Wind Turbine Fault Detection through Principal Component Analysis and Statistical Hypothesis Testing

Francesc Pozo[1,a], and Yolanda Vidal[1,b]

CoDAlab, Department of Mathematics, Escola Universitària d'Enginyeria Tècnica Industrial de Barcelona (EUETIB), Universitat Politècnica de Catalunya (UPC), Comte d'Urgell, 187, 08036 Barcelona, Spain

[a]francesc.pozo@upc.edu, [b]yolanda.vidal@upc.edu

Keywords: wind turbine; fault detection; principal component analysis; statistical hypothesis testing; FAST (Fatigue, Aerodynamics, Structures and Turbulence)

Abstract. This work addresses the problem of online fault detection of an advanced wind turbine benchmark under actuators (pitch and torque) and sensors (pitch angle measurement) faults of different type. The fault detection scheme starts by computing the baseline principal component analysis (PCA) model from the healthy wind turbine. Subsequently, when the structure is inspected or supervised, new measurements are obtained and projected into the baseline PCA model. When both sets of data are compared, a statistical hypothesis testing is used to make a decision on whether or not the wind turbine presents some fault. The effectiveness of the proposed fault-detection scheme is illustrated by numerical simulations on a well-known large wind turbine in the presence of wind turbulence and realistic fault scenarios.

Introduction

The past few years have seen a rapid growth in interest in wind turbine fault detection [1] through the use of condition monitoring and structural health monitoring (SHM) [2]. The SHM techniques are based on the idea that the change in mechanical properties of the structure will be captured by a change in its dynamic characteristics [3]. Existing techniques for fault detection can be broadly classified into two major categories: model-based methods and signal processing-based methods. For model-based fault detection, the system model could be mathematical—or knowledge-based [4]. Faults are detected based on the residual generated by state variable or model parameter estimation [5, 6]. For signal processing-based fault detection, mathematical or statistical operations are performed on the measurements (see, for example, [7]).

With respect to signal-processing-based fault detection, principal component analysis (PCA) is used in this framework as a way to condense and extract information from the collected signals. Following this structure, this work is focused on the development of a wind turbine fault detection strategy that combines a data driven baseline model—reference pattern obtained from the healthy structure—based on PCA and hypothesis testing.

Most industrial wind turbines are manufactured with an integrated system that can monitor various turbine parameters. These monitored data are collated and stored via a supervisory control and data acquisition (SCADA) system that archives the information in a convenient manner. These data quickly accumulates to create large and unmanageable volumes that can hinder attempts to deduce the health of a turbine's components. It would prove beneficial if the data could be analyzed and interpreted automatically (online) to support the operators in planning cost-effective maintenance activities [8]. This work describes a technique that can be used to identify incipient faults in the main components of a turbine through the analysis of this SCADA data. The strategy developed is based on principal component analysis and statistical hypothesis testing. The final section of the work shows the performance of the proposed techniques using an enhanced benchmark challenge for wind turbine fault detection, see [1].

Table 1: Fault scenarios.

Number	Fault	Type
F1	Pitch actuator	Change in dynamics: air content in oil
F2	Pitch actuator	Change in dynamics: pump wear
F3	Pitch actuator	Change in dynamics: hydraulic leakage
F4	Torque actuator	Offset (2000 Nm)
F5	Generator speed sensor	Scaling (1.2)
F6	Pitch angle sensor	Stuck (5^o)
F7	Pitch angle sensor	Stuck (10^o)
F8	Pitch angle sensor	Scaling (1.2)

Table 2: Assumed SCADA data

Number	Sensor Type	Units
1	Generated electrical power	kW
2	Rotor speed	rad/s
3	Generator speed	rad/s
4	Generator torque	Nm
5	first pitch angle	deg
6	second pitch angle	deg
7	third pitch angle	deg
8	fore-aft acceleration at tower bottom	m/s^2
9	side-to-side acceleration at tower bottom	m/s^2
10	fore-aft acceleration at mid-tower	m/s^2
11	side-to-side acceleration at mid-tower	m/s^2
12	fore-aft acceleration at tower top	m/s^2
13	side-to-side acceleration at tower top	m/s^2

