A Scalable Predictive Maintenance Model for Detecting Wind Turbine Component Failures Based on SCADA Data

Lorenzo Gigoni, Alessandro Betti *i-EM S.r.l*Livorno, Italy
lorenzo.gigoni@i-em.eu Mauro Tucci, Emanuele Crisostomi

University of Pisa

Pisa, Italy

mauro.tucci@unipi.it

Abstract—In this work, a novel predictive maintenance system is presented and applied to the main components of wind turbines. The proposed model is based on machine learning and statistical process control tools applied to SCADA (Supervisory Control And Data Acquisition) data of critical components. The test campaign was divided into two stages: a first two years long offline test, and a second one year long real-time test. The offline test used historical faults from six wind farms located in Italy and Romania, corresponding to a total of 150 wind turbines and an overall installed nominal power of 283 MW. The results demonstrate outstanding capabilities of anomaly prediction up to 2 months before device unscheduled downtime. Furthermore, the real-time 12-months test confirms the ability of the proposed system to detect several anomalies, therefore allowing the operators to identify the root causes, and to schedule maintenance actions before reaching a catastrophic stage.

Index Terms—Predictive Maintenance, SCADA, Wind Turbine

I. INTRODUCTION

Wind energy stands out nowadays as one of the most promising alternatives to conventional dispatchable energy sources. However, unscheduled downtime and components replacements represent an urgent issue to be addressed in order to mitigate the impact of operation and maintenance (O&M) costs over generation, and maintain wind energy an attractive and competitive choice [1]. In this context, SCADA-based Condition Monitoring [2], [3] recently emerged as a promising solution to shift from an expensive reactive maintenance strategy to a pre-emptive, predictive one. Traditional approaches include bivariate analysis based on power curve modeling [4] and condition parameter-based models [5], [6]. The latter approach, roughly speaking, consist in training over healthy wind turbine (WT) data samples and predicting the status of the WT by monitoring residuals between the forecasted and the measured parameters. However, while condition parameterbased models give some information of specific components affecting the WT operation, both such traditional methods fail to provide a comprehensive picture of the correlations among the component parameters. This work aims to overcome the drawbacks of the aforementioned methodologies; namely, the main objective is to accurately predict anomalies for the three most critical WT components (gearbox, generator bearing and main bearing) within a multivariate framework, with the final goal of leading to more informed maintenance decisions. The model exploits signals related to the energy conversion system and specific components. The proposed approach provides many novel contributions:

- (i) machine learning techniques and process control tools, i.e. the residual Hotelling T2 control chart [7], [8], are wisely combined;
- (ii) an innovative multivariate outliers removal (MOR) method, based on k-means clustering, is proposed and applied to eliminate the abnormal samples from training instances (the proposed MOR approach generalizes the procedures already presented in [4], [9]);
- (iii) the status of the components is monitored by means of an original probabilistic formula defining a Key Performance Indicator (KPI): warnings of different severity are triggered based on threshold crossing rules;
- (iv) the Plug and Play nature of the presented approach, i.e. the fast service scalability on wind farms of increasing size, is also an added value of this work;
- (v) the proposed systems was realised and tested on a large number of wind turbines in different farms and geographical areas, during three years of operation.

II. MODEL DESCRIPTION

The algorithmic core of the model is represented by the combination of an Auto-Associative Neural Network (AANN), used for achieving normally distributed signals, and a variant of the multivariate Hotelling T2 control chart, which supervises a specific subset of SCADA tags for each WT component (Fig. 1). The model is fed with signals related to both the energy conversion system (e.g. active power, wind speed, shafts speed, shaft torque, etc.) and component-specific (e.g. gear box oil temperature, gear bearing temperature, etc.). The subset of component tags was selected from the list of available plants signals (originally of the order of hundreds) by combining specific domain knowledge, as well as feature selection strategies.

The proposed multivariate control chart is built using historical data that correspond to healthy operation periods, according to on-site information. The historical dataset of multivariate instances undergoes several preprocessing operations, namely:

- (i) outliers are removed by means of a bivariate power curvebased approach;
- (ii) data samples are clustered using k-means clustering method, exploiting the classic Euclidean distance metric to group samples into clusters;
- (iii) component temperatures are seasonally adjusted by taking advantage of a least-squares fitting against outdoor temperature in the low load operative regime.

This operation allows to extend the lead time before incipient faults are predicted and reduces the rate of false positives.

