

Wind turbine gearbox multi-scale condition monitoring through operational data

Francesco Castellani^{1,*} , Matteo Vedovelli¹, Alessandro Canali¹, and Francesca Belcastro²

¹ University of Perugia, Department of Engineering, Via G. Duranti 93 Perugia, Italy

² FERA srl, Piazza Cavour 7 –20121 Milano, Italy

Received: 9 February 2024 / Accepted: 20 September 2024

Abstract. Since wind is expected to play a crucial role on the worldwide electricity production scenario, the reliability of the turbines is attracting attention of both industry and academia. New techniques for efficient condition monitoring of key components can be fundamental to optimising the performance and maintenance of a large fleet of turbines. The gearbox and bearings are the most critical mechanical components as they are responsible for a large proportion of the downtime of a wind turbine over its lifetime. However, the monitoring of wind turbine gearboxes is challenging due to the non-stationary nature of the operation and the lack of noise-free vibration measurements. In the present work, a new approach for efficient long to short term monitoring of wind turbine gearboxes has been developed based on real data. An turbine drivetrain failure was used as a test case to develop a new approach based on the use of multi-scale data sources. On the one hand, SCADA (Supervisory Control And Data Acquisition) data were used for general monitoring of the condition of the machine component on long to medium term time scales, while on the other hand, high resolution, triggered event data collected by a CMS (Condition Monitoring System) were used to refine the diagnosis and prognosis of the fault on a shorter time scale. Even though triggered spot events are very difficult to manage, the results show that the use of multi-scale high resolution CMS data can be fast and useful in fault diagnosis to classify a target machine with a healthy reference one. In the present work, the one-class SVM (Support Vector Method) was used for novelty detection. The approach, when applied to all available time scales, can be very precise in detecting the faulty machine and can therefore be proposed as a fast detection approach requiring less data compared to the classical data-driven regression normal behaviour model developed with continuously available SCADA data.

Keywords: Condition monitoring / wind turbine / gearbox bearings / fault detection

1 Introduction

The electricity generation from renewable sources has experienced a rapid acceleration in recent years. Projections suggest that the share of renewable electricity production in OECD countries will raise from 10.8% in 2019 to one-third by 2035 [1]. For example wind power emerged as the predominant source of electricity in the U.K in the first quarter of 2023. Meanwhile, during the same period, 42% of the electricity was derived from renewable sources, surpassing the 33% generated from fossil fuels. In this scenario, the profitability of renewable energy becomes crucial: in general, and in the particular case of wind energy, costs are mainly concentrated in expenses related to installation investment and operation and maintenance (O&M). Maintenance procedures often involve significant

costs and time, involving the transfer of labour, equipment, including the replacement of components [2]. This reality underlines the considerable attention paid in the literature to fault diagnosis in wind turbine gearboxes and bearings [3,4]. In particular, Figure 1 [5] illustrates that mechanical components such as gears and bearings have high failure rates and related downtime.

Modern industrial wind turbines are equipped with SCADA (Supervisory Control and Data Acquisition) systems that continuously collect a range of environmental, operational, mechanical, electrical and thermal measurements with a typical averaging time of ten minutes. This type of system was originally developed to allow remote control of wind turbines, but the ever-increasing number of monitored parameters has stimulated its use for condition/performance monitoring [6–10]. The high level of reliability and competitiveness required for modern wind turbines is driving the industry towards the application of Condition Based Maintenance [11] for a large number of key

* e-mail: francesco.castellani@unipg.it

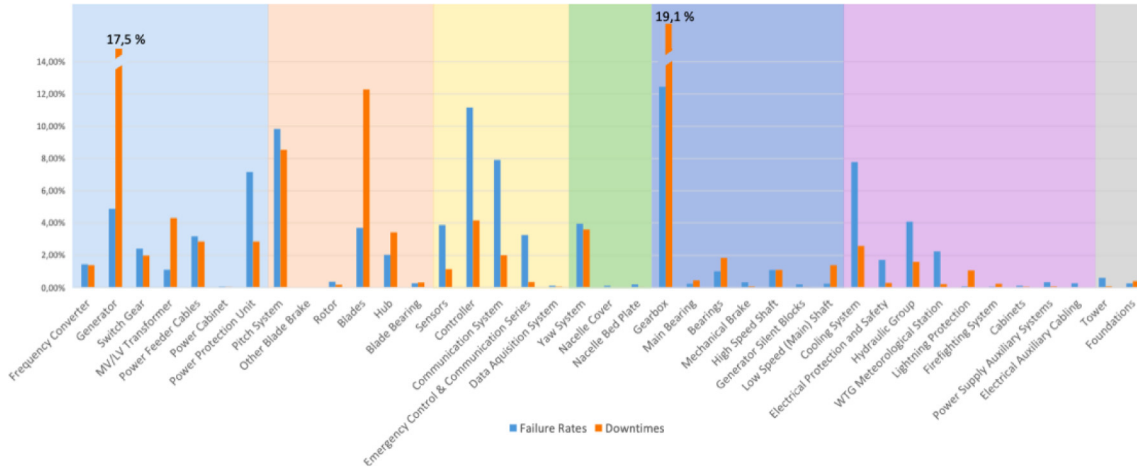


Fig. 1. Estimates of failure rates and associated downtime for the most important wind turbine components (from [5]).

components. This is fundamentally one of the most important ways to improve the performance of wind farms [12,13] and to optimise the life cycle of the machines with an optimal control of the ageing of components [14,15].

SCADA has given wind turbines a very practical and easy way of storing large amounts of data, but the main drawback is that the ten minute averaging time is too long to capture the overall dynamics of the system. Consequently, research on wind turbine fault detection based on SCADA data analysis often involves the use of medium-term [16] to long-term [17] trends derived from meaningful measurements, typically temperatures collected at rotating sub-components. The underlying idea is that the initiation of a fault should be accompanied by an abnormal release of heat, observable in the form of trends, spikes or increased variability. In practice, a model representing normal behaviour is constructed using data assumed to describe healthy operation, for example a regression [18] or a classification [19]. The statistical novelty of the measurements during the target period is then analysed. For the reasons outlined above, there is a growing trend towards the use of high frequency data for fault diagnosis in wind turbines, as can be seen in [20,21]. However, a key challenge associated with such data is that industrial systems typically only store it when specific triggered events occur, making continuous monitoring impractical. Consequently, a recent trend in the literature is to adopt an approach that combines detailed high-frequency analysis close to the fault with continuous monitoring based on SCADA, as done in references [22,23]. As such, multi-scale data analysis for wind turbine fault diagnosis represents an innovative research direction, and this study aims to contribute to the field through a discussion of a real-world test case. This study examines a case of gear bearing failure in an industrial wind farm employing a multi-scale approach, as outlined below:

- Long-term, through the analysis of hourly-averaged SCADA data for 6 years before the fault.
- Medium term, through the analysis of ten minutes averaged SCADA data for order of one year before the fault.
- Short term, through the analysis of spot events sampled with a frequency ranging from order of 1 Hz to thousands of Hz.

