Journal Paper

"Advanced analytics for detection and diagnosis of false alarms and faults: A real case study"

-Wind Energy-

July 2019

Alberto Pliego Marugan
CUNEF-Ingenium, Colegio Universitario de Estudios Financieros
Alberto.pliego@cunef.edu

Fausto Pedro García Márquez
Ingenium Research Group, Castilla-La Mancha University, Spain
FaustoPedro.Garcia@uclm.es

Cite as: Pliego Marugán, A., & García Márquez, F. P. (2019). Advanced analytics for detection and diagnosis of false alarms and faults: A real case study. Wind Energy.

DOI: https://doi.org/10.1002/we.2393

Advanced Analytics for Detection and Diagnosis of False Alarms and Faults: A Real Case Study

Alberto Pliego Marugán [1], Fausto Pedro García Márquez [2]

[1] CUNEF-Ingenium, Colegio Universitario de Estudios Financieros de Madrid, Madrid, Spain

[2] Ingenium Research Group, Universidad Castilla-La Mancha, 13071 Ciudad Real, Spain.

alberto.pliego@cunef.edu

faustopedro.garcia@uclm.es

Abstract

Onshore and offshore wind farms require a high level of advanced maintenance. Supervisory Control and Data Acquisition (SCADA) and condition monitoring systems are now being employed, generating large amounts of data. They require robust and flexible approaches to convert dataset into useful information. This paper presents a novel approach based on the correlations of SCADA variables to detect and identify faults and false alarms in wind turbines. A correlation matrix between all the SCADA variables is used for pattern recognition. A new method based on curve fittings is employed for detecting false alarms and abnormal behaviours or faults in the components. The study is done in a real case study, validated with false alarms.

Keywords: Analytics, Wind Turbine, Reliability, SCADA, False Alarms, Multivariable Analysis.

1 Introduction

Wind energy is the most developed renewable energy and it is in continuous growth. The cumulative global capacity rose from 3.5 GW in 1994 to more than 420 GW in 2016, and it is expected to reach 1000 GW in the 2030s (1)[1](Gómez Muñoz and García Márquez 2016)(1)[1](Gómez Muñoz and García Márquez 2016)[1](Gómez Muñoz and García Márquez, 2016).

This industry requires advanced maintenance strategies, both offshore and onshore. Operation and maintenance (O&M) tasks account for 12% of the total system costs for onshore wind farms, and up to 23% for offshore (2)[2](Tavner 2012)(2)[2](Tavner 2012)[2](Tavner, 2012). Offshore wind energy represents the most important growth in recent years. In 2014, the cumulative global offshore capacity was 8.8 GW, the European Union (EU) being the largest producer. More than $5 billion was invested by the EU in offshore windfarms in 2014 (1)[1](Gómez Muñoz and García Márquez 2016)(1)[1](Gómez Muñoz and García Márquez 2016)[1](Gómez Muñoz and García Márquez, 2016). The new technologies and information systems have led to improvements in design, production, installation and O&M tasks (3-5)[3-5](Jiménez et al. 2019b; Jiménez et al. 2019a; Gómez Muñoz et al. 2019)(3-5)[3-5](Jiménez, Muñoz, and Márquez 2019; Jiménez et al. 2019; Gómez Muñoz et al. 2019)[3-5](Jiménez et al., 2019b, Jiménez et al., 2019a, Gómez Muñoz et al., 2019).

Many research studies have been carried out to optimize these tasks for offshore (6)[6](Pliego Marugán et al. 2016)(6)[6](Pliego Marugán, García Márquez, and Pinar Pérez 2016)[6](Pliego Marugán et al., 2016) and onshore location (7)[7](Márquez et al. 2012)(7)[7](Márquez et al. 2012)[7](Márquez et al., 2012). The early detection of component faults enables the reduction of energy losses, downtimes, O&M costs and, consequently, the cost of energy (COE) (8)[8](Márquez et al. 2016)(8)[8](Márquez et al. 2016)[8](Márquez et al., 2016). New systems are being installed in Wind Turbines (WTs), e.g. condition monitoring systems (CMSs) or supervisory control and data acquisition (SCADA) (9)[9](Yang and Jiang 2013)(9)[9](Yang and Jiang 2013)[9](Yang and Jiang, 2013). The efficiency of these systems has been proved in several research studies, e.g. references (10-12)[10-12](McMillan and Ault 2007; Renewables 2007; Gómez Muñoz et al. 2018)(10-12)[10-12](McMillan and Ault 2007; Renewables 2007; Gómez Muñoz et al. 2018)[10-12](McMillan and Ault, 2007, Renewables, 2007, Gómez Muñoz et al., 2018).

SCADA systems can measure from a few tens to thousands of endogenous and exogenous variables. The data provided by the SCADA system could be classified into the following categories [13][13](Kusiak and Li 2011)(13)[13](Kusiak and Li 2011)[13](Kusiak and Li, 2011):

- *Weather parameters:* The parameters related to the wind, e.g. speed, intensity, direction, turbulences, etc., and other environmental parameters, e.g. pressure, humidity, temperature, etc. Additional parameters are collected when the SCADA system is in an offshore wind farm, e.g. salinity, wave height, etc.
- *Energy conversion parameters*: Parameters related to the energy conversion process, e.g. power output, voltage, pitch system, yaw system, etc.
- *Vibration parameters*: Vibrations from the drive train and tower.
- *Temperature parameters*: Temperatures of the turbine components and the air temperature close to the components.

There can be over 100 sensors on a single wind turbine. These sensors can generate data at the rate of 2000 values per minute resulting in about 4 Terabyte of data in one month [14][14](Kashyap 2014)(14)[14](Kashyap 2014)[14](Kashyap, 2014).

The growing number, size and complexity of the WTs, together with the SCADA, require advanced analytics for an optimal energy production. Many research studies are being carried out to improve WT reliability, availability, maintainability and safety (RAMS) by studying the influence of wind speed [15][15](Tavner et al. 2006)(15)[15](Tavner et al. 2006)[15](Tavner et al., 2006) or the gearboxes [16][16](Musial et al. 2007)(16)[16](Musial, Butterfield, and McNiff 2007)[16](Musial et al., 2007), estimating the performance from statistics [17][17](Billinton et al. 1996)(17)[17](Billinton, Chen, and Ghajar 1996)[17](Billinton et al., 1996), or using advanced algorithms such as wavelet transform [18,19][18,19](Gomez Munoz et al. 2014; de la Hermosa González et al. 2015)(18, 19)[18, 19](Gomez Munoz et al. 2014; de la Hermosa González, Márquez, and Dimlaye 2015)[18, 19](Gomez Munoz et al., 2014, de la Hermosa González et al., 2015).

A correct performance of the WT is essential to ensure the efficiency and the rentability of a wind farm. For this purpose, all the components and subsystems of the WT trend to be controlled. For instance, in the field of WT controlling, Kusiak et al. propose data mining, predictive control and evolutionary computation in a dynamic model [20][20](Kusiak et al. 2010)(20)[20](Kusiak, Li, and Song 2010)[20](Kusiak et al., 2010). Hand [21][21](Hand 2003)(21)[21](Hand 2003)[21](Hand, 2003) and Stol and Balas [22][22](Stol and Balas 2002)(22)[22](Stol and Balas 2002)[22](Stol and Balas, 2002) employ models for a disturbance accommodating control. Some studies present models that can analyze the variable speed of WTs with standard and adaptive algorithms by direct variable speed control (23,24)[23,24](Song et al. 2000; Johnson et al. 2006)(23, 24)[23, 24](Song, Dhinakaran, and Bao 2000; Johnson et al. 2006)[23, 24](Song et al., 2000, Johnson et al., 2006) or by pitch control (25,26)[25,26](Freeman and Balas 1999; Johnson and Fingersh 2008)(25, 26)[25, 26](Freeman and Balas 1999; Johnson and Fingersh 2008)[25, 26](Freeman and Balas, 1999, Johnson and Fingersh, 2008). New strategies are being used based on multivariable analysis from condition monitoring data (27,28)[27,28](Boukhezzar et al. 2007; Márquez and Muñoz 2012)(27, 28)[27, 28](Boukhezzar et al. 2007; Márquez and Muñoz 2012)[27, 28](Boukhezzar et al., 2007, Márquez and Muñoz, 2012) ; fuzzy logic based intelligent control (29)[29](Simoes et al. 1997)(29)[29](Simoes, Bose, and Spiegel 1997)[29](Simoes et al., 1997), (30)[30](Bououden et al. 2012)(30)[30](Bououden et al. 2012)[30](Bououden et al., 2012), dynamic control (31,32)[31,32](Muljadi et al. 1998; García Márquez et al. 2011)(31, 32)[31, 32](Muljadi, Pierce, and Migliore 1998; García Márquez et al. 2011)[31, 32](Muljadi et al., 1998, García Márquez et al., 2011) ; or non-linear models (33)[33](Boukhezzar and Siguerdidjane 2009)(33)[33](Boukhezzar and Siguerdidjane 2009)[33](Boukhezzar and Siguerdidjane, 2009), (34)[34](Kumar and Stol 2010)(34)[34](Kumar and Stol 2010)[34](Kumar and Stol, 2010). These strategies help to optimize electricity generation by establishing an adequate regulation of different systems, e.g. pitch, generator, yaw, etc. The SCADA systems and alarms studied by Qiu et al. (35)[35](Qiu et al. 2012)(35)[35](Qiu et al. 2012)[35](Qiu et al., 2012) by time-sequence and probability-based methods consider the alarms. The main contributions of this paper lie in the following aspects:

- This paper presents a new approach to determine the possibility of an error in the analysis of the data from SCADA system. There are many studies where fault prediction and diagnosis of WTs regarding the SCADA data are discussed, e.g. based on the frequency of occurrence of the failures (13)[13](Kusiak and Li 2011)(13)[13](Kusiak and Li 2011)[13](Kusiak and Li, 2011), using data mining algorithm to built time series models (36)[36](Kusiak et al. 2009)(36)[36](Kusiak, Song, and Zheng 2009)[36](Kusiak et al., 2009), studing the standard turbine performance parameters and the physics of the different failures (37)[37](Gray and Watson 2010)(37)[37](Gray and Watson 2010)[37](Gray and Watson, 2010), using graphical methods (1)[1](Gómez Muñoz and García Márquez 2016)(1)[1](Gómez Muñoz and García Márquez 2016)[1](Gómez Muñoz and García Márquez, 2016), or employing sensitivity analysis (38)[38](Muñoz et al. 2016)(38)[38](Muñoz, Márquez, and Tomás

2016)[38](Muñoz et al., 2016). However, these methods are not based on the multiple correlations that each parameter can have with rest of the parameter measured by the SCADA systems.

- The proposed methodology considers all the correlations between all the variables of the SCADA systems. Some of the research papers in the literature consider certain correlations between two variables, e.g. Yang and Jiang (9)[9](Yang and Jiang 2013)(9)[9](Yang and Jiang 2013)[9](Yang and Jiang, 2013) show the correlations windspeed-power, windspeed-temperature, power-temperature, generator speed- temperature; Gill et al. (39)[39](Gill et al. 2012)(39)[39](Gill, Stephen, and Galloway 2012)[39](Gill et al., 2012) studied the dependence between active energy and wind speed through power curve copula modelling; Igba et al. (40)[40](Igba et al. 2016)(40)[40](Igba et al. 2016)[40](Igba et al., 2016) studied the dependence between the vibration level and the power output. However, these studies do not use the complete correlation map obtained from the SCADA system.
- This paper addressed a validated method for identify a possible false alarm. There are some methods for fault detection that minimizes the probability of false alarms, e.g. counter-based residual thresholding (41)[41](Ozdemir et al. 2011)(41)[41](Ozdemir, Seiler, and Balas 2011)[41](Ozdemir et al., 2011), or provides statistical information about false alarms (42)[42](Chen et al. 2011)(42)[42](Chen et al. 2011)[42](Chen et al., 2011). However, the authors have not found a specific method to detect false alarms.

This paper is divided into 8 different sections. Section 1 is an introductory section where the context of wind energy is exposed, and the main contributions are highlighted. Section 2 details the different steps to follow to apply the proposed approach. Section 3 explains the different variables from the SCADA system that will be considered for the analysis. Section 4 presents the correlation analysis for the considered variables and section 5 determines the curve fitting models for some of these correlations. In section 6, the method is validated by comparing the estimated results with a real alarm report. Section 7 proposes a case study where the proposed method is applied and the different possible solutions are analyzed. Finally, in section 8 the main conclusions are extracted.

2 Dataset from the SCADA system

The SCADA system measure and store the parameters with a sampling rate of 10 minutes. Table I shows the 34 variables provided by the system. This table collects the ID number, the name and a brief description of each variable. It is important to remark that this database is provided by the Optimus European Project and, therefore, the measurements correspond to a specific SCADA system installed in a WT under specific conditions.

The SCADA system also provides a detailed alarm report. This information will be employed for validating the model proposed. The collected data are:

- Date of the alarms.
- Code of the alarm (confidential).
- Cause or description of the alarm.
- The state of the alarm (activation or deactivation).
- The severity.
- Required tasks.

Table I. Description and nomenclature of the SCADA variables.

Nº	Variable Name	Description
1	Gral accum. 1	They are used in hydraulic power units for blade pitch. In case of extreme wind, they set the blade to defence position. Each rotor blade has an independent actuator with an accumulator. The SCADA system measures the pressure of the accumulator of each blade.
2	Gral accum. 2	
3	Gral accum. 3	
4	Phi cosine	Power factor indicates the proportion of reactive power that is generated or consumed by the system. Depending on the type of generator used in the wind turbine, adjusting the power factor may require the installation of capacitors in the electrical system (43)[43](Chen et al. 2009)(43)[43](Chen, Guerrero, and Blaabjerg 2009)[43](Chen et al., 2009). The SCADA measures the power factor and the reactive power and provide the quantitative value of these parameters as an output.
5	Turbulence level	In this paper, a turbulence indicator is considered by evaluating the representative value of the turbulence standard deviation (σ_1), called turbulence level in this paper, obtained by the Normal Turbulence Model in IEC 61400-1 (44)[44](Carpman 2011)(44)[44](Carpman 2011)[44](Carpman, 2011): $$\sigma_1 = \frac{I_{15}(15 + a \cdot V_{hub})}{a + 1}$$ where I_{15} and a are tabulated parameter, and V_{hub} is the wind speed at hub height.
6	Oscillation level	The WT tower can be strongly affected by oscillations. They can be a symptom of faults. The oscillations can also affect other components such as the blades, and increase tower fatigue (45)[45](Knudsen et al. 2012)(45)[45](Knudsen, Bak, and Tabatabaeipour 2012)[45](Knudsen et al., 2012).
7	Vibration level	A sensor is installed into the nacelle for measuring the vibrations. The regulation of this parameter must be done under ISO 10816-21:2015 (4)[4](Jiménez et al. 2019a)(4)[4](Jiménez et al. 2019)[4](Jiménez et al., 2019a).
8	Pitch 1 angle	The largest WTs are usually equipped with blade pitch control. This mechanism has two main tasks: the first one is to adjust the pitch angle to control the speed of the rotor; the second task is to brake the rotor by setting the blades to the feathered position. The blade pitch system can be hydraulically or electrically driven (46)[46](Hau and von Renouard 2003)(46)[46](Hau and von Renouard 2003)[46](Hau and von Renouard, 2003).
9	Pitch 2 angle	
10	Pitch 3 angle	
11	Active power	Small variations in the wind speed can have a great effect on power generation. The electric power is measured before the grid to consider the consumption by the WT. Therefore, the effective power delivered to the grid is considered.
12	General accumulator pressure	The general accumulator provides the pressure to the hydraulic systems. The SCADA system measures the pressures of the hydraulic group and the general accumulator as the conditions of the hydraulic system.
13	Brake pressure	

14	Hydraulic group pressure	The objective of this system is to stop the rotor. Brakes must always be operative because they are one of the most important safety devices in the WT. The SCADA system measures the pressure of the hydraulic brake.
15	SP pitch angle	SP (System Parameter) is a measurement provided by the SCADA system to monitor the operation of the pitch system.
16	Hydraulic group oil temperature	The gearbox is one of the most critical and costly parts of a WT. The lubrication of the gearbox is essential. The analysis of this condition provides: lubricant condition, contaminants and machine wear (47)[47](Barrett and Stover 2013)(47)[47](Barrett and Stover 2013)[47](Barrett and Stover, 2013). The SCADA system measures the temperature of the gearbox and the hydraulic group oil.
17	Gearbox oil temperature	
18	Environmental temperature	Heat dissipation is dependent on the environmental temperature. There are some WTs that are subjected to extreme temperatures. Therefore, all the components, mainly electronics circuits, are designed to operate under these circumstances (48)[48](Burton et al. 2011)(48)[48](Burton et al. 2011)[48](Burton et al., 2011).
19	Drive end side generator bearing tª	The common bearing configuration is based on two bearings. The main shaft is supported by a self-aligning roller bearing (non-end side drive bearing), and the gearbox uses a cylindrical roller bearing (end side drive bearing) (49)[49](Yagi and NINOYU 2008)(49)[49](Yagi and NINOYU 2008)[49](Yagi and NINOYU, 2008). The temperature of the bearings must be controlled to improve reliability. The SCADA measures the temperature of both bearing systems.
20	Non-drive end side generator bearing tª	
21	Generator winding temperature	The generator windings must be monitored constantly. A fault in the insulation can cause production and cost losses. This temperature can be collected by sensors using a thermocouple (50)[50](Stone 2002)(50)[50](Stone 2002)[50](Stone, 2002).
22	Nacelle temperature	The nacelle is generally divided into a power cabinet and a control cabinet. The temperature of the nacelle is measured and transmitted to a central computer (46)[46](Hau and von Renouard 2003)(46)[46](Hau and von Renouard 2003)[46](Hau and von Renouard, 2003).
23	Lower gearbox radiator	The oil absorbs heat and it flows through the radiators to be cooled. A correct design and operation of the radiators is required to exchange heat and avoid possible critical failures due to overheating. The SCADA system measures the temperature of the gearbox radiators.
24	Upper gearbox radiator	
25	Gearbox bearing temperature	The temperature of the gearbox bearings must be controlled because a high temperature indicates excessive friction, i.e. it could be a fault in the bearings or cooling system.
26	Transformer 1 temperature	Transformer are usually located at the base of the tower for adapting, together with the power electronics, the electricity generated to the grid restrictions (51)[51](Jha 2010)(51)[51](Jha 2010)[51](Jha, 2010). The main purpose of the transformer is to increase the output voltage of the WT system. The SCADA system measures the temperature of the transformers.
27	Transformer 2 temperature	
28	Transformer 3 temperature	
29	Grid voltage	The WT in this case study is designed to generate 12 KV to the grid. The SCADA system measures the voltage injected to the grid and the possible fluctuations.