The pre-processed data is then used to train an AANN, which enforces an information compression on the hidden layer, and transforms data to be roughly normally distributed, which is known to be a desired feature for the subsequent T2 approach [9]–[11]. The multivariate output samples from the AANN are finally fed to the control chart, which detects changes in the underlying non-linear dynamics of the system by means of an original KPI formula. Such KPI, based on a probabilistic approach, monitors the population size housing in different T2 distribution sub-regions and penalizes deviations from the normal behavior, which corresponds to KPI values close to 1. Warning levels of different severity are finally triggered according to threshold crossing rules and shown

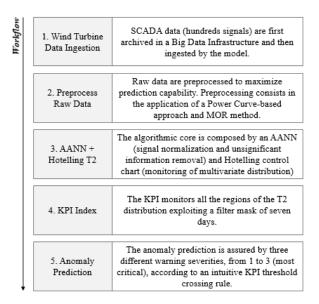


Fig. 1: Predictive Maintenance Model Workflow. The model is composed by five main steps. First, SCADA data are imported (1) then the dataset is preprocessed to maximize the model prediction capability (2). Later signals are transformed in normally-distributed tags and monitored by the Hotelling T2 chart (3), from which a KPI is finally computed (4) to uncover predictive insights(5)

on a Business Intelligence Analytics dashboard. The model workflow is shown in Fig.1.

III. RESULTS AND DISCUSSION

The model was tested on six wind farms located in Italy and Romania corresponding to four different WT manufacturers and two different sizes (1,5 MW and 2 MW), see Fig. 2. Globally, the model was evaluated over 150 wind turbines and 283 MW of nominal power. The positions of the wind farms are shown in Fig. 2. The large amount of SCADA data (Historical and Real-Time) have been archived and preliminary preprocessed by means of the Big Data infrastructure Microsoft Data Lake Analytics, before feeding the model.

A. Historical Validation

The software was trained and validated on historical data (2015-2016) using case studies (Tab. I) provided by plant operators. The obtained results highlighted an outstanding prediction strength of bearing failures from 2 weeks up to 1-2 months ahead, in good agreement with other research papers [2], [6].

Fig. 3 shows the output of the supervision model for the generator bearing of a Wind Turbine in Italy. On the left the KPI is shown as a function of time (blue curve), whereas on the right warnings of different severities (yellow, orange and red in ascending order of severity) are drawn, corresponding to different KPI threshold crossings. This clearly demonstrates the predictive power of the proposed KPI as explained below.



Fig. 2: Position of the 6 WFs, of which 4 are located in Southern Italy and Sardegna, and the remaining 2 in Romania. Globally, the model was evaluated over150 wind turbines and an overall nominal power of 283 MW. In particular, the bubble size in figure is proportional to the wind farm size.

TABLE I: Historical Cases exploited for model tuning and validation.

	Wind Fa	Anomalous Cases		
#	No. WT	WT Power	Manufacturer	Delivered
WF1	34	1.5	A	-
WF2	9	2.0	В	-
WF3	9	2.0	В	-
WF4	28	2.0	В	3 Gearbox
				5 Generator
WF5	35	2.0	С	1 Generator
WF6	35	2.0	D	3 Generator

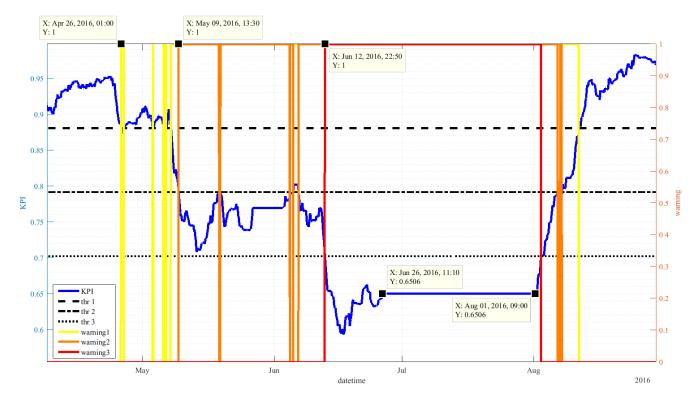


Fig. 3: Historical Validation: Supervision model output for the generator bearings of a Wind Turbine in Italy. On the left (right) the KPI (warnings of different severities) are shown as a function of time. The model early detected a generator bearing issue on 26th of April, i.e. at least 2 months before the operators detection. This problem involved a 1 month outage, which corresponds to the flat portion of the KPI in figure. In addition to the first warning level triggered on April, the model detected two further deviations from nominal condition: on the 9th of May it triggered warning level 2 and then on the 12th June it triggered the most severe warning (warning 3).

