Blind deconvolution based on cyclostationarity maximization and its application to fault identification[☆]

Marco Buzzoni[a,*], Jérôme Antoni[b], Gianluca D'Elia[a]

[a] *University of Ferrara, Department of engineering, via Saragat 1, 44122 Ferrara, Italy*
[b] *Université de Lyon, INSA-Lyon, Laboratoire Vibrations Acoustique, LVA EA677, F-69621 Villeurbanne, France*

Abstract

Blind deconvolution algorithms prove to be effective tools for fault identification, being able to extract excitation sources from noisy observations only. In this scenario, the present paper introduces a novel blind deconvolution method based on the generalized Rayleigh quotient and solved by means of an iterative eigenvalue decomposition algorithm. This approach is characterized by a weighting matrix that drives the deconvolution process, whereby it can be easily adapted to arbitrary criteria. Based on this framework, a novel criterion rooted on the cyclostationarity maximization of the excitation – as typically encountered with machine faults – is proposed and compared with other blind deconvolution methods existing in the literature. The comparisons involve both synthesized and real vibration signals, taking into account a gear tooth spall and an outer race bearing fault. The results reveal superior capability to recover impulsive cyclostationary sources with respect to other blind deconvolution methods, even in the presence of impulsive noise or under non-constant speed.

Keywords: Blind deconvolution, Rayleigh quotient optimization, Cyclostationary criterion, Cyclostationary signals, Fault identification

Nomenclature

AR	Auto-regressive
BD	Blind deconvolution
CS1	First-order cyclostationarity
CS2	Second-order cyclostationarity
CYCBD	Maximum second-order cyclostationarity blind deconvolution
EVA	Eigenvalue algorithm
ICS	Indicators of cyclostationarity
IQR	Interquartile range
IRF	Impulse response function
MCKD	Maximum correlated kurtosis deconvolution
MED	Minimum entropy deconvolution
MOMEDA	Multipoint optimal minimum entropy deconvolution adjusted
OFM	Objective function method
OMEDA	Optimal minimum entropy deconvolution adjusted
SIMO	Single-input-multi-output
SISO	Single-input-single-output

[☆] This document is a collaborative effort.

[*] Corresponding author
Email addresses: marco.buzzoni1@unife.it (Marco Buzzoni), jerome.antoni@insa-lyon.fr (Jérôme Antoni), gianluca.delia@unife.it (Gianluca D'Elia)

1. Introduction

The identification of impulsive faults is of major importance in the diagnosis of rotating machines, especially for gears and bearings which usually are the most critical components in many mechanical systems. The impulsive fault identification can be difficult, particularly in the early stage, since the impulsive pattern due the fault occurrence is often masked by background noise and other interferences. The situation is further worsen by the spreading effect of the unknown transmission path. For this purpose, blind deconvolution (BD) techniques can recover the impulsive pattern from noisy observations, even considering the effect of a unknown linear time-invariant system.

In the field of seismic signal processing, Wiggins [1] pioneered BD by developing an iterative algorithm based on the maximization of the kurtosis (called Varimax in his paper) in order to recover a spike-like source from a signal convolved with an unknown impulse response function (IRF). In the same field, Cabrelli [2] proposed another criterion, called D-Norm, geometrically equivalent to the Varimax norm, which poses a direct solution to BD. This method has been recently refined by McDonald and Zhao [3]. In the literature, these BD methods are known as minimum entropy deconvolution (MED) and optimal minimum entropy deconvolution adjusted (OMEDA), respectively. Other authors explored higher-order statistics as well as different optimization algorithms. Lee and Nandi [4] analyzed the performance of BD via higher-order statistics considering impacting signals from a vibrating cantilever beam. The same authors [5] demonstrated also that the objective function method (OFM), that has been exploited in MED and OMEDA, and the eigenvalue algorithm (EVA) give equivalent results considering the same experimental measurements. Another statistics that combines both skewness and kurtosis (called Jarque-Bera statistic) has been investigated by Obuchowski et al. [6] for the gear fault identification.

Among all these blind deconvolution methods, MED has been the most commonly used for machine fault identification. MED has been typically exploited in combination with other signal processing techniques in order to improve its performances for machine diagnosis since it recovers preferably a large single peak rather than train of impulses, as typically encountered with machine faults. Regarding the tooth fault detection, Endo and Randall [7] exploited MED in order to improve the gear fault detection based on auto-regressive (AR) models. This method has been further investigated by Endo et al. [8] for discriminating a gear tooth spall from a cracked tooth. A similar approach has been proposed by Sawalhi et al. [9] with regard to bearing fault diagnosis, taking advantage of the envelope spectrum driven by maximum spectral kurtosis. In a different way, the spectral kurtosis has been exploited also by He et al. [10] in order to extract multiple bearing faults.

The need of criteria dedicated to machine diagnosis led to the introduction of the correlated kurtosis and the Multi-Point D-Norm. The maximum correlated kurtosis deconvolution (MCKD), has been introduced by McDonald et al. [11] whereas the multipoint optimal minimum entropy deconvolution (MOMEDA), has been proposed by McDonald and Zhao [3]. Both the criteria try to enhance the vibration signal impulsiveness linked to a specified fault period overcoming the tendency of MED and OMEDA to recover a single dominant impulse.

