ANOVA and other statistical tools for bearing damage detection

ANOVA and other statistical tools for bearing damage detection

Alessandro Paolo Daga, Luigi Garibaldi, Alessandro Fasana, Stefano Marchesiello
DIRG Group, DIMEAS
Department of Mechanical and Aerospace Engineering,
Politecnico di Torino, Italy

Abstract

The aim of the paper is to exhaustively exploit and test some statistical tools, such as ANOVA and Linear Discriminant Analysis, to investigate a massive amounts of data collected over a rig available @DIRG Lab, specifically conceived to test high speed aeronautical bearings; the rig permits the control of rotational speed (6000 – 30000 RPM), radial load (0 to 1800 N) and temperature, and allows monitoring vibrations by means of 4 tri-axial accelerometers.

Fifteen different damages have been realised on the bearing but, for simplicity, this papers only treats those cases where simple identification methods have failed or not demonstrated to be fully affordable. The damages have been inferred on rolls or on the internal ring, with different severities, which are reported as a function of their extension, i.e. 150, 250, 450 µm. A total number of 17 combinations of load and speed have been analysed per each damaged bearing.

Although ANOVA rigorously applies when some conditions are respected on the probability distribution of the responses, such as Independence of observations, Normality (normal distribution of the residuals) and Homoscedasticity (homogeneity of variances – equal variances), the paper exploits the robustness of the technique even when data do not fully fall into the requisites.

Analyses are focused on the best features to be taken into account, trying to seek for the most informative, but also trying to extract a “best choice” for the acceleration direction and the most informative point to be monitored over the simple structure.

Wanting to focus on the classification of the single observation, Linear Discriminant Analysis has been tested, demonstrating to be quite effective as the number of misclassification is not very high, (at least considering the widest damages).

All these classifications have unfortunately the limit of requiring labelled examples. Acquisitions in un-damaged and damaged conditions are in fact essential to guarantee their applicability, which is quite often impossible for real industrial plants.

The target can be anyway reached by adopting distances from un-damaged conditions which, conversely, must be known as a reference. Advantages of the statistical methods are quickness, simplicity and full independence from human interaction.

Sommario

Introduction 3
The experimental setup 3
The data 4
The Analysis 4
ANOVA 5
Multi-comparison test: Fisher's Least Significant Difference (LSD) 6
Multivariate analysis: LDA classification 7
Principal Component Analysis (PCA) 8
Outlier Analysis 8
The Results 9
ANOVA 10
Classification 11
Outlier Analysis 12
Conclusions 15
References 16

Introduction

Diagnosis of rotating machinery is becoming every day more significant, thanks to many high-level techniques able to reveal information on the exact position of the fault, the amount of damage and allowing forecasts about the expected time to failure.

Unfortunately, these procedures often need to be human supervised and the computational burden makes them unsuitable for real-time implementation. In this respect, lower level algorithms, devoted to disclose a fault presence, could be a good alternative, for example as a continuous monitoring preliminary analysis.

With a particular focus on bearing damages, which are usually difficult to detect, some basic statistical tools will be then introduced and tested, to understand whether it's possible to effectively recognise the presence of a fault and with which efficiency and reliability.

The experimental data used in this discussion refers to the test rig at the Department of Mechanical and Aerospace Engineering of Politecnico di Torino, shortly introduced hereinafter.

The experimental setup

The considered test rig consists of a direct drive rotating shaft supported by two bearings, one of which (the farthest from the motor) will exhibit different damage levels. A third central bearing is used to load the shaft with an increasing force of 0, 1000, 1400 and 1800 N, while the speed is set at four different values of about 90, 180, 280, 370, 470 Hz for a total number of 17 combinations of load and speed.

The structure is equipped with four tri-axial accelerometers (positioned as reported in Figure 1) sampled at a frequency fs = 51200 Hz for a duration of T = 10 s.

In this paper, acquisitions form six differently damaged high-speed aeronautical bearings (Table 1) will be investigated, relying on the information from two accelerometers placed on the main supports (position 1 and 3).

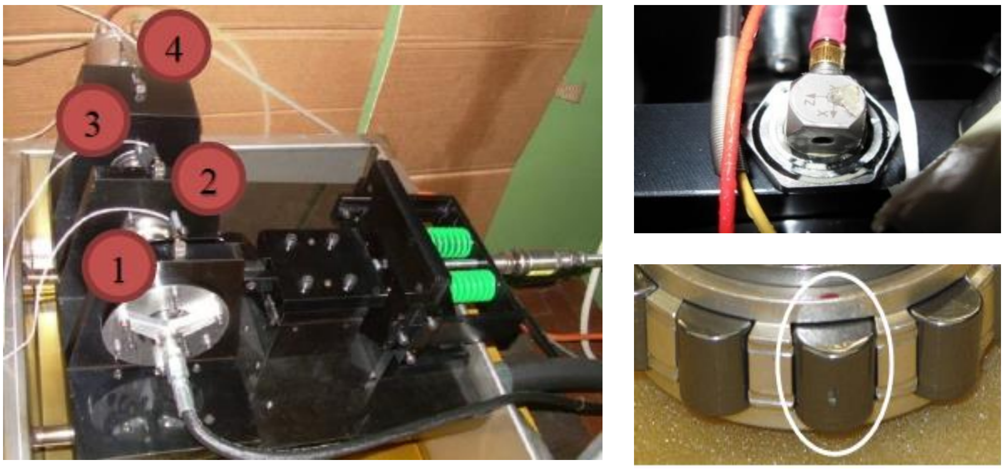


Figure 1: The test rig, the orientation of the triaxial accelerometers and the 4A damaged roller.

Table 1: Bearing codification according to damage type (Inner Ring or Rolling Element) and size.

