

PAPER • OPEN ACCESS

Artificial intelligence-based condition monitoring and predictive maintenance framework for wind turbines

To cite this article: Janine Maron *et al* 2022 *J. Phys.: Conf. Ser.* **2151** 012007

View the [article online](#) for updates and enhancements.

You may also like

- [Overview and Design of self-acting pitch control mechanism for vertical axis wind turbine using multi body simulation approach](#)
Prasad Chougule and Søren Nielsen
- [Effect of Inlet Section of Circular Section Spiral Case on Performance of Ultra-Low Specific Speed Diagonal Flow Turbine](#)
Yanpin Li and Xiaoyu Chen
- [Numerical analysis of flow interaction of turbine system in two-stage turbocharger of internal combustion engine](#)
Y B Liu, W L Zhuge, Y J Zhang et al.



The Electrochemical Society
Advancing solid state & electrochemical science & technology

242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

Abstract submission deadline: **April 8, 2022**

Connect. Engage. Champion. Empower. Accelerate.

MOVE SCIENCE FORWARD



Submit your abstract



Artificial intelligence-based condition monitoring and predictive maintenance framework for wind turbines

Janine Maron^{1*}, Dimitrios Anagnostos¹, Bernhard Brodbeck¹, Angela Meyer²

¹ WinJi AG, Badenerstrasse 808, 8048 Zurich, Switzerland ^{1*}

² Bern University of Applied Sciences, Quellgasse 21, 2501 Biel, Switzerland

*Corresponding author: janine.maron@win-ji.com

Abstract. The global wind power capacity continues to grow at a fast pace. However, the profit margins from wind power are being compressed in many countries. Thus, many wind farm owners seek to reduce their operational expenses, including those for maintenance work. In this study, an artificial intelligence-based condition monitoring and predictive maintenance framework for wind turbines is presented. The purpose of this framework is the automated early detection of operational faults in wind turbine systems and subsystems. The early detection of anomalies enables further diagnosis, condition-based maintenance and better planning of repairs. It can prevent consequential damage, lead to fewer turbine downtimes and extend the service lives of the monitored turbines. We present validation results from two onshore wind farms and demonstrate 97% accuracy for a 2-month detection horizon of developing fault events that require attention from maintenance staff.

1. Introduction

While the global wind power capacity is growing at astonishing rates, the guaranteed profit margins for owners and operators are constantly being compressed due to competitive energy markets [1]. Owners and managers of such assets are exploring additional revenue creation through the optimisation of their operation and maintenance activities. Especially for remote wind farms, such as offshore sites, predictive maintenance can greatly affect CAPEX spend, especially when performed frequently and remotely, avoiding expensive on-site visits [2-4].

The aim of this work is to propose and evaluate a framework for providing automated condition-based maintenance decision support for wind turbines and their electromechanical subsystems based on artificial intelligence algorithms. The purpose of this framework is to detect developing faults in wind turbine systems and subsystems as early and accurately as possible. The early detection of anomalies enables condition-based maintenance and better planning of repairs. This prevents consequential damage, leads to fewer turbine downtimes and extends the service life of a turbine [5-7].



This paper is organized as follows. Section 2 introduces the datasets used in this study and section 3 presents the methods developed. Our results are discussed in section 4. Finally, section 5 outlines our conclusions.

2. Datasets

In this work the condition monitoring data from the SCADA system of two commercial European onshore wind farms have been analyzed. These are variable-speed three-bladed horizontal axis turbines with pitch regulation. Table 1 provides an overview of the turbines' technical specifications. The data have been anonymized for this study. Ten-minute average values of condition variables from the supervisory control and data acquisition (SCADA) system served to train and test the model as specified below. The data represent approximately 30 months of SCADA data from a wind farm with four turbines (wind farm 1) and 4.5 years of SCADA data from a second wind farm with eight turbines (wind farm 2). The data comprise the rotor and generator speeds, active power production, currents and the temperatures of multiple drivetrain components, including gear bearings, generator windings and hydraulic oil temperatures, amongst other condition variables.

Additionally, independent information on repairs and technical problems of individual turbines in wind farms 1 and 2 has been evaluated to compare with the results of the proposed method.