Although the pivotal role of cyclostationarity in machine diagnosis has been widely recognized [12], it has not yet been applied explicitly in BD criteria for machine fault identification. In this scenario, just one cyclostationary criterion can be found in the literature devoted to vibration-based fault diagnosis, i.e. the MCKD [11], while the others are based on extracting the most impulsive contribution (MED [1] and OMEDA [2]) or a periodic impulse train (MOMEDA [3]). Despite MCKD is a cyclostationary criterion, it has been proposed empirically, without explicit mention of cyclostationarity. Moreover, the criterion at the base of MCKD (called correlated kurtosis) entails some disadvantages that limit its use in many real applications. Therefore, this research work tries to fill this gap proposing a simpler and more efficient criterion grounded on the cyclostationary framework.

A preliminary overview about standard BD criteria is given, pointing out some original considerations highlighting advantages and limits. In particular, with regard to MED, the relationship between kurtosis and differential entropy is clarified. OMEDA and MOMEDA have been reviewed providing an original interpretation connecting the non-iterative solutions of these two methods to (linear) least square solutions. Concurrently, a qualitative justification of why MCKD is based on a cyclostationary criterion is provided as well. Then, an iterative eigenvalue algorithm for BD of single-input-single-output (SISO) systems based on the generalized Rayleigh quotient is presented as well as its version for single-input-multi-output (SIMO) systems. This algorithm differs from the EVA introduced by Jelonnek et al. [13] by the fact that it's not restricted to the use of fourth-order (cross) cumulants. Furthermore, the deconvolution is guided by a weighting matrix that can be easily modified adapting the deconvolution algorithm to arbitrary criteria. The proposed BD method has been formulated considering the higher-order statistics maximization and con-

sequently the cyclostationarity maximization through the indicators of cyclostationarity (ICS). The latter formulation represents the core of this paper since it is the very original part of this research. The BD method based on the maximization of the second-order cyclostationarity, called CYCBD, is formulated for time-dependent signals and for angle-dependent signals, particularly useful for rotating machine diagnosis considering non-constant regimes. The CYCBD performances are compared with other BD methods, taking into account five cyclostationary synthesized signals. The simulated results demonstrate the capability of CYCBD to recover impulsive cyclostationary sources at constant and non-constant regimes. In support of the simulated results, two applications have been investigated regarding the identification of a gear tooth spall and an outer-race bearing fault. The experimental results highlight the capability of CYCBD to identify localized faults exhibiting cyclostationary behavior. In particular, the criterion maximized by the CYCBD (i.e. ICS_2) proves to be a robust and sensitive indicator in terms of early fault detection and identification.

The paper is organized as follows: the overview of common BD methods is presented in Section 2 including also some original interpretations; the proposed algorithm is introduced and explained in Section 3 providing also criteria based on higher-order statistics as well as on cyclostationarity indicators; Section 4 deals with the validation by means of simulated signals; Section 5 concerns the validation using real signals supported by an extended comparison with other BD methods; Section 6 gives the final remarks.

2. Overview about blind deconvolution techniques for machine diagnosis

This section provides the general formulation of (SISO) BD and a more specific formulation for rotating machine applications. Moreover, a general explanation of common BD algorithms used in rotating machine diagnosis is given, providing some original interpretations of these methods.

2.1. Problem statement

In general, BD aims to recover – i.e. deconvolve – an input signal $\mathbf{s}_0$ from a noisy observed signal $\mathbf{x}$, viz:

$$\mathbf{s} = \mathbf{x} * \mathbf{h} = (\mathbf{s}_0 * \mathbf{g}) * \mathbf{h} \approx \mathbf{s}_0 \tag{1}$$

where $\mathbf{g}$ is the unknown impulse response function (IRF), $\mathbf{h}$ is the inverse filter (from now assumed to be a FIR filter), $\mathbf{s}$ is the estimated input and $*$ refers to the convolution operation. Note that bold lowercase letters refer to vectors whereas bold capital letters refer to matrices. A convenient way to express the convolution for discrete signals in matrix form is:

$$\mathbf{s} = \mathbf{X}\mathbf{h} \tag{2a}$$

$$\begin{bmatrix} s[N-1] \\ \vdots \\ s[L-1] \end{bmatrix} = \begin{bmatrix} x[N-1] & \cdots & x[0] \\ \vdots & \ddots & \vdots \\ x[L-1] & \cdots & x[L-N-2] \end{bmatrix} \begin{bmatrix} h[0] \\ \vdots \\ h[N-1] \end{bmatrix} \tag{2b}$$

where L and N are the total samples of $\mathbf{s}$ and $\mathbf{h}$, respectively. Such expressions will be widely recalled in the next sections.