Code	***0A***	***1A***	***2A***	***3A***	***4A***	***5A***	***6A***
Damage type	***none***	***I.R.***	***I.R.***	***I.R.***	***R.E.***	***R.E.***	***R.E.***
Damage size [µm]	***-***	***450***	***250***	***150***	***450***	***250***	***150***

The data

In order to explore the available data in an efficient way, some features have been extracted. This allows to point out and summarize the hidden information using few characteristic parameters.

Wanting to cope with the need of speed and automation of the analysis, simple time-series features have been selected, such as *root mean square*, *skewness*, *kurtosis*, *peak value* and *crest factor* (peak/RMS).

Those features have been computed on shorter independent chunks (no overlap) of the original available data to ensure statistical reliability; the number of subdivisions has been chosen with a particular care, so as to balance the significance both on the features extraction, and on the further analysis.

According to these consideration, each of the 17 acquisitions (see Table 2) have been subdivided in one hundred 0,1 s parts.

Table 2: The operational conditions

Label:	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***13***	***14***	***15***	***16***	***17***
f (Hz)	***100***	***100***	***100***	***100***	***200***	***200***	***200***	***200***	***300***	***300***	***300***	***300***	***400***	***400***	***400***	***500***	***500***
F (N)	***0***	***1000***	***1400***	***1800***	***0***	***1000***	***1400***	***1800***	***0***	***1000***	***1400***	***1800***	***0***	***1000***	***1400***	***0***	***1000***

We obtained then 1700 points in a 30 dimensional space (6 channels, 5 features) per each of the 7 damage conditions, from healthy to 6A (Table 1).

The Analysis

In order to preliminary explore the available data, some established statistical tools have been exploited.

At first, a univariate Analysis Of Variance (ANOVA) was employed, together with the usual post hoc, multi-comparison tests, to infer the omnibus (*variance based*) null hypothesis (*presence of a relationship among groups – in this case damaged conditions*) from the data.

Then, the Linear Discriminant Analysis (LDA) algorithm was adopted, with the aim of understanding whether simple multivariate classification routines are able to distinguish among the different damage levels, and at which misclassification rates.

Being a multivariate problem of dimension 30 (6 channels, 5 features), a Principal Component Analysis (PCA) was also proposed with the aim of visualizing the data.

The main critical point of those analyses is that damaged acquisitions are not always available in real applications, in particular for big, expensive machines, whose damaged condition is very dangerous and cannot be investigated.

In this case, some unsupervised techniques (no labels) referring only to the healthy-normal condition are preferred. An outlier analysis was then performed, simply disclosing a fault presence as a deviation from normality, through Mahalanobis Distance.

ANOVA

ANOVA is a statistical inference tool able to deduce properties of the underlying distributions by an analysis of data. Practically it tests the omnibus null hypothesis H0 that all the different groups' populations have equal mean values, meaning that no significant difference is detectable.

Obviously, this could turn useful to test whether a damaged distribution differs from the healthy one, opening to a possibility of a discrimination at least in terms of distribution, if not for the single data points. Being ANOVA a univariate technique, it should be repeated per each channel and feature (30 combinations), so that it would be possible to make some considerations about the more relevant channels and features.

Unfortunately, even in case of H0 rejection, ANOVA is not able to provide additional information about which population differs the most and from which one of the others. Multiple two-sample tests (ANOVA reduces to a Student's t-test in this case) could be performed, but this would increase the chance of committing statistical type I error (false rejection of H0), so the ordinary procedure requires the more reliable post hoc tests, allowing multiple comparisons.

ANOVA assumes a linear model according to which, an observation of the j^{th} group will be found as a random draw from a normal distribution around the group mean μ_j (often called treatment).

$$y_{ij} = \mu_j + \varepsilon_{ij} \xrightarrow{\varepsilon_{ij} \sim N_j(0, \sigma_j) \qquad \sigma_j = \sigma \; \forall \, j = 1{:}G} \sigma_t^2 = \sigma_{wg}^2 + \sigma_{bg}^2$$

In the case of independent observations with homogeneous variances (homoscedasticity: equal $\sigma_j \forall$ group j), the overall variance σ_t^2 can be divided into a *within groups variance* σ_{wg}^2, approximately an average of the σ_j^2 weighted according to the numerousness n_j of each group, and a *between groups variance* σ_{bg}^2, the squared deviation of the groups means μ_j from the overall mean $\overline{y}$.

When σ_{bg}^2 and σ_{wg}^2 are statistically equal, their ratio is distributed according to the Fisher-Snedecor's $F_{(G-1,N-G)}$, so they can be compared with a *Fisher's F test*, which outputs a p-value, namely the probability that the statistical summary F, would be the same as, or more extreme than, the actual observed results.

$$\sigma_{bg}^2 = \frac{1}{G-1}\sum_{j=1}^{G}(\overline{y} - \mu_j)^2 \qquad \sigma_{bg}^2 \sim \chi^2_{(G-1)}$$

$$F = \frac{\sigma_{bg}^2}{\sigma_{wg}^2} \sim F_{(G-1,N-G)}$$

H0: same mean if
$$\boldsymbol{F \leq F^{\alpha}_{(G-1,N-G)}}$$
(critical value)

$$\sigma_{wg}^2 = \frac{1}{N-G}\sum_{j=1}^{G}\sum_{i=1}^{n_j}(y_{ij} - \mu_j)^2 \qquad \sigma_{wg}^2 \sim \chi^2_{(N-G)}$$

N: overall numerousness, G: number of groups;
ANOVA simplified formulas for equal $n_j = n$ in each group

Therefore, when the p-value is lower than $\alpha = 5\%$, the hypothesys of equal variances is rejected, together with the ANOVA H0 null hypothesis, at a confidence of $1 - \alpha = 95\%$ (Figure 2).

