Big Data management: A Vibration Monitoring point of view

Big Data management: A Vibration Monitoring point of view

Alessandro Paolo Daga
Politecnico di Torino
Torino, Italy
0000-0002-5341-7710
alessandro.daga@polito.it

Alessandro Fasana
Politecnico di Torino
Torino, Italy
0000-0001-5851-9890

Luigi Garibaldi
Politecnico di Torino
Torino, Italy
0000-0001-9415-1915

Stefano Marchesiello
Politecnico di Torino
Torino, Italy
0000-0001-6906-748X

***Abstract*—Vibration Monitoring is a particular kind of Condition Monitoring meant to infer the state of health of a machine from accelerometric measurements. From a practical point of view, the scope is then to extract from the acceleration data some valuable diagnostic information which could be used to detect the presence of possible damages (i.e., to produce knowledge about the state of health). When the monitoring is implemented online, in a continuous way, the raw accelerometric data sets can be very large and complex to be dealt with, as usually involve multiple channels (i.e., multiple locations and directions) and high sample rates (i.e., order of ksps - 10^3 samples per second), but the final knowledge about the state of health can, in principle, be summarized by a single binary information (i.e., healthy – 0 vs damaged – 1). This is commonly called Damage Detection. In this work, the big data management challenge is tackled from the point of view of statistical signal processing, so as to aggregate the multivariate data and condense them into single information of distance with respect to a healthy reference condition (i.e., the Novelty). When confounding influences (such as the work condition or the environmental condition) can be disregarded, the novelty information has a direct correspondence to the health information, so that an alarm indicating the detection of damage can be triggered upon exceeding a selected threshold for the limit novelty. Many different ways of solving such a binary classification problem can be found in the literature. Starting from the simplest, some of the more effective are compared in the present analysis, to finally select a reliable procedure for the big data management in vibration monitoring.**

Keywords—Vibration Monitoring, online monitoring, big data, lossy compression, novelty detection, confounders

I. INTRODUCTION

Nowadays industrial machines are expected to be operated at extreme operational conditions in a continuous way for years without maintenance [1]. Consequently, reliable tools able to assess the state of health of a machine while in operation are ever more important. Among the others, Vibration Monitoring (VM) is considered one of the more promising, both at the industrial level and at a research stage.

A particularly critical part of machines featuring rotating shafts is the transmission, in most of cases a gearbox, whose failure can prove both dangerous and expensive. Hence, this work will focus on such a fundamental component made of shafts, gearwheels and bearings.

Guidelines for measurement and evaluation of machine vibration can be found for example in [1]. By reading this industrial standard, interesting considerations regarding the most common analyses can be made:

- Vibration can be measured in terms of position, velocity or acceleration. Anyway, for rolling bearings the acceleration measurements are prescribed.
- Multiple locations and directions need to be measured.
- Even if in the annex B a brief description of higher-level techniques for rolling bearings can be found (e.g., Spectral Analysis, Envelope Analysis, Shock-Pulse techniques, etc.), the prescribed methodology is that of periodically extracting a single vibration level indicator, the broad-band root mean square (rms) acceleration, and to compare it to selected thresholds. In alternative, the peak acceleration values, the peak-to-rms ratio (i.e., the crest factor) and other features are suggested. Thresholds are set on the selected feature, so as to identify 4 zones (i.e., normal, acceptable, unsatisfactory, damaged) and trigger alarms (i.e., to provide a warning that a vibration level is reached or a significant change occurred) or trips (i.e., the magnitude of vibration is too high, the machine should be stopped).
- The reference condition should involve measurements in all the normal operational conditions (i.e., speed, load, pressure, temperature, etc.) at steady state, checking that the machine is well isolated from external vibration contributions.

