

Wind Turbine Fault Forensic Analysis of SCADA Data Using Machine Learning Techniques

Juan José Cárdenas^{2*}, Alejandro Blanco^{1,2} M., Jordi Solé-Casals¹, Pere Marti-Puig¹, Isaac Justicia², Jordi Cusidó²

¹ *Universitat de Vic - Universitat Central de Catalunya, Vic, Spain*

² *ITESTIT S.L., Terrassa, Spain*

* *Presenting author*

Abstract

Methodologies and tools that can support the task of **finding out the possible causes of a fault manifested by a specific alarm or a set of alarms** can benefit wind farm owners to increase availability and production and reduce costs. On the other hand, data availability from **SCADA of the wind park has a great potential of information that can support this specific task**, more when it is already available and using this data for fault diagnosis does not require any type of extra implementation or hardware installation in the wind turbine. However, due to the high number of available variables and data, analysing them can be a high time consuming task and when just well-known related variables are analysed hidden causes or not common causes cannot be or are hard to be found. For all these reasons, in this work, we present a **methodology and tool, part of Smartive platform**, that been fed by all available SCADA data and other source of information, **can support fault forensic analysis in order to find the main causes of faults.**

Introduction

One of the main tasks in O&M process is to find out the possible causes of a fault manifested by a specific alarm or a set of alarms that stops the wind turbine production [1]. This process is crucial to reduce time of repair or detect more critical faults in earlier stages. Methodologies and tools that can support this type of process can benefit wind farm owners to increase availability and production and reduce costs [2]. On the other hand, data availability from SCADA of the wind park has a great potential of information that can support this specific task, more when it is already available and using this data for fault diagnosis does not require any type of extra implementation or hardware installation in the wind turbine [3]. However, due to the high number of available variables and data, analysing them can be a high time consuming task and when just well-known related variables are analysed hidden causes or not common causes cannot be or are hard to be found. For all these reasons, in this work, we present a methodology and tool, part of **Smartive platform**, that been fed by all available SCADA data and other source of information, can support fault forensic analysis in order to find the main causes of faults.

Objectives

Improve efficiency of O&M tasks by mean of **“forensic” analysis** of SCADA data and alarm records in order to find the main “culprits” of failures and therefore:

- Reduce WTG down-time.
- Detect more critical faults in earlier stages.

For that we need:

- Find the **main related variable** with the fault in the current day of the failure and in previous days before failure.
- Build a **visual analysis** of the found relationships and depict the appropriated “history” that can explain the observations.
- Based on these results **plan the O&M tasks**

Methodology

The proposed methodology is based on statistical and machine learning techniques in order to find out the most related variables with the fault under analysis. Therefore there is a parallel process starting from the same alarms dataset which is split into two lanes. The first lane, Hypothesis testing, evaluates all the input variables in order to determine which variables changes its behaviour when an alarm is present of those which not, determining statistically which variable the expert must take into account.

The second lane, Predictive Model, implements a predictive model with a bench of classifiers algorithms each of those generates a variable importance metric which is described on caret library [5].

Finally, human expert analyses the results from the hypothesis testing with the variables that have change its behaviour prior a failure and compares with the most relevant variables reported at the model creation, merging both results he is able to find a new explanations of a past failure validated by a data analysis.

Methods

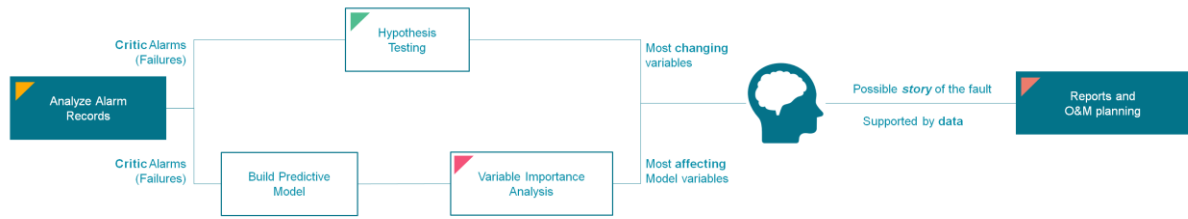


Figure 1: General Overview

Analyse Alarm Records

The methodology is supported by the analysis of the historical alarms; this involves the knowledge of a Human expert that knows how the turbine's power chain and mechanical parts works. This is because some alarms can produce another's alarms, as for example a low power alarm because grid degradation could produce a pitch failure alarm because the pitch engines cannot move accurately the blades due the sporadic power lost; this produces false alarms reports from the sensors about invalid blades position. So, the expert classifies the alarm hierarchy and determines each group of alarms for the analysis of one system.