According to the scheme proposed in Refs. [7, 11] valid for vibration signals belonging to gearboxes, the observed signal in Eq. (1) can be rearranged as summarized in Fig. 1. The measured signal $\mathbf{x}$ is assumed to be composed of: an impulsive part $\mathbf{s}_0$ due to a localized fault, a pure periodic component $\mathbf{p}$ (e.g. related to the gear mesh) and Gaussian background noise $\mathbf{n}$, such that:

$$\mathbf{x} = \mathbf{s}_0 * \mathbf{g}_s + \mathbf{p} * \mathbf{g}_p + \mathbf{n} * \mathbf{g}_n \tag{3}$$

where $\mathbf{g}_s$, $\mathbf{g}_p$ and $\mathbf{g}_n$ are the IRFs related to $\mathbf{s}_0$, $\mathbf{p}$ and $\mathbf{n}$, respectively. Substituting Eq. (3) into Eq. (1), the mathematical formulation of BD in the context of diagnostics is given by:

$$\mathbf{s} = \left(\mathbf{s}_0 * \mathbf{g}_s + \mathbf{p} * \mathbf{g}_p + \mathbf{n} * \mathbf{g}_n\right) * \mathbf{h} \approx \mathbf{s}_0 \tag{4}$$

where, in practice, $\mathbf{s}_0$ represents the excitation force due to an incipient fault. Thus, in this context, BD aims to estimate $\mathbf{h}$ such as to recover $\mathbf{s}_0$ linked to a machine faults minimizing the other contributions, i.e. $\mathbf{p}$ and $\mathbf{n}$.

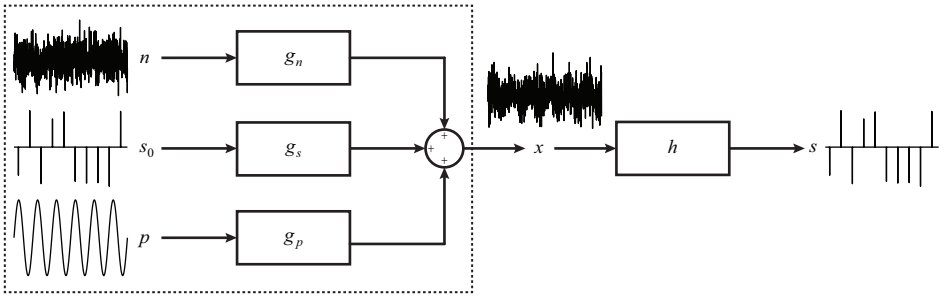

Fig. 1: General scheme of blind deconvolution for mechanical system diagnostics.

The assumptions of the mathematical problem in Eq. (4) are: the samples of $\mathbf{s}_0$ are independent identically distributed random variables having a distinctive statistical property (e.g. impulsiveness), $\mathbf{g}$ is a stable Linear Time Invariant system and the stationary Gaussian noise, $\mathbf{n}$, and the periodic contribution, $\mathbf{p}$, are additive and do not share the same characteristics of $\mathbf{s}_0$. As it is, the problem is ill-posed since the IRFs – namely $\mathbf{g}_s$, $\mathbf{g}_p$, $\mathbf{g}_n$ – are not available. However, an estimation of the solution can be achieved considering an arbitrary criterion that imposes an expected solution based on a prior assumption, e.g. assuming a certain statistical property is met by the desired estimated source. It should be noted that many BD methods are amplitude invariant, i.e. do not recover the actual source magnitude. However, BD deconvolution methods can provide useful information in terms of waveform, extracting a desired source that exhibits a certain statistical behavior.

2.2. Blind deconvolution algorithms for machine diagnosis

In the following section, four BD methods commonly used in the field of vibration-based machine diagnosis are briefly presented and discussed.

2.2.1. Minimum Entropy Deconvolution

The MED algorithm has been firstly introduced by Wiggins [1] in seismic applications, exploiting the kurtosis maximization in order to recover an impulse-like estimation of the source. A criterion based on kurtosis maximization represents a very reasonable choice also in machinery diagnostics since the kurtosis is widely used as a general purpose indicator for fault identification in bearings and gears.

With reference to Eq. (1), the method provides an estimation of $\mathbf{s}_0$ (and $\mathbf{h}$) given $\mathbf{x}$ which maximizes the following criterion:

$$\kappa_4 = \frac{\sum_{l=N}^{L-1} \mathbf{s}[l]^4}{\left(\sum_{l=N}^{L-1} \mathbf{s}[l]^2\right)^2}. \tag{5}$$

Note that this definition of kurtosis assumes zero-mean signals, as obtained after centering. It should be remarked that MED is based on kurtosis maximization rather than entropy minimization. The definition of the entropy can assume many possible declinations. However, in this context, it is convenient to recall the entropy definition given by the probability theory [14]: the entropy is a measure of the average amount of information needed to specify the state of a random variable. Thus, the probability distributions having peak-like shape would exhibit low entropy values. Since kurtosis is also a measure of the sharpness of a probability distribution, this perspective establishes the link between the entropy and kurtosis. Moreover, if the signal probability distribution is symmetric and slightly non-Gaussian, then the Gram-Charlier expansion of the differential entropy is related to the opposite of the kurtosis (see Appendix A).

Inverse filter $\mathbf{h}$ is the result of the following maximization problem:

$$\mathbf{h} = \underset{\mathbf{h}}{\operatorname{argmax}}(\kappa_4) \tag{6}$$

that can be estimated iteratively by the OFM. Analogous results can be achieved also using the EVA approach exploiting the fourth-order cross cumulant [5]. For the full demonstrations see: Refs. [1, 3] for the OFM and Ref. [5] for

the EVA. This criterion does not require any prior information about the mechanical systems involved; thus it can be considered more general than other criteria that need, for instance, the prior knowledge of the fault period.

