Wind turbine gearbox condition monitoring through a multi-scale data-driven approach.

Francesco Castellani[1], Matteo Vedovelli[1], Francesca Belcastro[2]

[1]University of Perugia, Department of Engineering, Perugia (ITALY)
francesco.castellani@unipg.it

[2]Fera srl, Piazza Cavour 7 - 20121 Milano (ITALY)

Abstract

Since wind is expected to play a crucial role on the worldwide electricity production scenario, the reliability of the turbines is attracting attention from industry as well from the scientific community. New techniques for efficient condition monitoring of the key components can be fundamental in order to optimize the performance and the maintenance of a large fleet of turbines. The gearbox and bearings represent the most critical mechanical components, as they are responsible for a large part of the wind turbine downtime during its overall life. Anyway, monitoring for wind turbine gears is challenging due to the non-stationarity of the operation and the lack of noise-free vibration measurements. In the present work, a new approach for long to short term efficient monitoring of wind turbine drivetrains has been developed basing on real-world data. An incipient fault on the drivetrain of a turbine has been used as a test case for developing a new approach based on the use of multi-scale data sources. On one side SCADA (Supervisory Control And Data Acquisition) data have been used for a general monitoring of the state of the machine's component on long to medium term time-scales, high multi-resolution data from triggered events collected by a CMS (Condition Monitoring System) were tested to refine the diagnosis and prognosis of the fault on a shorter scale. Even if triggered spots events are very difficult to be used when classifying a target machine with a healthy reference one, the results demonstrate that the use of CMS multi-scale high resolution data can be quick and useful in the fault diagnosis. In the present work, the one class-SVM (Support Vector Method) was used for novelty detection. The approach, when applied to all the available time scales, is able to detect the incoming fault also several years in advance and can therefore be proposed as quick detection approach requiring less data with respect to the classical data-driven regression normal behaviour model developed with continuously available SCADA data.

1 Introduction

The generation of electricity from renewable sources has been rapidly accelerating in the latest years. It is estimated that in the OECD countries the share of renewable electricity production will increase from 10.8% in OECD countries in 2019 to one-third by 2035 [1]. As an example, in the first quarter of 2023, wind has been the main source of electricity in U.K. and, in the same period, the 42% of the electricity has been produced from renewable sources against a 33% from fossil fuels.

In this context, it is fundamental to continuously improve the profitability of renewable energy technologies. In particular, for wind turbines, the highest share of costs comes from the Operation & Maintenance. Maintenance procedures are often expensive and time consuming, as they involve high quote transfer of equipment, workers and replacing parts [2]. As can be seen from Figure 1 [3], mechanical components (gears and bearings) have high failure rates and high associated downtime. This matter of fact motivates the vast attention in the literature to wind turbine gears and bearings fault diagnosis [4, 5].

Industrial wind turbines are commonly equipped with Supervisory Control And Data Acquisition (SCADA) control systems, which store with a typical averaging time of ten minutes a vast series of environmental, operational, mechanical, electrical and thermal measurements. This kind of systems has been developed for allowing remote control of wind turbines, but the increase in the number of monitored quantities has stimulated its use for condition/performance monitoring [6]. The advantage of SCADA systems is that they are practical and the data

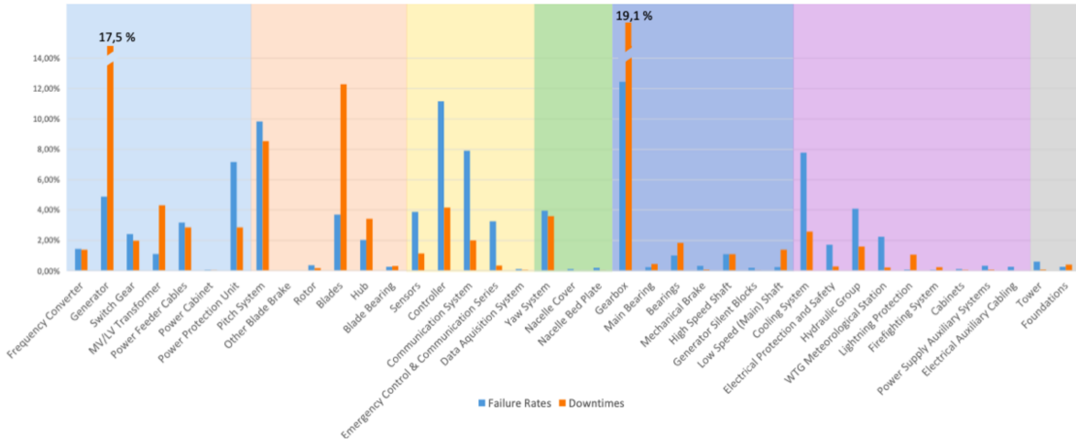

Figure 1: Estimates of failure rates and associated downtime for the most important wind turbine components.