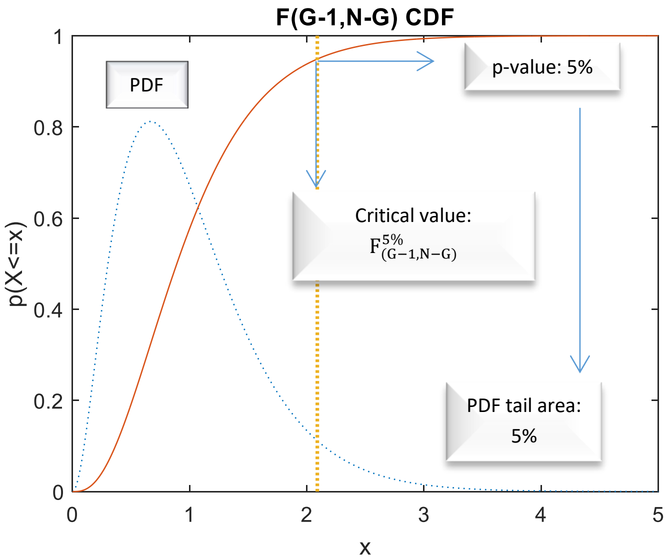


Figure 2: Fisher-Snedecor's $F_{(6,11839)}$ critical value for a p-value of 5% - one sided test

It is worth to remember that the ratio F can be used as a measure of separation among two groups distributions; in fact the farther two classes are, the bigger will be the between-groups variance, leading to more extreme values of F, which corresponds to lower p-values.

Multi-comparison test: Fisher's Least Significant Difference (LSD)

As previously introduced, ANOVA simultaneously infers the equality of the means for all the groups together. In order to decompose the analysis and compare the damaged conditions with the healthy, multiple comparisons are necessary. One of the most common tests to accomplish this task is the Fisher's Least Significant Difference (LSD). Considering that ANOVA reduces to a Student's t test if only two groups at a time are tested (in terms of distributios: $F_{(1,N-2)} \equiv t^2$), a simple idea is to use a set of individual t-tests. Fisher simply generalized these tests referring them to the pooled estimate of the standard deviation from all groups, so as to reduce the type I error (incorrect H0 rejection).

Student's t test: *Fisher's generalisation:* *Fisher's LSD:*

$$t = \frac{\mu_i - \mu_j}{\sqrt{\frac{\sigma_i^2 + \sigma_j^2}{n}}} \sim t_{(2n-2)} \longrightarrow t = \frac{\mu_i - \mu_j}{\sqrt{\frac{\sigma_{wg}^2}{n}}} \sim t_{(N-G)} \qquad LSD = |\mu_i - \mu_j| \leq t^{\alpha/2}_{(N-G)} \sqrt{\frac{\sigma_{wg}^2}{n}}$$

H0: same mean if $\quad t \leq t^{\alpha/2}_{(2n-2)}$

Simplified formulas for equal $n_j = n$ *in each group*

Therefore it's possible to build a limit range around the mean of each group of $\pm {}^{LSD}/_{2}$; intersecating groups will be then considered not significally distant, meaning that it will be hard to recognize them with enough confidence.

Multivariate analysis: LDA classification

In order to "fuse" the information from all the sensors and all the features, and then improving the classification, a multivariate analysis of variance (MANOVA) could be proposed. Unfortunately, it shows the same limitations of ANOVA, so instead of focusing on p-values of difficult interpretation, a multivariate classification has been preferred. Extending the considerations introduced for ANOVA, a Fisher's Linear Discriminant Analysis (LDA) have been proposed.

It must be kept in mind that such a classification, as previously introduced analyses, still requires labelled samples (it's still supervised); after building the classifier function anyway, the discrimination will be pointwise and no more in terms of distributions.

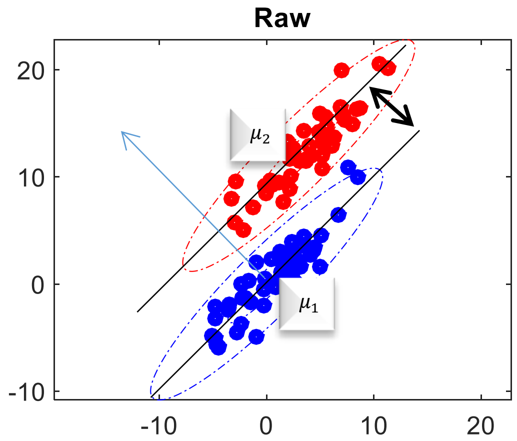


Figure 3: LDA idea - find the direction on which the projected samples will show the higher separation

LDA is simply a search for the projection ***w*** which maximizes the difference between the projected class means, normalized by a measure of the within-class scatter along the direction of ***w***. Thus the measure of separation is always given by the ratio $\sigma_{bg}^2/\sigma_{wg}^2$, generalizable for a multivariate space.

Considering only 2 groups, for the sake of simplicity:

Between class covariance matrix:

$$\overline{\overline{S_{bg}}} = (\overline{\mu_2} - \overline{\mu_1})(\overline{\mu_2} - \overline{\mu_1})'$$

Within class covariance matrix:

$$\overline{\overline{S_{wg}}} = \sum_{k \in C_1} (\overline{y^k} - \overline{\mu_1})(\overline{y^k} - \overline{\mu_1})' + \sum_{k \in C_2} (\overline{y^k} - \overline{\mu_2})(\overline{y^k} - \overline{\mu_2})'$$

$$J(w) = \frac{w'\overline{\overline{S_{bg}}}w}{w'\overline{\overline{S_{wg}}}w}$$

$$\arg\max_w J(w):$$

$$w \propto \overline{\overline{S_{wg}}}^{-1}(\overline{\mu_2} - \overline{\mu_1})$$

Extending it to multiple classes, it's possible to prove that, when ***w*** is an eigenvector of $\overline{\overline{S_{wg}}}^{-1}\overline{\overline{S_{bg}}}$, the separation will be equal to the corresponding eigenvalue:

$$w \propto PC\ of\ \overline{\overline{S_{wg}}}^{-1}\overline{\overline{S_{bg}}}$$

This algorithm, although very interesting from a theoretical point of view, does not usually perform well, as it expects linear separation among the classes.