Nevertheless, as pointed out in Ref. [3], MED tends to recover a single dominant impulse rather than a train of impulses, that actually is how the local faults of the rotating machines appear in vibration signals. This behavior is due to the nature of kurtosis since it theoretically reaches its maximum value for signals containing a unique (dominating) impulse. This shortcoming may occurs in particular when the FIR filter length is too long. The term "too long" is on purpose ambiguous; in fact, a priori estimation of the proper FIR filter length is not possible even if the FIR filter length is a critical parameter that strongly influences the final results. This limitation does not regard only MED but it is shared by all the blind deconvolution algorithms.

2.2.2. *Maximum Correlated Kurtosis Deconvolution*

The MED algorithm has been recently improved by McDonald et al. [11], which presented an iterative blind deconvolution method, called MCKD, based on a novel criterion called correlated kurtosis. The definition of correlated kurtosis reads:

$$CK_M = \frac{\sum_{l=N}^{L-1}\left(\prod_{m=0}^{M}\mathbf{s}[l-mT_s]\right)^2}{\left(\sum_{l=N}^{L-1}\mathbf{s}[l]^2\right)^{M+1}} \tag{7}$$

where T_s is the impulse period and M is the number of shifts. Note that the correlated kurtosis is equivalent to the kurtosis for $T_s = 0$ and $M = 1$. This criterion is a measure of signal impulsiveness connected with a given period T_s, taking advantage of two characteristics frequently encountered with machine faults: high kurtosis and repetitive occurrence of the impulses.

Despite correlated kurtosis has been proposed by intuition, it is worth noting that, de facto, is a cyclostationary criterion. A clarifying example is given in case of $M = 1$:

$$CK_1 = \frac{\sum_{l=N}^{L-1}(\mathbf{s}[l]\mathbf{s}[l-T_s])^2}{\left(\sum_{l=N}^{L-1}\mathbf{s}[l]^2\right)^2}. \tag{8}$$

Eq. (8) shows an evident similarity with the definition of kurtosis (see Eq. 5). However, it should be noted that the numerator of Eq. (8) is the discrete autocorrelation of the signal power, $\mathbf{s}^2$, at lag T_s. This quantity is nothing but the signal power contribution which changes cyclically with frequency $\frac{1}{T_s}$. De facto, since the correlated kurtosis is a measure of cyclostationarity according to a given cyclic frequency defined through $\frac{1}{T_s}$, MCKD allows for the estimation of a source exhibiting the maximum cyclostationarity at cyclic frequency $\frac{1}{T_s}$.

The number of shifts M represents a parameter to be carefully set when MCKD is applied to real vibration signals. From experience [11], low values of M may not encourage enough the deconvolution of periodic impulses. Moreover, high values of M (in general more than 8) can compromise the correct source estimation. The previous considerations can be used to provide a informal justification about how correlated kurtosis behaves with respect to M. The presence of strong background noise ($SNR < 0$) affects the estimation of correlated kurtosis in particular for low values of M. For instance, with regard to $M = 1$, the numerator of Eq. (8) is not a consistent estimator of discrete autocorrelation in presence of noise. Furthermore, it is frequent that many rotating machines, even at nominal constant speed, exhibit slight regime fluctuations which reflect in fluctuating values of T_s. This contributes further to a poor estimation of correlated kurtosis, in particular when $M \gg 1$.

.

2.2.3. *Optimal Minimum Entropy Deconvolution Adjusted*

The first direct method discussed in this work is the so called OMEDA, that is an improvement of the original algorithm, proposed by Cabrelli [2] and advanced by McDonald and Zhao [3]. This method is based on the following criterion:

$$\text{D-Norm} = \max_{l=N,1,\cdots,L-1}\left(\frac{|\mathbf{s}[l]|}{||\mathbf{s}||}\right) \tag{9}$$

where $\|\bullet\|$ is the Euclidean norm. The FIR filter that maximizes Eq. (9) is given by one of the columns of

$$\mathbf{H} = \left(\mathbf{X}\mathbf{X}^H\right)^{-1}\mathbf{X} \tag{10}$$

that corresponds to the output, $\mathbf{s}$, having maximum D-Norm. The OMEDA solution can be easily computed by pre-multiplying Eq. (10) by $\mathbf{X}^T$. The mathematical steps to obtain Eq. (10) can be found in Refs. [2, 3]. According to Refs. [2, 3], it has been proved by means of numerical simulations that the OMEDA returns outputs having simpler structure than the solution provided by MED.