are relatively easy to store continuously, while the drawback is that the averaging time is too long for capturing the dynamics of the machine. Consequently, the literature on wind turbine fault detection based on SCADA data analysis commonly involves the use of medium [7] to long-term [8] trends of meaningful measurements, which typically are temperatures collected at rotating sub-components. The rationale is that the onset of a fault should be accompanied by anomalous heat release in the form of trend, spike or increased variability. Thus, in practice, a normal behaviour model is constructed using data which can be reasonably assumed to describe healthy operation and the statistical novelty of the measurements in the target period is analysed. The normal behaviour model can be a regression [9] or a classification [10].

Given the above line of reasoning, it is comprehensible that there is a tendency towards the use of high-frequency data [11, 12] for wind turbine fault diagnosis. The main issue with that kind of data is that industrial systems typical store them only when some triggered events occur and therefore a continuous monitoring might be impracticable. Therefore, a recent trend in the literature is a multi-scale approach which interfaces the continuous monitoring based on SCADA data to a refined analysis in proximity of the fault, based on high-frequency data. Only a few examples of this co-integration approach are available in the literature, as [**?**, 14]. Therefore, the multi-scale data analysis for wind turbine fault diagnosis represents an innovative research direction, to which the present study aims at contributing through a real-world test case discussion.

In this work, a case of gear bearing failure in an industrial wind farm is analysed through the following multi-scale approach:

- long-term, through the analysis of hourly-averaged SCADA data for 6 years prior the fault;
- medium term, through the analysis of ten minutes averaged SCADA data for order of one year prior the fault;
- short term, through the analysis of spot events sampled with a frequency ranging from order of 1 Hz to order of thousands of Hz.

The added value of such approach is that it allows discussing critically the identification of the fault onset. Actually, in real-world test cases the classification of the data is complicated by the fact that a clear demarcation line between healthy and faulty measurements does not exist. The use of multiple time scales can thus be useful to clarify the patterns.

The structure of the work is the following. The test case and the data set are described in Section 2; the methods are briefly outlined in Section 3; the results are collected and discussed in Section 4; the conclusions are drawn in Section 5.

2 Test case and data sets

The analysed case is represented by an horizontal-axis multi-MW Wind Turbine that was affected by a recently discovered fault on a drive-train bearing.
The clear evidence of the fault was not revealed by the standard SCADA (Supervisory Control And Data Acquisition) data analysis and the final diagnosis was recently discovered by the presence of metal debris in the oil of the gearbox.
Generally the faults in bearings can have very different prognosis and failure modes [15]. They can dramatically bring quickly to severe faulty conditions (as the fatigue-driven faults) or affect the machine operation with a constant but slow evolution in time (as the wear or overload driven damages). The latter case is the one analysed in the present work and is generally more challenging to be correctly diagnosed. The machine in this case is able 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 system were used as they can provide a great amount of information on different time scales:

- 10 minutes averages for SCADA (continuous monitoring);
- 1 to 1000 Hz for high frequency SCADA (triggered events).

In Figure 2, the trend for the gear bearing 1 temperature on the last year of operation is represented. The standard SCADA analysis is not able to give a clear view of the fault, so that new approaches were used to unveil the damage evolution.