In this work the results for one wind turbine of each farm are being presented. In wind farm 1 all turbines show similar results as the one presented in this work. In wind farm 2, serious anomalies were only recognized for the turbine presented here, the analyzed technical components of all other turbines appear to be healthy.

Table 1. Turbine specifications

	Wind farm 1	Wind farm 2
OEM	Vestas	Nordex
Installed Power	2000 kW	2300 kW
Cut-in wind speed	4 m/s	4 m/s
rated wind speed	12 m/s	13 m/s
Cut-out wind speed	25 m/s	25 m/s
Hub height	80 m	100 m
Rotor diameter	90 m	90 m

3. Methods

The core idea of the proposed method is to create a digital model that predicts the normal behaviour of the system of interest based on its operating conditions and that detects significant deviations of predicted and actual behaviour. A deviation from this model can potentially indicate anomalies in the system [3,6,7]. The method can be separated into four steps: preprocessing, modeling, anomaly detection and postprocessing.

The first part of the algorithm involves preprocessing of the above datasets. Data preprocessing deals with rigid data cleaning and transformation of real measured data. This necessitates data cleaning and transformations. Filtering is applied to arrive at clean and meaningful datasets. Conditions of filtering relate to:

- data containing blank values for one of the input or the output variables of the model
- times when a turbine is not producing, comprising the following cases
 - the turbine is not producing due to low wind
 - the turbine is not producing despite sufficient wind. In this case the turbine is in a downtime, which can be caused by e.g. maintenance, grid or environmental curtailments.
- times when the power output of the turbine is reduced from the outside (e.g. grid curtailment, noise mode)
- times within two-hour phase after a shutdown, because technical parameters are not in their normal range shortly after a shutdown
- outliers and corrupted data.

After cleaning, the data is transformed to be comparable on the same scale. All technical parameters as well as power, wind speed and rotor speed have been log-transformed to this end.

In order to train the models, a training data set was defined based on a multi-months time period in which the turbines demonstrated no unusual operating behaviour and in which no maintenance needs occurred to the best of our knowledge. When there is a software update or another drastic change in the behaviour of the turbine, the models need to be retrained on the new setup. Furthermore, all important characteristics, for example seasonality, must be covered in the data and the dataset needs to contain enough information in order to prevent overfitting. This was ensured by the choice of the training dataset. The first year of the dataset was selected as training data. In case of a continuous anomaly detection process, a periodic retraining of the model is initiated.

Multiple machine learning algorithms were compared for training models of the normal operating behaviour of the turbines, including a random forest algorithm [8], support vector machines [9] and a gradient boosting algorithm [10]. The model that provided the best performance in terms of model training time and robustness was trained with a gradient boosting algorithm. Compared to the other training methods like the random forest algorithm and support vector machines, the model trained with the boosting algorithm achieved similar accuracies but was ten times faster in computation.

A separate model was trained for each turbine and subsystem sensor. We refer to a discussion of multi-target normal behaviour models in [11]. In the present study, the measures of wind speed and environmental temperature were taken as input variables to train the models of the considered turbine subsystems. With these two input variables, a model with sufficiently accurate fault detection accuracy could be created for the majority of the available sensors in the turbine subsystems. The input variables were limited to these two predictors due to their explanatory power and their high availability in most of the wind turbine data sets.

Finally, in order to declare data points as anomalies, two threshold conditions were introduced with regard to the residuals and the occurrence frequency of large residuals. First, the residual threshold was selected as the 99 %-quantile of the residual distribution to fit the distribution and following the assumption that the data in the training set represent the normal behaviour of the turbine. We focussed on the residuals as the difference between the prediction and the actual value of the target variable. Whenever a residual exceeded this threshold, it was flagged as a potential anomaly. A fault was flagged only if anomalies occurred frequently enough within a certain period of time. A frequency-related threshold was introduced to this end. It is used to distinguish actual long-term anomalies from single isolated deviations.

For a fault to be detected, this second threshold requires that a given minimum number of potential anomalies occur within a rolling time window of 24 hours. We found that setting this second threshold to at least 40 potential anomalies within 144 data points (corresponding to a 24-hour window) provides a satisfactory trade-off between too late detection and too many false positive alarms in practice. Based on the residual and frequency thresholds, anomalies could now be accordingly identified in the test set.

