On the use of Artificial Intelligence for Condition Monitoring in Horizontal-Axis Wind Turbines

On the use of Artificial Intelligence for Condition Monitoring in Horizontal-Axis Wind Turbines

Fabrizio Bonacina, Eric Stefan Miele, Alessandro Corsini

Department of Mechanical and Aerospace Engineering, Sapienza University of Rome, Via Eudossiana 118, I00184, Rome, Italy;

E-mail: fabrizio.bonacina@uniroma1.it; ericstefan.miele@uniroma1.it; alessandro.corsini@uniroma1.it

Abstract. Wind power is one of the fastest-growing renewable energy sectors and is considered instrumental in the ongoing decarbonization process. However, wind turbines (WTs) present high operation and maintenance costs caused by inefficiencies and failures, leading to everincreasing attention to effective Condition Monitoring (CM) strategies. Nowadays, modern WTs are integrated with sensor networks as part of the Supervisory Control and Data Acquisition (SCADA) system for supervision purposes. CM of wind farms through predictive models based on routinely collected SCADA data is envisaged as a viable mean of improving producibility by spotting operational inefficiencies. In this paper, we introduce an unsupervised anomaly detection framework for wind turbine using SCADA data. It involves the use of a multivariate feature selection algorithm based on a novel Combined Power Predictive Score (CPPS), where the information content of combinations of variables is considered for the prediction of one or more key parameters. The framework has been tested on SCADA data collected from an off-shore wind farm, and the results showed that it successfully detects anomalies and anticipates major bearing failures by outperforming a recent deep neural approach.

1. Introduction

Offshore Renewable Energy, and wind energy, in particular, are considered key ingredients in the EU strategy to meet the goal of climate neutrality by 2050. According to the EU Commission Europe's offshore wind capacity will grow from its current level of 12 GW to at least 60 GW by 2030 and to 300 GW by 2050 [1]. This increasing interest is motivated by a number of factors. To mention but a few: off-shore wind energy quality (even in the Mediterranean spots [2]), the new generations of multi-MW wind turbines (WTs) [3], the maturity of technology and know-how in the off-shore industries, and the cost competitiveness in terms of LCOE [3, 4, 5, 6] in view of the compensation of capital and erection costs with higher energy production.

In Europe, the cumulated 2019 offshore wind power capacity was 22 GW, almost all in the North Sea, with a plan to reach 450 GW by 2050 [7]. In Italy, with a present only on-shore wind power capacity of 10.5 GW, the National Integrated Climate and Energy Plan [8] foresees the starting of offshore wind sector with a target of 0.9 GW. Notably, to date the first 30 MW off-shore wind park has been fully commissioned in Taranto harbour area.

Concerning the technology trend, moving off-shore wind turbine size increase is recognized as one of the leading design factors, with floating offshore wind turbines particularly suited to

benefit from such a trend [9]. While this circumstance benefits from higher gross capacity factors as a consequence of larger and higher rotors, the impact of installation and operation costs per-kW could be largely affected. From an OM perspective, upscaled off-shore floating WTs are subjected to more complex loading patterns and to a number of parasite phenomena affecting the aero-hydro-servo-elastic behaviour of the rotors and resulting in even more severe operations of the drive-train. Among the possible configurations, the most common is the gearbox-operated WT, in which the main shaft is connected to the generator via a gearbox that increases the rotation speed typically up to 1500-1800 rpm. As reported in [10], the most critical components of a WT typically reside in the drivetrain, and among these the generator is particularly prone to failure due to its high speed and wide load fluctuation range. In addition, a WT consists of several bearings that allow rotary motion and support radial and axial loads. They include the rolling bearings of the drive system (such as main bearings, gearbox bearings and generator bearings) and the rotational bearings of the control system (such as blade and yaw bearings). As a result, WT bearings are subjected to heavy loads and large moments, so they are often subject to failure well before their expected designed useful life [11]. The statistics presented in [12] show that about 40% of the faults in induction motors was caused by bearings and the National Renewable Energy Laboratory (NREL) reports that 76% of gearboxes failed due to this components [13].

Therefore, diagnosis of incipient faults in drive-train components can effectively reduce the economic loss caused by downtime in power generation and unplanned maintenance, repairs and replacements. Remedial approaches to face OM challenges advocate Condition Monitoring (CM) strategies capable of early detection and isolation of incipient faults. CM is a key ingredient to enable condition-based maintenance, able to outperform the on-schedule state-of-the-art, in a view to identify (at early stages) component degradation and limit unnecessary outage of WTs. In WT mechanical components, CM is typically based on the acquisition of high-frequency data (e.g., vibrational and acoustic analysis), possibly processed through a variety of methods (see [14, 11] for recent reviews). However, this strategy suffers from several limitations as it demands the installation of additional sensors on WTs and specific data infrastructure, in fact discouraging the implementation [15, 16].

In contrast, common practice is to enable CM of wind power generation plants through analysis of data routinely collected by sensor networks in Supervisory Control and Data Acquisition (SCADA) systems for monitoring power-train status (e.g. bearing temperature, lube oil sub-system, etc.). SCADA data processing for condition monitoring is envisaged as a mean to enable wind turbine management in a predictive maintenance perspective [17, 18].