Principal Component Analysis (PCA)

The PCA is a technique that uses an orthogonal space transformation to convert a set of correlated quantities into uncorrelated variables called principal components. This transformation is basically a rotation of the space in such a way that the first principal component will explain the largest possible variance, while each succeeding component will show the highest possible variance under the constraint of orthogonality with the preceding ones. This is usually accomplished by eigenvalue decomposition of the data covariance matrix or singular value decomposition of the data matrix, usually after mean centering. PCA is sensitive to the relative scaling of the original variables, so a data normalization is often advisable (equivalent to use Z-scores). Alternatively, the data correlation matrix could be used.

In general, the main application of PCA is for reducing a complex data set to a lower dimension, revealing the sometimes hidden, simplified dynamics. This dimensionality reduction will be performed simply by focusing on the first few components that explain the majority of the variation, while neglecting the others.

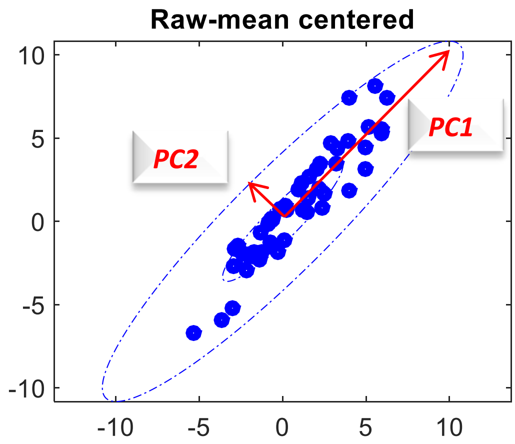


Figure 4: PCA

Although very useful for data visualization, for diagnostic purposes this variance-based dimensionality reduction is not really helpful as the condition-information is likely to be neglected, making the detection more challenging.

Outlier Analysis

Up to this point, all the methods proposed required labelled samples from all the damaged conditions (supervised techniques). Since this information is not always available in real applications, unsupervised options should be evaluated. The natural alternative to classification would be the *cluster analysis*, but, as the main purpose of condition monitoring is the early recognition of damages, possibly at the onset, the outlier analysis, trained only with data from the healthy-reference condition (training set), has been preferred.

In general, in a data set, a discordant measure is defined “outlier”, when, being inconsistent with the others, is believed to be generated by an alternate mechanism. The judgment on discordancy will depend on a measure of distance from the reference distribution, usually called Novelty Index (NI), on which a threshold can be defined [3].

The Mahalanobis distance (MD) is the optimal candidate for evaluating discordancy in a multi-dimensional space, because it is unitless and scale-invariant, and takes into account the correlations of the data set.

$$DM(X) = \sqrt{(X-\mu)^T S^{-1} (X-\mu)}$$

The Novelty Indices computed with Mahalanobis distance can be compared against some objective criterion (a threshold) to judge whether the corresponding data comes from the healthy distribution; furthermore, even for graphical purposes, these NI are the optimal univariate tool to display possible outliers of a multivariate dataset.

Unfortunately, the procedure to generate a suitable threshold is not trivial, as the distribution of the healthy data may be in general non-normal. In this respect, probability theory offers some good hints. For example, Chebyshev's inequality could be used to fix a limit, given that, for a wide class of probability distributions, "nearly all values are close to the mean — no more than a fraction of $1/k^2$ of the distribution's values can be more than k standard deviations away from the mean". This hypothesis in many cases overestimates the extreme values probabilities (for example for an ideal normal distribution, the tails decay more rapidly than that), but it helps understanding that in most of real-case distributions, tails are fatter than expected and particular attention should be kept.

To improve the thresholding operation, several repeated Monte Carlo simulations (MC) of a p-dimensional Gaussian distribution could be performed. Drawing n observations in p variables and computing the NIs, the maximum operator could be used to generate a robust threshold, for example taking the 99th percentile of the maxima distribution [3].

Furthermore, the use of Mahalanobis distance based Novelty Indices has even the advantage of compensating for limited operational and environmental conditions variation under linearity (or quasi linearity) assumption.

In fact, keeping in mind that the Mahalanobis distance is equivalent to an Eulerian distance on a transformed, standardized space, Novelty Indices computed in such a way will already account for a compensation, which is a consequence of the eigenvectors normalization on their own standard deviation.

The Results

After the feature extraction, the huge amount of data have been organized in 7 matrices (0 to 6A damages) of 1700 rows (17 speed and load combination, 100 samples each) and 30 columns (6 channels, 5 features each), corresponding to 1700 points in a 30 dimensional space.

The analysis started then with a preliminary visual exploration of the healthy data after PCA: the samples have been summarized in the 2D plane generated by the first two principal components, reported in Figure 5.

From the picture it is concluded that the effect of speed is much more relevant than the load, in fact, if we neglect the condition of zero loading (labels 1,5,9,13, in any case not very common), the data would be clustered in equal speed subgroups, almost regardless from the load (see 2-3-4, 6-7-8, 10-11-12 and 14-15 clusters).

This is obviously just a simplifying projection, neglecting a lot of information, anyhow it is useful to underline that the working conditions strongly affect the data distribution, so that it would be wiser to make comparisons at the same operational conditions.

Figure 6 shows for example the data referred to acquisition 12 (300 Hz, 1800 N), at different damage levels, after removing the mean value of the healthy data. It's easy to notice how the most damaged conditions (1A and 4A) will be the furthest from the healthy. This will be enhanced by the following outlier analysis.