A simpler interpretation of this approach can be given noting that Eq. (10) is similar to the solution of a Least Square problem. The product between the right side of Eq. (10) and the identity matrix yields to a set of solutions (namely a set of filter coefficients) where the l^{th} column of $\mathbf{H}$ is the Least Square solution of:

$$\mathbf{h}_k = \left(\mathbf{X}\mathbf{X}^H\right)^{-1}\mathbf{X}\boldsymbol{\delta}_l \tag{11}$$

where δ_l is the l^{th} column of the identity matrix. In other words, Eq. (11) is the solution of the following minimization problem:

$$\mathbf{h} = \underset{\mathbf{h}}{\operatorname{argmin}}\left(\|\mathbf{X}\mathbf{h} - \boldsymbol{\delta}_k\|\right) \tag{12}$$

where $\boldsymbol{\delta}_l$ represents an ideal impulsive signal with unitary amplitude and a delay equal to l samples. Eq. (11) is the Least Square solution of the filter coefficients based on a simple sharp target signal. Eq. (10) is the extension of this logic, being $\mathbf{H}$ the set of solutions obtained considering shifted Dirac impulses as target signals. This interpretation is in agreement with the formulation of OMEDA, since it is designed to recover the simplest source from a given observation. Indeed, OMEDA scans all the possible Least Square solutions keeping the one returning the highest D-Norm. From this standpoint, it is particularly simple to explain the tendency of OMEDA to deconvolve single peaks.

2.2.4. Multi-point Optimal Minimum Entropy Deconvolution Adjusted

An improvement of OMEDA has been proposed by McDonald and Zhao [3] by introducing a target vector $\mathbf{t}$ that promotes the deconvolution of a sequence of ideal periodic impulses instead of a single one. The modified version of D-Norm (Eq. (9)) is called Multi D-Norm and it reads:

$$\text{Multi D-Norm} = \frac{1}{\|\mathbf{t}\|}\frac{\mathbf{t}^T\mathbf{s}}{\|\mathbf{s}\|}. \tag{13}$$

The FIR filter $\mathbf{h}$ that maximizes Eq. (13) is:

$$\mathbf{h} = \left(\mathbf{X}\mathbf{X}^H\right)^{-1}\mathbf{X}\mathbf{t} \tag{14}$$

where $\mathbf{t}$ drives the deconvolution by imposing positions and weights of the target impulses. A valid solution can be obtained by selecting the FIR filter length greater than the fault period. In this way, the correct position of the impulses are automatically adjusted by the filter delay. As for OMEDA, it can actually be checked that maximizing Eq. (13) is equivalent to finding the Least Square solution of:

$$\mathbf{h} = \underset{\mathbf{h}}{\operatorname{argmin}}\left(\|\mathbf{X}\mathbf{h} - \mathbf{t}\|\right) \tag{15}$$

i.e. the filter coefficients are estimated in order to minimize the Least Square error between $\mathbf{s}$ and an ideal train of equispaced impulses, $\mathbf{t}$, having the same amplitude. Thus, by definition, MOMEDA overcomes the limitation of OMEDA and MED to preferably deconvolve only a single impulse and the restriction of MCKD to be dependent on the choice of M. However, MOMEDA is based on a criterion that allows for the extraction of periodic impulses, i.e. equispaced impulses having same amplitude. The periodic criterion at the base of MOMEDA is in contrast with the (typical) cyclostationary nature of many mechanical vibration signals [12]. For this reason, MOMEDA could lose effectiveness when is applied for machine fault identification.

3. Proposed method

In the following section, a BD method based on the generalized Rayleigh quotient and solved by means of an iterative eigenvalue decomposition algorithm is proposed. Firstly, a formulation of the proposed BD method through kurtosis maximization is obtained together with its extension to higher-order statistics. Consequently, the BD algorithm based on the cyclostationary maximization, which is the core of this research work, is derived as well for SISO ans SIMO models.

3.1. Blind deconvolution driven by kurtosis maximization and extension to higher-order statistics

Unlike the methods proposed in refs. [1, 5], BD with kurtosis maximization can be reformulated as follows. Let us first recall the definition of the kurtosis:

$$\kappa_4 = \frac{M_4}{M_2^2} = \frac{\mathbf{s}^H\mathbf{W}\mathbf{s}}{(\mathbf{s}^H\mathbf{s})^2}(L-N+1) \tag{16}$$

where M_4 and M_2 are the fourth and the second moments, respectively, and the weighting matrix $\mathbf{W}$ is defined as:

$$\mathbf{W} = diag\left(\frac{\mathbf{s}^2}{\mathbf{s}^H\mathbf{s}}\right)(L-1) = \begin{bmatrix} \ddots & & 0 \\ & s[l]^2 & \\ 0 & & \ddots \end{bmatrix} \frac{(L-N+1)}{\sum_{l=N}^{L-N+1} s[l]^2}. \tag{17}$$

Substituting Eq. (2a) in Eq. (16), the following expression can be obtained:

$$\kappa_4 = \frac{\mathbf{h}^H\mathbf{X}^H\mathbf{W}\mathbf{X}\mathbf{h}}{\mathbf{h}^H\mathbf{X}^H\mathbf{X}\mathbf{h}} = \frac{\mathbf{h}^H\mathbf{R}_{XWX}\mathbf{h}}{\mathbf{h}^H\mathbf{R}_{XX}\mathbf{h}} \tag{18}$$

where $\mathbf{R}_{XWX}$ and $\mathbf{R}_{XX}$ are the weighted correlation matrix and the correlation matrix, respectively. Coming to the kurtosis maximization, it should be noticed that Eq. (18) is a generalized Rayleigh quotient whose maximization with respect to $\mathbf{h}$ is equivalent to the eigenvector associated with the maximum eigenvalue λ of the following generalized eigenvalue problem:

$$\mathbf{R}_{XWX}\mathbf{h} = \mathbf{R}_{XX}\mathbf{h}\lambda \tag{19}$$

then λ corresponds to maximum κ_4. Since $\mathbf{R}_{XWX}$ and $\mathbf{R}_{XX}$ are real and symmetric by construction and $\mathbf{R}_{XX}$ is also semi-positive definite, it implies that λ must be non-negative. This property is in agreement with the fact that the kurtosis is always positive by definition. Furthermore, it should be noted that $\mathbf{W}$ has to be initialized with a guess; hence, the equivalence between the maximum λ and the maximum κ_4 is reached only by means of an iterative algorithm summarized by the following steps:

Step 1: assume a guess of $\mathbf{h}$;
Step 2: estimate $\mathbf{W}$ evaluating Eq. (2a) given $\mathbf{X}$ and guessed $\mathbf{h}$;
Step 3: solve Eq. (19) finding $\mathbf{h}$ associated to the maximum λ;
Step 4: return to Step 2 using $\mathbf{h}$ estimated in Step 3 until convergence.

Since BD is based on the hypothesis that the source is a independent identically distributed random variable, a good initialization of $\mathbf{h}$ is given by a whitening filter, according to Ref. [15]. In practice, the whitening filter can be computed by means of an auto-regressive model filter by using the Yule-Walker equations for the filter coefficients estimation. Moreover, in mechanical applications, the vibration signal spectra can be dominated by sharp peaks related to the gear mesh harmonics or other deterministic sources [16]. Thus, the inverse AR filter strongly attenuates all the predictable components, returning a signal with a flat spectral density, which is the expected shape, on the average, for a signal containing a series of impulses. In order to improve the algorithm speed, the complete evaluation of the generalized eigenvalue problem can be avoided taking advantage of the fact that the algorithm needs only the maximum value of λ. For this purpose, dedicated algorithms for the estimation of the greatest eigenvalue (e.g. the power method) can be exploited.

At this point, an extension of the proposed algorithm can be written considering an arbitrary p^{th} order normalized moment by properly modifying $\mathbf{W}$. Proceeding from the definition of the p^{th} order normalized moment, the following criterion is proposed:

$$\kappa_p = \frac{M_p}{M_2^{p/2}} = \frac{\mathbf{s}^H\mathbf{W}\mathbf{s}}{(\mathbf{s}^H\mathbf{s})^{p/2}}(L-N+1)^{\frac{p}{2}-2} \tag{20}$$

where the related weighting matrix is expressed as

$$\mathbf{W} = diag\left(\frac{\mathbf{s}^{p-2}}{(\mathbf{s}^H\mathbf{s})^{\frac{p}{2}-1}}\right)(L-N+1)^{\frac{p}{2}-2} = \begin{bmatrix} \ddots & & 0 \\ & s[l]^{p-2} & \\ 0 & & \ddots \end{bmatrix} \frac{(L-N+1)^{\frac{p}{2}-2}}{\sum_{l=N}^{L-1} s[l]^{p-2}}. \tag{21}$$

Hence, the BD algorithm via kurtosis maximization (see Eq. (18)) can be easily extended for any normalized moment just using the weighted function defined in Eq. (21). Note that in Eq. (21) only $p > 2$ is relevant in the interests of the detection of impulsive components since $\kappa_{p>2}$ is a measure of the impulsiveness.

3.2. Blind deconvolution driven by cyclostationarity maximization

Analogously to the algorithm proposed in Section 3.1, a novel criterion for BD is proposed hereafter based on the maximization of the cyclostationarity through the maximization of the ICS. This indicator has been proposed by Raad et al. [17] and its effectiveness on diagnostic purposes has been demonstrated in several applications such as gears [17], bearings [18] and IC engines [19].

First of all, it may be useful to introduce some basic notions about cyclostationarity. From a general standpoint, a cyclostationary process is a process exhibiting a periodic behavior of its statistical properties. Cyclostationarity plays a pivotal role in the vibration-based fault diagnosis since it has been demonstrated [12] that the rotating machine vibration signals are well modeled by cyclostationary processes. The real mechanical signals are often a mixture of first-order and second-order cyclostationary processes, called also CS1 and CS2, respectively. The CS1 part represents the perfectly deterministic part of the signal, which embodies all the periodic contributions, and the CS2 part is the random signal part which exhibits periodic fluctuations of its energy flow. On these grounds, it is clear that the cyclostationary approach turns out to be more realistic and general with respect to the approaches (as assumed in the MOMEDA) that consider only the periodic part. Furthermore, the concept of cyclic frequency should be given in order to clarify the further dissertation. In the cyclostationary scenario, the cyclic frequency can be seen as the frequency related to a certain (hidden) fluctuation of the signal energy, which can be related to physical phenomena as gear faults and bearing faults, for instance. In this context, the cyclic frequency for discrete-time signals is defined as

$$\alpha = \frac{k}{T_s} \tag{22}$$

where k is the sample index and T_s is the cycle (in samples), which can be related to a fault occurrence rate for instance.