Finally, if we consider the common vibration monitoring measurement setup, accelerometers are connected to analog-to-digital converters that sample the signals before storing them in digital data-storage devices. A high-level hardware of this kind, as the one used in this research, can easily sample 24-bit acquisitions at 51,2 kHz synchronously on multiple channels. In this condition, if a continuous monitoring is considered, the storage would be filled at a rate of:

$$m \cdot f_s \cdot n \tag{1}$$

given $n = 24\frac{bit}{sample} = 3\frac{B}{sample}$ (N.B., $1\,Byte = 8\,bit$), $m = 6$ channels (e.g. 2 triaxial accelerometers), $fs = 51200\frac{samples}{s}$, the storage filling rate would be then $921{,}6\,kB/s$ or $3{,}32\,GB/h$, which can be not easily sustainable for long times (i.e., finite storage) nor easily transmissible (i.e., transmission channels could have limited bandwidth).

Furthermore, in the simplest case, the scope is to perform Damage Detection, which consists of performing a binary classification: a "healthy" or a "damaged" label is assigned to a chunk of data. Then, it is common for industrial continuous monitoring equipment to store data in temporary buffers, compute the selected feature (e.g., the rms) and compare it to a given threshold which separate the "healthy" levels from the "anomalous" levels. Usually, in normal conditions, just the feature information is stored, while the vibration signal over time is saved only when some level is reached, to allow higher level analysis (e.g., spectra, envelopes, etc.)

In principle, this corresponds to an Anomaly Detection, or Novelty Detection, and can be used for Damage Detection when confounding influences such as the work condition or the environmental condition can be disregarded (N.B. the ISO standard prescribes to consider normal, steady state acquisitions), as the novelty information has then a direct correspondence to the health information. Anyway, such proposed Novelty Detection can be seen as a lossy (i.e., irreversible) compression of the data.

In information technology, lossy compression is commonly used to encode multimedia data (audio, video and images) so as to produce approximations which are close enough to the original data but require less storage. In this respect, one of the most effective compression algorithm is probably the discrete cosine transform (DCT), which, thanks to its compaction property, allows to find good approximations with few coefficients, and is used for example in baseline JPEG [2] and MP3 [3,4] standards. Despite commonly less effective in saving memory, lossless compression is also possible (e.g., see the FLAC audio standard [5] or the lossless JPEG [2]). This is usually performed in three steps [6]:

- Blocking: the signal is divided using rectangular time windowing (N.B., since the scope is not to perform spectral analysis, the rectangular window is selected for its simplicity)
- Linear Prediction: an autoregressive process is used to predict the signal from its past values, so that just few coefficients should be stored
- The residual from the prediction is entropy encoded to save storage space.

Notice that the first two points are conceptually identical to the ones used for lossy compression, as the linear prediction gives an autoregressive approximation of the signal. Furthermore, the first two points are also similar to the previously described industrial vibration monitoring procedure, as taking the rms of a sliding time-window corresponds to perform a blocking and to compute an approximation of the power envelope of the signal (N.B. the approximation is not of the signal itself, but of its power envelope, which generalizes the concept of signal amplitude into an instantaneous amplitude and is widely used in signal processing for diagnostics).

As highlighted in section B.2 of [1], various claims have been made in support of features different from the rms (i.e., the peak value, the crest factor, etc.). Nevertheless, even if one feature can outperform the others in some cases, no general rule can be found.

The idea pursued in this paper is then to turn the reliable, standard-compliant univariate analysis into a multivariate analysis able to condense the information from different channels and features into a single, optimal Novelty Index, improving both the diagnostics ability and the data compression rate.

II. METHODOLOGY

The proposed methodology aims to deal with multiple features extracted from multiple accelerometric acquisitions of sensors placed around a machine in several locations and directions after a blocking stage (i.e., time windowing to divide the signals in shorter chunks). Such a dataset will feature high redundancy, so that statistical tools can be used to decorrelate the data, highlighting hidden patterns which condense the information from the different sources.

In principle, from a mathematical point of view, the simplest way to do this is by using a linear projection operator that maps the variables of interest to a new variable which will be a linear combination of the others. The issue is then to find the direction along which the damage develops.

To better understand the idea, a 2-D visualization is reported in Fig. 1. A simulated bidimensional dataset is drawn from a bivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\sigma})$, with null mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\sigma}$ (the blue one in Fig.1), as if one vibration signal was acquired, divided in $n = 100$ chunks and two correlated features were extracted from each chunk.