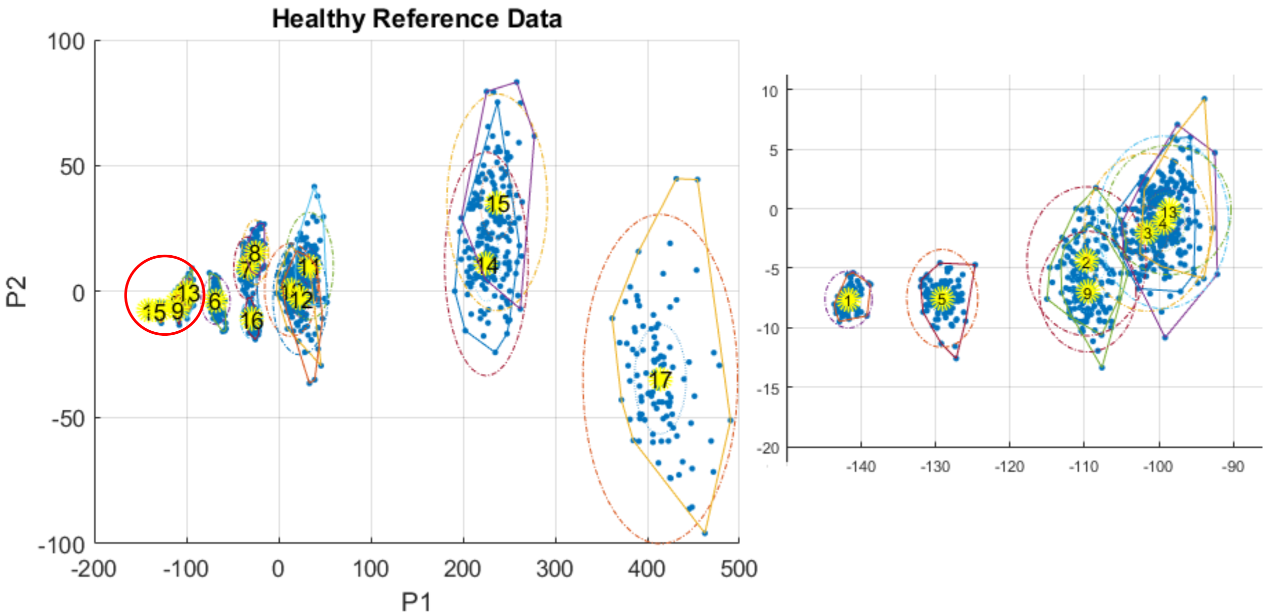


Figure 5: Healthy data for the 17 speed and load combinations (Table 2) _ 68% and 99% (1 and 3 sigma equivalent) ellipses compared to the convex hull.

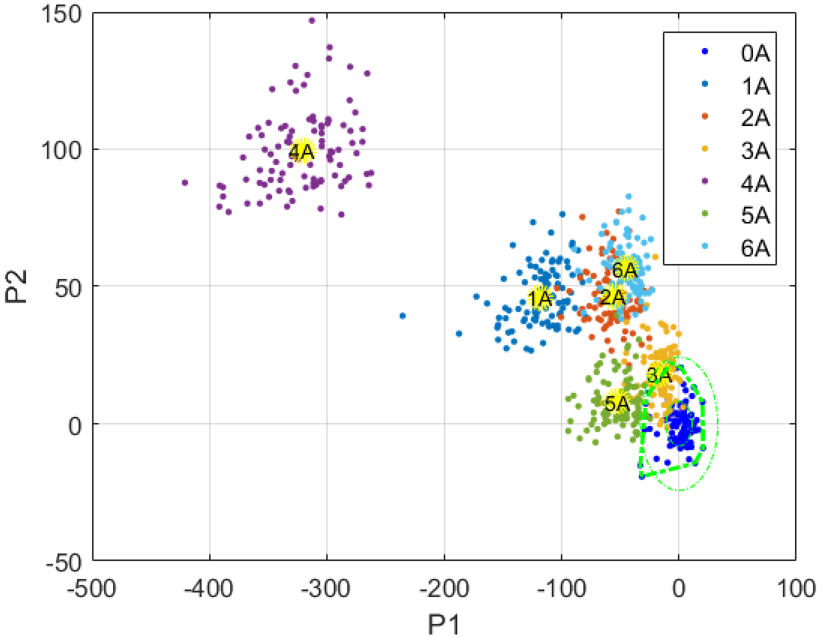


Figure 6: Healthy data compared to damaged acquisitions, centered on the same reference – work condition 12.

ANOVA

In order to preliminary evaluate the features performance and assess which sensor positioning and direction could enhance the damage detection, the Analysis Of Variance have been conducted on single channel, single feature distributions, for the entire set of 1700 points per 7 damaged conditions.

Although the assumptions of normality and homoscedasticity were not completely met, ANOVA is generally considered robust to those kind of violations, in particular for the case in which all the groups under analysis show equal numerousness, so the method was deemed to be accepted.

In all the 30 tests, the ANOVA p-values resulted almost negligible, so it was normal to focus on the multicomparison post-hoc test reported in Figure 7 which resulted much more informative.

In terms of features, it's easy to notice that *kurtosis* and *crest* are the best, as they are able to always discriminate 1A and 4A conditions (highest damage) from the others; furthermore, they seem more consistent with the damage. For example, focusing on channel 4, it's easy to notice a linear trend of distance and damage level.

However the other features seem to add some information, in particular for the less damaged conditions, so that it would not be wise to ignore them.