A number of research contributions appeared to date in this field with a focus on the use of Machine Learning (ML) approaches (see [19] for a recent review). ML-based methods involve the definition of reference models based on historical operational data by training a learning algorithm to predict one or more signals during periods when turbine components can be considered healthy. These models are then applied in real time to estimate the value of a specific signal and the prediction errors are used as indicators of anomalies and incipient failures. Such approaches have been successfully applied to predict faults across various wind turbine components such as turbine blades [20], generators [16], and gearboxes [21], while only very few publications can be found for bearings. Thus, as reported in [11], the latter can still be considered a research direction. However, the main problem of these ML-based approaches is the difficulty in isolating health conditions for training the reference model and also in handling prediction errors by means of control charts. In practice monitoring approaches are difficult to establish due to the various operational modes, which cause signals to widely fluctuate [22].

In addition, given the large number of variables monitored by SCADA systems, dimensionality reduction becomes crucial when predictive model are employed in real-time CM systems. To overcome these limitations, in this paper we propose the application of a feature selection method

directly on WT SCADA data to detect incipient faults. Specifically, we used a framework based on the Combined Predictive Power Score (CPPS) introduced in [23]. Unlike other standard approaches in the literature that only consider relationships between pairs of variables, CPPS takes into account the information content of combinations of variables to predict one or more key parameters. The CPPS augmentation resides in the selection of multi-variate wind turbine predictors for different targets. Moreover, the search in the multi-dimensional feature space is carried out in a SFS fashion explored using q expanding subsets of variable combinations. We have shown that variations in CPPS scores can be successfully used as indicators to detect anomalies and failures occurring on generator bearings in advance. One of the great advantages of this unsupervised approach applied for CM purposes lies in the fact that it does not require the isolation of a healthy condition period for training the model, and in addition, indicator monitoring is handled directly on the prediction scores, which do not suffer from the fluctuations and periodicity typical of signal residuals. The rest of the paper illustrates the steps of the proposed anomaly detection framework based on the CPPS. Then, a case study is described in which the method is applied to SCADA data collected from an off-shore wind farm. Finally, the results are shown and the performance of the CPPS is compared with that of a recent MTGCAE network-based deep neural fault detection system introduced in [24].

2. Methodology

In this section, we present the framework of the proposed methodology (see Figure 1 for details). In particular, we first discuss the data preparation phase including outlier removal, time series clustering and scaling, and the derivation of the Combined Predictive Power Score (CPPS)[24].

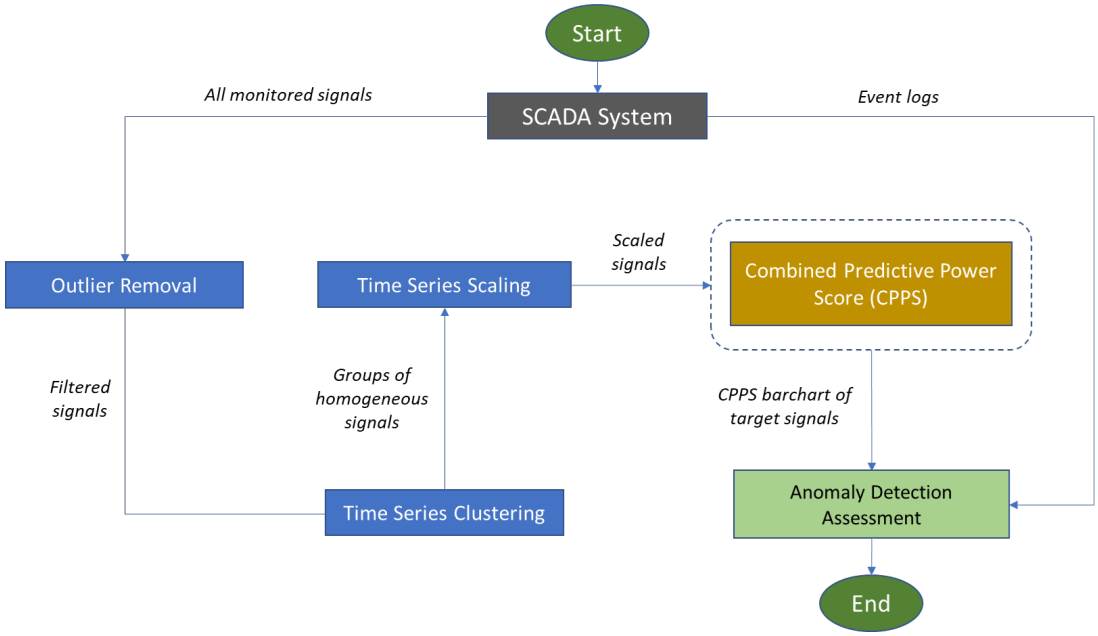

Figure 1. Framework of the proposed methodology for anomaly detection in WT based on CPPS.