4. Results

The extensive data collected by WinJi¹ through its digital services have been exploited in this work in order to develop and test an accurate fault detection scheme and demonstrate its performance in commercially operating wind turbines. We have compared and benchmarked a number of state-of-the-art algorithms for the creation of normal-operation models of the main subsystems of monitored wind turbines and for the early detection of developing operating faults. The validation was performed on two wind farms in the asset portfolio monitored by WinJi. The framework enables the detection of maintenance needs in gearbox and generator anomalies approximately two months before maintenance is actually performed, as shown in Figure 1. We could also verify that the maintenance work was effective in that the detected temperature anomalies disappeared after both monitored incidents (Figure 1). The developed anomaly detection framework was tested on the turbines of wind farms 1 and 2 in the portfolio of WinJi. The anomaly detection process described in section 3 was carried out for each turbine and monitored turbine condition variables separately.

As outlined in section 3, the fault alarm criterion is based on two different thresholds using the residuals computed as difference of the observed sensor measures and the predictions by the above normal behaviour model. An alarm was raised when the 99 % quantile of the residuals distribution was exceeded for more than 6.7 hours in the respective past 24 hours.

Table 2 details the subsystems sensor variables that were examined and the root mean square errors (RMSE) of the trained models per farm.

Table 2. Overview of subsystems sensor variables and RMSE of trained model (log-transformed test data) listed per farm. In case there was no data available or the trained normal behaviour model was not satisfactory, there is no RMSE value listed. As part of the preprocessing the variables have been log-transformed before model training.

	Wind farm 1	Wind farm 2
Converter Temp [°C]	0.039	-
Gear Bearing Temp [°C]	0.1	0.059
Gearbox Temp [°C]	0.036	-
Gear Oil Sump Temp [°C]	0.046	-
Generator Speed [rpm]	0.023	0.025
Generator Journal Temp A [°C]	-	0.086
Generator Journal Temp B [°C]	-	0.097
Generator Temp 1 [°C]	-	0.060
Generator Temp 2 [°C]	-	0.0

¹ www.win-ji.com

Hydraulic Oil Temp [°C]	0.104	-
Hydraulic Pressure [bar]	0.004	-
Power [kW]	0.104	0.139
Rotor Speed [rpm]	0.024	0.033
Spinner Temp [°C]	0.144	-
Transformer Phase Temp 1 [°C]	0.127	-
Transformer Phase Temp 2 [°C]	0.113	-
Transformer Phase Temp 3 [°C]	0.126	-

For the first wind farm, anomalies are predicted for the gearbox. Three out of four turbines were flagged in October to November 2019 and 2020, as presented in Fig. 1. Maintenance was carried out in November of both years, after which the anomalies ceased.

For the second wind farm, one out of eight turbines was diagnosed with unusually increased generator operating temperatures, as shown in Fig. 2. In this case the owner could be notified in time so a generator failure and possibly nacelle fire could be prevented. Both cases provided a time window of more than one month for early intervention and prevention.

Comparing the applied algorithms, we found that models trained based on the gradient boosting algorithm [10] performed most robustly in automated unsupervised detection of faults based on the component's normal operating behaviour. Moreover, while they achieved similar accuracy in describing the normal operating behaviour, the models trained with the gradient boosting algorithm outperformed the other benchmarked algorithms in terms of the model training times. The latter is a relevant and beneficial algorithm property for scaling our approach to a large number of monitored components and wind farms as in the WinJi portfolio. The RMSE of the compared models (Random forest, SVM, Boosting) and the times in seconds needed for training on an Apple M1 processor are shown in Table 3.

Table 3. Overview of multiple machine learning algorithms for training models of the normal operating behaviour of the turbines, the achieved RMSE of the trained model (log-transformed test data) and the model training time using the example of variable generator temperature 1.