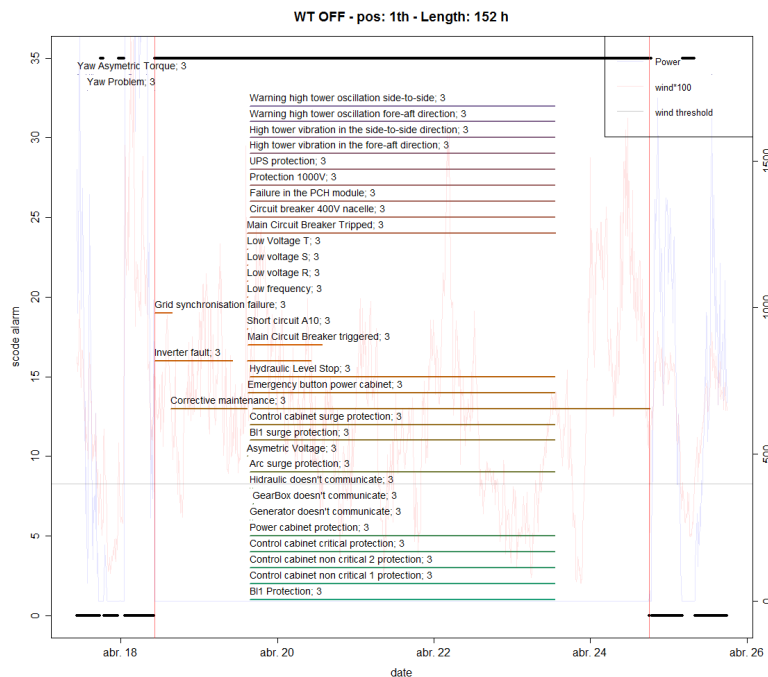


Figure 2: Alarm hierarchy, selection and grouping

Hypothesis testing

In order to choose the relevant variables a simply approach was selected. The hypothesis considers that when there is an alarm the variables doesn't change its values; this implies that the distribution of the average readings when there are alarms is the same as no alarm state. Then a test is done with p-value of $p=0.05$, only the variables that in alarm condition have extreme average values and are very probable will be selected [1].

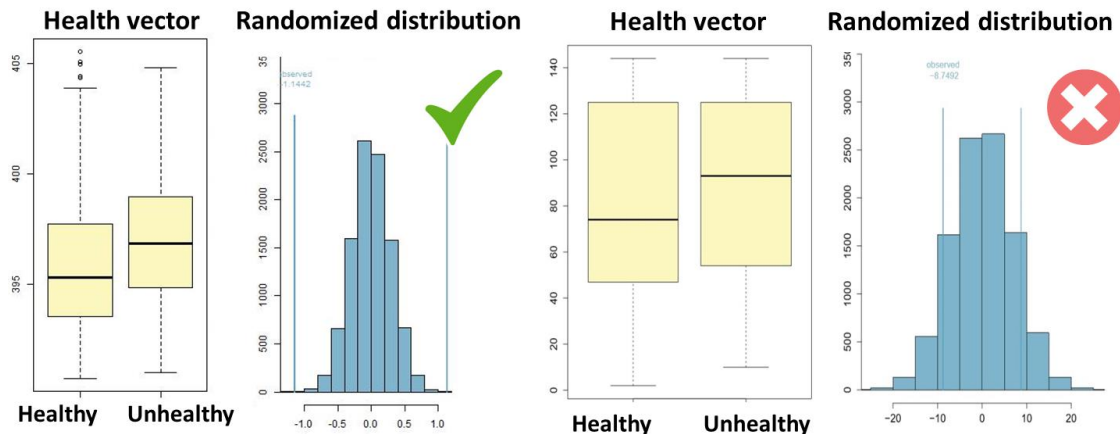


Figure 3: Hypothesis testing, determine if the variable behaviour is random or not.

Variable Importance analysis

The variable importance is calculated at each prediction algorithm (Naïve Bayes, decision tree, least squares...) [1, 4]. It is based on the ROC, measuring the area under the curve (AUC). The algorithms uses all the possible variables one time to determine a reference AUC, based on this reference, the algorithms runs exhaustively doing all the possible permutations of the input variables in order to measure the impact on AUC and evaluate with a score from 0 to 100.