Coming back to the proposed criterion, let us recall the general definition of the second-order ICS:

$$ICS_2 = \frac{\sum_{k>0} \left|{c_s}^k\right|^2}{\left|{c_s}^0\right|^2} \tag{23}$$

with

$${c_s}^k = \left\langle |\mathbf{s}|^2, e^{j2\pi\frac{k}{T_s}l} \right\rangle = \frac{1}{L-N+1}\sum_{l=N}^{L-1} |s[l]|^2 e^{-j2\pi\frac{k}{T_s}l} \tag{24a}$$

$${c_s}^0 = \frac{\|\mathbf{s}\|^2}{L-N+1}. \tag{24b}$$

Eqs. (24a) and (24b) may be expressed in a matrix form as follows:

$$c_s{}^k = \frac{\mathbf{E}^H|\mathbf{s}|^2}{L-N+1} \tag{25a}$$

$$c_s{}^0 = \frac{\mathbf{s}^H\mathbf{s}}{L-N+1} \tag{25b}$$

where

$$|\mathbf{s}|^2 = [|s[N]|^2, ..., |s[L-1]|^2]^T \tag{26a}$$

$$\mathbf{E} = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_k & \cdots & \mathbf{e}_K \end{bmatrix} \tag{26b}$$

$$\mathbf{e}_k = \begin{bmatrix} e^{-j2\pi\frac{k}{T_s}(N)} \\ \vdots \\ e^{-j2\pi\frac{k}{T_s}(L-1)} \end{bmatrix}. \tag{26c}$$

From Eqs. (24a) and (24b), Eq. (23) may be expressed as:

$$ICS_2 = \frac{|\mathbf{s}|^{2H}\mathbf{E}\mathbf{E}^H|\mathbf{s}|^2}{|\mathbf{s}^H\mathbf{s}|^2}. \tag{27}$$

At this juncture, it may be observed that the signal comprising the periodic components of $|\mathbf{s}|^2$, called $\mathbb{P}\left[|\mathbf{s}|^2\right]$, containing all the cyclic frequencies of interest k can be written as

$$\mathbb{P}[\mathbf{s}] = \frac{1}{L-N+1}\sum_k e_k\left({e_k}^H|s|^2\right) = \frac{\mathbf{E}\mathbf{E}^H|\mathbf{s}|^2}{L-N+1}. \tag{28}$$

Substituting Eq. (2a) and Eq. (28) into Eq. (27) after a simple manipulation returns the final outcome:

$$ICS_2 = \frac{\mathbf{h}^H\mathbf{X}^H\mathbf{W}\mathbf{X}\mathbf{h}}{\mathbf{h}^H\mathbf{X}^H\mathbf{X}\mathbf{h}} = \frac{\mathbf{h}^H\mathbf{R}_{XWX}\mathbf{h}}{\mathbf{h}^H\mathbf{R}_{XX}\mathbf{h}} \tag{29}$$

where the weighting matrix $\mathbf{W}$ reads:

$$\mathbf{W} = diag\left(\frac{\mathbb{P}\left[|\mathbf{s}|^2\right]}{\mathbf{s}^H\mathbf{s}}\right)(L-N+1) = \begin{bmatrix} \ddots & & 0 \\ & \mathbb{P}\left[|\mathbf{s}|^2\right] & \\ 0 & & \ddots \end{bmatrix}\frac{(L-N+1)}{\sum_{l=N}^{L-1}|\mathbf{s}|^2}. \tag{30}$$

As done in Section 3.1, this BD criterion based on ICS_2 can be generalized to the p^{th} order ICS by appropriately modifying Eq. (30):

$$\mathbf{W} = \begin{bmatrix} \ddots & & 0 \\ & \mathbb{P}[|\mathbf{s}|^p] & \\ 0 & & \ddots \end{bmatrix}\frac{(L-N+1)^{\frac{p}{2}}}{\sum_{l=N}^{L-1}|\mathbf{s}|^{\frac{p}{2}}}. \tag{31}$$

Eq. (29) is the core of the proposed cyclostationary BD method, namely CYCBD. By solving Eq. (29) through eq. (30), the proposed method extract the source exhibiting the maximum CS2 behavior according to the cyclic frequency k.

3.2.1. Extension to single-input multi-output systems

So far, this BD method has been developed for SISO systems. However, the proposed algorithm can be easily extended to SIMO systems when multiple responses are available. Indeed, this algorithm version allows for improving

the source estimation through a simultaneous blind deconvolution of multiple response signals. For the SIMO model, the deconvolution problem can be reformulated considering a number Q of responses $\mathbf{x}_q$. The deconvolution is carried out simultaneously for each $\mathbf{x}_q$ returning the estimation of $\mathbf{s}$, being the sum of the q^{th} contributions computed by the q^{th} inverse filters $\mathbf{h}_q$.