In particular, the aforementioned WT suffered a generator bearing fault during the year 2016. At that time no predictive strategy was applied, and the issue was detected by plant operators on the 26th of June 2016 and involved a device outage which lasted for 36 days (up to 1st August), as shown by the flat portion of the blue curve in Fig. 3. The ex-post application of supervision model anticipates the plant operator detection of about two months. Indeed the model triggers a first level warning on the 26th of April, where KPI falls down crossing warning level 1 threshold (dashed line). After fifteen days, the model detects a further deviation from the nominal

TABLE II: Anomalous cases delivered during Real-Time stage.

Wind Farm Characteristics				Anomalous Cases
#	No. WT	WT Power	Manufacturer	Delivered
WF1	34	1.5	A	2 Gearbox
				4 Generator
WF2	9	2.0	В	1 Generator
WF3	9	2.0	В	No Cases Delivered
WF4	28	2.0	В	2 Generator
				2 Main Bearings
WF5	35	2.0	С	3 Gearbox
				1 Generator
WF6	35	2.0	D	10 Gearbox

behavior of the generator bearing, which corresponds to the KPI crossing the warning level 2 threshold (dash-dot line) on the 9th of May. Finally, two weeks before the outage (i.e. 12th June) the model triggers the most severe warning (warning 3), corresponding to the most critical device condition. After maintenance intervention on the generator bearing, the KPI comes back to nominal condition and the model restarts to monitor the status of the device. Table I summarizes the characteristics of the six wind farms (number of WTs and their nominal power) and shows the anomalies detected during the offline campaign. The offline validation campaign has therefore demonstrated outstanding capabilities of early stage detection of WT main components anomalies, which basically corresponds, in the aforementioned case, to a prediction of generator bearing faults.

B. Real-Time Phase

Downstream of the the model tuning and validation campaign, the predictive service was installed on-site and tested on the year 2017, in order to verify its compatibility with plant operator activities. As a result, the model delivered 25 anomalous cases during one year of real time operation (Tab. II).

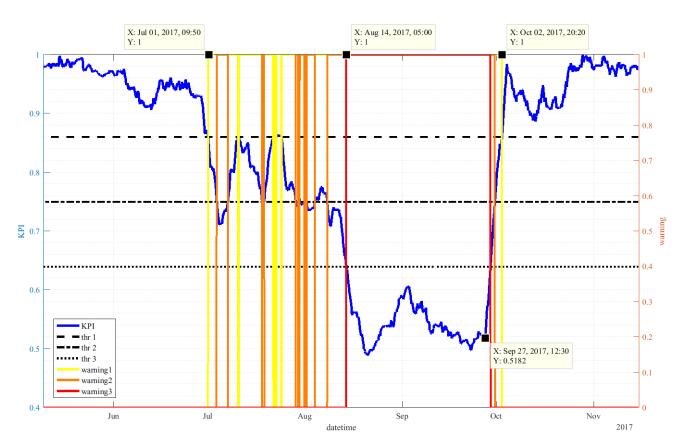


Fig. 4: Real-Time Phase: anomaly occurring at the gearbox of a WT in Romania. On the left (right) the KPI (warnings) are shown as a function of time. The model detected on the 1st of July an anomaly at the gearbox. The operators, thanks to the service provided, identified the root cause (gearbox cooling system issue) and scheduled maintenance activities avoiding further problem to the gearbox.

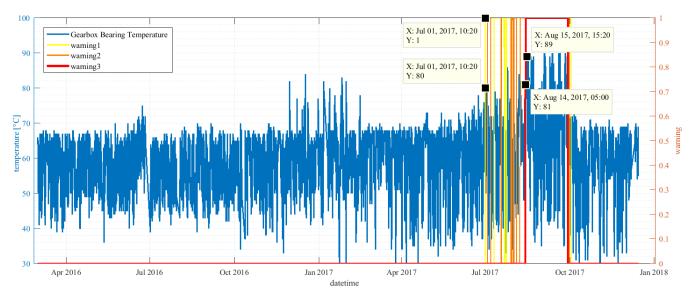


Fig. 5: Real-Time Phase: anomaly occurring at the gearbox of a WT in Romania. On the left (right) the gearbox bearing temperature (warnings) are shown as a function of time. On the 1st of July 2017 the model detected an anomaly at the gearbox, where the gearbox bearing temperature reached values over 80 C, rarely shown during the previous year. Later, the model triggered the most severe warning on the 14th of August, corresponding to a peak in the gearbox bearing temperature of almost 90 C.

Of these, we received a response from local plant operators for almost one half, and more than 90% of the latter corresponded really to true positives. The predicted were failures mainly related to gearbox, and in a smaller part to generator and main bearings (15, 8, and 2, respectively). Wind farm operators, exploiting model warnings, were able to apply efficient Predictive Maintenance strategies anticipating on-site actions from 2 weeks up to 1-2 months with respect to their traditional O&M activities. The capability of the model to reveal different fault classes was confirmed thanks to feedbacks received from local plant operators. In fact, the model also revealed cooling system issues, such as dirt accumulation, and the breaking of the engine fan, affecting generator and gearbox bearings. Furthermore, the model was capable of early detecting incipient heat dissipation issues after grease replacement in the generator bearings.