Outlier Removal We applied both monovariate and multivariate outlier removal. First, we define a 6-hours sliding window for each signal and filter samples outside the interquartile range of the window. Then, we perform a density-based clustering [25] on the power curve of each turbine, filtering samples not belonging to the main power curve cluster.

Time Series Clustering In order to avoid that highly correlated variables are selected as important features for each other, similar signals have been grouped together using a time series clustering algorithm. In particular, we compute the correlation matrix using the Pearson correlation coefficient between pairs of variables. Then, we build a weighted graph considering the correlation matrix as adjacency matrix and apply the Louvain method for community detection in order to cluster the original time series in homogeneous groups.

Moreover, the proposed methodology considers groups of homogeneous variables instead of single features for the process of feature selection.

Time Series Scaling Since the proposed algorithm is based on a neural network having multiple inputs and multiple outputs, it is necessary to scale the time series of the monitored variables. In particular, this was achieved by using the RobustScaler [26] which scales variables using statistics that are robust to outliers by removing the median and scaling data according to the interquartile range.

Predictive Power Score The Predictive Power Score (PPS) was originally proposed as an asymmetric, data-type-agnostic score to detect linear or non-linear relationships between two variables [27]. The score is calculated using only one variable trying to predict another one, and can be employed to select features as predictors for a target.

In the regression of real-valued variables, the PPS usually considers the mean absolute error (MAE) as evaluation metric. As baseline score, we calculate the MAE of a naive model ($MAE^{f_t}_{naive}$) that always predicts the median of the target variable f_t. Then, we select a candidate variable f_i as input for a model regressing the target variable f_t and compute the MAE ($MAE^{model}_{f_i,f_t}$). In this way, we can define the PPS for a candidate feature f_t and the target variable f_t as:

$$PPS(f_i, f_t) = 1 - \frac{MAE^{model}_{f_i,f_t}}{MAE^{naive}_{f_t}} \tag{1}$$

The authors of the PPS selected Decision Tree as regression model [28].

The PPS ranges from 0 to 1 and assumes values close to 1 when the candidate input feature of the model has a high predictive power over the target. Otherwise, values close to 0 indicate that the candidate feature has a low predictive power.

Combined Predictive Power Score We recently proposed in [23] a modified PPS method implemented with a feature selection algorithm, namely the Combined Predictive Power Score (CPPS), that extends PPS by considering combinations of candidate input features $F_g = \{f_i\} \subseteq F$ for the prediction of multiple target features $T = \{f_t\}$, where F is the set containing all input features, f_i is a candidate input and f_t a target.

In this way, the CPPS can detect linear or non-linear relationships between different groups of variables. The model employed for the definition of the CPPS has, therefore, to regress a multivariate output by considering multivariate inputs. We used a Multilayer Perceptron (MLP) having one hidden layer with $5 \cdot |F_g|$ units and ReLu activation [29].

Similarly to Equation 1, the CPPS for a candidate group of variables F_g and a target group T is defined as:

$$CPPS(F_g, T) = 1 - \frac{MAE^{model}_{F_g,T}}{MAE^{naive}_{T}} \tag{2}$$

where $MAE^{model}_{F_g,T}$ is the MAE of the MLP regression model and MAE^{naive}_{T} is the MAE of the naive model that always predicts the median of each target variable in T.

The pseudocode in Algorithm 1 shows how the combination of input variables as predictors for a specific target group T are selected by maximizing the CPPS.

Data: F, T, q, ϵ
Result: F_{opt}
$C \leftarrow \{\{f_i\}, f_i \in F\}$;
do
  $C_{new} \leftarrow \{\}$;
  for $k \leftarrow 1$ **to** q **do**
    $F_k = CPPS\_k(C, T, k)$;
    foreach $f_i \in F$ **do**
      $F_g = F_k \cup \{f_i\}$;
      $C_{new} = C_{new} \cup \{F_g\}$;
    end
  end
  $C_{old} \leftarrow C$;
  $C \leftarrow C_{new}$;
while $CPPS(CPPS\_k(C, T, 1), T) - CPPS(CPPS\_k(C_{old}, T, 1), T) > \epsilon$;
$F_{opt} = CPPS\_k(C, T, 1)$;

Algorithm 1: The algorithm shows how the predictors for a target group are selected by maximizing the CPPS.

Being C a set containing combinations of variables, $CPPS\_k(C, T, k)$ computes the score $CPPS(F_g, T)$ for each combinations of variables $F_g \in C$ with respect to the target group T, sorts them in descending order and selects the k-th group.

The inputs of the algorithm are F, T, q and ϵ, namely the set of original input features, the set of target features, the number of best combinations of variables to consider during each iteration and the minimum increase of the CPPS required to avoid early stopping.

At each iteration we consider the k combinations of variables in C that achieve the highest CPPS and merge them with the original features one at the time, thus generating new combinations F_g for the next iteration. At the beginning of the algorithm, C is initialized with combinations that coincide with the single variables in F.

The algorithm stops when an iteration achieves a CPPS improvement with respect to the previous iteration lower that ϵ.