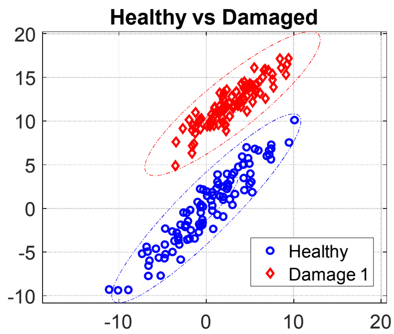

Fig. 1. Simulated dataset: draw from a bivariate normal distribution. Helathy: N([0;0],[13,12;12,13]), Damage 1: N([3;12],[11.2,8.2;8.2,7.5]).

Immediately it can be found that tilting these two axes approximately 45 degrees, it is possible to capture most of the variability along a single axis, as shown in Fig. 2. In fact, if the two features were perfectly correlated (e.g. they form perfect line) it could be possible to discard the second of the two tilted axes while still capturing the full distribution.

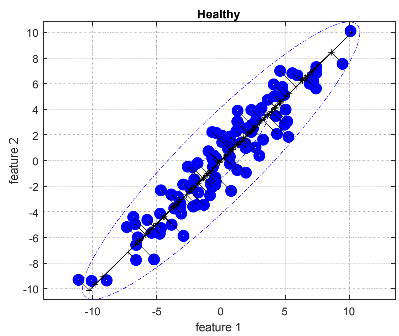

Fig. 2. Healthy dataset with a possible linear projection highlighted.

If confounders are negligible, the main axis must correspond to the direction along which damage develops; on the contrary, if the main axis pictures a variability related to a confounding influence, it may happen that damage will develop in a direction orthogonal or nearly orthogonal to the main axis. In both cases, a projection (i.e., a linear combination of features) in the damage evolution direction will be more informative than the two original features.

From a mathematical point of view, the Principal Component Analysis (PCA) can be used to find the linear projection operator mapping the original variables to a new coordinate frame where the axes represent maximal variability. This decorrelates the dataset, allowing the discovery of latent information. Furthermore, as the first principal components explain most of the variability, PCA can be used to reduce the data dimensionality to few dimensions which can be easily visualized in graphs. This corresponds to a lossy compression.

If acquisitions from a damaged state are available (N.B., this is not always possible for safety reasons), an additional step can be taken: the direction featuring maximum separation of the two classes (i.e., healthy or damaged) can be found at a training stage using Fisher's Linear Discriminant Analysis (LDA). LDA can be seen again as a lossy compression, as its projection results into a single variable obtained as a linear combination of the original variables.

The remaining problem is that damage is not bounded to develop according to a single particular direction. The proposal is then to use the Mahalanobis Distance to produce a Novelty Index which is sensitive to the distance from the healthy distribution centroid but not to the direction.

The involved algorithms are briefly described hereinafter and applied to the simulated dataset for demonstration.

A. Principal Component Analysis

Principal component analysis (PCA), also known as the Karhunen-Loeve transform, is a technique meant to find an orthogonal projection able to decorrelate a dataset. In the principal space in fact, the resulting features are uncorrelated, and ordered according to the explained variance. In this regard, PCA corresponds to a Singular Value Decomposition of the covariance matrix of the dataset, which is then diagonalized by the transform [7]. Indeed, given a d-dimensional dataset of n observations in matricial form $X \in R^{d \times n}$, the corresponding covariance matrix can be found by its centered version X_0 (i.e., the mean value of each of the d rows is removed) as

$$S = \frac{1}{n-1} X_0 X_0' \tag{2}$$

PCA corresponds then to the solution of the eigenproblem:

$$SV = V\lambda \tag{3}$$

where V is the orthogonal matrix whose columns are the d eigenvectors v_j while Λ is the diagonal matrix of the d eigenvalues λ_j of the matrix S, sorted to have descending magnitude.

The matrix V can be used as a linear transform to decorrelate the dataset X, that is, to rotate the coordinate system toward the principal directions identified by the eigenvectors of matrix S:

$$Z = V' X_0 \tag{4}$$

Z is then a $d \times n$ matrix whose rows contains the so-called principal scores, linear combinations of the original multivariate data, ordered according to the explained variance. The first principal scores $Z_1 = v_1' X_0$ can be then used as a first order approximation of the whole dataset.