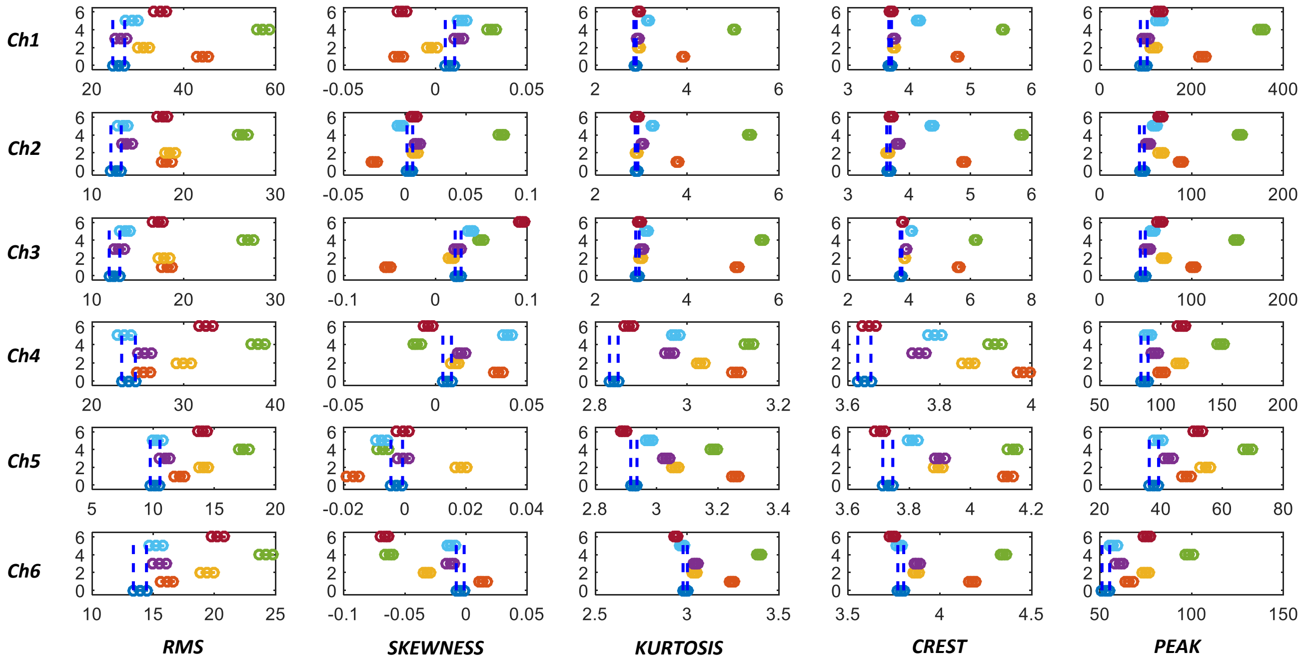


Figure 7: ANOVA post-hoc, Multicomparison result. For different channels and features, all the 6 damage conditions are compared to the healthy reference (0) through LSD limits.

Classification

In order to test if it is possible to discriminate the different damage conditions using the information from the 6 channels and the 5 features altogether, a multivariate classification have been proposed.

The overall data set has been divided in two parts: the first 60 samples per each of the 17 operational conditions were used to train a classifier, which have been tested on the same set (in sample), but also on the remaining points used as a validation set (out of sample).

The performances have been evaluated through confusion matrices showing the approximated percentage of classified samples against the true, target class (read by rows). The further this matrix is from the identity, the worst the classification.

Focusing on Table 3, it’s possible to notice that a linear discriminant, although quite good in recognizing 1A and 4A conditions, shows some troubles in the other cases, in particular for correctly classifying undamaged acquisitions (only 50% of “healthy points” were really un-damaged).

Table 3: LDA confusion matrices

LDA		*in sample (training)*							*out of sample (validation)*						
		Target Class							Target Class						
		0A	1A	2A	3A	4A	5A	6A	0A	1A	2A	3A	4A	5A	6A
Output Class	0A	51	0	10	12	0	9	15	48	0	9	15	0	12	14
	1A	12	73	0	4	1	6	0	12	73	1	4	1	6	0
	2A	14	0	59	13	0	4	7	16	0	52	17	0	5	7
	3A	25	0	11	43	0	11	7	22	0	12	44	0	13	7
	4A	1	4	0	0	81	10	0	1	4	0	1	80	11	0
	5A	17	1	6	5	0	63	5	15	1	8	8	0	61	4
	6A	13	0	8	12	0	7	57	14	0	8	12	0	7	57

Outlier Analysis

As previously introduced, in many cases damaged acquisitions are not available, so an Outlier Analysis (based only on a healthy training set) seems ideal. This procedure, based on Mahalanobis distance, is very fast and effective in pointing out discordant measures, but it must be kept in mind that novelty does not necessarily imply the presence of a damage, as even variations of operational or environmental conditions would lead to similar deviations.

To underline this fact, a first analysis on the entire data set has been conducted, as reported in Figure 8. In this case the strong effect of the different operational conditions can be seen both on the healthy, and on the damaged cases. Even if the biggest damages can be recognized quite effectively, almost all the other conditions will show a high rate of Missed Alarms. Furthermore many outliers will trigger False Alarms even in the training set.