Condition monitoring of wind turbine gearboxes can be challenging due to the large number of components and the wide range of speeds involved. In this context, the ability to use a large amount of data covering key parameters over a wide range of time scales can be fundamental.

The time horizon for the development of a mechanical failure in a wind turbine gearbox can vary from a few hours to many months depending on:

- The speed of rotation of the components and their design characteristics;
- The operating conditions in terms of variations in speed and loads;
- The nature and cause of faults.

In this context, all available data should be used for optimal monitoring of the components in order to improve the reliability and performance of the wind turbines; obviously, the monitoring analysis should cover all possible time scales, requiring an integrated multi-time-scale approach. The merit of this approach lies in its ability to provide a critical discussion on the identification of the onset of failure. In real test cases, the classification of data is complicated by the lack of a clear demarcation line between healthy and faulty measurements: the use of multiple time scales proves valuable in clarifying patterns. The structure of manuscript is as follows: [Section 2](#) describes the test case and the dataset; [Section 3](#) provides a brief overview of the methods used; [Section 4](#) presents and discusses the results; and [Section 5](#) draws conclusions.

2 Testcase and dataset

The case analysed is a multi-MW horizontal axis wind turbine affected by a recently discovered fault in a drive train bearing.

Although standard SCADA (Supervisory Control And Data Acquisition) data analysis gave no clear indication of the fault, the final diagnosis was made by the discovery of metal debris in the gearbox oil.

In general, bearing failures can have very different prognoses and failure modes [24]. They can lead quickly

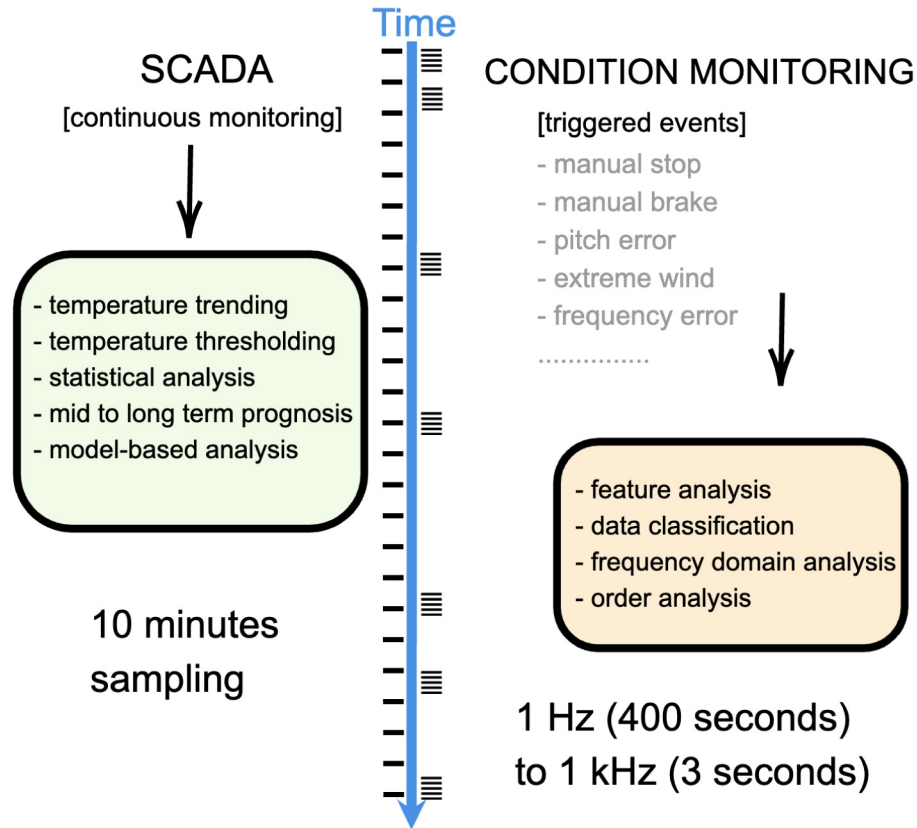


Fig. 2. Schematic of available datasets and how they can be used.

and dramatically to severe failures (as in the case of fatigue failures), or they can affect machine operation gradually, with a slow progression over time, as in the case of wear or overload failures. The latter case, analysed in the present work, is generally more difficult to diagnose accurately. In this case, the machine may continue to operate for a long time in faulty conditions and the evolution of the damage is not clearly visible due to the weak symptoms. Data from industrial SCADA and condition monitoring systems were used simultaneously to provide a large amount of information on different time scales. The SCADA system collects statistics on a 10 minute time scale (min, max, avg...); the time scale for acquisition is poor, but it guarantees continuous monitoring of a large number of machine parameters, ideal for long term data sets. The condition monitoring system is capable of collecting high resolution data with a frequency of 1 Hz to 1 kHz. Due to the large amount of data, monitoring in this case is not continuous and only selected events (not evenly spaced in time) with a limited number of parameters are available. [Figure 2](#) summarises the main sources of data that have been used in this work and some of the possible approaches for each of the time scales.

The events collected by the condition monitoring system provide a high-resolution picture of an abnormal event (i.e. an error or extreme condition) or an external forced intervention, such as a manual stop or remote external command, that causes a rapid dynamic change in the machine's state.

Table 1. Characteristics of the data from the CMS.

Frequency of acquisition	Number of points	Acquisition time (s)
1 kHz	3000	3
100 Hz	2000	20
20 Hz	1000	50
10 Hz	1000	100
2 Hz	400	200
1 Hz	400	400

In this case, the data are stored on different timescales to allow efficient monitoring of both the fast and slow components; in any case, the shorter the timescale (higher acquisition frequency), the shorter the total observation time as detailed in [Table 1](#).