Wind turbine and fault: e100_1_yaw2 Best time windows: 1

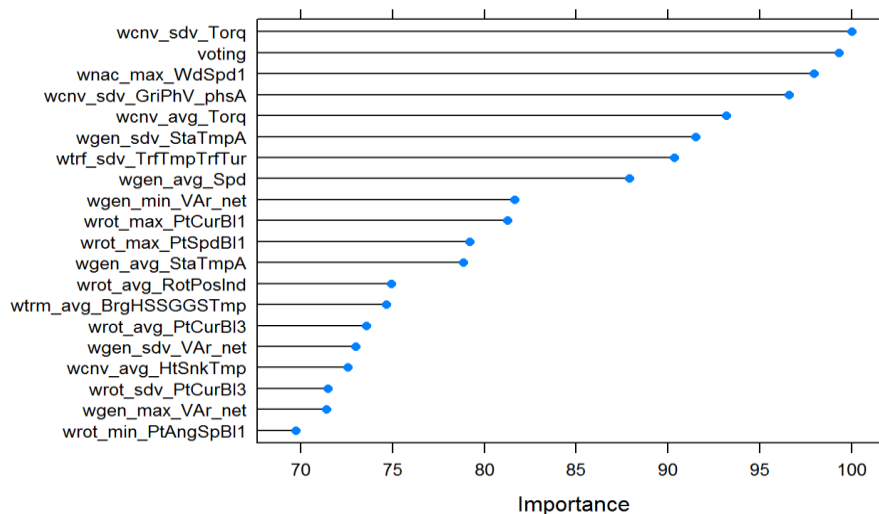


Figure 4: Importance results variable for the YAW alarms (grid stability problems)

Results

The results obtained with this methodology were evaluated by an expert on the field. This expert refused the result of alarms at yaw motors where reflected on grid frequency and voltage variations variables, but with further analysis with information of the grid stability of the historical data, the expert accepted the results from the variable importance because he found a reasoning explanation, once the turbines start producing and this implies deliver high power to the grid on the wind plant where are connected , the grid cannot absorb these differences and is reflected on others wind turbines that are still moving to start producing.

- Unexpected culprit: torque unbalance in yaw motors due to grid instability

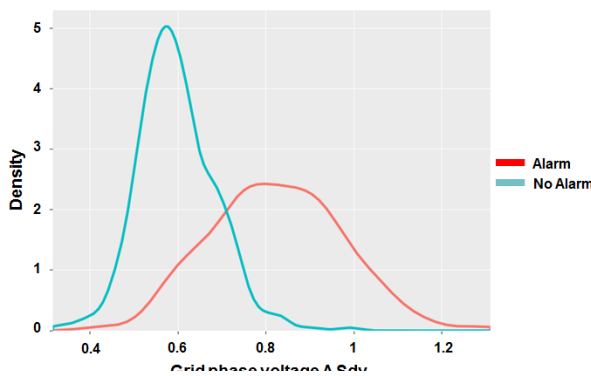


Figure 6: Grid phase instability healthy vs alarm

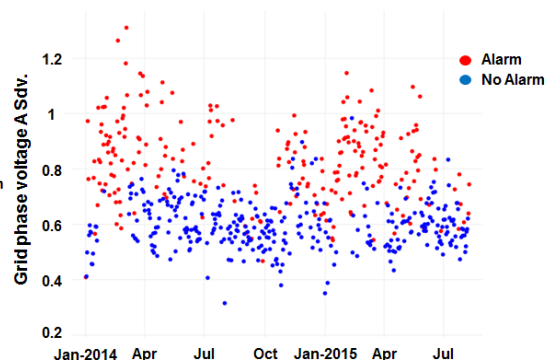


Figure 5: Sdv. of grid, alarm and healthy state.

- Confirmation of possible culprit in Gear Box fault: possible cause – mechanical stress of the Wind Turbine because high intensity production