30	Total reactive power	See Variable 4
31	Generator speed	The generator studied is designed to work between 750- 1500 rpm. The gearbox adapts the speed of the high-speed shaft. The speed of the generator and the rotor are monitored by the SCADA system.
32	Rotor speed	
33	Wind speed	This is one of the most relevant factors for establishing the different modes of operation. The wind speed, the density of the air and the blades define the energy resource. The SCADA system uses an anemometer located on top of the nacelle for measuring the wind speed.
34	Yaw	This system ensures that the nacelle is oriented correctly. A yaw system can be controlled by hydraulic or electric motors (46)[46](Hau and von Renouard 2003)(46)[46](Hau and von Renouard 2003)[46](Hau and von Renouard, 2003). The SCADA system measures the position angle of the yaw system.

3 Proposed approach

SCADA systems have a set of logical control laws. The method proposed supports these laws to improve decisions and extract information that is not being considered. It verifies that the system does not generate false alarms, i.e. alarms generated by a fault in the SCADA system. The method also detects and identifies abnormal behaviours that the SCADA cannot identify. Figure 1 shows the flowchart of the novel method.

A real database from the OPTIMUS European Project has been employed (52)[52](project 2014)(52)[52](project 2014)[52](project, 2014). The database is composed of a set of parameter measurements and an alarm report. A Pearson correlation coefficient is employed to set the interrelationship between the different parameters. The parameters are placed in a dynamic correlation matrix. Different groups will be set to classify the significant correlations. This approach has not been found in literature.

Figure 1 shows the flowchart of the proposed method.

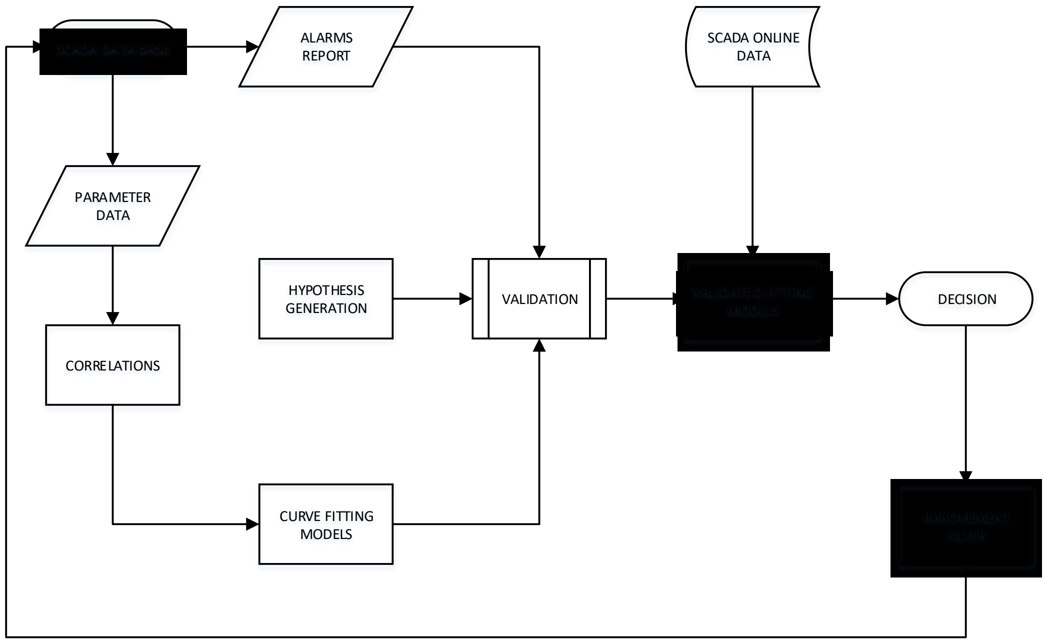

Figure 1. Approach flowchart

A real SCADA database is employed in this paper. This database has been provided by the consortium of the Optimus European Project (52)[52](project 2014)(52)[52](project 2014)[52](project, 2014). The database is composed of a set of parameters measurements and a report of the historical alarms. The measurements will be processed setting (Pearson) correlations between the different parameters.

Once the correlations are identified, fitting models are created for determining the goodness of the measurements. Upper and lower confidence bounds will be stablishing for placing the measurements into two different categories: correlations or miscorrelations. Many miscorrelations could indicate an abnormal condition.

The fitting models are validated by setting the relationships between the real alarms and the outcomes of the fitting model. Then, different hypothesis are done to analyse the model accuracy in the scenarios. Finally, the model is employed online to the inputs to provide the condition of the system. The models are used to provide a complementary response to the SCADA system. A decision can be made according to the correlations pattern when the online SCADA data is inputted into the validated fitting models.

The main advantage of the approach is the analysis of the interrelationship between variables. It is known that faults can be caused by a single cause or several causes. These causes can be monitoring by a signal, or variable, to set the condition. In other words, any fault can be characterized by a specific behavior of the variables related to the causes of the fault. This paper aims at distinguishing between those cases in which the anomalies of the variables are consistent with the generation of a fault, and the cases that present "abnormal anomalies" of variables that can induce to a false alarm. The paper states that there are some variables that, because of their nature, cause an impact on the value of others. A possible error in the data acquisition system can be occurring when this effect is not read in the dataset and, therefore, a false alarm or a missed alarm could appear.

This interpretation can generate different scenarios. Table II shows the scenarios and the decisions that will be make in each one.

Table II. Correlation Matrix (**R**) for SCADA dataset

	SCADA values ok	SCADA alarm
Correlations ok	No operation is required.	SCADA is providing correct information. Operation is required according to the alarm.
Miscorrelation	The variables have correct values but there may be an abnormal behaviour that is not being considered.	False alarm. If the alarm is caused by the miscorrelated variable, then it is necessary to check the SCADA sensors.

The approach proposed in this paper aims to facilitate a decision-making process by generating complementary information to the SCADA system. Table I proposes a general idea about the type of information that the approach will generate. The coincidence between the SCADA system and the outcomes of the approach will be used to validate the SCADA outcome, however, the discrepancy between both parts can be an indicator of a false or a missed alarm.

The models are dynamic due to the changing conditions of the WT, e.g. weather, life cycle, maintenance tasks, failures, etc. The final stage consists in the evaluation of the approach outcomes once the real conditions of the system can be checked. According to this evaluation, a filter will determine if new data is added to the database or is discarded [(53)][53](García Márquez and García‐Pardo 2010)(53)[53](García Márquez and García‐Pardo 2010)[53](García Márquez and García‐Pardo, 2010).

It is not possible to identify a false alarm with 100% accuracy. The most critical alarms need to be studied in detail although they have been identified as false alarms. The identification of false alarms can be useful for setting priorities and detecting sensor faults. Therefore, the main objective in this paper is to support maintenance decision making.

3.1 Correlation Analysis

The correlation is a statistical feature that determines the relationship between two variables. The novel method proposed uses the Pearson correlation coefficient to correlate the variables of the SCADA. The coefficient r between two signals $\mathbf{x}$ and $\mathbf{y}$ is the covariance of the two signals divided by the product of their standard deviations. This coefficient is calculated by:

$$r = \frac{N(\sum_{n=1}^{N} xy) - (\sum_{n=1}^{N} x)(\sum_{n=1}^{N} y)}{\sqrt{(N\sum_{n=1}^{N} x^2 - (\sum_{n=1}^{N} x)^2)(N\sum_{n=1}^{N} y^2 - (\sum_{n=1}^{N} y)^2)}}$$

Where N is the total number of samples

The coefficient can vary from -1 to 1. It represents the degree of correlation between the signals, 1 being a perfect positive correlation; e.g. if $\mathbf{x}$ increases, then $\mathbf{y}$ increases, and -1 a perfect negative correlation; e.g. if $\mathbf{x}$ increases, then $\mathbf{y}$ decreases. Figure 2 shows an example of the different correlation coefficients.