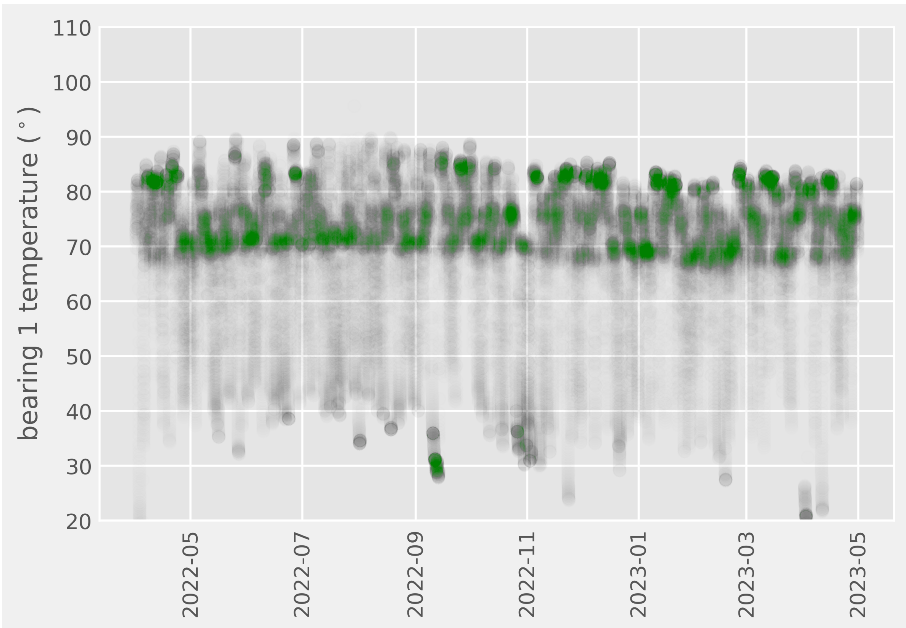

Figure 2: Trend of gear bearing 1 temperature over the last year.

3 Methods

A mechanical fault is generally detectable by the temperature rise or by the increase of the level of the vibrations. Both of them were used in this work, but on different time scales. Namely, three different approaches have been developed to study the evolution of the damage:

1. A differential SCADA data analysis for the temperature of the bearings;
2. A Random Forest Regression data-driven model for all the temperatures of the components;
3. A feature classification approach with the One-Class Support Vector Machine for the high resolution triggered events on the drive-train and tower vibrations.

The differential analysis was used to investigate if the machine under investigation is displaying a rise in the temperature of the bearing. The parameter analysed over different time horizons is the one calculated for each time step with Equation 1, where T_{tar} is the temperature of the bearing of the turbine under investigation and T_{ref} is the temperature of one or more reference healthy units.

$$\Delta T = T_{T_{tar}} - \frac{i=1}{N} \sum_{1}^{N} T_{T_{ref,i}} \tag{1}$$

The same kind of analysis can also be performed using a data-driven approach by means of a machine-learning algorithm. In this case, the reference is represented by the results obtained by a normal-behaviour model trained on an healthy data-set of the same machine. For this analysis, a Random Forest Regression model was used to be trained as normal-behaviour model and the used input parameters are listed in table 1. The different bearings temperatures were considered as output in order to calculate for each component the model residuals.
The One-Class SVM was the selected feature classification approach used for processing the high resolution triggered events. In this case, the tower vibration components, the drive-train oscillation and the rotor speed were used for elaborating the features used for classification. The reference classification model was trained on healthy events. The features were calculated scaling the Power Spectral Density of the parameters on different time scales from 1 to 20 Hz.

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

Table 1: Input parameters for the Random Forest Regression Model

4 Results

The differential SCADA data analysis was performed with different time resolutions and different time horizons:

- a short-term analysis was developed using 10 minutes averages SCADA parameters over a period of 18 months;
- a long-term analysis was developed using hourly averages over a period of more than 6 years.

The real symptom of the fault was discovered only when running the long-term analysis. As clearly highlighted in Figure 3, a raising trend of the differential temperature of gear bearing 1 was discovered for the turbine under investigation. The most clear evidence of the damage was unveiled when analysing on the long-term the trend for monthly averages of ΔT for bearing 1. The threshold for the alarm was defined as:

$$Thr = \mu \pm \chi\sigma \tag{2}$$

where μ is the mean of the reference parameter, σ is the standard deviation and χ is a constant tuned in agreement with the time resolution. For the long-term analysis and the monthly means a $\chi = 4$ was selected so that the alarm was triggered on the beginning of May 2021 as can be observed in Figure 3. The hourly data are much noisier and the same threshold brings to possible early false alarms.
The same processing on the residuals for the normal behaviour model with the Random Forest Regression gave similar results triggering an alarm in June 2021 (see figure 4). The results from the same analysis on the short-term period were not able to detect the fault. This occurs because we are dealing with a weak slow-evolving fault on an high speed shaft bearing.

