ingEnium Research Group
www.ingeniumgroup.eu

Journal Paper

"Multivariable Analysis for Advanced Analytics

of Wind Turbine Management"

Advances in Intelligent Systems and Computing

Accepted 24 August 2016

Alberto Pliego Marugán
Ingenium Research Group, Universidad de Castilla-La Mancha
Alberto.Pliego@uclm.es

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

Mayorkinos Papaelias
University of Birmingham, United Kingdom

Cite as: Marugán, A.P.; Márquez, F.P.G.; Papaelias, M. In *Multivariable analysis for advanced analytics of wind turbine management*, Proceedings of the Tenth International Conference on Management Science and Engineering Management, 2017; Springer: pp 319-328. DOI: 10.1007/978-981-10-1837-4_28

Contents

Chapter 1
Multivariable Analysis for Advanced Analytics of Wind Turbine Management

Alberto Pliego Marugán, Fausto Pedro García Márquez and Mayorkinos Papaelias

Abstract Operation and maintenance tasks on the wind turbines have an essential role to ensure the correct condition of the system and to minimize losses and increase the productivity. The condition monitoring systems installed on the main components of the wind turbines provide information about the tasks that should be carried out over the time. A novel statistical methodology for multivariable analysis of big data from wind turbines is presented in this paper. The objective is to analyse the necessary information from the condition monitoring systems installed in wind farms. The novel approach filters the main parameters from the collected signals and uses advanced computational techniques for evaluating the data and giving meaning to them. The main advantage of the approach is the possibility of the big data analysis based on the main information available.

Keywords Condition monitoring systems · Multivariable analysis · Wind turbine maintenance · Neural networks

1.1 Introduction

Nowadays, wind energy is the renewable energy source with the highest growth. Figure 1 shows the cumulative global capacity, that it has risen from 3.5 GW in 1994 to more than 420 GW in 2016. Therefore, the global capacity has grown by

A. Marugán
Ingenium Research Group, Universidad Castilla-La Mancha, 13071 Ciudad Real, Spain

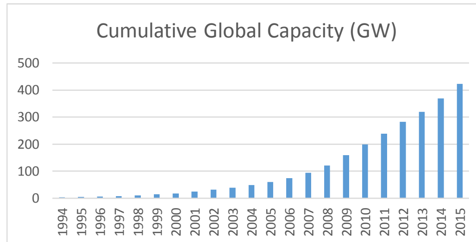
F. Márquez (✉)
Ingenium Research Group, Universidad Castilla-La Mancha, 13071 Ciudad Real, Spain
e-mail: alberto.pliego, faustopedro.garcia@uclm.es

M. Papaelias
University of Birmingham, United Kingdom

J. Xu et al. (eds.), *International Conference on Management Science and Engineering Management 2015*,
Advances in Intelligent Systems and Computing XXX, DOI: XXXXXXXXXX

more than 10,000 percent after two decades. Moreover, the wind energy is estimated to increase at an annual average rate of 6% until 2035 [21].

Fig. 1.1 Cumulative global capacity. Source: Global Wind Energy Council [21]



The offshore wind farms are being the wind energy source with more importance. The main advantages of the offshore wind farms are [5]:

- The wind power captured by wind turbines (WTs) is more than onshore.
- The size of offshore wind farms can be larger than onshore.
- The environmental impact for offshore is less than in onshore.
  The principal disadvantages are:
- It is more complex to evaluate the wind feature.
- Larger investment costs. The offshore installation cost is 1.44 million €/MW, where the onshore is 0.78 million €/MW [11].
- Operation and maintenance (O&M) tasks are more complex and expensive than onshore. The offshore O&M costs are from 18% to 23% of the total system costs, being 12% for onshore wind farms [23].

