed and produce highly accurate results. The majority of the work is spent during the preprepossessing, aggregation, and categorization of the different data. The tool's performance is also high due to the fact that the results are averaged for the wind farm and are not shown for individual turbines.

Previous work has proven the relationship between wind speed and failure rates for onshore and offshore wind turbines, indicating that higher wind speeds will result to higher turbine failure rates.[4,6,8] This paper took those results a step further and quantified them on a component level and also used the environmental data to help predict future failures. These results can then be used in operational strategy and cost prediction tools, such as ECUME O&M.[13]

The use of such a tool would allow prescriptive maintenance to be applied at the offshore wind assets, allowing the right personnel and materials to be in place, minimizing delay and downtime.

5 | CONCLUSION

This study has presented a novel alarm estimation tool for the offshore wind industry, by considering the wind speed and TI parameters. Five years of operational data have been used in order to create a robust model. The data were collected by the individual SCADA systems of the turbines at Teesside offshore wind farm, categorized by failure type and subassembly, and integrated into an RDB. They were then binned into different wind speeds and TIs by alarm type, and a generic turbine model was created. A model was built, and it was tested by using 2 years of data from the site's met mast. The results are very promising, showing an overall 99.3% accuracy of the model for the 2 years of data. The results are less accurate for short-term forecasting, making the tool more suitable for longer term strategy and operational cost estimations.

The study has indicated that such a tool can be built by offshore wind farm operators by just incorporating maintenance logs, turbine alarms, and wind speed measurements, which are available at all operational sites. Moreover, due to the small size of the inputs and the generic nature of the tool, it can run very quickly and predict future alarms in a matter of seconds.

The application of such a tool can be expanded to other operational onshore and offshore wind farms and tested by wind farm operators. This will allow wind farm operators to prioritize their task effectively, have the relevant personnel and spare parts available, and reduce the downtime of the assets due to a potential fault.

Future work will focus on the implementation of the results using an O&M cost prediction model to better understand the effect that a failure can have on the turbine's location and how that affects the overall downtime and cost of the repair for the asset. Future work could also focus in understanding the sensitivity of the inputs and outputs of such a tool. It could also be interesting to further investigate the lower failure alarm instances experienced at the turbine's rated wind speed. Finally, future work could investigate the influence of the wind speed forecast accuracy in the prediction accuracy, by simulating wind speed time series with different accuracy for the same period.

AKNOWLEDGEMENTS

This research was made possible with support through the Industrial Doctorate Centre for Offshore Renewable Energy (IDCORE) funded by the Energy Technology Institute and the RCUK Energy Programme (grant number EP/J500847/1) and EDF Energy. The authors would like to thank EDF Renewables UK for providing access to the data.

ORCID

Alexios Koltsidopoulos Papatzimos https://orcid.org/0000-0003-1129-4801

REFERENCES

1. Ioannou A, Angus A, Brennan F. A lifecycle techno-economic model of offshore wind energy for different entry and exit instances. *Appl Energy*. 2018;221:406-424.
2. Reder M, Yurusen NY, Melero JJ. Data-driven learning framework for associating weather conditions and wind turbine failures. *Reliab Eng Syst Saf*. 2018;169:554-569.
3. Tavner P, Edwards C, Brinkman A, Spinato F. Influence of wind speed on wind turbine reliability. *Wind Eng*. 2006;30(1):55-72.
4. Carroll J, McDonald A, McMillan D. Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines. *Wind Energy*. 2016;19(6):1107-1119.
5. Kourkoulis V, Carroll J, Leithead W. Effect of site conditions on offshore wind turbine failures. In: International Conference on Renewable Power Generation (RPG 2015); 2015; Beijing, China:17-18. IET Conference Publications 679, pp. 837–842.
6. Qiu Y, Feng Y, Tavner P, Richardson P, Erdos G, Chen B. Wind turbine SCADA alarm analysis for improving reliability. *Wind Energy*. 2012;15(8):951-966.
7. Gonzalez E, Reder M, Melero JJ. SCADA alarms processing for wind turbine component failure detection. *J Phys Conf Ser*. 2016;753(7):072019.
8. Leahy K, Gallagher CV, Bruton K, O'Donovan P, O'Sullivan DT. Automatically identifying and predicting unplanned wind turbine stoppages using SCADA and alarms system data: case study and results. *Journal of Physics: Conference Series*. 2017;926(1):012011.
9. Nguyen TH, Prinz A, Friisø T, Nossum R, Tyapin I. A framework for data integration of offshore wind farms. *Renew Energy*. 2013;60:150-161.
10. Koltsidopoulos Papatzimos A, Dawood T, Thies PR. An integrated data management approach for offshore wind turbine failure root cause analysis. In: ASME International Conference on Offshore Mechanics and Arctic Engineering (OMAE); 2017; Trondheim, Norway:25-30. V03BT02A012.
11. Koltsidopoulos Papatzimos A, Thies PR, Dawood T. Data-driven operational reliability assessment for offshore wind farms. Reliability Engineering & System Safety (Under Review).
12. Koltsidopoulos Papatzimos A, Dawood T, Thies PR. Data insights from an offshore wind turbine gearbox replacement. *J Phys Conf Ser*. 2003;1104(01):2018.
13. Douard F, Domecq C, Lair W. A probabilistic approach to introduce risk measurement indicators to an offshore wind project evaluation improvement to an existing tool ECUME. *Energy Procedia*. 2012;24:255-262.

How to cite this article: Koltsidopoulos Papatzimos A, Thies PR, Dawood T. Offshore wind turbine fault alarm prediction. *Wind Energy*. 2019;1–10. https://doi.org/10.1002/we.2402