Wind Turbine Benchmark Model

A complete description of the wind turbine benchmark model, as well as the used baseline torque and pitch controllers, can be found in [1]. This benchmark proposes a set of realistic fault scenarios considered in an aeroelastic computer-aided engineering tool for horizontal axes wind turbines called FAST, see [9]. The numerical simulations use the onshore version of a large wind turbine that is representative of typical utility-scale land- and sea-based multimegawatt turbines described by [10]. This wind turbine is a conventional three-bladed upwind variable-speed turbine of 5 MW. In this work we deal with the full load region of operation (also called region 3). That is, the proposed controllers main objective is that the electric power follows the rated power.

All the studied faults originate from actual faults in wind turbines [1]. Table 1 summarizes all the considered fault scenarios.

Table 2 presents the assumed available measurements. These sensors are representative of the types of sensors that are available on a MW-scale commercial wind turbine.

Fault Detection Strategy

The overall fault detection strategy is based on principal component analysis and statistical hypothesis testing. A baseline pattern or PCA model is created with the healthy state of the wind turbine in the presence of wind turbulence. When the current state of the wind turbine has to be diagnosed, the collected data is projected using the PCA model. The final diagnosis is performed using statistical hypothesis testing.

The main paradigm of vibration based structural health monitoring is based on the basic idea that a change in physical properties due to structural changes or damage will cause detectable changes in dynamical responses. Usually, the healthy structure is excited by a signal to create a pattern. Subsequently, the structure to be diagnosed is excited by the *same* signal and the dynamic response is compared with the pattern.

However, in our application, the only available excitation of the wind turbines is the wind turbulence. Thus, the key idea behind the detection strategy is the assumption that a change in the behavior of the overall system, even with a *different* excitation, has to be detected. The results presented in the numerical results Section confirm this hypothesis.

Data Driven Baseline Modeling Based on PCA Let us start the PCA modeling by measuring, from a healthy wind turbine, a sensor during $(nL-1)\Delta$ seconds, where Δ is the sampling time and $n, L \in \mathbb{N}$. The discretized measures of the sensor are a real vector

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1L} & x_{21} & x_{22} & \cdots & x_{2L} & \cdots & x_{n1} & x_{n2} & \cdots & x_{nL} \end{pmatrix} \in \mathbb{R}^{nL} \tag{1}$$

where the real number $x_{ij},\ i=1,\ldots,n,\ j=1,\ldots,L$ corresponds to the measure of the sensor at time $((i-1)L+(j-1))\,\Delta$ seconds. This collected data can be arranged in matrix form as follows:

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1L} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{iL} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nL} \end{pmatrix} \in \mathcal{M}_{n\times L}(\mathbb{R}) \tag{2}$$

where $\mathcal{M}_{n\times L}(\mathbb{R})$ is the vector space of $n \times L$ matrices over $\mathbb{R}$. When the measures are obtained from $N \in \mathbb{N}$ sensors also during $(nL-1)\Delta$ seconds, the collected data, for each sensor, can be arranged in a matrix as in Equation (2). Finally, all the collected data coming from the N sensors is disposed in a matrix $\mathbf{X} \in \mathcal{M}_{n\times(N\cdot L)}$ as follows:

$$\begin{aligned} \mathbf{X} &= \left(\begin{array}{cccc|ccc|c|ccc} x^1_{11} & x^1_{12} & \cdots & x^1_{1L} & x^2_{11} & \cdots & x^2_{1L} & \cdots & x^N_{11} & \cdots & x^N_{1L} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x^1_{i1} & x^1_{i2} & \cdots & x^1_{iL} & x^2_{i1} & \cdots & x^2_{iL} & \cdots & x^N_{i1} & \cdots & x^N_{iL} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x^1_{n1} & x^1_{n2} & \cdots & x^1_{nL} & x^2_{n1} & \cdots & x^2_{nL} & \cdots & x^N_{n1} & \cdots & x^N_{nL} \end{array}\right) \\ &= \left(\mathbf{X}^1 \,\middle|\, \mathbf{X}^2 \,\middle|\, \cdots \,\middle|\, \mathbf{X}^N \right) \end{aligned} \tag{3}$$

where the superindex $k=1,\ldots,N$ of each element x^k_{ij} in the matrix represents the number of sensor.