The rapid growth of the wind energy generation and, specially, the fast expansion of the offshore windfarms, it is due to the new and larger wind turbines that present a more complex system. The use of Condition Monitoring Systems (CMSs) is becoming very useful to reduce the lost energy, the downtimes and the required outlays for O&M tasks [2]. There are several studies that prove the economic benefit of CMSs [15]. Different techniques and methods of condition monitoring (CM) are employed for detection and diagnosis of faults of WTs [9, 13, 16, 19]. Most of the research papers consider the CM in WTs referred to blades [24], gearboxes [8], electrical or electronic components [4] and tower [10]. CM leads to improve reliability, availability, maintainability, and safety (RAMS) to increase the productivity of wind farms [6, 18].

The first step of the CM process is the choice of an adequate technique for data acquisition, including electronic signals with the measurement of the required conditions, e.g., sound, vibration, voltage, temperature, speed [7]. Then, a correct signal processing method is applied, e.g. fast Fourier transform, wavelet transforms, hidden Markov models, statistical methods and trend analysis [10, 20]. In this paper a statistical methodology that can be used to improve the cost effectiveness of the CMSs is presented. The procedure is based on the idea that it is useful to collect the important information provided by the data and exclude the data that do not provide relevant information.

1.2 Methodology Proposed

As aforementioned, the methodology presented hereby is based in a statistical use of the data of different CMSs. There are different monitoring techniques, being the most employed: vibrations (vibration-based damage detection) using sensors such as piezoelectric accelerometers; oil analysis to determine viscosity and levels of contaminants; acoustic emission; motor current signature analysis [25]. The CMSs are aimed to provide a continuous monitoring of the condition of dynamic parts and power electronics of the wind turbines. The methodology proposed facilitates the determination of the condition of some components of the wind turbines using historical data and comparing them with the current information. The main assumption is that the wind turbine is working properly most of the time. The approach proposed in this paper, detailed in section 1.2.1, gathers different feature parameters from certain parts of the received signal and carry out comparisons.

1.2.1 Gathering Feature Parameters from Signals

The CMS considered generates data each ten seconds every three hours. The amount of data is usually proportional to the accuracy of the analysis. However, the method is based on the idea that it is not necessary to storage the entire signal. This has the advantage of reducing the computational cost for analyzing the signals that can be received from a large number of CMS. Once the signals have been collected, the approach generated the main parameters that characterise the signals. The approach analyses big data, and therefore is it important to reduce the signal of thousands of samples to the main parameters, being the main:

- Average of the signal. This value can be useful when the signal has not a lot of abrupt changes. It can be used as a feature parameter for temperature or humidity signals.
- Energy (E) and power (W) of the signal. The expressions used to calculate such parameters are:

$$E = \sum_{n=-\infty}^{\infty} |x(n)|^2,$$

$$W = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} |x(n)|^2,$$

where N is the total number of samples and $x(i)$ the data x for the sample i. This parameters can be very representative for electric signals or vibration signals.

- Maximum peaks of the signal. The maximum peaks of the signal represent the highest or the lowest amplitude reached by the signal. A peak can be very important in those signals that are almost constant all the time.
- Maximum peaks of the Fast Fourier Transform (FFT). The FFT represents the signal in the frequency domain. The maximum peaks are referred to the amplitude of the main frequency and the most important harmonics.
- Pearson correlation coefficient (r). The correlation coefficient between two discrete signals x and y can be expressed as:

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

A case study is shown in Fig. 1.2 in order to clarify the calculation of the Pearson correlation coefficient for the collected signals, where 4 signals are considered. The date of each signal is different but the 4 signal are collected under the same conditions. Signal 1 (dark blue line) and 2 (green line) are increasing, whereas signals 3 (red line) and 4 (light blue) contain some noise. This function correlates