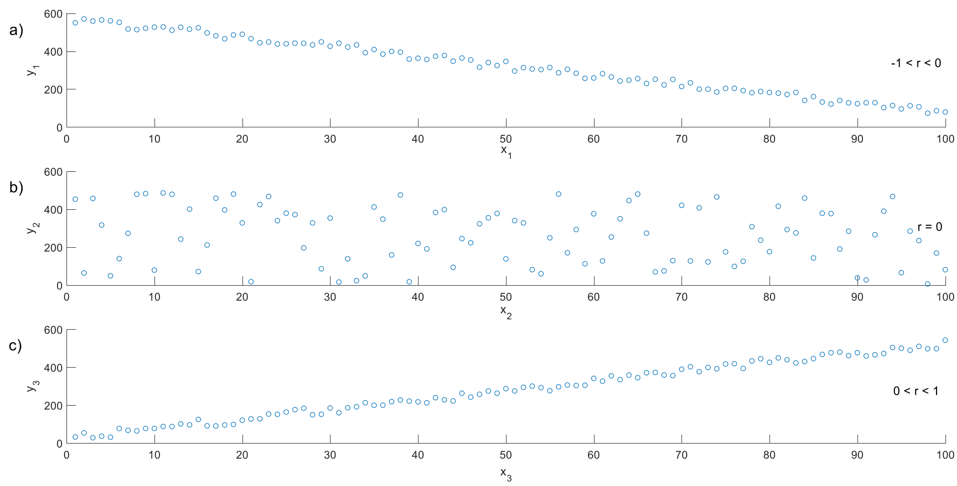

Figure 2. Possible correlation coefficients: a) negative correlation; b) no correlation; c) positive correlation.

Figure 2 shows that there is a positive correlation between x_3 and y_3, an inverse correlation between $x_1$and y_1, and no correlation between x_2 and y_2. This measurement will be used as an indicator of the accuracy of data from a SCADA system and as a detector of abnormal behaviours of the components.

The variables in Table II are represented in detail in Table III. The rows represent a total of n different variables measured by the SCADA. The columns indicate a total of m dates when the measurement was taken (every 10 minutes).

Table III. Data structure from SCADA

	Date 1	Date 2	…	Date j	…	Date m-1	Date m

Variable 1	$P_{1,1}$	$P_{1,2}$		$P_{1,j}$		$P_{1,m-1}$	$P_{1,m}$
Variable 2	$P_{2,1}$	$P_{2,2}$		$P_{2,j}$		$P_{2,m-1}$	$P_{2,m}$
…							
Variable i	$P_{i,1}$	$P_{i,2}$		$P_{i,j}$		$P_{i,m-1}$	$P_{i,m}$
…							
Variable n-1	$P_{n-1,1}$	$P_{n-1,2}$		$P_{n-1,j}$		$P_{n-1,m-1}$	$P_{n-1,m}$
Variable n	$P_{n,1}$	$P_{n,2}$		$P_{n,j}$		$P_{n,m-1}$	$P_{n,m}$

A correlation matrix (**R**) has been created to determine the correlations between the collected variables. Table IV shows that **R** is a *nxn* matrix, where the main diagonal always has value 1 due to the correlation of a variable with itself.

Table IV. Correlation Matrix (R) for SCADA data

	Variable 1	Variable 2	…	Variable i	…	Variable n-1	Variable n
Variable 1	1	$r_{1,2}$		$r_{1,i}$		$r_{1,n-1}$	$r_{1,n}$
Variable 2	$r_{2,1}$	1	…	$r_{2,i}$	…	$r_{2,n-1}$	$r_{2,n}$
…	…	…	1	…	…	…	…
Variable i	$r_{i,1}$	$r_{i,2}$	…	1	…	$r_{i,n-1}$	$r_{i,n}$
…	…	…	…	…	1	…	…
Variable n-1	$r_{n-1,1}$	$r_{n-1,2}$	…	$r_{n-1,i}$	…	1	$r_{n-1,n}$
Variable n	$r_{n,1}$	$r_{n,2}$	…	$r_{n,i}$	…	$r_{n,n-1}$	1

The following 4 categories have been set in order to classify the obtained Pearson correlation coefficients (54)[54](Martínez Ortega et al. 2009)(54)[54](Martínez Ortega et al. 2009)[54](Martínez Ortega et al., 2009), (55)[55](Statutor)(55)[55](Statutor)[55](Statutor), (56)[56](Levy and Wolf 2012)(56)[56](Levy and Wolf 2012)[56](Levy and Wolf, 2012):

- Weak correlation $0.3 \leq |r| < 0.5$
- Moderate correlation: $0.5 \leq |r| < 0.7$
- Strong correlation: $|r| \geq 0.7$
- Perfect correlation $|r| = 1$

Figure 3 presents the correlations between the different variables. Each cell determines if the correlation is weak (w, yellow), moderate (m, green), strong (s, red) or perfect (p, blue). The correlations of each variable with itself are perfect but not representative (p, black). A blank cell indicates that the correlation coefficient is less than 0.3. The matrix has been built with a dataset of a year. This analysis considers all the parameters and has not been found in literature.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	p	s	s	w							w	w		w																				
2	s	p	s	w	w		w				w	w		w							w												w	
3	s	s	p	w																										w				
4	w	w	w	p	m		m	w	w	w	m	w		w	w	w			m	m	m		m	m		w	w	w	w	m			s	
5		w		m	p		s				s	w		w			m		m	m	s		m	m	m	m	m	m	w		m	m	s	
6						p																												

7		w		m	s		p				s	w		w			w		m	m	m		m	m	m	w	w	w			m	m	s	
8				w				p	p	p					s										w						m	m		
9				w				p	p	p					s										w						m	m		
10				w				p	p	p					s										w						m	m		
11	w	w		m	s		s				p	w		w			m		m	m	s		m	s	m	m	m	m	w		m	m	s	
12	w	w		w	w		w				w	p		s					w	w	w		w	w						w			w	
13					w	w	w						p				w								w						w	w	w	
14	w	w		w	w		w				w	s		p					w	w	w		w	w						w			w	
15				w				s	s	s					p										w						m	m		
16				w												p		m	m	m	w	m	m	m		m	m	m						
17					m		w				m		w				p		m	m	m		m	m	s	w	w	w			m	m	w	w
18																m		p				s	w	w		m	m	m						
19				m	m		m				m	w		w		m	m		p	s	s		s	s	m	s	s	s	w		m	m	m	
20				m	m		m				m	w		w		m	m		s	p	s		s	s	m	s	s	s	w		m	m	m	
21		w		m	s		m				s	w		w		w	m		s	s	p		s	s	m	m	m	m	w		m	m	s	
22																m		s				p	w	w		m	m	m						
23				m	m		m				m	w		w		m	m	w	s	s	s	w	p	s	m	s	s	s			m	m	m	
24				m	m		m				s	w		w		m	m	w	s	s	s	w	s	p	m	s	s	s			m	m	m	
25					m		m	w	w	w	m		w		w		s		m	m	m		m	m	p	w	w	w			s	s	m	
26				w	m		w				m					m	w	m	s	s	m	m	s	s	w	p	s	s			w	w	w	
27				w	m		w				m					m	w	m	s	s	m	m	s	s	w	s	p	s			w	w	m	
28				w	m		w				m					m	w	m	s	s	m	m	s	s	w	s	s	p			w	w	m	
29				w	w						w								w	w	w								p	w			w	
30			w	m								w		w															w	p				
31					m		m	m	m	m	m		w		m		m		m	m	m		m	m	s	w	w	w			p	s	m	
32					m		m	m	m	m	m		w		m		m		m	m	m		m	m	s	w	w	w			s	p	m	
33		w		s	s		s				s	w	w	w			w		m	m	s		m	m	m	w	m	m	w		m	m	p	
34																	w																	p

Figure 3. Case study correlation matrix

Figure 3 shows that there are 80 weak correlations, 94 moderate correlations, 46 strong correlations, and 3 "true" perfect correlations (the perfect correlations in black cells are not representative). These numbers are calculated based on the fact that the correlation matrix is symmetric, i.e. all correlations are repeated twice, but they are counted once.

It is demonstrated that the correlations can be used to identify abnormal behaviors of the WT or possible false alarms [9][9](Yang and Jiang 2013)(9)[9](Yang and Jiang 2013)[9](Yang and Jiang, 2013). This matrix will be used in this paper to identify the variables that can be adapted to a curve fitting model.

3.2 Curve fitting models

The correlated variables are employed for establishing curve fittings to determine whether a certain variable has a normal behaviour or not. Only the strong correlations will be considered because they are the most representative. A fitting model is built for each red cell in Figure 3.