The objective of the principal component analysis is to find a linear transformation orthogonal matrix $\mathbf{P} \in \mathcal{M}_{(N\cdot L)\times(N\cdot L)}(\mathbb{R})$ that will be used to transform or project the original data matrix $\mathbf{X}$ according to the subsequent matrix product:

$$\mathbf{T} = \mathbf{X}\mathbf{P} \in \mathcal{M}_{n\times(N\cdot L)}(\mathbb{R}) \tag{4}$$

where $\mathbf{T}$ is a matrix having a diagonal covariance matrix.

Group Scaling

Since the data in matrix $\mathbf{X}$ is affected by diverse wind turbulence, come from several sensors and could have different scales and magnitudes, it is required to apply a preprocessing step to rescale the data using the mean of all measurements of the sensor at the same column and the standard deviation of all measurements of the sensor [11].

More precisely, for $k = 1, 2, \ldots, N$ we define

$$\mu_j^k = \frac{1}{n} \sum_{i=1}^{n} x_{ij}^k, \; j = 1, \ldots, L, \tag{5}$$

$$\mu^k = \frac{1}{nL} \sum_{i=1}^{n} \sum_{j=1}^{L} x_{ij}^k, \tag{6}$$

$$\sigma^k = \sqrt{\frac{1}{nL} \sum_{i=1}^{n} \sum_{j=1}^{L} (x_{ij}^k - \mu^k)^2} \tag{7}$$

where μ_j^k is the mean of the measures placed at the same column, that is, the mean of the n measures of sensor k in matrix $\mathbf{X}^k$ at time instants $((i-1)L + (j-1))\,\Delta$ seconds, $i = 1, \ldots, n$; μ^k is the mean of all the elements in matrix $\mathbf{X}^k$, that is, the mean of all the measures of sensor k; and σ^k is the standard deviation of all the measures of sensor k. Therefore, the elements x_{ij}^k of matrix $\mathbf{X}$ are scaled to define a new matrix $\check{\mathbf{X}}$ as

$$\check{x}_{ij}^k := \frac{x_{ij}^k - \mu_j^k}{\sigma^k}, \; i = 1, \ldots, n, \; j = 1, \ldots, L, \; k = 1, \ldots, N. \tag{8}$$

When the data are normalized using Equation (8), the scaling procedure is called variable scaling or group scaling [12].

For the sake of clarity, and throughout the rest of the paper, the scaled matrix $\check{\mathbf{X}}$ is renamed as simply $\mathbf{X}$. The mean of each column vector in the scaled matrix $\mathbf{X}$ can be computed as

$$\frac{1}{n} \sum_{i=1}^{n} \check{x}_{ij}^k = \frac{1}{n} \sum_{i=1}^{n} \frac{x_{ij}^k - \mu_j^k}{\sigma^k} = \frac{1}{n\sigma^k} \sum_{i=1}^{n} \left(x_{ij}^k - \mu_j^k\right) \tag{9}$$

$$= \frac{1}{n\sigma^k} \left[\left(\sum_{i=1}^{n} x_{ij}^k \right) - n\mu_j^k \right] \tag{10}$$

$$= \frac{1}{n\sigma^k} \left(n\mu_j^k - n\mu_j^k \right) = 0 \tag{11}$$

Since the scaled matrix $\mathbf{X}$ is a mean-centered matrix, it is possible to calculate its covariance matrix as follows:

$$\mathbf{C_X} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X} \in \mathcal{M}_{(N \cdot L) \times (N \cdot L)}(\mathbb{R}) \tag{12}$$

The covariance matrix $\mathbf{C_X}$ is a $(N \cdot L) \times (N \cdot L)$ symmetric matrix that measures the degree of linear relationship within the data set between all possible pairs of columns. At this point it is worth noting that each column can be viewed as a virtual sensor and, therefore, each column vector $\mathbf{X}(:, j) \in \mathbb{R}^n, \; j = 1, \ldots, N \cdot L$, represents a set of measurements from one virtual sensor.