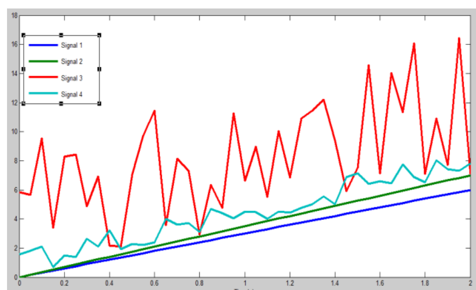


Fig. 1.2 Example for correlating signals

the 4 signals placing the different correlation coefficients into a correlation matrix R as follows:

$$R = \begin{bmatrix} 1 & 1 & 0.52 & 0.96 \\ 1 & 1 & 0.52 & 0.96 \\ 0.52 & 0.52 & 1 & 0.43 \\ 0.96 & 0.96 & 0.43 & 1 \end{bmatrix}.$$

Each element r_{ij} of the matrix R contains the result of correlating the signals i and j. It is a symmetric matrix where the diagonal is filled with ones due to those elements represent the correlation of a signal with itself. The correlation between signals 1 and 2 is perfect because both are linear sequences. Table 1.1 shows the average of each column avoiding the 1 of the diagonal. This value will be used as a feature parameter of each signal.

Table 1.1 Average of correlation coefficients

	Signal 1	Signal 2	Signal 3	Signal 4
Correlation coefficient average	0.82	0.82	0.49	0.78

Table 1.1 shows that an abnormal behavior of the system could be due from the Signal 3 because the wind turbine is assumed to work properly most of the time. The possibility that a value can represent an abnormal condition of the system is because the different from the rest one value is bigger.

- Other time-domain feature parameters [3]. Other measurements can provide additional information of the signals, e.g. RMS, Standard Deviation, Skewness, Kurtosis, Crest Factor, Clearance Factor, Shape Indicator, Impulse Indicator, etc.

The approach therefore organizes chronologically the data and associates them with a certain condition of the system given by an alarm (section 1.2.2).

1.2.2 Data Management

The approach does an ordination of all the data when the main parameters have been obtained from each signal. The data are chronologically ordered as shown in Table 1.2:

Table 1.2 Chronological Ordination of parameters

	Signal 1			Signal 2		
	Pearson Corr.Coef	Maximum FFT peak	Energy Energy	Pearson Corr.Coef	Maximum FFT peak	Energy Energy
Date 1	R^1_{11}	M^1_{12}	E^1_{13}	R^2_{11}	M^2_{12}	E^2_{13}
Date 2	R^1_{21}	M^1_{22}	E^1_{13}	R^2_{21}	M^2_{22}	E^2_{13}
Date 3	R^1_{31}	M^1_{32}	E^1_{13}	R^2_{31}	M^2_{32}	E^2_{13}
Date 4	R^1_{41}	M^1_{42}	E^1_{13}	R^2_{41}	M^2_{42}	E^2_{13}

The element e^k_{ij} corresponds to the j parameter of the k signal collected at the time (date) i. The parameters placed in the same row should correspond to a signal collected at the same time.

The parameter are then associated with the corresponding condition of the system when every parameters have been ordered. The conditions of the wind turbine, given by the alarms, are collected for the period of time that are analysed the dataset. The case study considers a database from the European Project called OPTIMUS [17]. The main objective is to co-relate each row of the matrix to a specific condition of the wind turbine. Table 1.3 shows this correlation.

Table 1.3 Association of data with condition of the WT

	Param. Signal 1	Param. Signal 2	Param. Signal i	Condition of the WT
Date 1	e^1_{1j}	e^2_{1j}	e^k_{1j}	C1
Date 2	e^1_{2j}	e^2_{2j}	e^k_{2j}	C2
Date 3	e^1_{3j}	e^2_{3j}	e^k_{3j}	C3
Date 4	e^1_{4j}	e^2_{4j}	e^k_{4j}	C4

Table 1.2 shows the data (rows) and the associated condition of the WT. These conditions can be related to a normal state of the system, specific failures and different alarms. The main objective of this procedure is to establish a relationship between the main parameters extracted from the signal and the condition of the WT at a certain time. Considering the big data characteristics, e.g there are thousands of signals and, therefore, millions of parameters, the analysis should be established employing statistical and heuristic methods. A novel pattern recognition through neural networks (NN) is employed in the approach.