Fig. 4 shows an example of an anomaly detected by the model and occurring at a gearbox component of a WT located in Romania. During July 2017, the WT suffered an issue at the gearbox cooling system. In particular, the dirt accumulation involved an inefficient heat dissipation. On the 1st July the model early detected an abnormal gearbox temperature and triggered the first level warning to the wind farm operators. The operators benefited from the real-time service, receiving continuous feedback of the anomaly by monitoring the KPI time series, which show further deviations after the 1st of July, up to triggering the most severe warning on the 14th of August. Finally, they identified the root cause of the anomaly and scheduled the maintenance activities at the end of September, avoiding any further problems to the gearbox, which restored its full operation after the 2nd of October (Fig. 4). A qualitative analysis of the model output for the case described above is shown in Fig. 5, where the gearbox bearing temperature (blue curve, left axis) and warnings (right axis) are shown as a function of time. In correspondence of warning level 1, the gearbox bearing temperature reaches over 80 C, rarely seen during the previous year. Later, when the model triggered the most severe warning on the 14th of August, the gearbox bearing arrived at abnormal temperatures of 90 C or higher, well above the nominal operational temperatures.

The Real-Time service confirmed therefore the validity of the proposed approach and its contribution to extend the component lifetimes, as well as to move from a traditional reactive or time-based maintenance activity towards a predictive maintenance strategy.

IV. CONCLUSION

In this paper a novel predictive maintenance model for WT sub-components failures was presented, based on the integration of Machine Learning and statistical process control tools, and applying a multivariate approach at all the processing levels. The model is fed with SCADA tags and is composed mainly by four processing phases: (i) an outliers removal phase based on power curve modeling and k-means clustering, (ii) a seasonally adjusting procedure for temperature signals, (iii) an AANN-based residual approach to normalize tags, (iv) and

finally a residual control chart based on a novel KPI formula to trigger components warnings. Results from a vast campaign on 150 wind turbines demonstrate the ability of the proposed methodology to predict failures at gearbox, generator and main bearing levels, with lead times up to 1-2 months, and ensuring early detection at least.

The availability of a large amount of data from different spatial, temporal and technological points of view, and the fast service deployment on new wind farms represent, at the best of our knowledge, a distinct feature of the proposed system with respect to other published studies.

REFERENCES

- [1] D. Milborrow, Operation and maintenance costs compared and revealed, Wind Stats 19 (3), pp. 3 (2006)
- [2] D. Coronado, K. Fischer, Condition Monitoring of Wind Turbine: State of Art, User Experience and recommendations, Fraunhofer Institute for Wind Energy and Energy System Technology IWES Northwest (2015)
- [3] A. Kusiak, Z. Zhang, A. Verma, Prediction, operations, and condition monitoring in wind energy, Energy 60, pp. 1-12 (2013)
- [4] A. Kusiak, Monitoring Wind Farms with Performances Curves, IEE Transactions of Sustainable Energy 4, pp. 192-199 (2013)
- [5] P. Sun, J. Li, C. Wang, X. Lei, A Generalized Model for Wind Turbine Anomaly Identification Based on SCADA data, Applied Energy 168, pp. 550-567 (2016)
- [6] M. Schlechtingen, I. Santos, Comparative Analysis of Neural Network and Regression Based Condition Monitoring Approaches for Wind Turbine Fault Detection, Mechanical System and Signal Processing 25, pp. 1849-75 (2010)
- [7] Process or Product Monitoring and Control, Engineering Statistics Handbook, www.itl.nist.gov/div898/handbook/pmc/pmc.htm
- [8] A.T. Muzalwana, M. Susila., The application of Hotelling's T2 Control Chart in an Automotive Stamped Parts Manufacturing Plant, Statistics in Research, University of Malaya Press, ISBN 978-983-100-847-8 (2015)
- [9] H. Yang, M. Huang, S. Yang, Integrating Auto-Associative Neural Networks with Hotelling T2 Control Charts for Wind Turbine Fault Detection, Energies 8, pp. 12100-12115 (2015)
- [10] M. A. Kramer, Nonlinear Principal Component Analysis Using Autoassociative Neural Networks, AIChE Journal 37, No. 2, pp. 233-243 (1991)
- [11] H. Yang, M. Huang, P. Huang, Detection of Wind Turbine Faults Using a Data Mining Approach, Journal of Energy Engineering 142 (3), pp. 04015021:1-7 (2016)