The subspaces in PCA are defined by the eigenvectors and eigenvalues of the covariance matrix as follows:

$$\mathbf{C_X} \mathbf{P} = \mathbf{P} \Lambda \tag{13}$$

where the columns of $\mathbf{P} \in \mathcal{M}_{(N \cdot L) \times (N \cdot L)}(\mathbb{R})$ are the eigenvectors of $\mathbf{C_X}$. The diagonal terms of matrix $\Lambda \in \mathcal{M}_{(N \cdot L) \times (N \cdot L)}(\mathbb{R})$ are the eigenvalues $\lambda_i, \; i = 1, \ldots, N \cdot L$, of $\mathbf{C_X}$ whereas the off-diagonal terms are zero, that is,

$$\Lambda_{ii} = \lambda_i, \; i = 1, \ldots, N \cdot L \tag{14}$$

$$\Lambda_{ij} = 0, \; i, j = 1, \ldots, N \cdot L, \; i \neq j \tag{15}$$

The eigenvectors $p_j,\ j = 1, \ldots, N \cdot L$, representing the columns of the transformation matrix $\mathbf{P}$ are classified according to the eigenvalues in descending order and they are called the principal components of the data set. The eigenvector with the highest eigenvalue, called the first principal component, represents the most important pattern in the data with the largest quantity of information.

Matrix $\mathbf{P}$ is usually called the principal components of the data set and matrix $\mathbf{T}$ is the projected matrix to the principal component space. Using all the $N \cdot L$ principal components, that is, in the full dimensional case, the orthogonality of $\mathbf{P}$ implies $\mathbf{P}\mathbf{P}^{\mathrm{T}} = \mathbf{I}$, where $\mathbf{I}$ is the $(N \cdot L) \times (N \cdot L)$ identity matrix. Therefore, the projection can be inverted to recover the original data as

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T \tag{16}$$

However, the objective of PCA is, as said before, to reduce the dimensionality of the data set $\mathbf{X}$ by selecting only a limited number $\ell < N \cdot L$ of principal components, that is, only the eigenvectors related to the ℓ highest eigenvalues. Thus, given the reduced matrix

$$\hat{\mathbf{P}} = (p_1|p_2|\cdots|p_\ell) \in \mathcal{M}_{N \cdot L \times \ell}(\mathbb{R}) \tag{17}$$

matrix $\hat{\mathbf{T}}$ is defined as

$$\hat{\mathbf{T}} = \mathbf{X}\hat{\mathbf{P}} \in \mathcal{M}_{n \times \ell}(\mathbb{R}) \tag{18}$$

Note that opposite to $\mathbf{T}$, $\hat{\mathbf{T}}$ is no longer invertible. Consequently, it is not possible to fully recover $\mathbf{X}$ although $\hat{\mathbf{T}}$ can be projected back onto the original $N \cdot L-$dimensional space to get a data matrix $\hat{\mathbf{X}}$ as follows:

$$\hat{\mathbf{X}} = \hat{\mathbf{T}}\hat{\mathbf{P}}^T \in \mathcal{M}_{n \times (N \cdot L)}(\mathbb{R}) \tag{19}$$

The difference between the original data matrix $\mathbf{X}$ and $\hat{\mathbf{X}}$ is defined as the residual error matrix $\mathbf{E}$ or $\tilde{\mathbf{X}}$ as follows:

$$\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}} \tag{20}$$

or, equivalenty,

$$\mathbf{X} = \hat{\mathbf{X}} + \mathbf{E} = \hat{\mathbf{T}}\hat{\mathbf{P}}^T + \mathbf{E} \tag{21}$$