1.2.3 Novel Pattern Recognition and Classification by Neural Network

In the former section, a way of manipulating and organizing data has been exposed. The purpose is to prepare the data to be subjected to a pattern recognition analysis. There are numerous models and procedures for pattern recognition analysis, i.e. statistical model, structural model, template matching model, NN based model, fuzzy based model, hybrid models, etc. [12, 22]. Each method has some advantages in function of the information available and the outcomes desired.

NN are used in problems that cannot be formulated as an exact method or an analytical solution. NN learns by itself and provides a good solution for the problem simulating the biological neurons in a reasonable time. An artificial NN consists of neurons that are simple processing units and weighted connections between those neurons [1]. The typical structure of a NN is showed in Fig 1.3.

The variables presented in Fig. 1.3 are:

- Input layer: The nodes that forms the input layer are passive. They only duplicate the received value and propagate it through the multiple outputs.
- Hidden layer: The nodes of this layer are active because they modify the data in function of the weights.
- Output layer: The nodes of this layer are active due to they combine and modify the data in order to produce the k outputs of Fig 1.3.
- or represents the strength of a connection (the connecting weight) between two neurons i and j, or between j and k.
- u or u nodes represents the simple processing units called neurons.

Basically, the NN receives a dataset and enter into a training process in order to recognise several patterns. The training process fits the different weights in order to provide the output. If the output is known, then the training is defined as supervised, otherwise, it is called unsupervised training. In this paper it is assumed that there is information available about the condition of the wind turbine all the time. The condition of the WT corresponds to the desired outputs of the NN, where the inputs are set in each row of the Table II. The data used to design the NN is divided in the following groups:

- Training set: Around 75% of the total amount of data.
- Validation set: Around 15% of the total amount of data.
- Testing set: Around 15% of the total amount of data.

The NN is expected to be able to learn from data and predict the output when a generic input is considered. The advantage of this model is that whatever the input, the NN will provide a predefined output.

1.3 Real Case Study

The data considered for this real case study is obtained from the European Project entitled OPTIMUS [17]. The datasets are come from the following CMSs installed in a WT:

- Vibration data from a Train Driver. The data available are signals of 1 second collected every three hours in 8 different points of the drive train from 01/04/2014 to 31/01/2015. Fig. 1.4 shows the eight different regions of the drive train where vibration signals are collected. The signal will be identified regarding to this numeration, i.e. V1, V2 ⋯ V8.

The data available consists of a Report of Alarms generated by a SCADA registered from 01/10/2012 to 15/05/2015 where there are different types of alarms

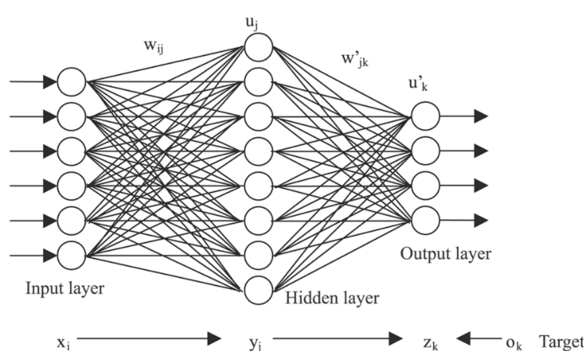


Fig. 1.3 Structure of an Artificial NN [23]

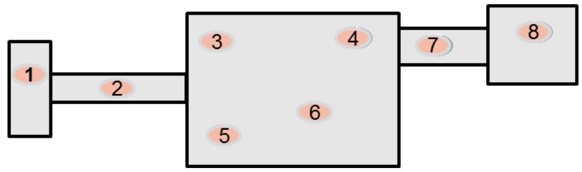


Fig. 1.4 Location of the different sensors in the train system of the WT

identified by a specific code. The codes are not indicated for reasons of confidentiality.

The vibration data from the Drive Train CM module has been considered and then it has been calculated the main parameter of the signal where they have been correlated by the FFTs.