PCA is visualized in Fig. 3 for the healthy dataset.

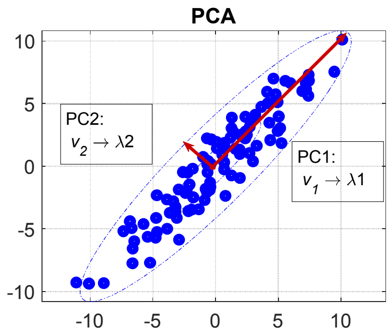

Fig. 3. PCA visualized on the healthy simulated data.

B. Fisher's Linear Discriminant Analysis

When acquisitions from a damaged state are available, it makes sense to exploit this additional information to look for the actual direction along which the damage develops in the multivariate space. In 1-D cases, the parameter which characterizes the distance between two distributions featuring the same variance σ, but different mean values μ_1 and μ_2, is called effect size and can be found as

$$d^* = \frac{\mu_1 - \mu_2}{\sigma} \tag{5}$$

This idea was used by Fisher in its Liner Discriminant Analysis (LDA) [8]. In short, collecting the multivariate features in the rows of a matrix X, LDA searches for the optimal linear projection

$$y = w' X \tag{6}$$

namely the direction w which maximizes the difference between the projected class-means distance (i.e., the projected the between-class covariance matrix S_b), normalized by a measure of scatter (i.e., the projected within-class covariance matrix S_w) along the same direction.

In general, in case of multiple classes, it is possible to prove that the direction of maximum separation w corresponds to the first principal component of the matrix $S_w^{-1} S_b$, so that PCA can be used for LDA (N.B., this holds for a dimensionality larger than the number of classes). If the problem is reduced to two classes (i.e., 1 and 2), the formulation simplifies to

$$S_b = (\mu_2 - \mu_1)(\mu_2 - \mu_1)' \tag{7}$$

$$S_w = \sum_{h=1}^{n_1} (x_h - \mu_1)(x_h - \mu_1)' + \sum_{k=1}^{n_2} (x_k - \mu_2)(x_k - \mu_2)' \tag{8}$$

$$d(w) = \frac{w' S_b w}{w' S_w w} \tag{9}$$

$$\arg\max_w d(w) : w \propto S_w^{-1}(\mu_2 - \mu_1) \tag{10}$$

Hence, the maximization of the measure of separation $d(w)$ results in a direction w that can be computed as the inverse within-class covariance matrix by the distance of the two classes centroids. LDA direction is visualized in Fig.4 for the simulated dataset.

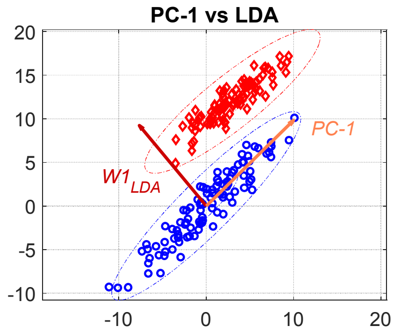

Fig. 4. LDA visualised on the healthy and the damaged simulated data and compared to the first principal component.

C. Mahalanobis Distance

The Mahalanobis Distance (MD) is a multivariate standardization tool able to produce unitless, scale invariant measures of the distance of some points from a distribution: the distance is zero if the point is at the centroid of the distribution, while it increases as the point moves away. MD accounts for correlation, so that if the data is whitened (i.e., decorrelated and rescaled on each principal component to obtain unit variance, so that the covariance matrix becomes the identity matrix), the Mahalanobis distance is equivalent to the Euclidean [9]. The essence of MD is pictured in Fig. 4.

The MD for the centred dataset X_0 is formulated as:

$$MD = \sqrt{X_0' S^{-1} X_0} \tag{11}$$

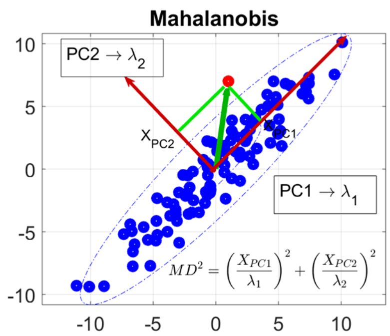

Fig. 5. Mahalanobis distance visualised on the healthy data.