The residual error matrix $\mathbf{E}$ describes the variability not represented by the data matrix $\hat{\mathbf{X}}$, and can also be expressed as

$$\mathbf{E} = \mathbf{X}(\mathbf{I} - \hat{\mathbf{P}}\hat{\mathbf{P}}^T) \tag{22}$$

Even though the real measures obtained from the sensors as a function of time represent physical magnitudes, when these measures are projected and the scores are obtained, these scores no longer represent any physical magnitude [13]. The key aspect in this approach is that the scores from different experiments can be compared with the reference pattern to try to detect a different behavior.

Fault Detection Based on Hypothesis Testing The current wind turbine to diagnose is subjected to a wind turbulence. When the measures are obtained from $N \in \mathbb{N}$ sensors during $(\nu L - 1)\Delta$ seconds, a new data matrix $\mathbf{Y}$ is constructed as in Equation (3):

$$\mathbf{Y} = \left(\begin{array}{cccc|ccc|c|ccc} y_{11}^1 & y_{12}^1 & \cdots & y_{1L}^1 & y_{11}^2 & \cdots & y_{1L}^2 & \cdots & y_{11}^N & \cdots & y_{1L}^N \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ y_{i1}^1 & y_{i2}^1 & \cdots & y_{iL}^1 & y_{i1}^2 & \cdots & y_{iL}^2 & \cdots & y_{i1}^N & \cdots & y_{iL}^N \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ y_{\nu 1}^1 & y_{\nu 2}^1 & \cdots & y_{\nu L}^1 & y_{\nu 1}^2 & \cdots & y_{\nu L}^2 & \cdots & y_{\nu 1}^N & \cdots & y_{\nu L}^N \end{array}\right) \in \mathcal{M}_{\nu \times (N \cdot L)}(\mathbb{R}) \tag{23}$$

Before the collected data arranged in matrix $\mathbf{Y}$ is projected into the new space spanned by the eigenvectors in matrix $\mathbf{P}$ in Equation (13), the matrix has to be scaled to define a new matrix $\check{\mathbf{Y}}$ as in Equation (8):

$$\check{y}_{ij}^k := \frac{y_{ij}^k - \mu_j^k}{\sigma^k},\ i = 1, \ldots, \nu,\ j = 1, \ldots, L,\ k = 1, \ldots, N, \tag{24}$$

where μ_j^k and σ^k are defined in Equations (5) and (7), respectively.

The projection of each row vector $r^i = \check{\mathbf{Y}}(i,:) \in \mathbb{R}^{N \cdot L}, i = 1, \ldots, \nu$ of matrix $\check{\mathbf{Y}}$ into the space spanned by the eigenvectors in $\hat{\mathbf{P}}$ is performed through the following vector to matrix multiplication:

$$t^i = r^i \cdot \hat{\mathbf{P}} \in \mathbb{R}^\ell \tag{25}$$

For each row vector $r^i, i = 1, \ldots, \nu$, the first component of vector t^i is called the first score or score 1; similarly, the second component of vector t^i is called the second score or score 2, and so on.

In a standard application of the principal component analysis strategy in the field of structural health monitoring, the scores allow a visual grouping or separation [14]. In some other cases, as in [15], two classical indices can be used for damage detection, such as the Q index (also known as SPE, square prediction error) and the Hotelling's T^2 index. In this case, however, a visual grouping, clustering or separation cannot be performed either by projection onto the two first principal components or by plotting the natural logarithm of indices Q and T^2 of samples coming from the healthy and faulty wind turbines. Therefore, a more powerful and reliable tool is needed to be able to detect a fault in the wind turbine.

The Random Nature of the Scores

Since the turbulent wind can be considered as a random process, the dynamic response of the wind turbine can be considered as a stochastic process and the measurements in r^i are also stochastic. Therefore, each component of t^i acquires this stochastic nature and it will be regarded as a random variable to construct the stochastic approach in this work.