	RMSE on test dataset for variable generator temp 1 [°C]	Model training time [sec] for variable generator temp 1
Support vector machine	0.062	1320
Random Forest	0.064	2040
Gradient boosting	0.060	130

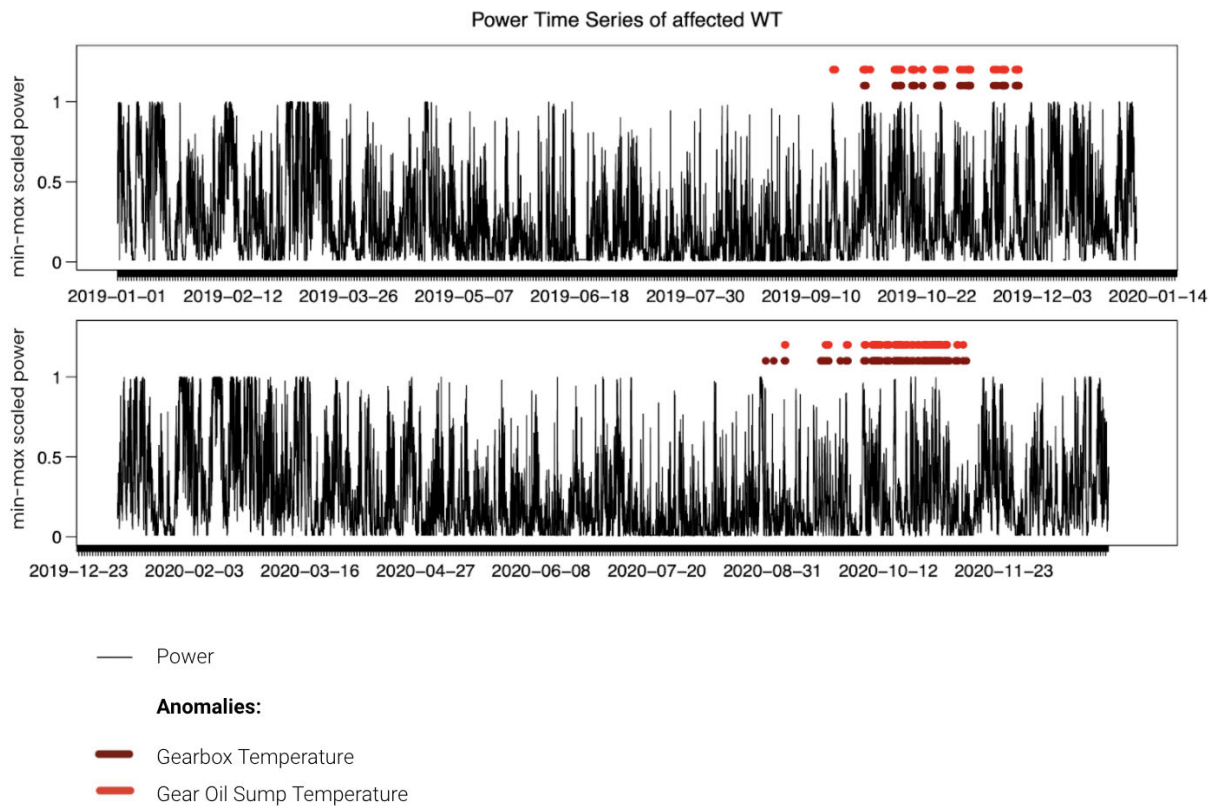


Figure 1. Active power produced by the turbine that developed a gearbox fault in wind farm 1 scaled with the min-max method to a range in $[0, 1]$. Anomalies detected with the developed models for a subset of 13 monitored WT condition variables, namely the gearbox temperature and gear oil sump temperature, based on 10-minute values from the SCADA system in wind farm 1.

As presented in Fig. 1, an alarm was triggered when the residuals indicated unusual operating behaviour of the respective WT component based on the defined alarm criterion. Excessive gearbox temperatures first start to evolve in Sept. 2019, indicating maintenance need in the gearbox. The alarm switches off immediately after the work has been performed on 17 Nov. 2019, indicating the effectiveness of the performed maintenance activities. A similar situation recurred in 2020 in the same gearbox, with gearbox temperature starting to rise unusually in Sept. 2020 and maintenance performed to successfully resolve the condition on 15 Nov. 2020.