The date of the alarms do not coincide with the data of the signals, therefore it has been associated the closest previous vibration data to each alarm.

The main parameters of the vibration data will be the input of the NN and the output will be defined in function of the number of alarms taken into account. In order to clarify this, Table 1.4 presents the data used for designing the NN.

Table 1.4 Inputs and outputs for NN

	Input data			Output data		
	V1	V2	Vi	Alarm 1	Alarm 2	Alarm 3
Date 1	P^1_{1j}	P^2_{1j}	P^k_{1j}	1	0	0
Date 2	P^1_{2j}	P^2_{2j}	P^k_{2j}	0	0	0
Date 3	P^1_{3j}	P^2_{3j}	P^k_{3j}	0	1	0
Date 4	P^1_{4j}	P^2_{4j}	P^k_{4j}	0	0	1

The NNs have been designed attending to the geometric pyramid rule [14], i.e. the number of neurons of the hidden layer should be calculated by using the following expression:

$$h = \sqrt{m \times n},$$

where h is the number of neurons in the hidden layer, m is the number of neurons in the input layer and n is the number of neurons in the output layer.

In this case study, a total of three parameters of each signal Vi have been considered as inputs because the number of signal is 8. Therefore the number of neurons in the input layer is 8×3=24. Two different networks have been developed according to two kinds of outputs:

- NN1: The possible outputs are "Alarm generated or not generated". The number of neurons in the hidden layer is 5.
- NN2: It has been determined that the possible outputs are Alarm 1, Alarm 2, Alarm 3 or none of them. The number of neurons in the hidden layer is 10.

Fig. 1.5 presents the outcomes of the NNs through two confusion matrixes. The

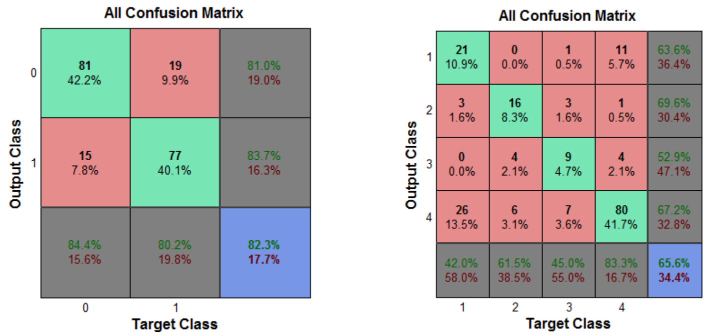


Fig. 1.5 Outcomes of the NN1 (left) and NN2 (right)

confusion matrixes presented in Fig. 1.5 summarise the results from training, validation and test data process. A confusion matrix is a comparison between the desired result (target class) and the outcome provided by the NN (output class). The main objective is to maximise the value of the diagonal of the matrix.

The NN1 informs that it is possible to determine from the vibration data whether an alarm will occur or not with an accuracy of 82.3%. The NN2 is able to distinguish the type of alarm with a success of 65.6 %.

1.4 Conclusions

This paper has presented a novel approach in order to analyse big data from condition monitoring systems and SCADA of a wind turbine. The methodology analyses the condition of the system. The main information of the data is obtained with the purpose of simplifying the analysis of the big data. The approach is based on three steps: To extract feature parameters; order, and; association of the data and pattern recognition. A real case study is presented in order to apply the procedure and demonstrate and validate the methodology. The data used hereby correspond to different condition monitoring systems installed in a wind turbine. Neural networks have been employed in order to analyse the main parameters gathered. Two different networks have been developed according to two kinds of outputs: NN1 the outputs are “Alarm generated or not generated”; NN2 considers Alarm 1, Alarm 2, Alarm 3 or none of them. The NN1 reports that it is possible to determine from the vibration data whether an alarm will occur or not with a high accuracy. The NN2 is able to distinguish the type of alarm with a medium accuracy.