Figures 4, 5 and 6 show some examples of the fitting models used for modelling the correlations of variable 5 (turbulence level). Several methods can be selected for maximizing the quality of the fitting, e.g. R-square, sum of square error (SSE), root mean square error (RMSE), etc., for instance, by choosing the correct test [57][57](Paternoster et al. 1998)(57)[57](Paternoster et al. 1998)[57](Paternoster et al., 1998), comparing different models [58][58](Clogg et al. 1995)(58)[58](Clogg, Petkova, and Haritou 1995)[58](Clogg et al., 1995), using simple means (59)[59](Schielzeth 2010)(59)[59](Schielzeth 2010)[59](Schielzeth, 2010) or using spatially autocorrelated errors terms (60)[60](Dubin 1988)(60)[60](Dubin 1988)[60](Dubin, 1988). In this paper, the polynomial models have been employed to maximizing the R-square. These models have been calculated by the curve fitting tool of MATLAB. The confidence bounds are also statistically calculated by this tool. Figure 4 shows the correlation between the turbulence level and the vibration level. The mathematical model chosen for this fitting is a third-degree polynomial model:

$$f_{5,6}(x) = p_1x^3 + p_2x^2 + p_3x + p_4$$

where x is the variable 5 (turbulence level) and the coefficients with 95% confidence bounds are:

$$p_1 = -0.006414 \ (-0.008557, -0.00427)$$
$$p_2 = 0.1072 \ (0.07252, 0.1419)$$
$$p_3 = 0.4391 \ (0.2672, 0.6109)$$
$$p_4 = 0.2211 \ (-0.03152, 0.4737)$$

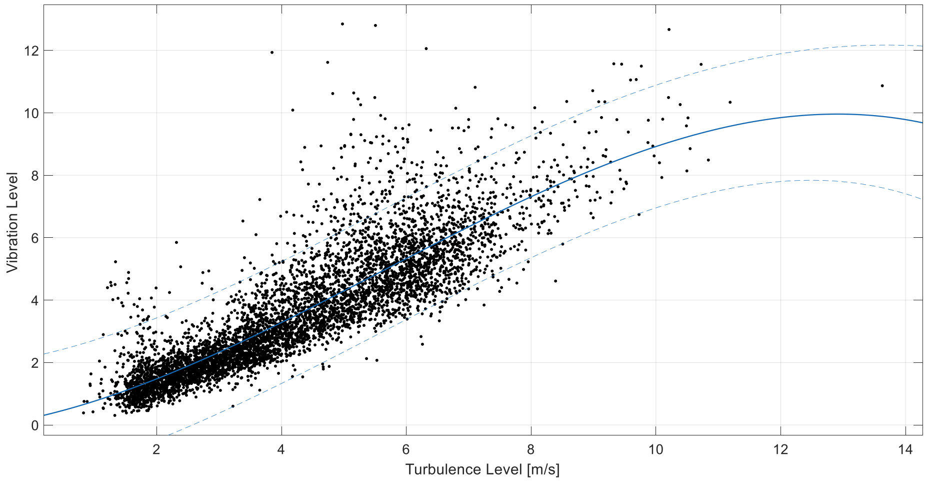

Figure 4. Turbulence vs Vibration

Figure 5 shows the relationship between the turbulence level and the power generated. The mathematical model chosen for this fitting is the seventh-degree polynomial model:

$$f_{30,5}(x) = p_1x^7 + p_2x^6 + p_3x^5 + p_4x^4 + p_5x^3 + p_6x^2 + p_7x + p_8$$

where x is the variable 5 (turbulence level) and the coefficients with 95% confidence bounds are:

$$p1 = 0.002816 \ (-0.0004231, 0.006054)$$
$$p2 = -0.1734 \ (-0.3206, -0.02621)$$
$$p3 = 3.951 \ (1.254, 6.647)$$
$$p4 = -42.37 \ (-68.06, -16.67)$$
$$p5 = 218.2 \ (81.32, 355.1)$$
$$p6 = -486.1 \ (-891.8, -80.48)$$

$$p7 \ = 598.2 \ \ (-17.81, 1214)$$
$$p8 \ = -291.6 \ \ (-660.6, 77.45)$$

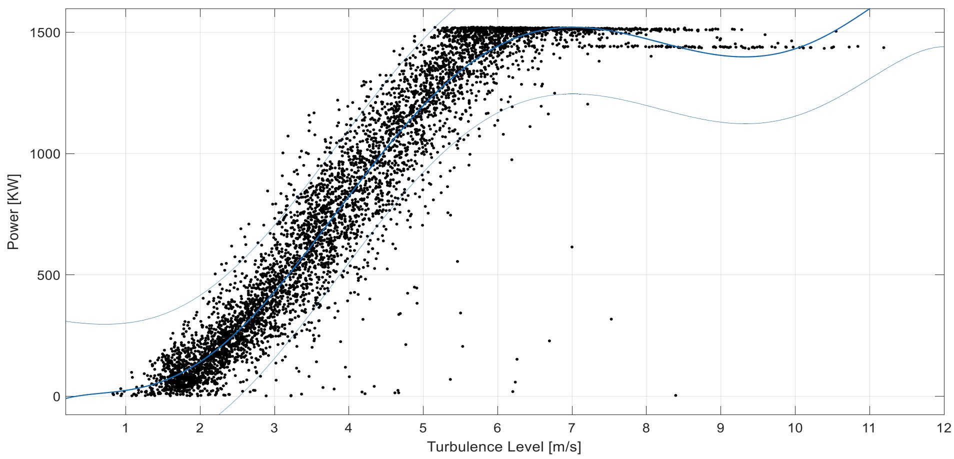

Figure 5. Turbulence vs Power

Figure 6 shows the relationship between the turbulence level and wind speed. In this case a first-degree polynomial model has been set as:

$$f_{5,33}(x) \ = \ p_1 x \ + \ p_2$$

Being x the variable 5 (turbulence level) and p_i the coefficients with 95% confidence bounds:

$$p_1 \ = 1.864 \ \ (1.851, 1.878)$$
$$p_2 \ = 0.852 \ \ (0.7907, 0.9133)$$

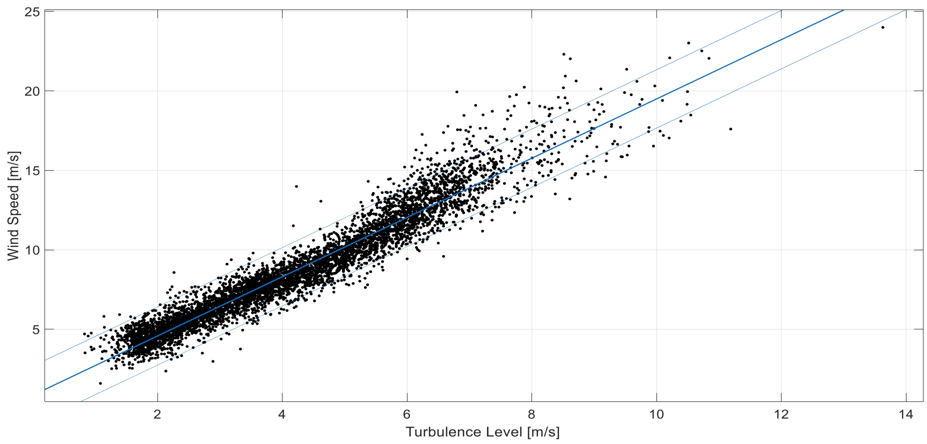

Figure 6. Turbulence vs Wind Speed

The confidence bounds enable the points under control, and those that are not, to be identified. Those points that are not under control are considered as miscorrelations. Hereinafter, a miscorrelation is a point out of the confidence bounds in which two variables that should be strongly correlated according to Figure 3 lose their dependence. This paper considers the idea that the ideal condition of the WT will maintain the correlations shown in Figure 3. However, it is also based on a real case study, so finding ideal scenarios is complex. In this case, all the miscorrelations are considered in order to determine if there are any abnormal events.

4 Quantification and validation of the model

Figure 7 shows the total number of miscorrelations (blue line) over time. These number of miscorrelations corresponds to the total number of points out of the confidence bounds of the fitting models. Since only the strong correlations are being considered, the maximum number of miscorrelations corresponds to the total amount of red cells in Figure 3. The abscises axis corresponds to the different samples that the SCADA system provides every ten minutes. The tags of this axis represent the date of each sample provided by function <<*datenum*>> of MATLAB. This function assigns a serial data number to each sampling date and time. The period shown in Figure 7 corresponds to 5781 samples, i.e. a time window of around 40 days.

The activations and deactivations of alarms are also represented by a small green circle and a red cross. The number of miscorrelations is a discrete value, but they are represented by a continuous line to differentiate them clearly from the alarms. The grey circles indicate periods in which the conditions can be considered as normal.