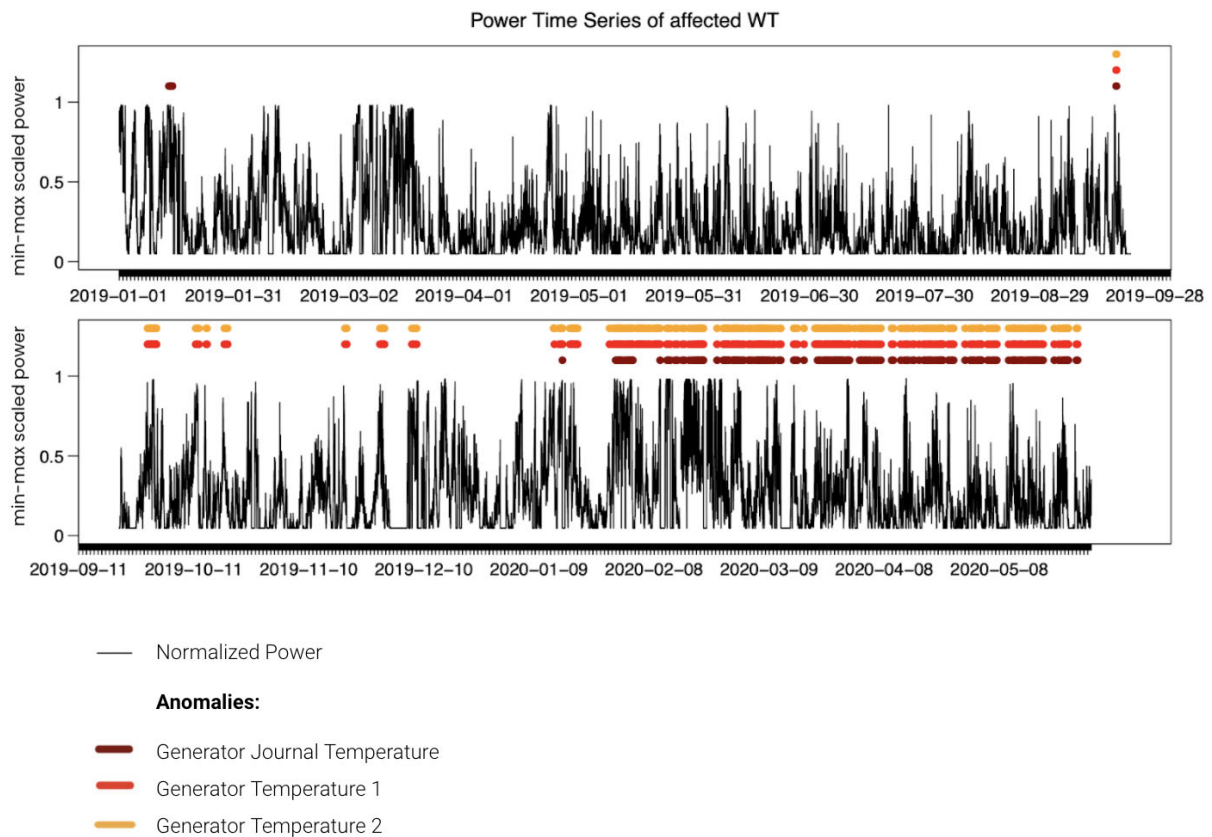


Figure 2. Demonstration of the fault detection algorithm for drivetrain subsystems of a turbine of wind farm 2.

Another result from the validation of the developed framework is the estimation of the probability that a fault causing downtime will actually appear. For this we used the following methodology:

- For each detected anomaly, we checked the error code information k days into the future for downtimes. k is the number of days following an anomaly that we examined for downtimes.
- We only considered downtimes connected to technical failure and unplanned maintenance.
- k was chosen as a rolling window with values [5, 10, 20, 30, 60]
- The Positive Predictive Value (PPV) of the framework was assessed, based on the following equation:

$$PPV = (\text{Number of True Positives}) / (\text{Number of True Positives} + \text{Number of False Positives})$$

Fig. 3 provides an overview of the PPV values for the main systems that exhibited useful correlations for this dataset. When the proposed method detects an anomaly for the gearbox or the converter, it is almost certain (>97%) that within a month ($k = 30$), the wind turbine will experience a fault and stoppage. PPV for transformers is lower but still above 75% and interestingly the gear bearing temperatures do not produce a relevant result with PPV never exceeding 20%.