Test for the Equality of Means

The objective of the present work is to examine whether the current wind turbine is healthy or subjected to a fault as those described in Table 1. To achieve this end, we have a PCA model (matrix $\hat{\mathbf{P}}$ in Equation (17)) built as in Section with data coming from a wind turbine in a full healthy state. For each principal component $j = 1, \ldots, \ell$, the baseline sample is defined as the set of n real numbers computed as the $j-$th component of the vector to matrix multiplication $\mathbf{X}(i,:) \cdot \hat{\mathbf{P}}$. Note that n is the number of rows of matrix $\mathbf{X}$ in Equation (3). That is, we define the baseline sample as the set of numbers $\{\tau_j^i\}_{i=1,\ldots,n}$ given by

$$\tau_j^i := (\mathbf{X}(i,:) \cdot \hat{\mathbf{P}})(j) = \mathbf{X}(i,:) \cdot \hat{\mathbf{P}} \cdot \mathbf{e}_j,\ i = 1, \ldots, n, \tag{26}$$

where $\mathbf{e}_j$ is the $j-$th vector of the canonical basis.

Similarly, and for each principal component $j = 1, \ldots, \ell$, the sample of the current wind turbine to diagnose is defined as the set of ν real numbers computed as the $j-$th component of the vector t^i in Equation (25). Note that ν is the number of rows of matrix $\mathbf{Y}$ in Equation (23). That is, we define the sample to diagnose as the set of numbers $\{t_j^i\}_{i=1,\ldots,\nu}$ given by

$$t_j^i := t^i \cdot \mathbf{e}_j,\ i = 1, \ldots, \nu. \tag{27}$$

Let us consider that, for a given principal component, (a) the baseline sample is a random sample of a random variable having a normal distribution with unknown mean μ_X and unknown standard deviation σ_X; and (b) the random sample of the current wind turbine is also normally distributed with unknown mean μ_Y and unknown standard deviation σ_Y. Let us finally consider that the variances of these two samples are not necessarily equal. As said previously, the problem that we will consider is to determine whether these means are equal, that is, $\mu_X = \mu_Y$, or equivalently, $\mu_X - \mu_Y = 0$. This statement leads immediately to a test of the hypotheses

$$H_0 : \mu_X - \mu_Y = 0 \text{ versus} \tag{28}$$

$$H_1 : \mu_X - \mu_Y \neq 0 \tag{29}$$

that is, the null hypothesis is "the sample of the wind turbine to be diagnosed is distributed as the baseline sample" and the alternative hypothesis is "the sample of the wind turbine to be diagnosed is not distributed as the baseline sample". In other words, if the result of the test is that the null hypothesis is not rejected, the current wind turbine is categorized as healthy. Otherwise, if the null hypothesis is rejected in favor of the alternative, this would indicate the presence of some faults in the wind turbine.

The test is based on the Welch-Satterthwaite method [16], which is outlined below. When random samples of size n and ν, respectively, are taken from two normal distributions $N(\mu_X, \sigma_X)$ and $N(\mu_Y, \sigma_Y)$ and the population variances are unknown, the random variable

$$\mathcal{W} = \frac{(\bar{X} - \bar{Y}) + (\mu_X - \mu_Y)}{\sqrt{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{\nu}\right)}} \tag{30}$$

can be approximated with a t-distribution with ρ degrees of freedom, that is

$$\mathcal{W} \hookrightarrow t_\rho \tag{31}$$

where

$$\rho = \left\lfloor \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{\nu}\right)^2}{\frac{(s_X^2/n)^2}{n-1} + \frac{(s_Y^2/\nu)^2}{\nu - 1}} \right\rfloor \tag{32}$$

and where $\bar{X}, \bar{Y}$ is the sample mean as a random variable; S^2 is the sample variance as a random variable; s^2 is the variance of a sample; and $\lfloor \cdot \rfloor$ is the floor function.