D. Comparison over the simulated test dataset

The simulated dataset shown in Fig. 2 and integrated with an alternative damaged set can be used to test for the effectiveness of the proposed algorithms. The idea is to account for two main conditions:

a) The healthy dataset is not affected by confounders, so that the principal direction corresponds to the direction along which damage develops (i.e., PC1). Hence, when the damage is fully developed, a likely hypothesis for its distribution is that of being aligned or nearly aligned to the same direction (i.e., Damage 2, magenta in Fig. 6(a)).

b) The healthy dataset is affected by a confounder, whose effect is predominant (i.e., along PC1). When damage develops, a likely hypothesis for its distribution is the red one in Fig.6(a) (i.e., Damage 1), meaning that damage develops in a direction which can be orthogonal or nearly orthogonal to PC1.

Fig.6(b) shows the result in terms of Novelty Indices (NIs) computed by projecting the dataset along the first principal component (PC1). As can be noticed, condition a) can be well distinguished, so that the dimensionality can be reduced to 1 without much loss off information. Nevertheless, for case b) the separation is not so good, meaning that some information is lost (N.B., from the original 2D dataset in Fig.6(a) it is clear that the classes are all linearly separable).

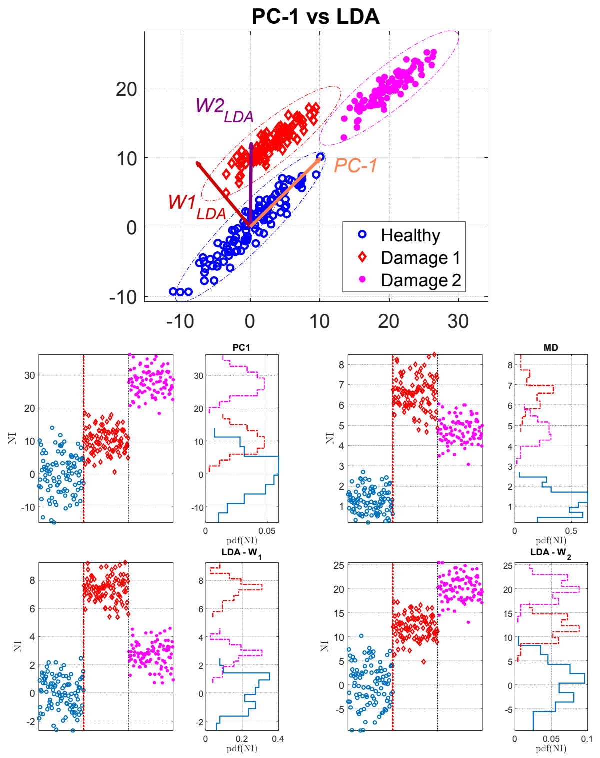

Fig. 6. Novelty Indices computed using the simulated dataset composed by a healthy reference and 2 possible damages. (a) Simulated dataset: draw from a bivariate normal distribution. Helathy: N([0;0],[13,12;12,13]), Damage 1: N([3;12],[11.2,8.2;8.2,7.5]) , Damage 2: N([20;20],[11.2,8.2;8.2,7.5]). The first pricnipal direction, the LDA direction for damage 1 (w_1) and the LDA direction for damage 2 (w_2) are added. (b) projection on PC1. (c) Mahalanobis distance from the healthy distribution. (d) projection on w_1 (e) projection on w_2. (e) Mahalanobis distance from the healthy distribution.

Focusing on Fig.6(d) and Fig.6(e), the NIs for the projection along the LD directions are shown. In the first case, the direction of maximum separation of the healthy distribution and damage 1 is used. The second plot on the contrary is produced projecting along the direction of maximum separation of the healthy distribution and damage 2. As it is easy to notice, both cases perform very well in highlighting the damaged distribution on which they were trained but have poorer performances when the other damage is considered.

Finally, Fig.6(c) shows the Mahalanobis Distance NIs. In this case it can be noticed that a perfect separation is possible for both the damaged conditions at the same time with the advantage that the training does not use information about the damaged acquisitions (i.e., the learning is not completely supervised, but semi-supervised, as for PC method).