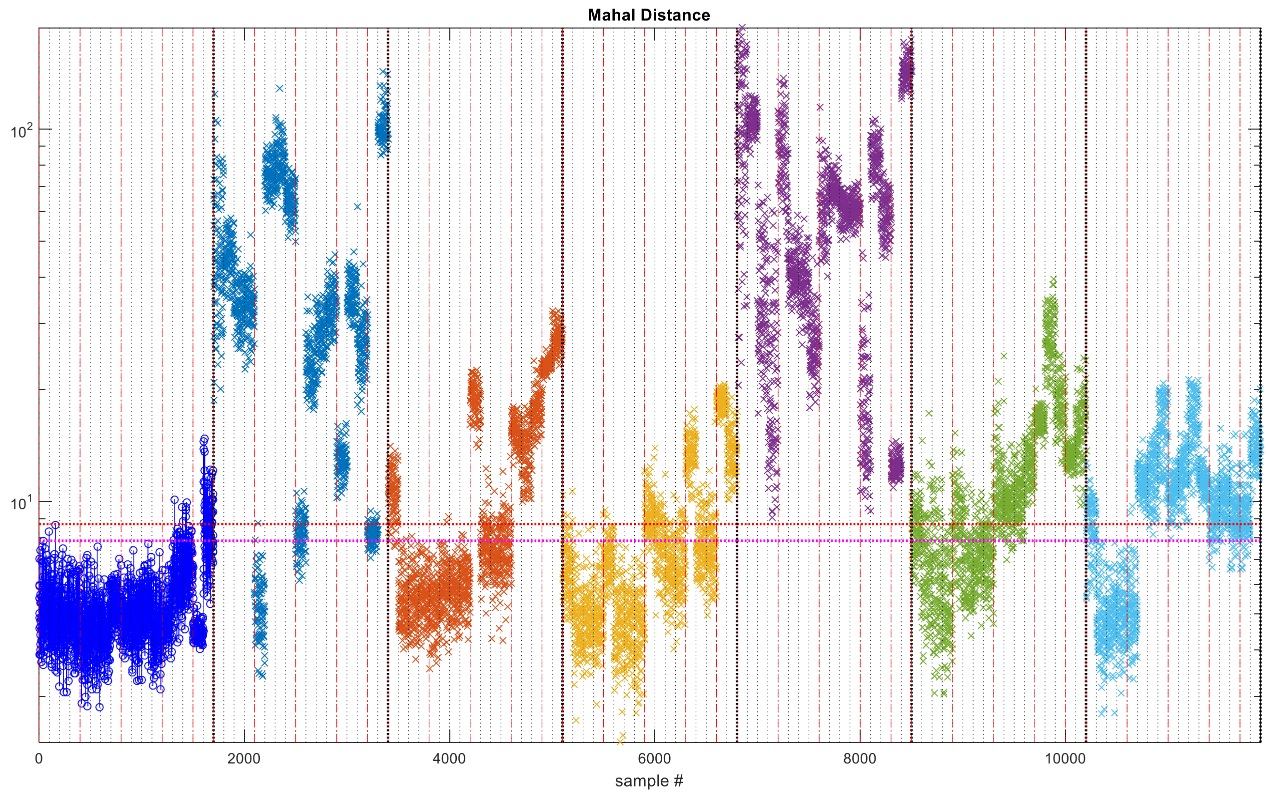
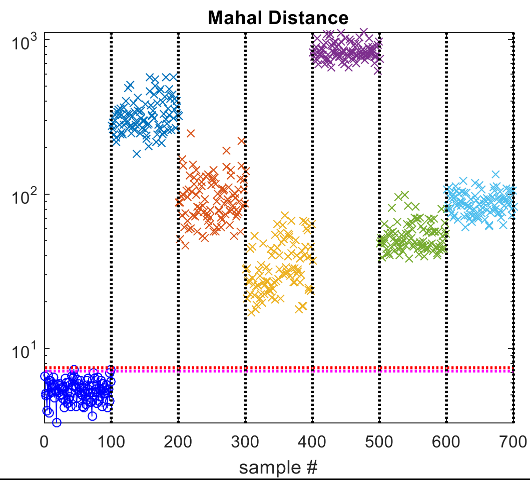
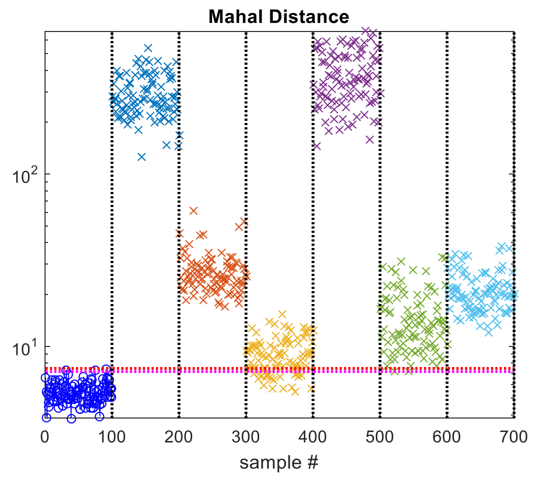


Figure 8: Mahalanobis distance for the entire available data altogether, without considering the different operational conditions; in red the 99th percentile of the maxima distribution from 1000 Monte Carlo repetitions, used as a threshold.

It is worth then to repeat the analysis on each of the 17 available work conditions separately. The results in terms of False and Missed Alarms are reported in Table 4 and graphically in Figure 9. In this case it's easy to notice that the performances are really improved, in fact, the alarm rates is very satisfying for almost all the conditions (just at low speed, condition 3 and 4, seem to have difficulty to diagnose 3A condition). Additionally, in many cases (look at Figure 9 - condition 12) the damaged condition show a wide distance from the healthy, so that the threshold could be increased to annihilate the FA rate, without increasing the MA. It's interesting to notice that these Novelty Indices are even consistent with the damage, as they change almost monotonically with the severity. This could be used to diagnose not only the presence, but even the size of a damage.





Work condition 12	*Work condition 3*

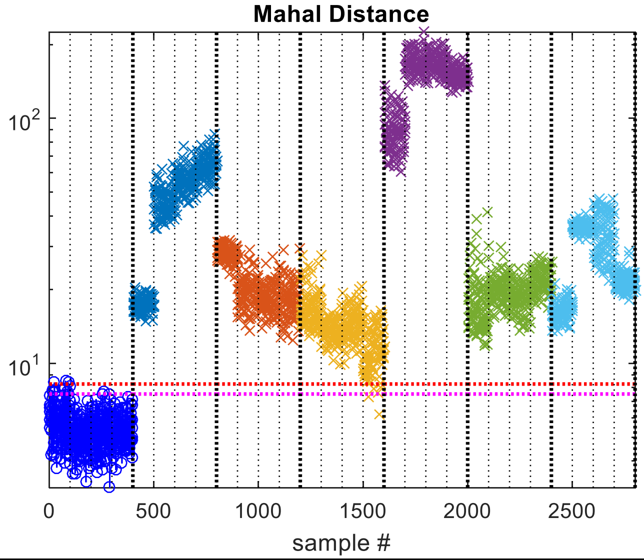
Figure 9: Mahalanobis distance for certain operational conditions.