The reference time for acquisition is the time of the event ($t=0$); for each time scale, the system collects 70% of the data before and 30% of the data after the reference time. In this way only the acquisitions from 1 up to 20 Hz can give a good representation of the machine's transitory (see [Fig. 3](#)). It should be also noted that for very high resolution time scales (1 kHz and 100 Hz) the system collect data only for few parameters (mainly the gearbox speed). As a consequence in present work it was not possible to work directly in the frequency domain for analysing the fault.

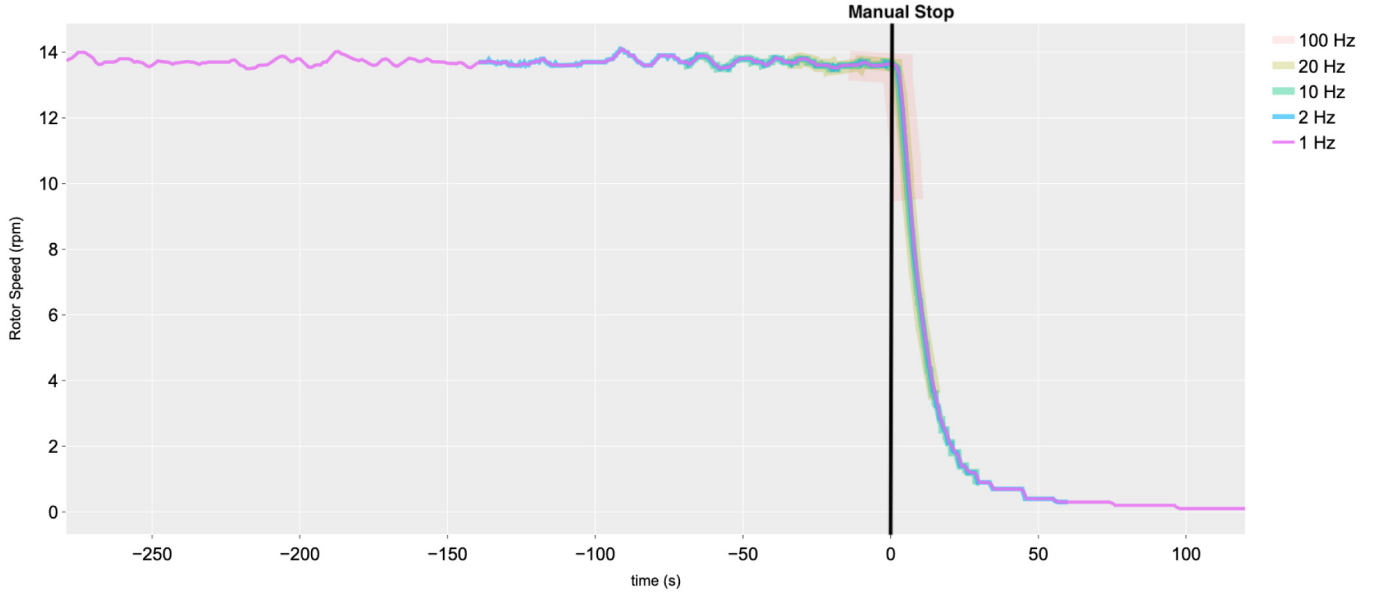


Fig. 3. Time histories of the rotor speed in the time scales from 1 to 100 Hz for a “Manual Stop” event.

Among the different triggered events in the present work, only the “manual stop” events have been considered, as they can represent similar transient time histories, making it easier to classify the machine health states.

The long-term trend for the temperatures of the gearbox components are summarised in Figure 4. The seasonal trend is clearly visible from the plots but any abnormal behaviour can be inferred for the damaged turbine T_{tar} compared to the reference healthy one T_{ref} .

The picture presented by the roughly averaged SCADA temperature time histories of Figure 4 is indeed very noisy due to the effects of highly variable environmental and operating conditions. The alternating oscillations between T_{tar} and T_{ref} may also be caused by the different maintenance level of the cooling fans of each unit. In any case, there is no clear and stable difference between the two machines.

Figure 5 shows the trend on a shorter time scale (the raw 10 minute averages from the SCADA) for the temperature of the gearbox bearing 1 (the damaged component) over the last year of operation. The scatter plot is opaque to show the wide scatter of the raw SCADA data, which varies over a wide range depending on operating and environmental conditions. Standard SCADA analysis alone does not provide a clear view of the fault, so new approaches were used to reveal the progression of the damage.

3 Methods

In this work the increase level of vibration and the temperature’s rise are used on different time scale to detect the mechanical fault.

Specifically, three different approaches have been developed to investigate the damage’s evolution; the first two approaches were based on the use of SCADA data while the last one was based on high-frequency triggered events.

3.1 Differential SCADA temperatures analysis

A differential analysis of the SCADA data for the temperature of the bearings was used to reveal a possible heating caused by the evolution of the fault. In terms of SCADA analysis, it is important to highlight these main concepts:

- Data must be pre-filtered to obtain a long-term database with all turbines operating simultaneously.
- A large and long data set is required.
- The analysis cannot be applied to a single turbine and a healthy reference is required.

The aim of the differential analysis was to investigate whether the turbine under consideration showed an increase in bearing temperature. The parameter studied over different time horizons is derived for each time step using the equation (1), where T_{tar_k} is the bearing temperature of the turbine under study at each k^{th} time step and T_{ref_k} is the temperature of one or more reference units operating in a healthy state at the same time step.

$$\Delta T_k = T_{tar_k} - \frac{1}{N} \sum_{i=1}^N T_{ref_k,i}, \quad (1)$$

In equation (1) N is the total number of turbines considered for healthy behaviour.