III. Test on Experimental Data from PHM '09 Contest Gearbox Test Rig

In order to make a better comparison, the gearbox test rig acquisitions from the 2009 Prognostics and Health Management Data Challenge (PHM 2009 Data Challenge) were selected. The challenge in fact was targeted on fault classification of both gears and bearings at multiple speeds and loads, so that it covers a wide range of damages and operational variability. Many conference challenges and open dataset can be found in the literature (e.g., some are described in [10,11,12,13]), but it was decided to use the PHM '09 as it was particularly meant for damage classification [14].

The gearbox used for the measurement campaign is shown in Fig.7 and features a 3 shafts reducer with 6 rolling element bearings and 4 spur gearwheels with 32, 96, 48 and 80 teeth (T). The location for the two accelerometers is also highlighted.

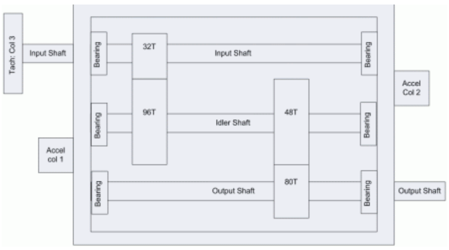

Fig. 7. PHM'09 challenge gearbox.

The overall dataset is composed by 560 measurements acquired at a sampling rate of $200/3\ kHz$ per each of the three channels (input side accelerometer, output side accelerometer and tachometer) for a duration of $4\ s$. Nevertheless, this particular analysis will focus on just 20 acquisitions from 4 health states while the gearbox was operated at 5 different input shaft speeds (30, 35, 40, 45, 50 Hz). Only the two accelerometric acquisitions will be considered. The selected health states are:

0. Completely healthy
1. 32T wheel featuring a chipped tooth and 48T wheel mounted with eccentricity
2. 32T wheel featuring a chipped tooth, 48T wheel mounted with eccentricity, 80T wheel broken, Input Shaft - Input Side bearing with inner race damage, Idle Shaft - Input Side bearing with ball damage
3. Input Shaft – Input Side bearing with inner race damage, sheared keyway on output side.

170 MB of storage are required for this reduced dataset. Having five $4\ s$ acquisitions from each health condition, it was decided to perform the blocking with a rectangular window $0.2\ s$ long, so as to produce 100 samples per each health condition (i.e., 5 subgroups at increasing rotational speed featuring 20 samples each; N.B. notice that this time window cannot be too long or too short, otherwise the features could lose their significance). From each chunk 5 common time-series features were extracted: root-mean-square, skewness, kurtosis, peak-value and crest-factor. Hence, a 2 channel acquisition that originally counted a total of $2 \cdot 266656\ samples = 533312\ samples$ was reduced to a total $20 \cdot 2 \cdot 5 = 200\ samples$, leading to a compression ratio of roughly 2667 times (N.B., a lossy compression obviously). Considering that the multivariate analysis that will be applied hereinafter are meant to condense the information of the 10 dimensional space composed by the 5 features extracted from each of the 2 channels into a single Novelty Index, the final acquisition will count $20\ samples$, for an overall compression ratio of 26667 times.

The dataset used to test the dimensionality reduction algorithms is reported graphically in Fig.8. It is relevant to remember that each of the four health conditions feature 100 samples from 5 different rotational speeds (from 30 Hz to 50 Hz), so that the speed will play the role of a confounder for the damage detection.

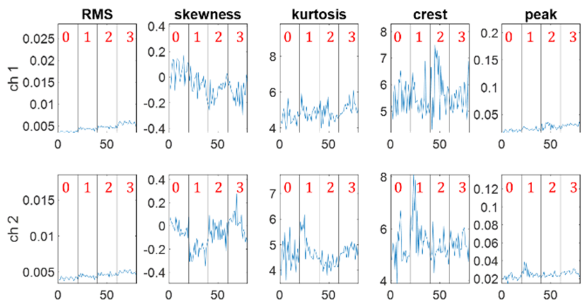

Fig. 8. The extracted featuers from the two channels in the 4 health conditions (N.B., each condition features 100 samples from 5 different rotational speeds from 30 Hz to 50 Hz.)