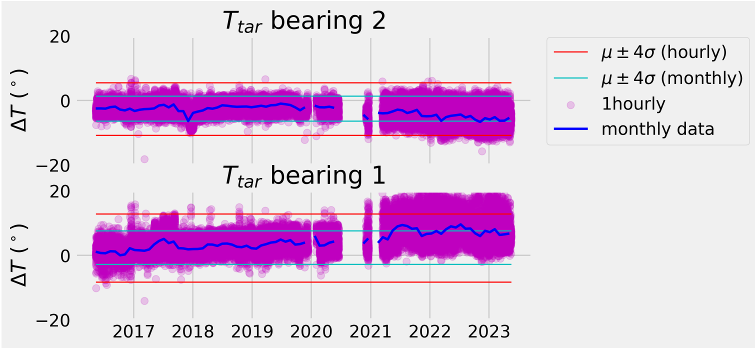

Figure 3: Results from the SCADA analysis.

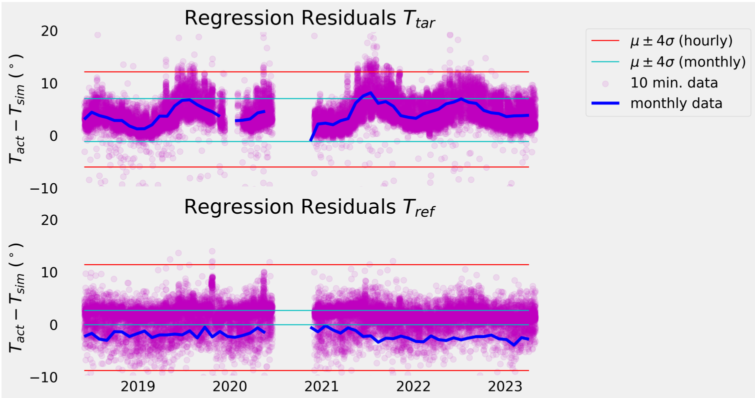

Figure 4: Random Forest Regression residuals analysis for gear bearing 1.

The results from the One-Class SVM classification revealed that, with this approach, only triggering a general alarm was possible, without classifying the prognosis of the damage. The best results were obtained when using 20 Hz data without considering the rotor speed (figure 5). When using also the rotor speed, differences between faulty target and healthy reference are boosted at 20 Hz but with the other time scales the fault is not clearly distinguishable (figure 6).

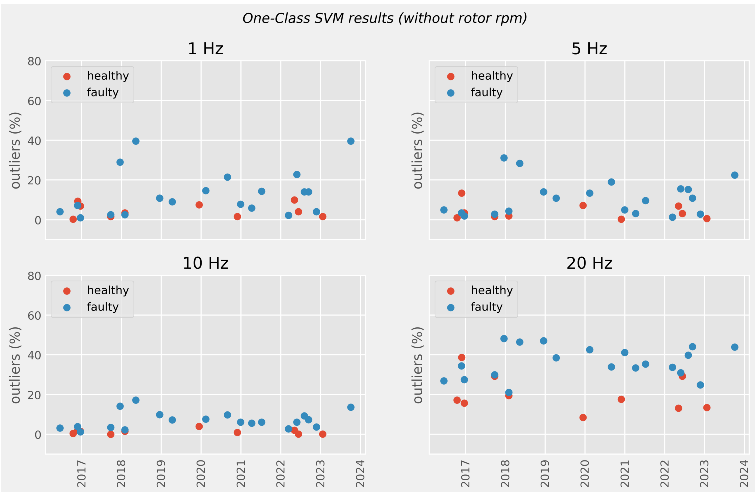

Figure 5: One-Class SVM results without rotor rpm.

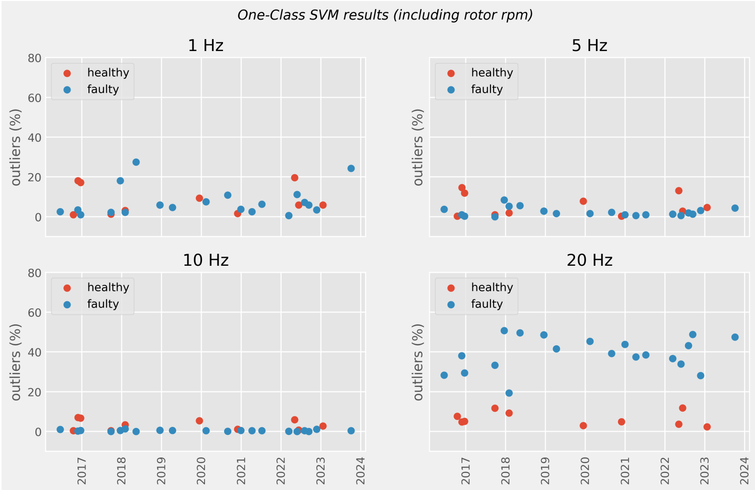

Figure 6: One-Class SVM results with rotor rpm.