In order to consider the output of the time series clustering algorithm previously discussed, for each feature f_i added to a new combination F_g, all variables belonging to the same homogeneous groups are appended to the combination. The selection process of predictors presented in Algorithm 1 provided an average speedup of 18x with respect to trying random combinations of features up to triplets.

For the specific purpose of anomaly detection, we proposed a dynamic application of this approach where the algorithm is trained on groups of observations covering a time window of one month with an overlap of two weeks. Finally, the CPPS results are represented in the form of bar charts for each target signal.

3. Case Study

The proposed framework is validted in the detection of anomalies in a EDP (Energias de Portugal) off-shore wind farm located in the West African Gulf of Guinea available through

the open dataset [30].
Data were collected from four WTs belonging to the same off-shore wind farm, each having three blades with a diameter of 90 m and an hub height of 80 m. The maximum rotor speed is 14.9 rpm, and the maximum rated power is 2 MW at a nominal wind speed of 12 m/s. The main shaft is connected to an asynchronous generator via a three-stage planetary gearbox that increases the rotation speed to a maximum of 2016 rpm. The wind farm is ranked class 2 according to the standard IEC 61400 [31].
All WTs are equipped with a SCADA system for the monitoring of multiple parameters collected from the main components together with ambient measurements. A list of the monitored parameters is presented in Table 1. The dataset covers a time span of about 20 months (from January 1, 2016 to September 1, 2017) with a sampling rate of 10 minutes.

Table 1. Parameters monitored by WT SCADA system.

Signal ID	Description	Component
Gen_Bear_Temp	Temperature in generator bearing 1 (Drive End)	Generator Bearings
Gen_Bear2_Temp	Temperature in generator bearing 2 (Non-Drive End)	Generator Bearings
Gen_RPM	Generator rpm	Generator
Gen_Phase1_Temp	Temperature inside generator in stator windings phase 1	Generator
Gen_Phase2_Temp	Temperature inside generator in stator windings phase 2	Generator
Gen_Phase3_Temp	Temperature inside generator in stator windings phase 3	Generator
Gen_SlipRing_Temp	Temperature in the split ring chamber	Generator
Hyd_Oil_Temp	Temperature oil in hydraulic group	Hydraulic
Gear_Oil_Temp	Temperature oil in gearbox	Gearbox
Gear_Bear_Temp	Temperature in gearbox bearing on high speed shaft	Gearbox
Nac_Temp	Temperature in nacelle	Nacelle
Nac_Direction	Nacelle direction	Nacelle
Rtr_RPM	Rotor rpm	Rotor
Amb_WindSpeed	Wind speed	Ambient
Amb_WindDir_Relative	Wind relative direction	Ambient
Amb_WindDir_Abs	Wind absolute direction	Ambient
Amb_Temp	Ambient temperature	Ambient
Prod_TotActPwr	Total active power	Production
Prod_TotReactPwr	Total reactive power	Production
Grd_Prod_PsblPwr	Grid Power Request	Grid
HVTrafo_Phase1_Temp	Temperature in HV transformer phase L1	Transformer
HVTrafo_Phase2_Temp	Temperature in HV transformer phase L2	Transformer
HVTrafo_Phase3_Temp	Temperature in HV transformer phase L3	Transformer
Cont_Top_Temp	Temperature in the top nacelle controller	Controller
Cont_Hub_Temp	Temperature in the hub controller	Controller
Cont_VCP_Temp	Temperature on the VCP-board	Controller
Cont_VCP_ChokcoilTemp	Temperature in the choke coils on the VCS-section	Controller
Cont_VCP_WtrTemp	Temperature in the VCS cooling water	Controller
Spin_Temp	Temperature in the nose cone	Spinner
Blds_PitchAngle	Blades pitch angle	Blades

For this study, we focused in particular on WT T09, as it showed a series of failure events during the reporting period that led to unplanned maintenance interventions with a total of 15 days of outages of the WT (see Table 2 for details).

Looking at the event logs of the SCADA system shown in Table 2, it is possible to notice a first high temperature alarm reported on the generator bearings of WT T09 on 07 June, 2016, followed by a second anomaly of the same nature recorded about two months later. Subsequently, on 17 October, 2016, a major fault was reported on the generator bearings, which led to a downtime on T09 lasting about 26 hours for the repair of the damaged components. After about three months of operation, another serious failure of the generator bearings occurs, probably related

Table 2. Main events reported in the maintenance log file available in ref. [30].

Turbine ID	Event ID	Event timestamp	Assembly/Sub-assembly	Type of Event
T09	A09GB1	07/06/2016; 16:59	Generator bearings	Alarm: high temperature generator bearing
T09	A09GB2	22/08/2016; 18:25	Generator bearings	Alarm: high temperature generator bearing
T09	R09GB3	17/10/2016; 09:19	Generator bearings	Generator bearings repair
T09	R09GB4	25/01/2017; 12:55	Generator bearings	Generator bearings replacement

to thermal cracks, leading to a long downtime for component replacement lasting about two weeks.