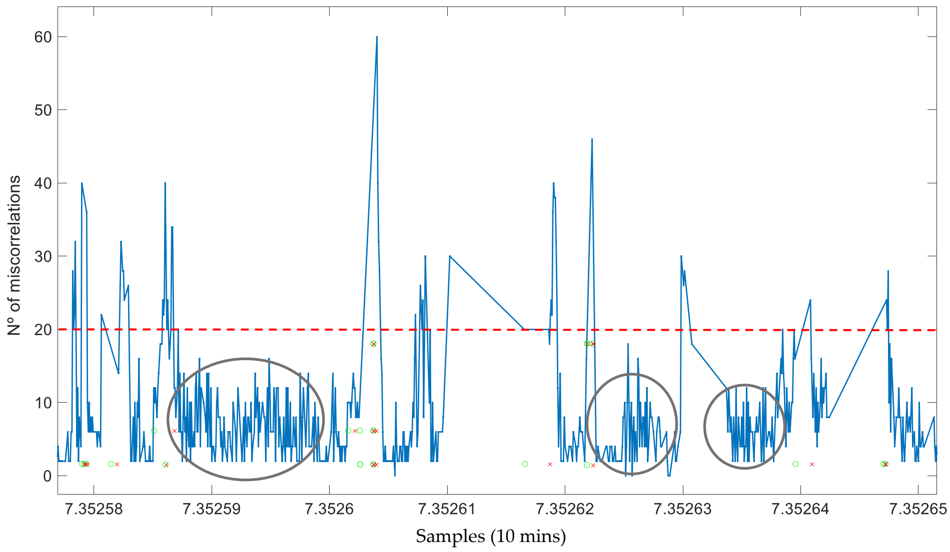

Figure 7. Number of miscorrelations and alarms

The real database from OPTIMUS European Project (52)[52](project 2014)(52)[52](project 2014)[52](project, 2014) employed in this paper shows that most of alarms appear when there are many miscorrelation, and the miscorrelations disappear when the alarms are deactivated. Therefore, the proposed curve fittings are providing additional information to detect an abnormal behavior in the system. The quantification of these model can be carried out by measuring the total number of miscorrelations and the period of time that the variables remain miscorrelated. In this case, the authors propose a maximum threshold (dashed red line) to determine the existence of an abnormal behavior. This threshold is calculated by adapting the data to a statistic distribution and using a statistical process control. The Anderson–Darling test has been carried out to find the most adequate distribution. This test resulted that a half normal distribution should be used to model the number of miscorrelations.

Figure 8 shows the SCADA histogram and the fit created.

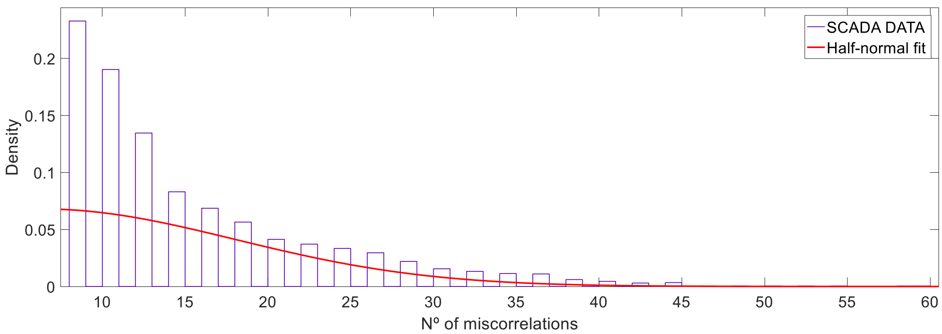

Figure 8. Distribution of number of miscorrelations

The threshold has been set by using the upper control limit of a x-bar control chart (61)[61](Chakraborti 2000)(61)[61](Chakraborti 2000)[61](Chakraborti, 2000). In this case, the upper control limit (UCL) results 20.1 miscorrelations. This is the control limit that will be employed to detect abnormal conditions.

Besides the number of miscorrelations, it has been observed that the duration of the miscorrelation is other important factor to determine the existence of an alarm. These miscorrelations usually do not disappear until the alarm is deactivated, i.e. until the failure is corrected. Although the number of miscorrelations is not high, an alarm can be detected due to their persistence in the time. This persistence factor allows the probability of success to be increased. Since each SCADA carried out every ten minutes, the probability of success due to the persistence factor has been quantified and stablished by different confusion matrix. Four different sceneries have been considered to evaluate the accuracy of the approach. The following outputs classes have been considered in the confusion matrix:

- Class 1: True Alarm. The number of miscorrelations is above the threshold and the SCADA activates an alarm.
- Class 2: False alarm. The number of miscorrelations is below the thresholds. A false alarm could be identified if the SCADA generates an alarm and: the number of miscorrelations is below the threshold or; the

number of miscorrelations surpass the threshold, but it returns below the threshold in the next measurement.

- Class 3: Missed alarm. The number of miscorrelations is above the threshold and the SCADA does not activates an alarm. A missed alarm can be identified when the number of miscorrelations is above the threshold indefinitely and the SCADA system has not activated an alarm.
- Class 4: True No alarm. The number of miscorrelations is below the thresholds and the SCADA does not activates an alarm.

Figure 9 shows the accuracy of the proposed methods through the comparison between the estimated outcomes and the real condition. The validation is done considering the real historic data provided by the European Project Optimus (52)[52](project 2014)(52)[52](project 2014)[52](project, 2014). The dataset contains the information of the occurrence of false alarms that are compared with the outcomes of the model. The results are presented in 4 different confusion matrices, where the accuracy of the proposed method depends of the miscorrelation persistence.

Approach output

PERSISTENCE 10 MINS					
C1	223	4	8	28	84%
C2	6	21	2	3	65%
C3	5	2	24	4	68%
C4	15	3	1	5212	99%
	C1	C2	C3	C4	Accuracy

Real condition

Approach output

PERSISTENCE 20 MINS					
C1	227	1	6	17	90%
C2	6	26	2	3	70%
C3	5	2	26	3	72%
C4	11	1	1	5224	99%
	C1	C2	C3	C4	Accuracy

Real condition

Approach output

PERSISTENCE 30 MINS					
C1	233	1	3	11	93%
C2	4	27	2	2	77%
C3	1	2	28	3	82%
C4	11	1	1	5231	99%
	C1	C2	C3	C4	Accuracy

Real condition

Approach output

PERSISTENCE 40 MINS					
C1	236	1	3	11	94%
C2	3	27	2	0	84%
C3	0	2	28	3	85%
C4	10	0	1	5233	99%
	C1	C2	C3	C4	Accuracy

Real condition

Figure 9. Confusion matrices. Accuracy of the method and influence of persistence

Artificial scenarios have been considered to obtain the footprint that a specific variable would provide in case of abnormal value. The identification of this footprints in a real sample of the SCADA can reveal an error in the measurement and, therefore, a false alarm. It is because of the variable value is abnormal due to a true fault, then the strongly correlated variable should also present abnormal values.

Figure 10 shows the number of miscorrelations for each variable when a specific variable has an abnormal value. The abscissa axis represents the set of variables and the ordinate axis shows the number of miscorrelations. Figure 10(b) shows a scenario where the variable 5 (turbulence level) has an abnormal value, in the same way for variables 4, 11, 21 and 33 (Figures 10(a), (c), (d) and (e), respectively).

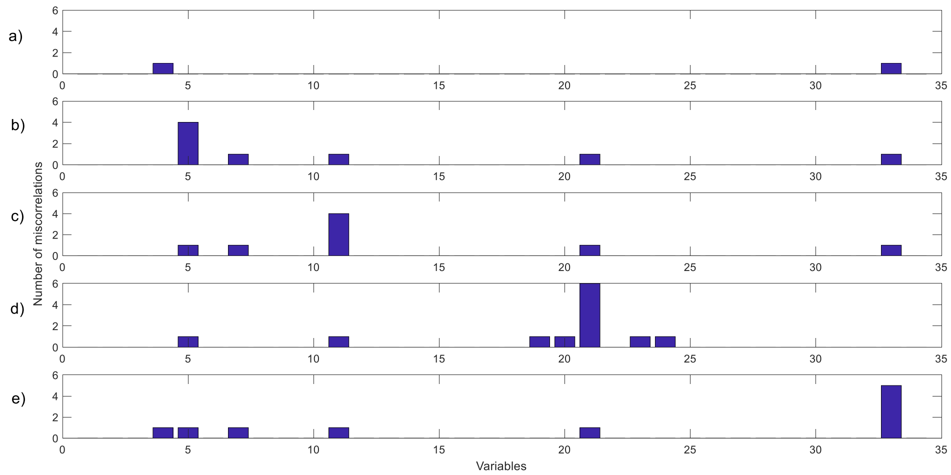

Figure 10. Patterns obtained from miscorrelations. a) Incorrect value of variable 4, b) Incorrect value of variable 5, c) Incorrect value of variable 11, d) Incorrect value of variable 21, e) Incorrect value of variable 33

It is observed that the variable related to an abnormal behaviour indicates a specific pattern, e.g. according to Figure 3, variable 5 is strongly correlated with variables 7, 11, 21 and 33. Therefore, an abnormal value of 5 causes four miscorrelations in 5, and one miscorrelation in the rest of the correlated variables. Same conclusions can be extracted from Figures 10(a), (c), (d) and (e). These patterns are used for identifying the specific variable which is causing the possible false alarm.

5 Example

Figure 11 shows the turbulence level over three months. The dash line indicates the threshold. Measurements above this threshold will produce a "High turbulence level" alarm. Different inputs have been selected to apply the proposed model and evaluate the response. 4 points have been highlighted (J, Q, R and Z) in Figure 11. These points have been selected to consider all the possible scenarios of the SCADA responses and the proposed model.