The performance of the algorithms will be compared not only in terms of Novelty Indices, but also in terms of receiver operating characteristic (ROC) curves generated by increasing the level of a threshold and computing each time the true positive rate (TPR: the percentage of damaged points correctly identified over the total damaged) and the false positive rate (FPR: the percentage of healthy points identified as damaged over the total healthy). The farthest the ROC curve is from the 45° line (i.e., a random classification), the better.

The results are collected inf Fig.9. In particular, Fig.9(a) shows the NIs resulting from the projection along the first principal component of the healthy reference set. As expected, the confounder has a strong impact on the result, so that the classification is very poor. Obviously, it is possible that a principal component other than the first could be able to better picture the damage level, but considering it is always a linear projection, it cannot be better than the one found by LDA. Focusing then on Fig.9(b), the NIs computed by projecting the dataset on the LD direction of maximum separation are shown (N.B., in this case the three different damage conditions 1,2,3 are considered as a single larger damage class used for training the LDA). In this case, damage condition 1 and 2 can be properly separated from the healthy, but is not the same for condition 3, which can be barely distinguished. Finally, Fig.9(c) shows the MD-NIs. In this case the third damage condition seems again to be not completely distinguishable, but the separation is better than the one obtained with LDA. Furthermore, the MD training is made on the healthy acquisition alone, without the need of using the damaged acquisition (as for LDA). The better performances of MD-NIs are highlighted in the last picture (Fig.9(d)) where the ROC curve proves to be the farthest from the 45° line.

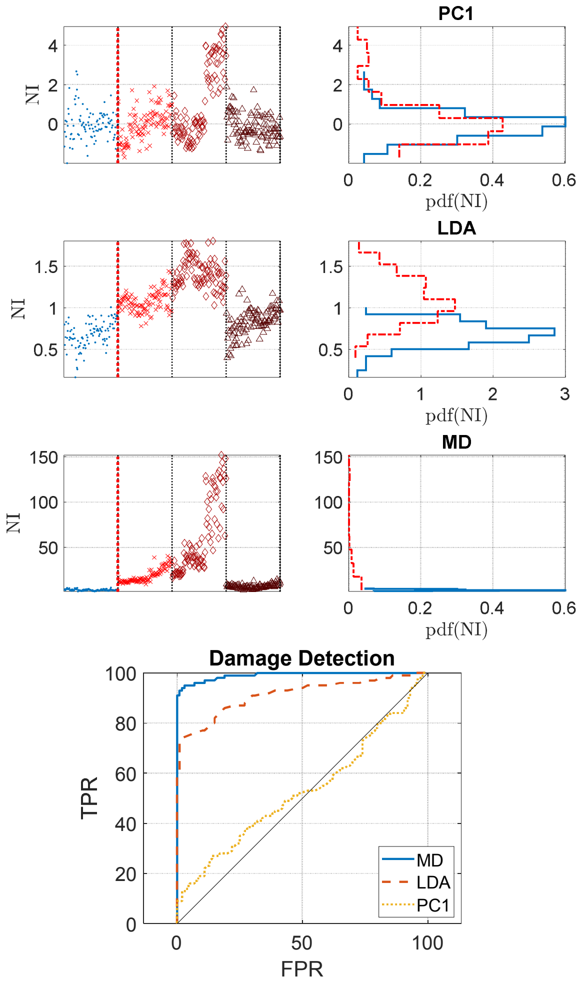

Fig. 9. Novelty Indices computed on the extracted featuers shown in Fig.8. The four health sates are separated by dashed lines. (a) projection on PC1. (b) projection on LDA direction. (c) Mahalanobis distance from the healthy distribution. (d) Damage Detection performance as a receiver operating characteristic (ROC) curves in terms of True Positive Rate vs False Positive Rate.

IV. Conclusions

The aim of this paper was to foster the vibration monitoring creating a procedure to shrink the big datasets using a lossy compression able to retain most of the diagnostic information. In fact, not only the information can be retained, but the damage related patterns can also be unveiled at the same time, so that the stored data will be both lighter and more meaningful from a damage detection point of view. The analysis is composed by two fundamental steps:

- The extraction of features on blocked time signals (N.B., simple time-series features were computed).
- The multivariate information condensation to a single Novelty Index.