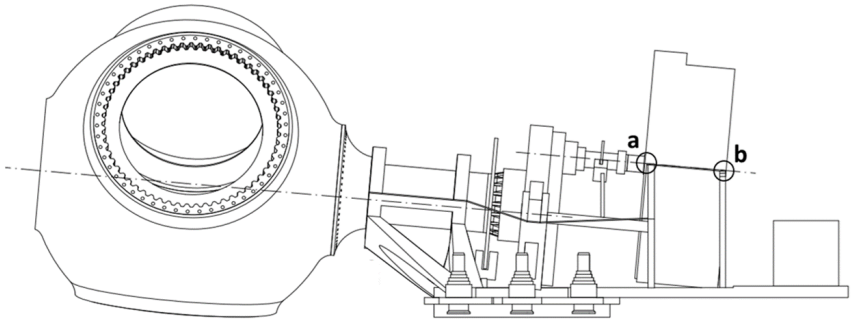

Figure 2. Longitudinal view of a typical drive train and locations of Gen_Bear_Temp (a) and Gen_Bear2_Temp (b) sensors.

4. Results and Discussion

In the following, the anomaly detection analysis focuses on signals from temperature sensors Gen_Bear_Temp and Gen_Bear2_Temp, namely the drive-end and non-drive end bearing sensors, as illustrated in Figure 2. Figure 3 shows the trend of the two signals Gen_Bear_Temp and Gen_Bear2_Temp of WT T09 as recorded in the SCADA system during the 9-month sampling period. Notably, signals are scaled using the Robustscaler technique. Figure 3 also includes the time stamp of events included in Table 2.

From the trend analysis of the scaled signals, it is possible to note that, after the two high temperature alarms (event IDs: A09GB1 and A09GB2) reported in the maintenance log in 2016 (i.e. on 07 June, and 26 August respectively), an anomalous period of about 3 months occurs where the two temperatures lose their typical correlation and show divergences in terms of magnitude.

Figure 4 provides a detail of this anomalous period, where the downtime due to the two unscheduled maintenance operations can also be observed. Concerning the event ID R09GB3, the maintenance log reports the downtime for replacement on 25 January, 2017, while the signal analysis in Figure 4 shows an initial downtime from 15 January, 2017 to 20 January, 2017, followed by a restart attempt and subsequent outages between 24 January, 2017 and 2 February, 2017.

The first key step in the anomaly detection data pipeline is the reduction of features starting from the available SCADA data set. To this end, the time series are clustered using the Louvain community detection method. Figure 5, instead, shows the result of the clustering applied to

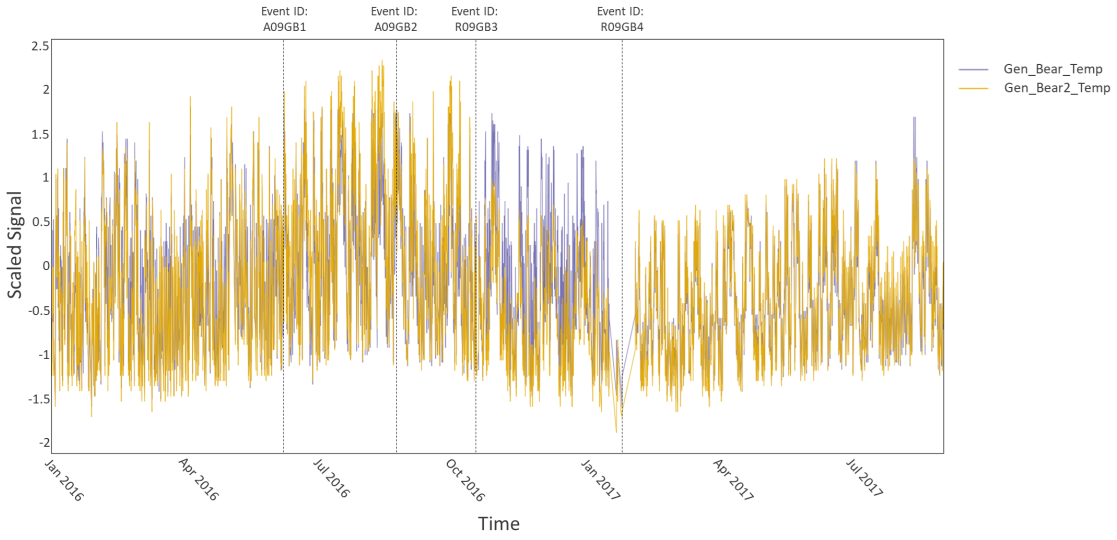

Figure 3. Trend of the scaled temperature signals Gen_Bear_Temp and Gen_Bear2_Temp of T09 from 1 January, 2016 to 1 September 2017. The plot also shows the occurrences of the four events as reported in the maintenance log.