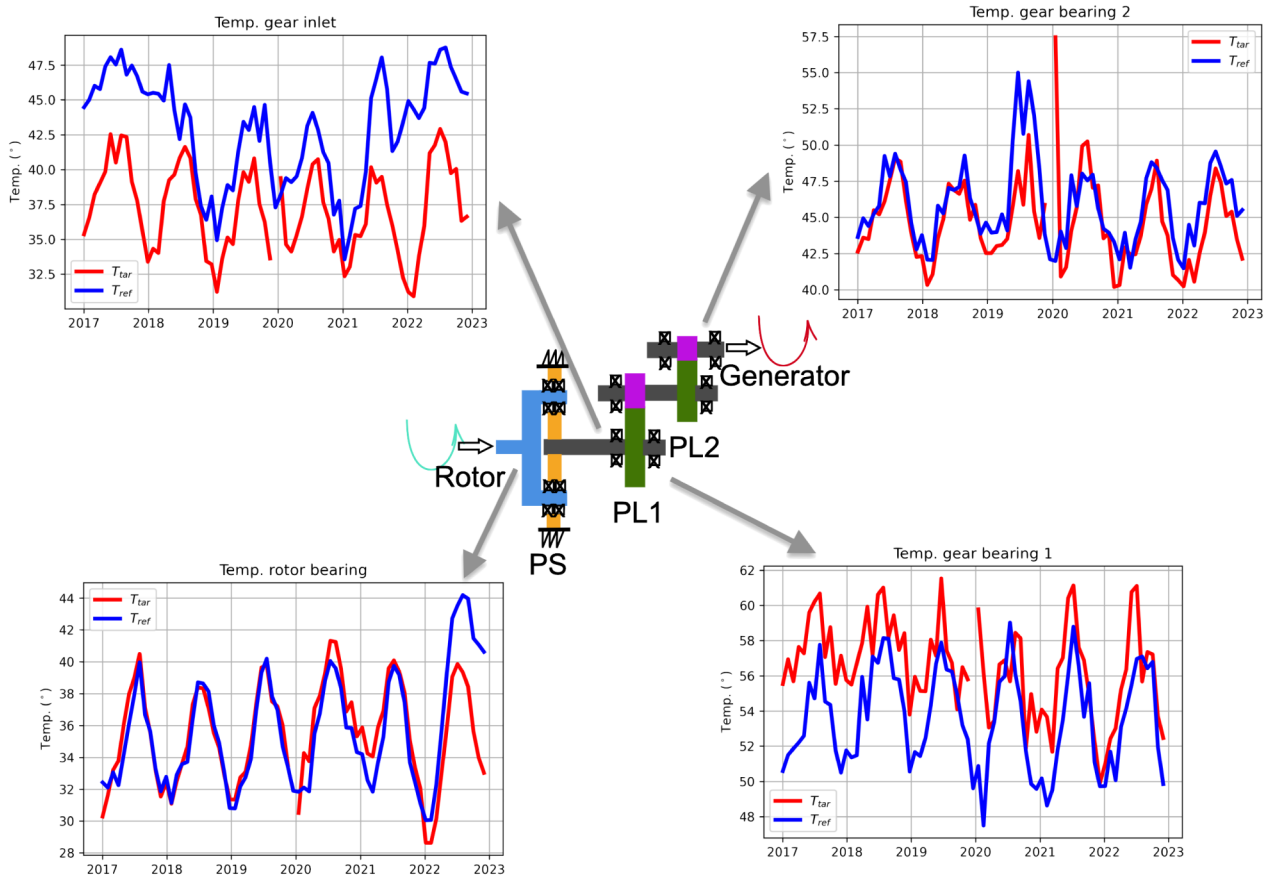


Fig. 4. Summary of the long-term trend for the gearbox temperature measurements within the SCADA database.

3.2 Regressive temperatures analysis

The use of data-driven models to detect temperature anomalies in wind turbine drive-train components can give useful results, especially when applied to SCADA data [25,26].

In present work a Random Forest Regression [27] data-driven model for all component temperatures was used to train a “Normal Behaviour Model” (NBM) to be used as a reference for each turbine. In this scenario, the reference is established by the results generated by a normal behaviour model trained on a dataset representing the healthy state of the same turbine (for all the turbines considered, the first year of operation was considered). For this type of analysis, the input parameters listed in Table 2 were used. The different bearing temperatures were considered as outputs to calculate model residuals for each component (the difference between the actual temperature and the predicted temperature).

3.3 Spot events classification

A feature classification approach using the One-Class Support Vector Machine [28,29] for the high resolution triggered events based on the vibration of the tower and the drive train has been developed to detect the faulty events. In this case, the selected feature classification parameters are:

Table 2. Input parameters for the random forest regression model.

Symbol	Parameter	Units
V	Wind Speed	m/s
P	Active Power	kW
β	Blade Angle	°
P_{oil}	Gear oil Pressure	bar
ω	Rotor Speed	rpm
T_{amb}	Ambient Temperature	°C

- Drive train vibration.
- Rotor speed.
- Tower vibration components.

Torque is not used because it is derived from electrical parameters and is not a “pure” mechanical signal.

The SVM-based one-class classification method relies on identifying the smallest hyper-sphere (with radius r and centre c) that incorporates all the data points ($x_i \forall i = 1, 2, \dots, n$). This approach is referred to as Support Vector Data Description and can be expressed by equation (2) in the following constrained optimisation form:

$$\min_{r,c} \left\{ r^2 \mid \|\Phi(x_i) - c\|^2 \leq r^2 \right\} \forall i = 1, 2, \dots, n \quad (2)$$

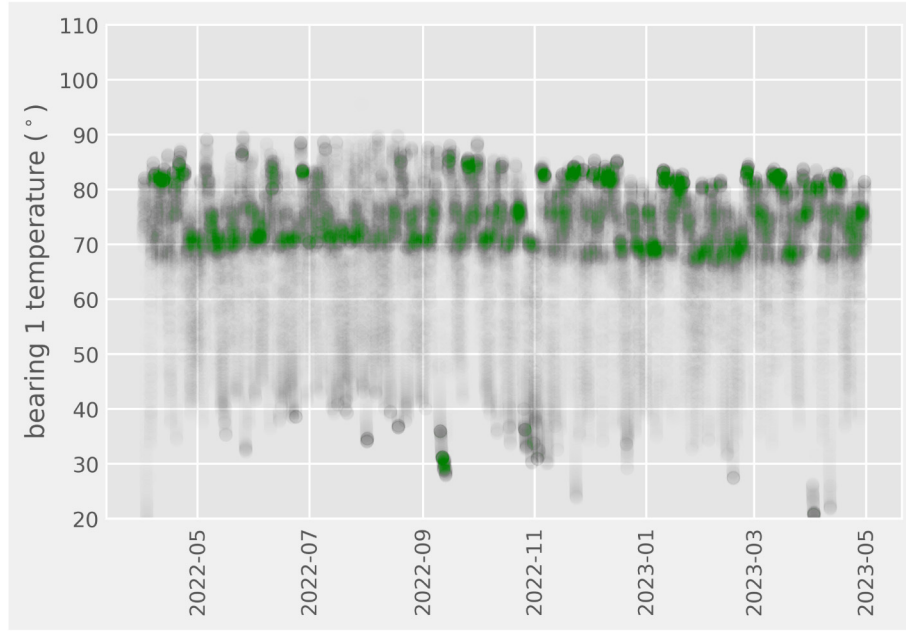


Fig. 5. Trend of gear bearing 1 temperature over the last year.

where $\Phi(x_i)$ is a function to map x_i into a vector of higher dimensionality.