In this regard, the proposed analysis can be seen as a multivariate extension of the univariate vibration monitoring methodology prescribed by [1].

In this work three algorithms were considered, the PCA, the LDA and the MD. The projection along the first PC proved to be effective only in the absence of confounders. When confounding influences are present on the contrary, it is likely that the damage related information will be pictured by a principal component other than the first. When acquisitions from damaged states are also available during the training stage, LDA can be used to find the direction of best separation. If the dataset is complex anyway (e.g., multiple damages evolving in different directions, with distributions not linearly separable but still distinguishable) LDA can have troubles. The final solution is then to use the MD to produce a sort of nonlinear transform to polar coordinates where the radial distance from the centroid (rescaled along the principal components according to the corresponding principal values) is used as NI.

As demonstrated on the simulated dataset and on the PHM '09 gearbox experimental dataset, MD-NIs prove to be the best solution in terms of damage detection ability, with the advantage that the training is based only on healthy acquisitions. Test of a similar procedure were also conducted on wind turbines gearboxes, leading to satisfactory results [15] as MD-NIs proved to be robust to quasi-linear confounding influences. The reliable, standard-compliant univariate analysis in [1] was then efficiently and effectively turned into a multivariate analysis able to condense the information from different channels and features into a single, optimal Novelty Index, improving the diagnostics ability while reducing the required volume for data storage.

References

[1] ISO 20816-1:2017: Mechanical Vibration - Measurement And Evaluation Of Machine Vibration - General Guidelines.

[2] G. Hudson, A. Léger, B. Niss and I. Sebestyén, "JPEG at 25: Still Going Strong," in IEEE MultiMedia, vol. 24, no. 2, pp. 96-103, Apr.-June 2017, doi: 10.1109/MMUL.2017.38.

[3] ISO/IEC 11172-3:1993: Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 3: Audio

[4] MPEG official website: https://mpeg.chiariglione.org/standards/mpeg-1/audio

[5] FLAC official website: https://xiph.org/flac/format.html

[6] T. Robinson, SHORTEN: Simple lossless and near-lossless waveform compression, 1997.

[7] I.T. Jolliffe, "Principal Component Analysis", Springer, 2002. DOI: 10.2307/1270093.

[8] C. Bishop, "Pattern Recognition and Machine Learning", Springer-Verlag New York, 2006. ISBN: 978-0-387-31073-2.

[9] A. P. Daga; L. Garibaldi. Machine Vibration Monitoring for Diagnostics through Hypothesis Testing. Information 2019, 10, 204.

[10] X. Jia; B. Huang; J. Feng; H. Cai; J. Lee. A Review of PHM Data Competitions from 2008 to 2017. 2018.

[11] J. Antoni et al., "Feedback on the Surveillance 8 challenge: Vibration-based diagnosis of a Safran aircraft engine", Mechanical Systems and Signal Processing, Volume 97, Pages 112-144, 2017, DOI: 10.1016/j.ymssp.2017.01.037.

[12] A. P. Daga; L. Garibaldi. GA-Adaptive Template Matching for Offline Shape Motion Tracking Based on Edge Detection: IAS Estimation from the SURVISHNO 2019 Challenge Video for Machine Diagnostics Purposes. Algorithms 2020, 13, 33.

[13] A. P. Daga, A. Fasana, S. Marchesiello, L. Garibaldi, "The Politecnico di Torino rolling bearing test rig: Description and analysis of open access data", Mechanical Systems and Signal Processing, Volume 120, 2019. DOI: 10.1016/j.ymssp.2018.10.010.

[14] H. Al-Atat; D. Siegel; J. Lee. A Systematic Methodology for Gearbox Health Assessment and Fault Classification. International Journal of Prognostics and Health Management. 2011.

[15] F. Castellani; L. Garibaldi; A. P. Daga; D. Astolfi; F. Natili. Diagnosis of Faulty Wind Turbine Bearings Using Tower Vibration Measurements. *Energies* **2020**, *13*, 1474.