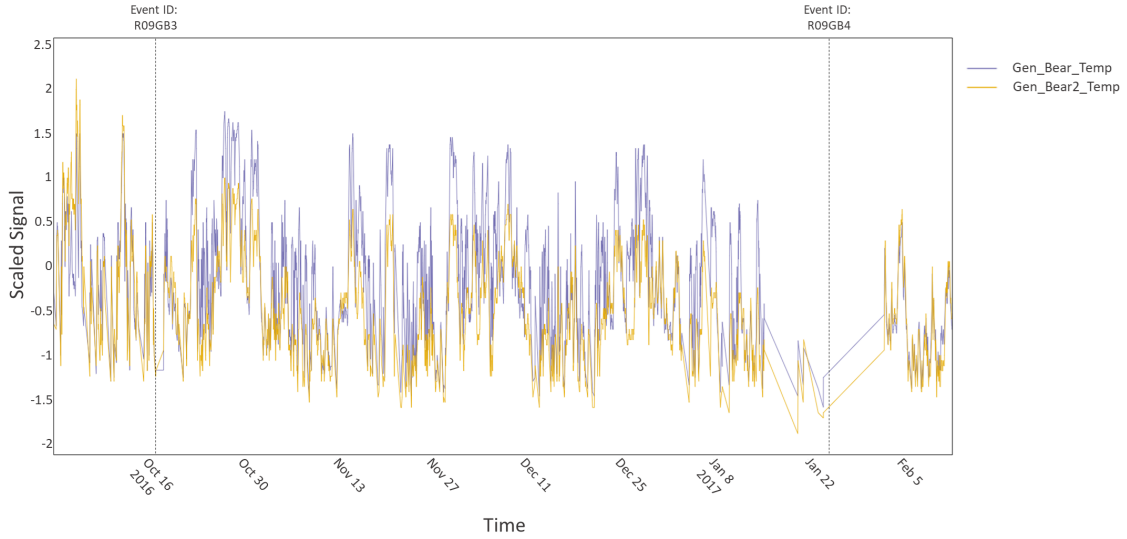

Figure 4. Detail of the anomalous period included between the two unscheduled maintenance operations (event IDs R09GB3 and R09GB4).

signals included in Table 1, after removing the outliers as described in Section 2. Specifically, Figure 5 presents the output of time series clustering using a force-directed algorithm, namely the Fruchterman–Reingold layout [32], where each node is associated with a variable, while edges represent the mutual correlation among signals. Notably, 20 highly correlated signals were clustered into 8 different homogeneous groups which, in addition to the other 10 independent signals, were used as input for the CPPS model. It is also worth noting that the clustering isolates homogeneous variables (e.g. generator and rotor rpm, or wind and nacelle directions)

but it also distinguish temperature signals from the diverse electrical components (generator phases, trafo and inverter).

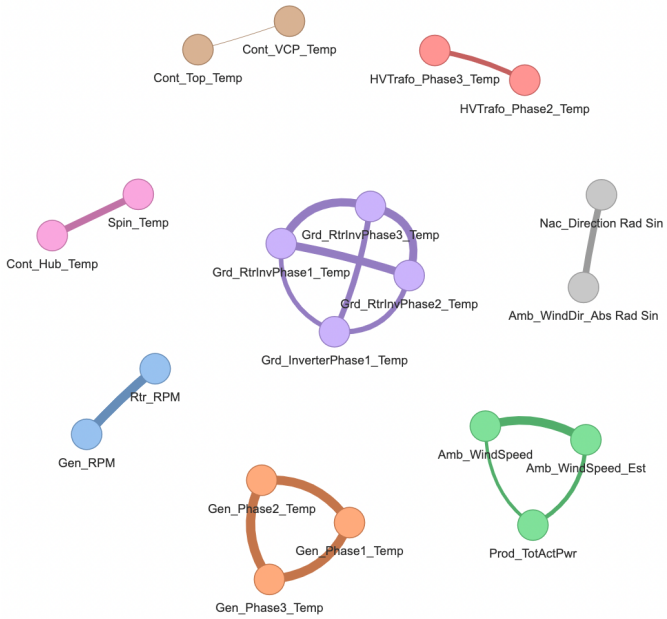

Figure 5. Clusters of time series, Fruchterman-Reingold layout.

The results of the anomaly detection model are detailed in Figure 6. In particular, the figure shows the CPPS bar charts, where each histogram represents the maximum score obtained from the combination of predictors for the two target variables Gen_Bear_Temp and Gen_Bear2_Temp. Each bar represents a two week time window used by the multivariate regressor. Different clusters of variables contribute the CPPS of generator bearing temperatures. Gen_Bear_Temp predictor depends on generator phase temperature signals, while Gen_Bear2_Temp is scored principally by Gen_Bear_Temp.

Interestingly, bar chart in Figure 6, from March 2016 to March 2017, give evidence of a reduction in the scores of the models for both the bearing whereas the remaining period of observation featured CPPS above the 0.8 threshold.

To this end the CPPS trends (and level) seem to anticipate the occurrence of alarms documented in the SCADA log. For instance, CPPS indicators drop during March 2016 well in advance with respect to the event ID A09GB1, i.e. high temperature alarm of generator bearings dated June 2016. In addition, when comparing the CPPS trends against the timestamp of the SCADA log, it is evident that Gen_Bear_Temp CPPS linearly descended before the maintenance interventions and it reached the minimum immediately before the event R09GB4, i.e. indicated as bearing replacement in the log.