The value of the standardized test statistic using this method is defined as

$$t_{\text{obs}} = \frac{\bar{x} - \bar{y}}{\sqrt{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{\nu}\right)}} \tag{33}$$

where $\bar{x}, \bar{y}$ is the mean of a particular sample. The quantity t_{obs} is the fault indicator. We can then construct the following test:

$$|t_{\text{obs}}| \leq t^\star \implies \text{Fail to reject } H_0 \tag{34}$$

$$|t_{\text{obs}}| > t^\star \implies \text{Reject } H_0 \tag{35}$$

where $t^\star$ is such that

$$P(t_\rho < t^\star) = 1 - \frac{\alpha}{2} \tag{36}$$

and α is the chosen risk (significance) level for the test. More precisely, the null hypothesis is rejected if $|t_{\text{obs}}| > t^\star$ (this would indicate the existence of a fault in the wind turbine). Otherwise, if $|t_{\text{obs}}| \leq t^\star$ there is no statistical evidence to suggest that both samples are normally distributed but with different means, thus indicating that no fault in the wind turbine has been found.

Table 3: Scheme for the presentation of the results in Table 4.

	Undamaged Sample (H_0)	**Damaged Sample (H_1)**
Fail to reject H_0	Correct decision	Type II error (missing fault)
Reject H_0	Type I error (false alarm)	Correct decision

Table 4: Categorization of the samples with respect to the presence or absence of damage and the result of the test for each of the four scores when the size of the samples to diagnose is $\nu = 50$.

	score 1		**score 2**		**score 3**		**score 4**	
	H_0	H_1	H_0	H_1	H_0	H_1	H_0	H_1
Fail to reject H_0	16	0	12	1	11	5	9	1
Reject H_0	0	8	4	7	5	3	7	7

Simulation Results

To validate the fault detection strategy presented, we first consider a total of 24 samples of $\nu = 50$ elements each, according to the following distribution: 16 samples of a healthy wind turbine; and 8 samples of a faulty wind turbine with respect to each of the eight different fault scenarios described in Table 1.

In the numerical simulations in this Section, each sample of $\nu = 50$ elements is composed by the measures obtained from the $N = 13$ sensors detailed in Table 2 during $(\nu \cdot L - 1)\Delta = 312.4875$ seconds, where $L = 500$ and the sampling time $\Delta = 0.0125$ seconds. The measures of each sample are then arranged in a $\nu \times (N \cdot L)$ matrix as in Equation (23).

For the first four principal components (score 1 to score 4), these 24 samples plus the baseline sample of $n = 50$ elements are used to test for the equality of means, with a level of significance $\alpha = 0.36$. Each sample of $\nu = 50$ elements is categorized as follows: (i) number of samples from the healthy wind turbine (healthy sample) which were classified by the hypothesis test as 'healthy' (fail to reject H_0); (ii) faulty sample classified by the test as "faulty" (reject H_0); (iii) samples from the faulty structure (faulty sample) classified as "healthy"; and (iv) faulty sample classified as "faulty". The results for the first four principal components presented in Table 4 are organized according to the scheme in Table 3. It can be stressed from each principal component in Table 4 that the sum of the columns is constant: 16 samples in the first column (healthy wind turbine) and 8 more samples in the second column (faulty wind turbine).

In Table 4, it is worth noting that two kinds of misclassification are presented which are denoted as follows. Type I error (false positive or false alarm), when the wind turbine is healthy but the null hypothesis is rejected and therefore classified as faulty. The probability of committing a type I error is α, the level of significance. On the other hand, Type II error (false negative or missing fault), when the structure is faulty but the null hypothesis is not rejected and therefore classified as healthy. The probability of committing a type II error is called γ.

It can be observed from Table 4 that, in the numerical simulations, Type I errors (false alarms) and Type II errors (missing faults) appear only when scores 2, 3 or 4 are considered, while when the first score is used all the decisions are correct. The better performance of the first score is an expected result in the sense that the first principal component is the component that accounts for the largest possible variance.