Acknowledgements The work reported herewith has been financially supported by the Spanish Ministerio de Economła y Competitividad, under Research Grant DPI2015-67264, and the FP7 Research project with reference FP-7-Energy-2012-TREN-1:322430.

References

1. Arbib MA (2003) The handbook of brain theory and neural networks. MIT Press
2. Azevedo HDM, Araújo AM, Bouchonneau N (2016) A review of wind turbine bearing condition monitoring: State of the art and challenges. Renewable and Sustainable Energy Reviews 56:368–379
3. Chen D, Wang W (2002) Classification of wavelet map patterns using multi-layer neural networks for gear fault detection. Mechanical systems and signal processing 16(4):695–704
4. Chen Z, Guerrero JM, Blaabjerg F (2009) A review of the state of the art of power electronics for wind turbines. Power Electronics, IEEE Transactions on 24(8):1859–1875
5. Esteban MD, Diez JJ, López JS, Negro V (2011) Why offshore wind energy? Renewable Energy 36(2):444–450
6. Gomez Munoz C, De la Hermosa et al (2014) A novel approach to fault detection and diagnosis on wind turbines. Global Nest Journal 16(6):1029–1037
7. Gómez Muñoz CQ, García Márquez FP (2016) A new fault location approach for acoustic emission techniques in wind turbines. Energies 9(1):40
8. Guo P, Bai N (2011) Wind turbine gearbox condition monitoring with AAKR and moving window statistic methods. Energies 4(11):2077–2093
9. de la Hermosa González RR, Márquez FPG et al (2015) Maintenance management of wind turbines structures via mfcs and wavelet transforms. Renewable and Sustainable Energy Reviews 48:472–482
10. Jiang W, Fan Q, Gong J (2010) Optimization of welding joint between tower and bottom flange based on residual stress considerations in a wind turbine. Energy 35(1):461–467
11. Lloyd G (2007) Guideline for the certification of condition monitoring systems for wind turbines. Hamburg, Germany
12. Márquez FPG, Muñoz JMC (2012) A pattern recognition and data analysis method for maintenance management. International Journal of Systems Science 43(6):1014–1028
13. Márquez FPG, Pérez JMP, Marugán AP, Papaelias M (2016) Identification of critical components of wind turbines using FTA over the time. Renewable Energy 87:869–883
14. Masters T (1993) Practical neural network recipes in C++. Morgan Kaufmann
15. McMillan D, Ault GW (2007) Quantification of condition monitoring benefit for offshore wind turbines. Wind Engineering 31(4):267–285
16. Novaes Pires de G, Alencar E, Kraj A (2010) Remote conditioning monitoring system for a hybrid wind diesel system-application at fernando de naronha island, brasil
17. OPTIMUS (2014) Demonstration of methods and tools for the optimisation of operational reliability of large-scale industrial wind turbines, optimus project
18. Papaelias M, Cheng L et al (2016) Inspection and structural health monitoring techniques for concentrated solar power plants. Renewable Energy 85:1178–1191
19. Pliego Marugán A, García Márquez F (2014) System management for remote condition monitoring in railway systems. In: Railway Condition Monitoring (RCM 2014), 6th IET Conference on, IET, pp 1–10
20. Pliego Marugán A, García Márquez FP, Pinar Pérez JM (2016) Optimal maintenance management of offshore wind farms. Energies 9(1):46
21. Pullen A (2015) Global wind report annual market update 2014. Tech. rep., Global Wind Report Annual Market
22. Asht S, Dass R (2012) Pattern recognition techniques: A review. International Journal of Computer Science and Telecommunications 3(8)
23. Tavner P (2012) Offshore wind turbine reliability
24. Tsai CS, Hsieh CT, Huang SJ (2006) Enhancement of damage-detection of wind turbine blades via cwt-based approaches. Energy Conversion, IEEE Transactions on 21(3):776–781
25. Wymore ML, Van Dam JE et al (2015) A survey of health monitoring systems for wind turbines. Renewable and Sustainable Energy Reviews 52:976–990