5. Conclusions

In this paper, we present an original unsupervised anomaly detection framework in the context of horizontal axis wind turbines (WTs) based on SCADA data. The core of the method is a feature selection algorithm, namely a Combined Power Predictive Scores (CPPS). Unlike classical PSS approaches that consider the contribution of individual features, the CPPS allows to select a combination of features considering their predictive power with respect to a group of target variables. In particular, the CPPS is based on the error of a neural model that regresses the target variables by considering a combination of the input features.

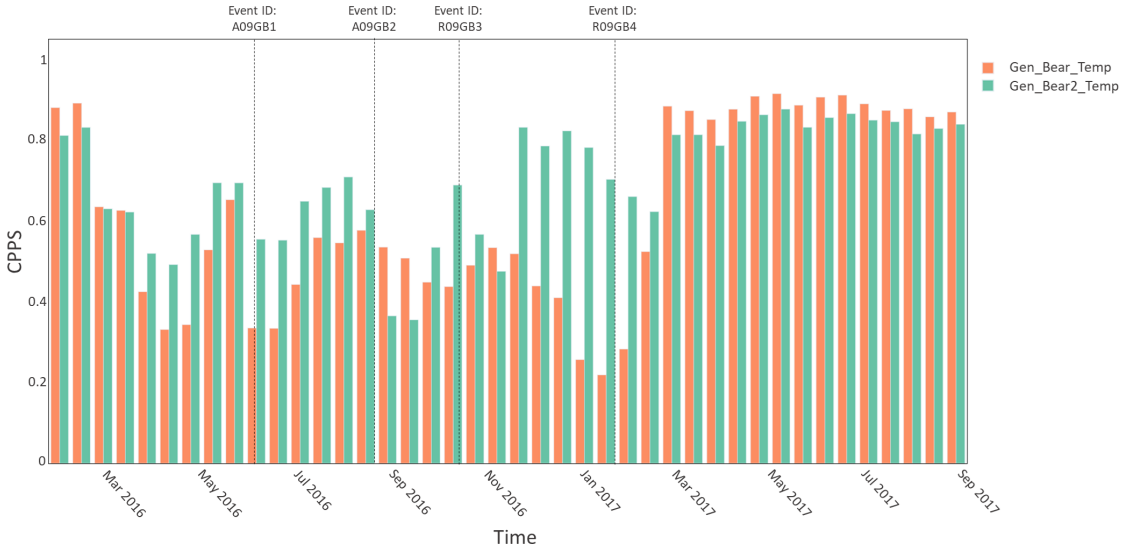

Figure 6. Bar chart of the CPPS scores obtained for the two target variables Gen_Bear_Temp and Gen_Bear2_Temp from 1 January, 2016 to 1 September 2017. The plot also shows the occurrences of the four events as reported in the maintenance log.

We proposed a dynamic application of this approach, where an MLP network is trained on groups of observations covering a window lenght of one month with an overlap of two weeks.
The framework has been tested on SCADA data sampled from a 2MW WT installed in an off-site wind farm. The dataset covers a total of 30 parameters monitored for 20 months with a sampling rate of 10 min. The model inputs are represented by 20 homogeneous groups of signals plus 10 independent signals.
The results showed that, from the analysis of the evolution of the CPPS over time, it is possible to identify a significant reduction in the scores linked to the two generator bearings as early as 8 months before a repair intervention on these components reported in the maintenance log. This intervention was then followed about 3 months later by the complete replacement of the damaged bearings. A control chart (e.g. based on a 3-sigma rule) can be defined to automate the identification of anomalies and to signal an alarm when a reduction in scores on a specific target variable occurs.
The model outperforms another recent deep anomaly detetcion approach based on an MTGCAE model applied to the same dataset, where the advance obtained on the same event is about 18 days. Moreover, the proposed framework resolves the main problems typically associated with ML-based approaches, because it does not require a preliminary isolation of the healty conditions for training the reference model, and also simplifies the definition of the control charts since they are applied to the model scores rather than to the typically noisy and seasonal signals.
Since the model is unsupervised and completely data-driven, we expect it to be independent of the specific use case and potentially applicable to any WT equipped with a SCADA system.

References

[1] EU An. Strategy to harness the potential of offshore renewable energy for a climate neutral future. *European Commission: Brussels, Belgium*, 2020.
[2] Laura Serri, Lisa Colle, Bruno Vitali, and Tullia Bonomi. Floating offshore wind farms in italy beyond 2030

and beyond 2060: Preliminary results of a techno-economic assessment. *Applied Sciences*, 10(24):8899, 2020.

[3] Marco Liserre, Roberto Cardenas, Marta Molinas, and Jose Rodriguez. Overview of multi-mw wind turbines and wind parks. *IEEE Transactions on Industrial Electronics*, 58(4):1081–1095, 2011.

[4] K. T. Fung, R. L. Scheffler, and J. Stolpe. Wind energy - a utility perspective. *IEEE Transactions on Power Apparatus and Systems*, PAS-100(3):1176–1182, 1981.

[5] Ezio Sesto and Claudio Casale. Exploitation of wind as an energy source to meet the world's electricity demand. *Journal of Wind Engineering and Industrial Aerodynamics*, 74:375–387, 1998.