However, the above formulation is highly restrictive, and can be strongly affected by outliers. Therefore also flexible formulations are proposed for practical applications; further details on the basis of the approach can be found in [30].

The reference classification model was trained on healthy events and features were computed by scaling the power spectral density of the parameters over different time scales from 1 to 20 Hz. The final monitored parameter was the percentage of outliers detected by the algorithm: events were classified as faulty if a large number of time steps within each spot event were detected as outliers.

4 Results

In this section, the results of the considered test case are summarised and the main advantages and disadvantages of the different approaches are discussed. Finally, an analysis of the evolution of the failure in the long term horizon is discussed.

4.1 Multi-scale approach

The analysis of SCADA data differentials was performed using different time resolutions and horizons:

- A short-term analysis using 10-minute averages of SCADA parameters over a period of 18 months.
- A long-term analysis using hourly averages over a period of more than 6 years.

The manifestation of the fault only became clear during the long-term analysis. As shown in Figure 6, a trend in the differential temperature of gear bearing 1 was detected for

the analysed turbine. The most pronounced evidence of damage was revealed when examining the long-term trend of the monthly averages of ΔT for bearing 1. The alarm level was defined as follows:

$$Thr = \mu \pm \chi\sigma \quad (3)$$

where μ is the mean of the reference parameter, χ is a constant adapted to the time resolution and σ is the standard deviation. For the application of equation (3) in the long-term analysis and the monthly averages, a value of $\chi = 4$ has been chosen, leading to the alarm at the beginning of May 2021, as shown in Figure 6. The hourly data are much noisier, and the same threshold leads to possible early false alarms.

Applying the same processing to the residuals for the normal behaviour model using Random Forest regression produced similar results, triggering an alert in June 2021 (see Fig. 7). However, the results of the same analysis for the short term failed to detect the error. This is due to the nature of the weak, slow-evolving damage on a high-speed shaft bearing.

The results of the one-class SVM classification showed that with this method it was only possible to raise a general alarm, without being able to formulate a specific prognosis and define the location of the damage. Optimal results were obtained by using 20 Hz data while excluding the rotor speed from the analysis (as in Fig. 8). This is because when rotor speed is also introduced, the differences between the faulty target and the healthy reference are increased at 20 Hz, but with the other time scales the fault was not clearly discernible (Fig. 9).

Getting an idea of the location of the fault is indeed much more difficult using this approach.

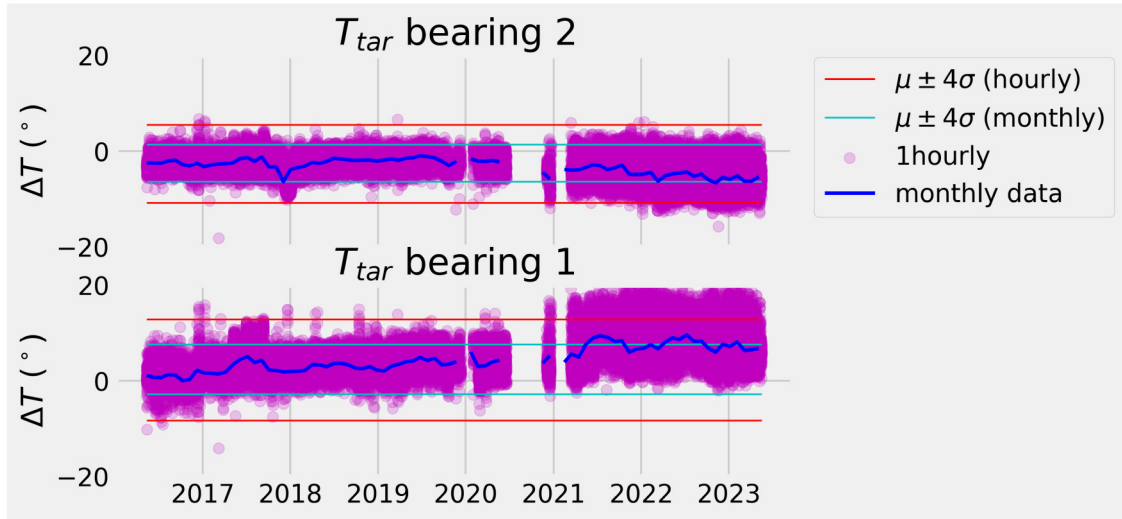


Fig. 6. Results from the SCADA analysis.

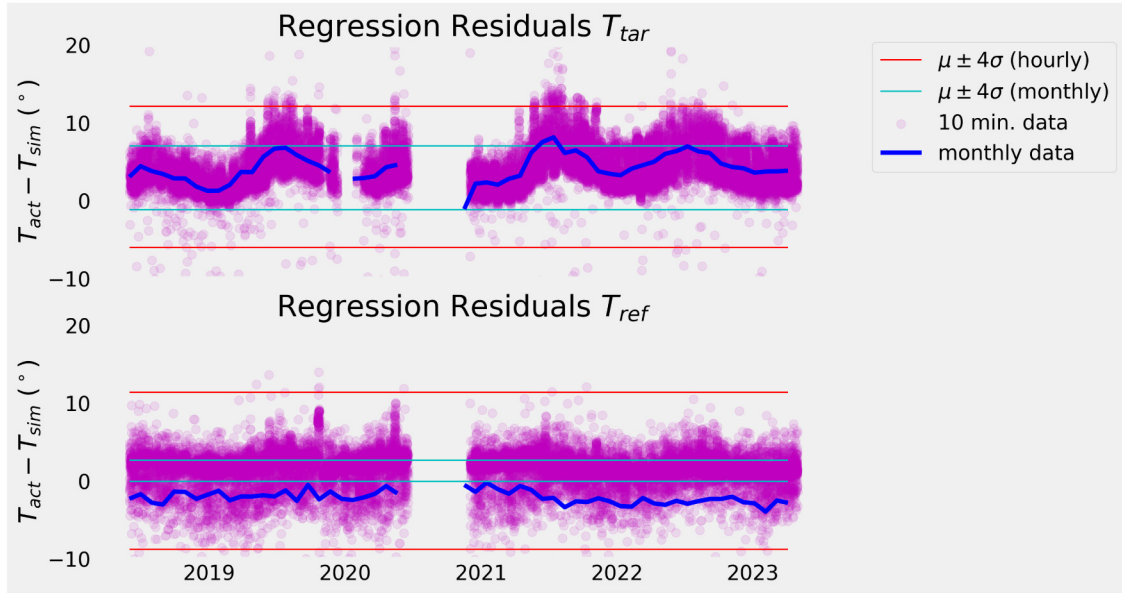


Fig. 7. Random Forest Regression residuals analysis for gear bearing 1.