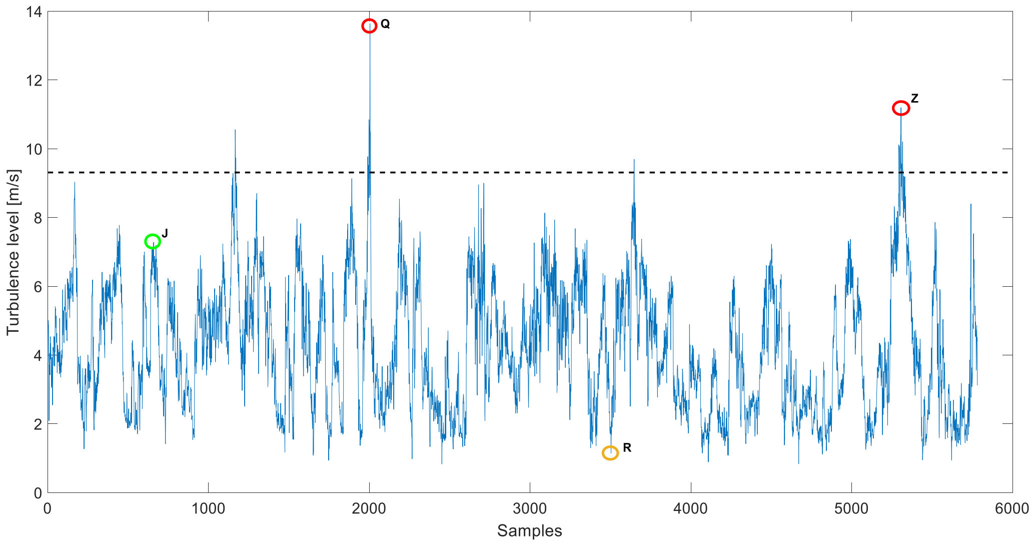

Figure 11. Example. Turbulence signal and threshold

Figure 12 shows the miscorrelations between variables at the same time for J (Figure 12(a)), Q (Figure 12(b)), R (Figure 12(c)) and Z (Figure 12(d)). The correlation of variable 5 (turbulence level) is considered in this example.

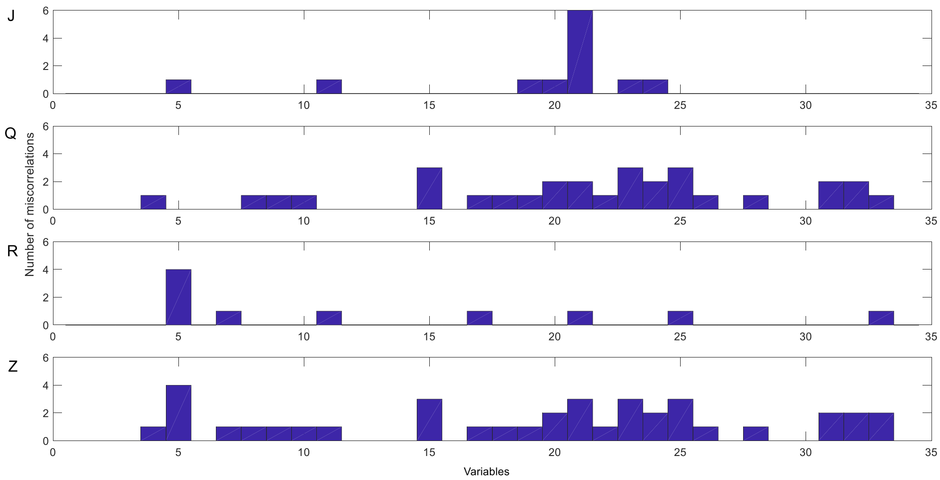

Figure 12. Example. Miscorrelations in points J, Q, R and Z (Figure 9)

A miscorrelation in a certain variable is not necessarily caused by an abnormal value of such variable. For example, variable 5 presents a miscorrelation in Figure 12(J), but the miscorrelation is caused because of an abnormal value of variable 21. Otherwise, variable number 5 would indicate at least 4 miscorrelations (see Figure 10). Table V summarizes the decision taken in each case.

Table V. Parameters of the SCADA system

Measurement	SCADA Alarm	Miscorrelation of the variable	Persistence	Decision
J	No	No	-	No action required
Q	Yes	No	-	The variable has reached the threshold, but it maintains the correlations. The measurement of the SCADA is ok. "False alarm" (Class 2)
R	No	Yes	Yes	The threshold is not reached but the variable presents an abnormal behavior. The miscorrelation persists in time. "Missed alarm" (Class 3)
			No	None significative anomaly. (Class 4)
Z	Yes	Yes	Yes	The measurement of the SCADA is ok. "High turbulence level" alarm should be considered. (Class 1)
			No	The turbulence reached the threshold, but it is miscorrelated for a short period of time. This could indicate that a measurement mistake has occurred. "False alarm" (Class 2)

False alarms will be identified when the SCADA activates an alarm, the variable is uncorrelated with the rests of variables and this miscorrelation persists in the time. To determine the accuracy of the estimation, it would be necessary to achieve the persistence of miscorrelations of each point J, Q, R, Z. The probability of success in the decision depends on the persistence of the miscorrelation, according to Figure 9.

Figure 3 shows that this method could be applied to all the variables, taking into account also medium and weak correlations.

6 Conclusions

Condition monitoring systems and SCADA are being employed in the wind energy industry and require advanced analytics. This paper provides a novel approach to identify false alarms and abnormal behaviours that standard SCADA systems cannot detect. The method uses the Pearson correlation coefficient for all the variables of the SCADA. A correlation matrix has been employed to analyse all the correlations together. This matrix classifies the correlations into 4 levels, and it is employed for pattern recognition. The values of the matrix are employed for creating a set of curve fittings. These fittings are used for detecting anomalies in the outputs of the SCADA system. They have been validated using a real alarm report and considering several hypotheses.

Correlations between variables have been demonstrated to be a good indicator of the condition of the system. The persistence of an unjustified miscorrelation is a symptom of an abnormal condition. The number of miscorrelations and the its persistence has been quantified for a specific database provided by the Optimus European Project.

A real case study has been considered in this research work. Two new types of alarms are generated by the approach: the first alarm indicates that the system receives a normal value of a certain parameter; the second one indicates a false alarm or a possible fault in the SCADA system.

Acknowledgements

The work reported herewith has been applied to the Spanish Ministerio de Economía y Competitividad, under Research Grants DPI2015-67264-P.

References

1. Gómez Muñoz, C.Q.; García Márquez, F.P. A new fault location approach for acoustic emission techniques in wind turbines. *Energies* **2016**, *9*, 40.
2. Tavner, P. Offshore wind turbines: Reliability. *Availability and Maintenance, The Institution of Engineering and Technology, London, UK* **2012**.
3. Jiménez, A.A.; Muñoz, C.Q.G.; Márquez, F.P.G. Dirt and mud detection and diagnosis on a wind turbine blade employing guided waves and supervised learning classifiers. *Reliability Engineering & System Safety* **2019**, *184*, 2-12.
4. Jiménez, A.A.; Márquez, F.P.G.; Moraleda, V.B.; Muñoz, C.Q.G. Linear and nonlinear features and machine learning for wind turbine blade ice detection and diagnosis. *Renewable Energy* **2019**, *132*, 1034-1048.
5. Gómez Muñoz, C.Q.; García Marquez, F.P.; Hernandez Crespo, B.; Makaya, K. Structural health monitoring for delamination detection and location in wind turbine blades employing guided waves. *Wind Energy* **2019**.
6. Pliego Marugán, A.; García Márquez, F.P.; Pinar Pérez, J.M. Optimal maintenance management of offshore wind farms. *Energies* **2016**, *9*, 46.
7. Márquez, F.P.G.; Tobias, A.M.; Pérez, J.M.P.; Papaelias, M. Condition monitoring of wind turbines: Techniques and methods. *Renewable Energy* **2012**, *46*, 169-178.
8. Márquez, F.P.G.; Pérez, J.M.P.; Marugán, A.P.; Papaelias, M. Identification of critical components of wind turbines using fta over the time. *Renewable Energy* **2016**, *87*, 869-883.
9. Yang, W.; Jiang, J. Wind turbine condition monitoring by the approach of scada data analysis. *Renewable Energy* **2013**, *53*, 365-376.
10. McMillan, D.; Ault, G.W. Quantification of condition monitoring benefit for offshore wind turbines. *Wind Engineering* **2007**, *31*, 267-285.
11. Renewables, G. Guidelines for the certification of condition monitoring systems for wind turbines. 2007.