Summary

In this work, numerical simulations (with a well-known benchmark wind turbine) show that the proposed PCA plus statistical hypothesis testing is a valuable tool in fault detection for wind turbines. It is noteworthy that, in the simulations, when the first score is used all the decisions are correct (there are no false alarms and no missing faults). We believe that PCA plus statistical hypothesis testing has tremendous potential in decreasing maintenance costs.

Acknowledgments

This work has been partially funded by the Spanish Ministry of Economy and Competitiveness through the research project DPI2014-58427-C2-1-R, and by the Generalitat de Catalunya through the research project 2014 SGR 859.

References

[1] Odgaard, P.; Johnson, K. Wind Turbine Fault Diagnosis and Fault Tolerant Control—An Enhanced Benchmark Challenge. In Proceedings of the 2013 American Control Conference (ACC), Washington, DC, USA, 17–19 June 2013; pp. 1–6.

[2] Soman, R.N.; Malinowski, P.H.; Ostachowicz, W.M. Bi-axial neutral axis tracking for damage detection in wind-turbine towers. *Wind Energy* **2015**, doi:10.1002/we.1856.

[3] Griffith, D.T.; Yoder, N.C.; Resor, B.; White, J.; Paquette, J. Structural health and prognostics management for the enhancement of offshore wind turbine operations and maintenance strategies. *Wind Energy* **2014**, *17*, 1737–1751.

[4] Ding, S.X. *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*; Springer Science & Business Media: London, UK, 2008.

[5] Shi, F.; Patton, R. An active fault tolerant control approach to an offshore wind turbine model. *Renew. Energy* **2015**, *75*, 788–798.

[6] Vidal, Y.; Tutiven, C.; Rodellar, J.; Acho, L. Fault Diagnosis and Fault-Tolerant Control of Wind Turbines via a Discrete Time Controller with a Disturbance Compensator. *Energies* **2015**, *8*, 4300–4316.

[7] Dong, J.; Verhaegen, M. Data driven fault detection and isolation of a wind turbine benchmark. In Proceedings of the International Federation of Automatic Control (IFAC) World Congress, Milano, Italy, 28 August–2 September 2011; Volume 2, pp. 7086–7091.

[8] Kusiak, A.; Li, W.; Song, Z. Dynamic control of wind turbines. *Renew. Energy* **2010**, *35*, 456–463.

[9] Jonkman, J. NWTC Information Portal (FAST). https://nwtc.nrel.gov/FAST. Last modified 19-March-2015; Accessed 18-December-2015.

[10] Jonkman, J.M.; Butterfield, S.; Musial, W.; Scott, G. *Definition of a 5-MW Reference Wind Turbine for Offshore System Development*. National Renewable Energy Laboratory, Golden, CO, USA, 2009.

[11] Anaya, M.; Tibaduiza, D.; Pozo, F. A bioinspired methodology based on an artificial immune system for damage detection in structural health monitoring. *Shock Vibration* **2015**, *2015*, 1–15.

[12] Anaya, M.; Tibaduiza, D.; Pozo, F. Detection and classification of structural changes using artificial immune systems and fuzzy clustering. *Int. J. Bio-Inspired Comput.*, in press.

[13] Mujica, L.E.; Ruiz, M.; Pozo, F.; Rodellar, J.; Güemes, A. A structural damage detection indicator based on principal component analysis and statistical hypothesis testing. *Smart Mater. Struct.* **2014**, *23*, 1–12.

[14] Mujica, L.E.; Rodellar, J.; Fernández, A.; Güemes, A. Q-statistic and T^2-statistic PCA-based measures for damage assessment in structures. *Struct. Health Monit.* **2011**, *10*, 539–553.

[15] Odgaard, P.F.; Lin, B.; Jorgensen, S.B. Observer and data-driven-model-based fault detection in power plant coal mills. *IEEE Trans. Energy Convers.* **2008**, *23*, 659–668.

[16] Ugarte, M.D.; Militino, A.F.; Arnholt, A. *Probability and Statistics with R*; CRC Press (Taylor & Francis Group): Boca Raton, FL, USA, 2008.