4.2 Damage progression analysis

A final analysis of the damage progression was developed using the long term time histories of the residuals of the Random Forest Regression NBM and the particle count signal, which detects metal debris in the gearbox oil. The particle counting system is very useful to detect the damage in its final stage and to follow the final progression; this measurement provided the “ground truth” reference for the test case.

The trend of the residuals as well as the trend of the particle counts were organised on a daily basis as cumulative sums over a time horizon of 33 months before

the final component failure (TTF=Time To Failure). The progression was described by the normalised cumulative integral according to Eq. (4):

$$P(t) = \frac{\int_0^t R(\tau) d\tau}{\int_0^{TTF} R(\tau) d\tau} \quad (4)$$

where

- P(t) is the normalised progression at time t.
- TTF is the Time To Failure.
- R is the instant progression represented by the temperature residual

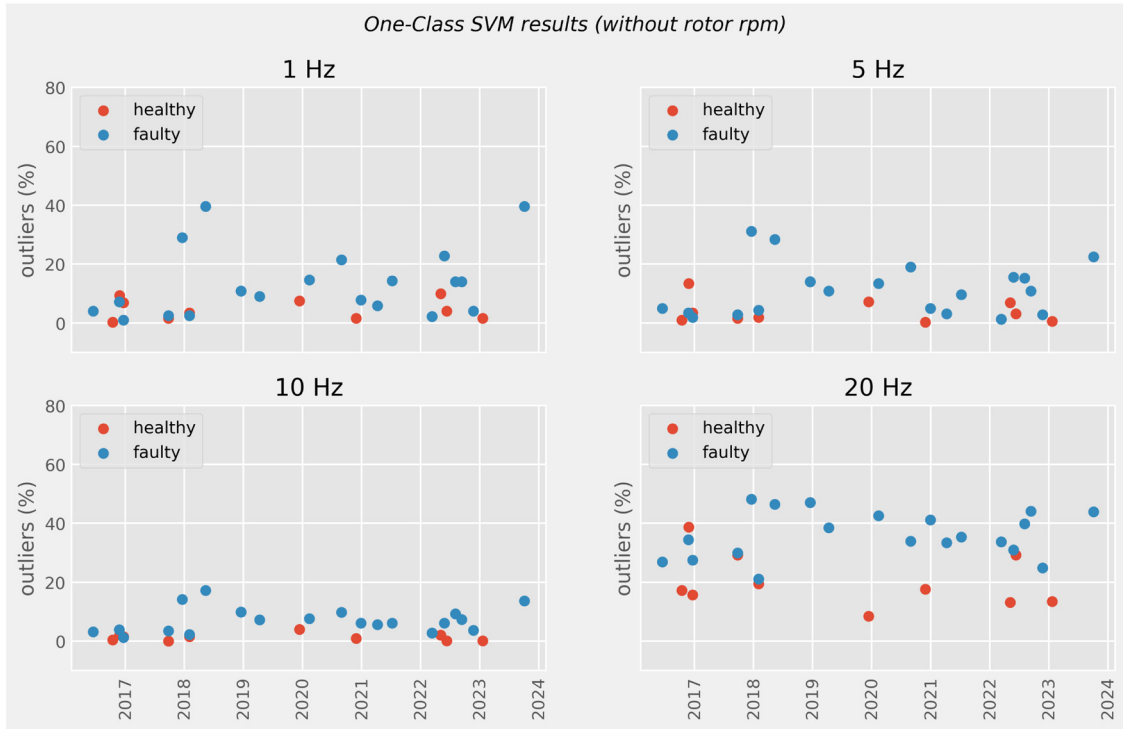


Fig. 8. One-Class SVM results without rotor rpm.

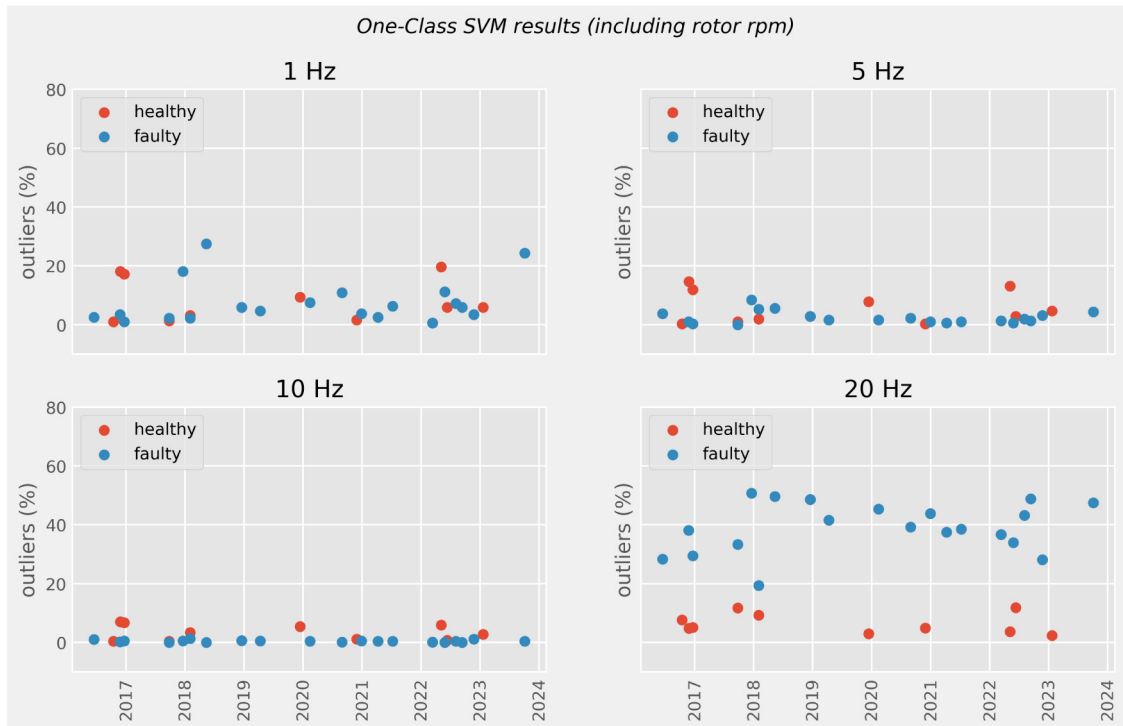


Fig. 9. One-Class SVM results with rotor rpm.