With this approach it is also much more difficult to have an idea of the location of the fault.

5 Conclusions

Early fault diagnosis can be a complex task, especially for low evolving faults as the one analysed for the test-case and observing the right time horizon is fundamental. Standard SCADA trending analysis is not able in this case to give a useful insight on the fault evolution, so that challenging multi-scale approaches have been considered. Machine Learning approaches can give an important contribution to understand the location of the faulty components and the evolution of the damage as demonstrated with the application of the Random Forest Regression for developing a normal behaviour model. Spot high resolution events are too noisy and too random so that only with the use of a classifier can help only in triggering a general alarm, but cannot be used to analyse the evolution of the fault. The future work will include:

- analysing more failures involving different components with a multi-scale approach;
- exploring, in the case of high resolution CM data, a wider use of frequency domain analysis and machine learning;
- implementing a wider use of machine's information (i.e. alarm and counters) for optimally filtering the operational conditions of interest.

Acknowledgements

The author acknowledges the FERA srl company for providing all the data and the technical information for the study.

References

[1] Wang, Q., Dong, Z., Li, R., & Wang, L. *Renewable energy and economic growth: new insight from country risks.* Energy, 238, 122018.

[2] Tusar, M. I. H., & Sarker, B. R. *Maintenance cost minimization models for offshore wind farms: A systematic and critical review.* International Journal of Energy Research, 46(4), 3739-3765.

[3] Reder, M. D., Gonzalez, E., & Melero, J. J. *Wind turbine failures-tackling current problems in failure data analysis.* In Journal of Physics: Conference Series (Vol. 753, No. 7, p. 072027). IOP Publishing.

[4] Liu, Z., & Zhang, L. *A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings.* Measurement, 149, 107002.

[5] Gu, H., Liu, W. Y., Gao, Q. W., & Zhang, Y. *A review on wind turbines gearbox fault diagnosis methods.* Journal of Vibroengineering, 23(1), 26-43.

[6] Pandit, R., Astolfi, D., Hong, J., Infield, D., & Santos, M. *SCADA data for wind turbine data-driven condition/performance monitoring: A review on state-of-art, challenges and future trends.* Wind Engineering, 47(2), 422-441.

[7] Tutiven, C., Encalada-Davila, A., Vidal, Y., & Benalcazar-Parra, C. *Detecting bearing failures in wind energy parks: A main bearing early damage detection method using SCADA data and a convolutional autoencoder.* Energy Science & Engineering, 11(4), 1395-1411.

[8] Murgia, A., Verbeke, R., Tsiporkova, E., Terzi, L., & Astolfi, D. *Discussion on the Suitability of SCADA-Based Condition Monitoring for Wind Turbine Fault Diagnosis through Temperature Data Analysis.* Energies, 16(2), 620.

[9] Encalada-Davila, A., Puruncajas, B., TutivÃ©n, C., & Vidal, Y. *Wind turbine main bearing fault prognosis based solely on scada data.* Sensors, 21(6), 2228.

[10] Vidal, Y., Pozo, F., & TutivÃ©n, C. *Wind turbine multi-fault detection and classification based on SCADA data.* Energies, 11(11), 3018.

[11] Gonzalez, E., Stephen, B., Infield, D., & Melero, J. J. *Using high-frequency SCADA data for wind turbine performance monitoring: A sensitivity study.* Renewable energy, 131, 841-853.

[12] Verma, A., Zappala', D., Sheng, S., & Watson, S. J. *Wind turbine gearbox fault prognosis using high-frequency SCADA data.* In Journal of Physics: Conference Series (Vol. 2265, No. 3, p. 032067). IOP Publishing.

[13] Natili, F., Daga, A. P., Castellani, F., & Garibaldi, L. *Multi-scale wind turbine bearings supervision techniques using industrial SCADA and vibration data.* Applied Sciences, 11(15), 6785.

[14] Turnbull, A., Carroll, J., & McDonald, A. *Combining SCADA and vibration data into a single anomaly detection model to predict wind turbine component failure.* Wind Energy, 24(3), 197-211.

[15] Jin, X., Chen, Y., Wang, L., Han, H., & Chen, P. (2021). *Failure prediction, monitoring and diagnosis methods for slewing bearings of large-scale wind turbine: A review.* Measurement, 172, 108855.