12. Gómez Muñoz, C.Q.; Arcos Jimenez, A.; García Marquez, F.P.; Kogia, M.; Cheng, L.; Mohimi, A.; Papaelias, M. Cracks and welds detection approach in solar receiver tubes employing electromagnetic acoustic transducers. *Structural Health Monitoring* **2018**, *17*, 1046-1055.
13. Kusiak, A.; Li, W. The prediction and diagnosis of wind turbine faults. *Renewable Energy* **2011**, *36*, 16-23.
14. Kashyap, S. Data generated by wind and solar plants. Real big data. http://algoengines.com/2014/07/29/data-generated-by-wind-and-solar-plan/ (23/07/2018),
15. Tavner, P.; Edwards, C.; Brinkman, A.; Spinato, F. Influence of wind speed on wind turbine reliability. *Wind Engineering* **2006**, *30*, 55-72.
16. Musial, W.; Butterfield, S.; McNiff, B. In *Improving wind turbine gearbox reliability*, European Wind Energy Conference, Milan, Italy, 2007; pp 7-10.
17. Billinton, R.; Chen, H.; Ghajar, R. Time-series models for reliability evaluation of power systems including wind energy. *Microelectronics Reliability* **1996**, *36*, 1253-1261.
18. Gomez Munoz, C.; De la Hermosa Gonzalez-Carrato, R.; Trapero Arenas, J.; Garcia Marquez, F. A novel approach to fault detection and diagnosis on wind turbines. *Global Nest Journal* **2014**, *16*, 1029-1037.
19. de la Hermosa González, R.R.; Márquez, F.P.G.; Dimlaye, V. Maintenance management of wind turbines structures via mfcs and wavelet transforms. *Renewable and Sustainable Energy Reviews* **2015**, *48*, 472-482.
20. Kusiak, A.; Li, W.; Song, Z. Dynamic control of wind turbines. *Renewable energy* **2010**, *35*, 456-463.
21. Hand, M.M. *Mitigation of wind turbine/vortex interaction using disturbance accommodating control*. National Renewable Energy Laboratory Boulder, CO: 2003.
22. Stol, K.; Balas, M. In *Periodic disturbance accommodating control for speed regulation of wind turbines*, ASME 2002 Wind Energy Symposium, 2002; American Society of Mechanical Engineers: pp 310-320.
23. Song, Y.; Dhinakaran, B.; Bao, X. Variable speed control of wind turbines using nonlinear and adaptive algorithms. *Journal of Wind Engineering and Industrial Aerodynamics* **2000**, *85*, 293-308.
24. Johnson, K.E.; Pao, L.Y.; Balas, M.J.; Fingersh, L.J. Control of variable-speed wind turbines: Standard and adaptive techniques for maximizing energy capture. *IEEE control systems* **2006**, *26*, 70-81.
25. Freeman, J.; Balas, M. In *An investigation of variable speed horizontal-axis wind turbines using direct model-reference adaptive control*, 37th Aerospace sciences meeting and exhibit, 1999; p 28.
26. Johnson, K.E.; Fingersh, L.J. Adaptive pitch control of variable-speed wind turbines. *Journal of Solar Energy Engineering* **2008**, *130*, 031012.
27. Boukhezzar, B.; Lupu, L.; Siguerdidjane, H.; Hand, M. Multivariable control strategy for variable speed, variable pitch wind turbines. *Renewable Energy* **2007**, *32*, 1273-1287.
28. Márquez, F.P.G.; Muñoz, J.M.C. A pattern recognition and data analysis method for maintenance management. *International Journal of Systems Science* **2012**, *43*, 1014-1028.
29. Simoes, M.G.; Bose, B.K.; Spiegel, R.J. Fuzzy logic based intelligent control of a variable speed cage machine wind generation system. *IEEE transactions on power electronics* **1997**, *12*, 87-95.

30. Bououden, S.; Chadli, M.; Filali, S.; El Hajjaji, A. Fuzzy model based multivariable predictive control of a variable speed wind turbine: Lmi approach. *Renewable Energy* **2012**, *37*, 434-439.
31. Muljadi, E.; Pierce, K.; Migliore, P. In *Control strategy for variable-speed, stall-regulated wind turbines*, Proceedings of the American control conference, 1998; AMERICAN AUTOMATIC CONTROL COUNCIL: pp 1710-1714.
32. García Márquez, F.P.; Villasalero, M.; Pérez, P.; María, J. Technological innovation and dynamic capabilities in the spanish wind energy business. **2011**.
33. Boukhezzar, B.; Siguerdidjane, H. Nonlinear control with wind estimation of a dfig variable speed wind turbine for power capture optimization. *Energy Conversion and Management* **2009**, *50*, 885-892.
34. Kumar, A.; Stol, K. Simulating feedback linearization control of wind turbines using high-order models. *Wind Energy* **2010**, *13*, 419-432.
35. Qiu, Y.; Feng, Y.; Tavner, P.; Richardson, P.; Erdos, G.; Chen, B. Wind turbine scada alarm analysis for improving reliability. *Wind Energy* **2012**, *15*, 951-966.
36. Kusiak, A.; Song, Z.; Zheng, H. Anticipatory control of wind turbines with data-driven predictive models. *IEEE Transactions on Energy Conversion* **2009**, *24*, 766-774.
37. Gray, C.S.; Watson, S.J. Physics of failure approach to wind turbine condition based maintenance. *Wind Energy* **2010**, *13*, 395-405.
38. Muñoz, C.Q.G.; Márquez, F.P.G.; Tomás, J.M.S. Ice detection using thermal infrared radiometry on wind turbine blades. *Measurement* **2016**, *93*, 157-163.
39. Gill, S.; Stephen, B.; Galloway, S. Wind turbine condition assessment through power curve copula modeling. *IEEE Transactions on Sustainable Energy* **2012**, *3*, 94-101.
40. Igba, J.; Alemzadeh, K.; Durugbo, C.; Eiriksson, E.T. Analysing rms and peak values of vibration signals for condition monitoring of wind turbine gearboxes. *Renewable Energy* **2016**, *91*, 90-106.
41. Ozdemir, A.A.; Seiler, P.; Balas, G.J. In *Wind turbine fault detection using counter-based residual thresholding*, Proceedings of IFAC world congress, 2011; pp 8289-8294.
42. Chen, W.; Ding, S.X.; Haghani, A.; Naik, A.; Khan, A.Q.; Yin, S. Observer-based fdi schemes for wind turbine benchmark. *IFAC Proceedings Volumes* **2011**, *44*, 7073-7078.
43. Chen, Z.; Guerrero, J.M.; Blaabjerg, F. A review of the state of the art of power electronics for wind turbines. *IEEE Transactions on power electronics* **2009**, *24*, 1859-1875.
44. Carpman, N. Turbulence intensity in complex environments and its influence on small wind turbines. 2011.
45. Knudsen, T.; Bak, T.; Tabatabaeipour, S. In *Detection of excessive wind turbine tower oscillations fore-aft and sideways*, American Control Conference (ACC), 2012, 2012; Citeseer: pp 5795-5800.
46. Hau, E.; von Renouard, H. *Wind turbines: Fundamentals, technologies, application, economics*. Springer: 2003.
47. Barrett, M.P.; Stover, J. Understanding oil analysis: How it can improve the reliability of wind turbine gearboxes. *Gear Technology: Elk Grove Village, IL, USA* **2013**, 104-111.
48. Burton, T.; Jenkins, N.; Sharpe, D.; Bossanyi, E. *Wind energy handbook*. John Wiley & Sons: 2011.

49. Yagi, S.; NINOYU, N. Technical trends in wind turbine bearings. *TECHNICAL REVIEW* **2008**.
50. Stone, G. Advancements during the past quarter century in on-line monitoring of motor and generator winding insulation. *IEEE Transactions on Dielectrics and Electrical Insulation* **2002**, *9*, 746-751.
51. Jha, A.R. *Wind turbine technology*. CRC press: 2010.
52. project, E. Demonstration of methods and tools for the optimisation of operational reliability of large-scale industrial wind turbines, optimus project. . **2014**.
53. García Márquez, F.P.; García-Pardo, I.P. Principal component analysis applied to filtered signals for maintenance management. *Quality and Reliability Engineering International* **2010**, *26*, 523-527.
54. Martínez Ortega, R.M.; Tuya Pendás, L.C.; Martínez Ortega, M.; Pérez Abreu, A.; Cánovas, A.M. El coeficiente de correlación de los rangos de spearman caracterización. *Revista Habanera de Ciencias Médicas* **2009**, *8*, 0-0.
55. Statutor. Community project encouraging academics to share statistics support resources all stcp resources are released under a creative commons licence http://www.statstutor.ac.uk/resources/uploaded/pearsoncorrelation3.pdf.
56. Levy, N.; Wolf, L. In *Minimal correlation classification*, European Conference on Computer Vision, 2012; Springer: pp 29-42.
57. Paternoster, R.; Brame, R.; Mazerolle, P.; Piquero, A. Using the correct statistical test for the equality of regression coefficients. *Criminology* **1998**, *36*, 859-866.
58. Clogg, C.C.; Petkova, E.; Haritou, A. Statistical methods for comparing regression coefficients between models. *American Journal of Sociology* **1995**, *100*, 1261-1293.
59. Schielzeth, H. Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution* **2010**, *1*, 103-113.
60. Dubin, R.A. Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *The Review of Economics and Statistics* **1988**, 466-474.
61. Chakraborti, S. Run length, average run length and false alarm rate of shewhart x-bar chart: Exact derivations by conditioning. *Communications in Statistics-Simulation and Computation* **2000**, *29*, 61-81.