Table 4: False and Missed Alarms for the 17 operational conditions, considered independently and compared to their own reference healthy acquisitions (see Figure 9); the 99th percentile of the maxima distribution from 1000 Monte Carlo repetitions, used as a threshold, is reported as well.

	FA	*MA*						*MC's 99%*
	0A	***1A***	***2A***	***3A***	***4A***	***5A***	***6A***	*threshold*
1	2	0	0	0	0	0	0	7,42
2	2	0	0	0	0	0	0	7,37
3	3	0	0	21	0	2	0	7,38
4	4	0	0	7	0	0	0	7,46
5	3	0	0	0	0	0	0	7,42
6	1	0	0	0	0	0	0	7,43
7	1	0	0	0	0	0	0	7,37
8	3	0	0	0	0	0	0	7,41
9	3	0	0	0	0	0	0	7,44
10	1	0	0	0	0	0	0	7,37
11	4	0	0	0	0	0	0	7,41
12	2	0	0	0	0	0	0	7,39
13	4	0	0	0	0	0	0	7,43
14	2	0	0	0	0	0	0	7,49
15	4	0	0	0	0	0	0	7,42
16	1	0	0	0	0	0	0	7,40
17	0	0	0	0	0	0	0	7,38
referred to a 100 points sample: values can be considered as %						*average threshold:*		7,41

In order to show the ability of MD Outlier Analysis of compensating for reduced variations around a nominal condition (linear or quasi linear effect), a test have been conducted on 4 sets at constant speed but with changing load (acquisitions 9 to 12, at 300 Hz, while load ranges from 0 to Max).

The results are reported in Figure 10.



Alarms %	
0A	3 %
1A	0 %
2A	0 %
3A	0,5%
4A	0 %
5A	0 %
6A	0 %
MC's 99% threshold: 8,2	

Figure 10: Mahalanobis distance for operational conditions 9-12 (constant speed 300 Hz).

Conclusions

Different supervised and unsupervised techniques were tested in this paper, to compare their performances in extracting information from a big, high dimensional data set.

As a preliminary analysis, PCA has been used to reduce the dimensionality and visualize the entire data set in a 2D representation.

Then 30 univariate analyses of variance have been conducted, together with their corresponding multicomparison tests. This highlighted that all the channels and all the features were able to give different kind of information for different damages, so it was wise to fuse all this intelligence with multivariate tools.

LDA was then tested, proving that in the 30 dimensional space, the different speed-load and damage conditions were quite well recognizable even if not really linearly separated.

A Mahalanobis distance based outlier analysis was then finally proposed as an unsupervised technique to detect deviations from the healthy condition. Selecting as reference a single operational condition, and comparing its relative damaged sets, very good results were obtained in terms of False and Missed Alarms rate, and the damage severity was deducible as well. Taking advantage of the intrinsic ability of compensating for simple linear or quasi linear hidden effects, this analysis proved to be good even for groups of acquisitions at constant speed but variable load. Unfortunately, if the entire 17 operational conditions were considered altogether, the method could well recognize only 4A and 1A damages (the more severe), reducing its reliability.

These results are anyway very good if considering also the quickness of the algorithms, the simplicity and the full independence from human interaction, which make them suitable for real time implementation.

References

[1] Camila Gianini Gonsalez, Samuel da Silva, Michael J. Brennan, Vicente Lopes Junior, "*Structural damage detection in an aeronautical panel using analysis of variance*", Mechanical Systems and Signal Processing 52-53 (2015).

[2] A.-M. Yan, G. Kerschen, P. De Boe, J.-C. Golinval, "*Structural damage diagnosis under varying environmental conditions—Part I: A linear analysis*", Mechanical Systems and Signal Processing 19 (2005).

[3] K. Worden, G. Manson, N. R. J. Fieller, "*Damage detection using outlier analysis*" , Journal of Sound and Vibration (2000).

[4] K. Worden , S. G. Pierce , G. Manson , W. R. Philp , W. J. Staszewski & B. Culshaw , "*Detection of defects in composite plates using Lamb waves and novelty detection*", International Journal of Systems Science, 31:11, 1397-1409, (2000).

[5] K. Worden, D. W. Allen, H. Sohn, C. R. Farrar., "*Damage detection in mechanical structures using extreme value statistics*" Proceedings of SPIE - The International Society for Optical Engineering, 4693, 289–299, (2002).

[6] D. Toshkova, N. Lieven, P. Morrish, P. Hutchinson, "*Applying Extreme Value Theory for alarm and warning levels setting under variable operating conditions*", EWSHM, (2016).

[7] A. Bellino, A. Fasana, L. Garibaldi, S. Marchesiello "*PCA-Based detection of damage in time-varying systems*", Mechanical Systems and Signal Processing vol. 24 (2010).

[8] Daga A. P., Marchesiello S., Fasana A., Garibaldi L. "*Bearing damage detection techniques and their enhancement: comparison over real data*", Surveillance 8 Conference, (2015).

[9] *Decisions under Uncertainty-Probabilistic Analysis for Engineering Decisions*, Ian Jordaan, Cambridge University Press, (2005).

[10] *Principal Component Analysis - Second Edition*, I.T. Jolliffe, Springer

[11] *Exploratory Data Analysis with MATLAB®,* Wendy L. Martinez, Angel R. Martinez, Jeffrey L. Solka, CRC Press