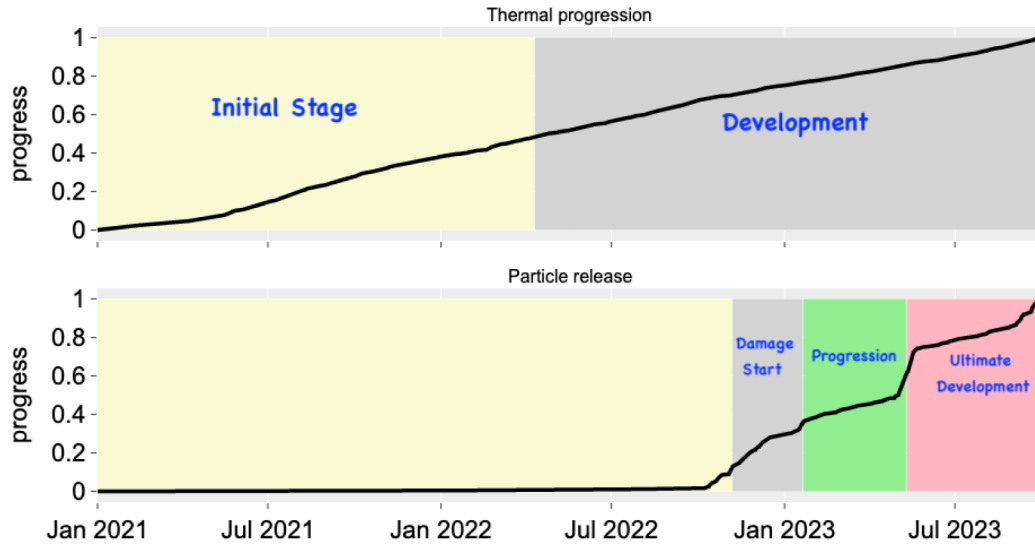


Fig. 10. Fault progress analysis using the residual of the temperature estimated using the NBM (Thermal progression) and the oil particle counting system (particle release).

$R(\tau) = T_{act} - T_{NBM}$ (difference between the actual temperature and the one estimate with the NBM) or the number of particle released (when analysing the particle counting).

The obtained time histories were finally classified using the Pruned Exact Linear Time (PELT) approach [31] for change points detection. The main objective of this approach is to describe the damage progression by optimal data partitioning in order to find the fundamental change points that describe the evolution of the machine health over time. The approach involves a pruning step within a search method that aims to minimize the expression in Eq. (5):

$$\sum_{i=1}^{m+1} [C(y_{\tau_{i-1}+1}, \dots, y_{\tau_i}) + \beta] \quad (5)$$

where C is a cost function for the i^{th} segment, m is the number of change points and β is a penalty to guard against over fitting.

The PELT method combines optimal partitioning and pruning to achieve the optimal segmentation; more details on the principles of this approach can be found in [32]. Using this approach it was possible to detect changes in the evolution of the defect over time. As can be seen in Figure 10, the cumulative sum of the residual temperature (thermal release) is able to follow the slow progression of the defect and can be used for early detection, but it is not able to characterise the final stages of the defect. Particle counting, on the other hand, is only able to analyse the very last stages of damage.

It is clear that the severe operating conditions that led to the failure began to generate heat long before the damage actually progressed.

5 Conclusions

Detecting defects in their early stages can be challenging, especially when dealing with slowly evolving problems such as the test case under investigation. However, it is essential to pay close attention to the appropriate time horizon during the analysis. In this scenario, conventional SCADA trend analysis does not provide meaningful insight into the evolution of the fault, so multi-scale approaches were explored. An important contribution to a deeper understanding of the fault (e.g. the location of the faulty components or the evolution of the damage) could come from machine learning approaches: this is evident in the application of Random Forest Regression to develop a model that captures normal system behaviour. This type of approach has given good results when applied to 10 minute SCADA data. The analysis of high resolution events from CMS has been much more challenging because they are generally too noisy and too random. In this case, only the use of a classifier can raise a general alarm, but without providing further details about the fault and its evolution.

Future efforts to develop the multi-scale machine monitoring approach will include:

- Investigating additional failures involving different components to assess the robustness of the method.
- Explore the possibility of combining frequency domain feature extraction and machine learning, especially when dealing with high resolution CM data;
- Extend the use of machine information such as counters and alarms to optimise the filtering of operating conditions of interest.

Funding

The Article Processing Charges for this article are taken in charge by the French Association of Mechanics (AFM).

Conflicts of interest

The authors have no relevant financial or non-financial interests to disclose.

Data availability statement

Data used in this paper are not public.

Author contribution statement

Conceptualization, FC, AC and MV; Methodology, FC, AC and MV; Validation, FC, AC and FB; Formal Analysis, FC and AC; Investigation, FC, AC and FB; Data Curation, FC and FB; Writing Original Draft Preparation, FC and AC; Writing Review & Editing, FC and AC; Supervision, FC and FB.