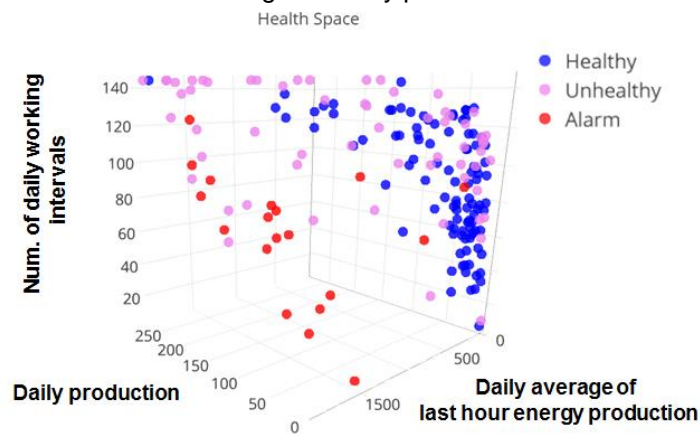
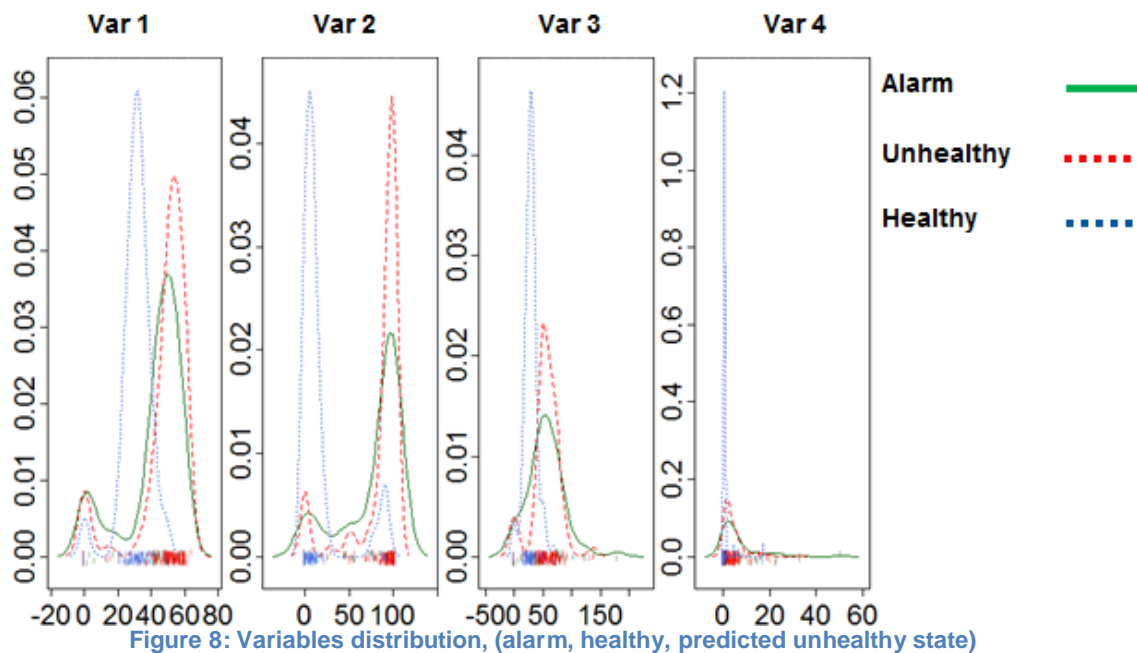


Figure 7: High mechanical stress alarms have a direct relation with the production

The following graph shows the difference of a set of four variables that the states are clearly separated. The curves in green shows its values once there is a real alarm, the blue dashed one when the machine is working properly and finally the unhealthy state that is the prediction made by the classifiers (only the best one).

It's clear that in **Var 1** and **Var 2** the difference is noticeable, so pure statistical methods like hypothesis-testing will do good job here, but the case of **Var 3** and **Var 4** is not as clear, so another methods have to be used, like this case classification methods.



Conclusions

It has selected and assessed a set of algorithms for feature selection in machine learning classifiers when used for wind-turbine failure prediction. The best of them have been evaluated versus an exhaustive method.

The results reveals that the variables have enough information combined that makes possible a pre-selection based of mutual information between a subset of them defining a specifically alarm as target. Also, there are evidences that for a large number of inputs, in this case six or more, search exhaustively all the possible combinations of variables versus use a suboptimal solution based on feature selection algorithms will deliver almost the same performance with less computation time.

References

1. Cárdenas, Juan J. et al. 2016. "Cloud-based Platform for Wind Turbine's Health Estimation and Failure Prediction." *EWEA Workshop 2016: Analysis of Operating Wind Farms 2016 – 3rd edition, Bilbao-Spain*, 14-15 April 2016.
2. Coronado, Diego and Fisher, Katharina, "Condition Monitoring of Wind Turbines: State of the Art, User Experience and Recommendations" *Fraunhofer Institute for Wind Energy and Energy System Technology IWES*, Project Report, January 2015.
3. Tchakoua, Pierre et al. 2014. "Wind Turbine Condition Monitoring: State-of-the-Art Review, New Trends, and Future Challenges." *Energies* 7(4): 2595–2630. <http://www.mdpi.com/1996-1073/7/4/2595/htm>.
4. Schlechtingen, Meik, and I Santos. 2011. "Condition Monitoring with Ordinary Wind Turbine SCADA data—A Neuro-Fuzzy Approach." *Proceedings of the European Wind Energy Association (EWEA) Offshore*.
5. Kuhn, Max. "Building Predictive Models in R Using the Caret Package." *Journal of Statistical Software*. N.p., Nov. 2008. Web. 26 Sept. 2016. <https://www.jstatsoft.org/article/view/v028i05/v28i05.pdf>
6. Smartive Platform: <http://smartive.eu/>