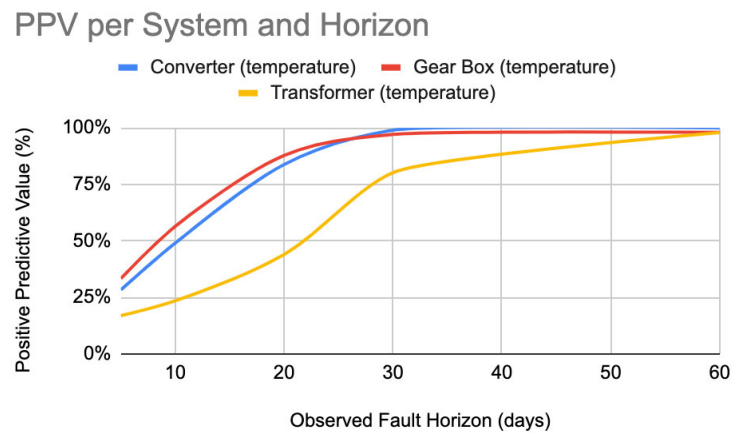


Figure 3. PPV graph vs. fault horizon. In parentheses the type of signal that was used per system.

5. Conclusions

We have presented a condition-based maintenance framework for wind turbines that can successfully detect developing faults at an early stage and assess the effectiveness of performed maintenance activities using artificial intelligence techniques. The automated early detection enables condition-based maintenance work and informed planning and scheduling of repair activities. This helps prevent consequential damage, leads to fewer turbine downtimes and can extend the service life of the monitored turbines. We have presented validation results from two commercial onshore wind farms. We also demonstrated that the framework enables verifying the effectiveness of performed maintenance work.

The presented framework is capable of manufacturer-agnostic early fault detection and allows notification of the wind farm managers and owners. This can lead to optimised operations and generation of additional value from the asset while also avoiding unplanned expenses and downtimes. Furthermore, this framework is part of a digital platform, providing transparency and insights fully remotely with minimal visits on-site. As demonstrated, such systems are highly beneficial for efficient management of large portfolios and value creation in the assets themselves.

6. Acknowledgements

This work has been partially funded by Innosuisse, project nr. 31533.1 IP-EE.

References

- [1] McInnis, D., M. Capezzali, 2020, Managing Wind Turbine Generators with a Profit Maximized Approach, Sustainability 12, doi: 10.3390/su12177139.
- [2] International Renewable Energy Agency, 2019, Renewable power generation costs in 2019, International Renewable Energy Agency.
- [3] Meyer, A., B. Brodbeck, 2020, Data-driven Performance Fault Detection in Commercial Wind Turbines, Proceedings of the 5th European Conference of the Prognostics and Health Management Society 2020, Turin, Italy, ISBN 978-1-93-626332-5.
- [4] Fischer, K., D. Coronado, 2015, Condition monitoring of wind turbines: State of the art, user experience and recommendations, VGB PowerTech 2015(07):51-56.
- [5] Hameed, Z., Y. Hong, Y. Cho, S. Ahn, C. Song, 2009, Condition monitoring and fault detection of wind turbines and related algorithms: A review, Renewable and Sustainable Energy Reviews 13, doi: 10.1016/j.rser.2007.05.008.
- [6] Zaher, A., S. McArthur, D. Infield, Y. Patel, 2009, Online wind turbine fault detection through automated SCADA data analysis, Wind Energy, 2009, doi: 10.1002/we.319.
- [7] Tautz-Weinert, J., S. Watson, 2017, Using SCADA data for wind turbine condition monitoring – a review, IET Renew. Power Gener. 11, doi: 10.1049/iet-rpg.2016.0248.
- [8] Breiman, L., 2001, Random Forests, Machine Learning 45: 5–32, doi: 10.1023/A:1010933404324.
- [9] Drucker, H., C. Burges, L. Kaufman, A. Smola, V. Vapnik, 1996, Support Vector Regression Machines, Proceedings of the 9th International Conference on Neural Information Processing Systems, Denver, CO, USA, December 2-5, 1996.
- [10] Chen, T., C. Guestrin, 2016, XGBoost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), Association for Computing Machinery, New York, NY, USA, 785–794, doi: 10.1145/2939672.2939785.
- [11] Meyer, A., 2021, Multi-target normal behaviour models for wind farm condition monitoring, Applied Energy, Elsevier, doi: 10.1016/j.apenergy.2021.117342.