References

- [1] Q. Wang, Z. Dong, R. Li, L. Wang, Renewable energy and economic growth: new insight from country risks, *Energy* **238**, 122018 (2022)
- [2] M.I.H. Tusar, B.R. Sarker, Maintenance cost minimization models for offshore wind farms: A systematic and critical review, *Int. J. Energy Res.* **46**, 3739–3765 (2022)
- [3] Z. Liu, L. Zhang, A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings, *Measurement* **149**, 107002 (2020)
- [4] H. Gu, W.Y. Liu, Q.W. Gao, Y. Zhang, A review on wind turbines gearbox fault diagnosis methods, *J. Vibroeng.* **23**, 26–43 (2022)
- [5] M.D. Reder, E. Gonzalez, J.J. Melero, Wind turbine failures-tackling current problems in failure data analysis, *J. Phys.: Conf. Ser.* **753**, 072027 (2016)
- [6] F. Castellani, A. Garinei, L. Terzi, D. Astolfi, M. Gaudiosi, Improving windfarm operation practice through numerical modelling and supervisory control and data acquisition data analysis, *IET Renew. Power Generat.* **8**, 367–379 (2014)
- [7] F. Harrou, K.R. Kini, M. Madakyaru, Y. Sun, Uncovering sensor faults in wind turbines: an improved multivariate statistical approach for condition monitoring using SCADA data, *Sustain. Energy Grids Netw.* **35**, 101126 (2023)
- [8] D. Astolfi, F. Castellani, L. Terzi, Mathematical methods for SCADA data mining of onshore wind farms: Performance evaluation and wake analysis, *Wind Eng.* **40**, 69–85 (2016)
- [9] X. Chesterman, T. Verstraeten, P.J. Daems, A. Nowé, J. Helsens, Overview of normal behavior modeling approaches for SCADA-based wind turbine condition monitoring demonstrated on data from operational wind farms, *Wind Energy Sci.* **8**, 893–924 (2023)
- [10] R. Pandit, D. Astolfi, J. Hong, D. Infield, M. Santos, SCADA data for wind turbine data-driven condition/performance monitoring: a review on state-of-art, challenges and future trends, *Wind Eng.* **47**, 422–441 (2023)
- [11] S.Y. Oh et al., Condition-based maintenance of wind turbine structures: a state-of-the-art review, *Renew. Sustain. Energy Rev.* **204**, 114799 (2024)
- [12] F. Castellani, A. Garinei, L. Terzi, D. Astolfi, M. Gaudiosi, Improving windfarm operation practice through numerical modelling and supervisory control and data acquisition data analysis, *IET Renew. Power Generat.* **8**, 367–379 (2014)
- [13] D. Astolfi, F. Castellani, L. Terzi, Mathematical methods for SCADA data mining of onshore wind farms: Performance evaluation and wake analysis, *Wind Eng.* **40**, 69–85 (2016)
- [14] Y. Liu, L. Zhang, Data-driven fault identification of ageing wind turbine, in *2022 UKACC 13th international conference on Control (CONTROL)*. IEEE (2022)
- [15] D. Astolfi, R. Byrne, F. Castellani, Analysis of wind turbine aging through operation curves, *Energies* **13**, 5623 (2020)
- [16] C. Tutiven, A. Encalada-Davila, Y. Vidal, C. Benalcázar-Parra, Detecting bearing failures in wind energy parks: a main bearing early damage detection method using SCADA data and a convolutional autoencoder, *Energy Sci. Eng.* **11**, 1395–1411 (2023)
- [17] A. Murgia, R. Verbeke, E. Tsiporkova, L. Terzi, D. Astolfi, Discussion on the suitability of SCADA-based condition monitoring for wind turbine fault diagnosis through temperature data analysis, *Energies* **16**, 620 (2023)
- [18] A. Encalada-Davila, B. Puruncajas, C. Tutivén, Y. Vidal, Wind turbine main bearing fault prognosis based solely on scada data, *Sensors* **21**, 2228 (2021)
- [19] Y. Vidal, F. Pozo, C. Tutivén, Wind turbine multi-fault detection and classification based on SCADA data, *Energies* **11**, 3018 (2018)
- [20] E. Gonzalez, B. Stephen, D. Infield, J.J. Melero, Using high-frequency SCADA data for wind turbine performance monitoring: a sensitivity study, *Renew. Energy* **131**, 841–853 (2019)
- [21] A. Verma, D. Zappala', S. Sheng, S.J. Watson, Wind turbine gearbox fault prognosis using high-frequency SCADA data, *J. Phys.: Conf. Ser.* **2265**, 032067 (2022)
- [22] F. Natili, A.P. Daga, F. Castellani, L. Garibaldi, Multi-scale wind turbine bearings supervision techniques using industrial SCADA and vibration data, *Appl. Sci.* **11**, 6785 (2021)
- [23] A. Turnbull, J. Carroll, A. McDonald, Combining SCADA and vibration data into a single anomaly detection model to predict wind turbine component failure, *Wind Energy* **24**, 197–211 (2021)
- [24] X. Jin, Y. Chen, L. Wang, H. Han, P. Chen, Failure prediction, monitoring and diagnosis methods for slewing bearings of large-scale wind turbine: a review, *Measurement* **172**, 108855 (2021)
- [25] A. Meyer, SCADA-based fault detection in wind turbines: data-driven techniques and applications, in: *Non-Destructive Testing and Condition Monitoring Techniques in Wind Energy*. Elsevier, 2023, pp. 1–13. <https://doi.org/10.1016/B978-0-323-99666-2.00001-0>

- [26] M. Jankauskas, A. Serackis, M. Šapurov, R. Pomarnacki, A. Baskys, V.K. Hyunh, T. Vaimann, J. Zakis, Exploring the limits of early predictive maintenance in wind turbines applying an anomaly detection technique, *Sensors* **23**, 5695 (2023)
- [27] P.F. Smith, S. Ganesh, P. Liu, A comparison of random forest regression and multiple linear regression for prediction in neuroscience, *J. Neurosci. Methods* **220**, 85–91 (2013)
- [28] H.J. Shin, D.H. Eom, S.S. Kim, One-class support vector machines—an application in machine fault detection and classification, *Comput. Ind. Eng.* **48**, 395–408 (2008)
- [29] J. Saari, D. Strömbergsson, J. Lundberg, A. Thomson, Detection and identification of windmill bearing faults using a one-class support vector machine (SVM), *Measurement* **137**, 287–301 (2019)
- [30] B. Schölkopf et al., Estimating the support of a high-dimensional distribution, *Neural Computat.* **13**, 1443–1471 (2001)
- [31] G. Dorcas Wambui, G.A. Waititu, A. Wanjoya, The power of the pruned exact linear time (PELT) test in multiple changepoint detection, *Am. J. Theor. Appl. Stat.* **4**, 581 (2015)
- [32] R. Killick, P. Fearnhead, I.A. Eckley, Optimal detection of changepoints with a linear computational cost, *J. Am. Stat. Assoc.* **107**, 1590–1598 (2012)

Cite this article as: F. Castellani, M. Vedovelli, A. Canali, F. Belcastro, Wind turbine gearbox multi-scale condition monitoring through operational data, *Mechanics & Industry* **25**, 28 (2024)