[6] Hui-Ming Wee, Wen-Hsiung Yang, Chao-Wu Chou, and Marivic V Padilan. Renewable energy supply chains, performance, application barriers, and strategies for further development. *Renewable and Sustainable Energy Reviews*, 16(8):5451–5465, 2012.

[7] Wind Europe. Our energy our future: How offshore wind will help europe go carbon-neutral. *Wind Eur., Brussels, Belgium, Tech. Rep*, 2019.

[8] Ministry of Economic Development, Ministry of the Environment, Protection of Natural Resources, the Sea, Ministry of Infrastructure, and Transport. Integrated national energy and climate plan, italy, 2019.

[9] F Papi and A Bianchini. Technical challenges in floating offshore wind turbine upscaling: A critical analysis based on the nrel 5 mw and iea 15 mw reference turbines. *Renewable and Sustainable Energy Reviews*, 162:112489, 2022.

[10] Xiaojia Kong, Tongle Xu, Junqing Ji, Fanghao Zou, Wei Yuan, and Leian Zhang. Wind turbine bearing incipient fault diagnosis based on adaptive exponential wavelet threshold function with improved cpso. *IEEE Access*, 9:122457–122473, 2021.

[11] Zepeng Liu and Long Zhang. A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings. *Measurement*, 149:107002, 2020.

[12] Motor Reliability Working Group et al. Report of large motor reliability survey of industrial and commercial installations, part i. *IEEE Trans. Industrial Applications*, 1(4):865–872, 1985.

[13] F Oyague, D Gorman, and S Sheng. Nrel gearbox reliability collaborative experimental data overview and analysis: Preprint.

[14] Adrian Stetco, Fateme Dinmohammadi, Xingyu Zhao, Valentin Robu, David Flynn, Mike Barnes, John Keane, and Goran Nenadic. Machine learning methods for wind turbine condition monitoring: A review. *Renewable energy*, 133:620–635, 2019.

[15] Z Hameed, YS Hong, YM Cho, SH Ahn, and CK Song. Condition monitoring and fault detection of wind turbines and related algorithms: A review. *Renewable and Sustainable energy reviews*, 13(1):1–39, 2009.

[16] ASAE Zaher, SDJ McArthur, DG Infield, and Y Patel. Online wind turbine fault detection through automated scada data analysis. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, 12(6):574–593, 2009.

[17] Alexis Lebranchu, Sylvie Charbonnier, Christophe Bérenguer, and Frédéric Prévost. A combined mono-and multi-turbine approach for fault indicator synthesis and wind turbine monitoring using scada data. *ISA transactions*, 87:272–281, 2019.

[18] Diogo Menezes, Mateus Mendes, Jorge Alexandre Almeida, and Torres Farinha. Wind farm and resource datasets: A comprehensive survey and overview. *Energies*, 13(18):4702, 2020.

[19] Jorge Maldonado-Correa, Sergio Martín-Martínez, Estefanía Artigao, and Emilio Gómez-Lázaro. Using scada data for wind turbine condition monitoring: A systematic literature review. *Energies*, 13(12):3132, 2020.

[20] Andrew Kusiak and Anoop Verma. A data-driven approach for monitoring blade pitch faults in wind turbines. *IEEE Transactions on Sustainable Energy*, 2(1):87–96, 2010.

[21] Andrew Kusiak and Anoop Verma. Analyzing bearing faults in wind turbines: A data-mining approach. *Renewable Energy*, 48:110–116, 2012.

[22] Meik Schlechtingen and Ilmar Ferreira Santos. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mechanical systems and signal processing*, 25(5):1849–1875, 2011.

[23] Eric Stefan Miele, Fabrizio Bonacina, Alessandro Corsini, Arianna Peruch, Marco Cannarozzo, Daniele Baldan, and Francesco Pennisi. Unsupervised feature selection of multi-sensor scada data in horizontal axis wind turbine condition monitoring.

[24] Eric Stefan Miele, Fabrizio Bonacina, and Alessandro Corsini. Deep anomaly detection in horizontal axis wind turbines using graph convolutional autoencoders for multivariate time series. *Energy and AI*, 8:100145, 2022.

[25] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.

[26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer,

R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[27] Florian Wetschoreck, Tobias Krabel, and Surya Krishnamurthy. 8080labs/ppscore: zenodo release, October 2020.

[28] Xindong Wu, Vipin Kumar, J Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J McLachlan, Angus Ng, Bing Liu, S Yu Philip, et al. Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37, 2008.

[29] Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

[30] EDP. Wind farm 1 - failures, 2016. Data retrieved from EDP Open Data, https://opendata.edp.com/explore/dataset/htw-failures-2016/information/.

[31] Peter Hauge Madsen and DTU Risø. Introduction to the iec 61400-1 standard. *Risø National Laboratory, Technical University of Denmark*, 2008.

[32] Lin Bi, Yue Wang, Jian-ping Zhao, Hui Qi, and Ying Zhang. Social network information visualization based on fruchterman reingold layout algorithm. In *2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA)*, pages 270–273. IEEE, 2018.