Eqs. (1a), (29) and (19), that have been formulated for SISO systems, can be extended also for SIMO systems. $\mathbf{R}_{XWX}$ and $\mathbf{R}_{XX}$ can be expressed as cross-correlation matrices, being $\mathbf{R}_{qq}$ the auto-correlation matrix of $\mathbf{x}_q$ and $\mathbf{R}_{q\tilde{q}}$ the cross-correlation matrix between $\mathbf{x}_q$ and $\mathbf{x}_{\tilde{q}}$ (where $\tilde{q} = 1, \cdots, Q$ and $q \neq \tilde{q}$). Thus, $\mathbf{R}_{XX}$ can is given by:

$$\mathbf{R}_{XX} = \begin{bmatrix} \ddots & & \mathbf{R}_{q\tilde{q}} \\ & \mathbf{R}_{qq} & \\ \mathbf{R}_{\tilde{q}q} & & \ddots \end{bmatrix} \tag{32}$$

where the diagonal elements are composed of the auto-correlation matrices of $\mathbf{x}_q$ and the off-diagonal elements are composed of the cross-correlation matrices of $\mathbf{x}_q$ and $\mathbf{x}_{\tilde{q}}$. The weighted correlation matrix, $\mathbf{R}_{XWX}$, can be defined analogously. Again, the deconvolution problem can be iteratively through by the generalized eigenvalue problem reported in Eq. (19) which returns $\mathbf{h}$, that is the concatenation of vectors of filter coefficients related to the q^{th} response

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_q \\ \vdots \\ \mathbf{h}_Q \end{bmatrix}, \tag{33}$$

and λ that is is the final value of the maximized criterion. From Eq. (33), the q^{th} contribution to the source $\mathbf{s}$ can be calculated as:

$$\mathbf{s}_q = \mathbf{X}_q \mathbf{h}_q \tag{34}$$

where $\mathbf{X}_q$ is the Toeplitz matrix estimated as in Eq. (1b) related to the q^{th} response $\mathbf{x}_q$. Finally, the overall source, $\mathbf{s}$, is computed by the sum of all the q^{th} sources calculated using Eq. (34):

$$\mathbf{s} = \sum_{q=1}^{Q} \mathbf{s}_q. \tag{35}$$

3.2.2. Weighting matrix in the angle domain

It is a matter of fact that vibration signatures of rotating machines often exhibit periodicity locked in the angle domain rather than in the time domain [12, 20]. Hence, the source estimation can be improved by re-defining the weighting matrix reported in Eq. (31), that is composed of periodic components in the time domain, in a different form consisting of periodic components in the angle domain. Recalling $\mathbb{P}[|\mathbf{s}|^p]$ in Eq. (31), this term is obtained by extracting the Fourier coefficients related to a certain set of cyclic frequencies and then returning the filtered signal in the time domain; viz:

$$\mathbb{P}[|\mathbf{s}|^p] = \sum_k c_k e^{-j2\pi \frac{k}{T_s} l} \quad with \quad c_k = \frac{1}{L-N+1} \sum_{l=N}^{L-1} |s[l]|^p e^{-j2\pi \frac{k}{T_s} l}. \tag{36}$$

Exploiting the measured instantaneous speed (if available), the periodic component locked in the angular domain may be extracted avoiding resampling [21] by a change of variable in Eq. (36). Therefore, under the assumptions that the considered signal is time/angle dependent, i.e. $\mathbf{s} = \mathbf{s}(\theta(t))$, Eq. (31) can be expressed in the time/angle domain as:

$$\mathbb{P}[|\mathbf{s}|^p] = \sum_k c_k e^{jk\theta[l]} \quad with \quad \begin{cases} c_k = \frac{1}{\Theta} \sum_{l=N}^{L-1} |s[l]|^p e^{-jk\theta[l]} \dot{\theta}[l] \\ \theta[l] = \sum_{\tilde{l}=N}^{l-1} \dot{\theta}[\tilde{l}] \\ \Theta = \sum_{l=N}^{L-1} \dot{\theta}[l] \end{cases} \tag{37}$$

where Θ is the normalization term.

The cyclostationary BD method based on Eq. (37) is hereafter referred as CYCBDang. The periodic component extracted by using Eq. (37) should lead to better results for the diagnostics of rotating machines since it improves the estimation of the cyclostationary signal part also in presence of speed fluctuations.

3.2.3. Further considerations

It should be observed that the weighting matrix reported in Eq. (30) is reminiscent. De facto, **W** expressed for the maximum kurtosis BD (see Eq. (17)) is very similar to the weighting function obtained in Eq. (30). From this similarity it can be deduced that the maximization of ICS_2 is equivalent to maximize the kurtosis of the CS2 part according to the cyclic frequency k and its multiples. From the physical point of view, this criterion drives the deconvolution (based on maximum kurtosis) with respect to only one phenomenon characterized by a specific cycle.

This property is particularly relevant for the diagnosis of rotating machines. Indeed, as pointed out in Refs. [3, 11], it may be exploited to highlight a specific fault as well as to discriminate the faulty rotating component from the healthy ones. However, the proposed criterion adds something more than the others presented in the specialized literature. On the one hand, MOMEDA deals with the extraction of an impulsive source approaching a Dirac comb, which is deterministic. On the other hand, MCKD exploits a cyclostationary criterion - i.e. the correlated kurtosis - as the CYCBD but owns some limitations (see also Section 2.2.2). The first one is that the MCKD can be used just for short signals involving a limited number of consecutive impulses [11]. Indeed, the selected signal length is related to the shift number, M, that cannot be too large, as explained in Section 2.2.2. The second one regards the definition of the impulse period T_s. Being T_s a constant, it is formally valid only for systems operating at constant speed while CYCBD can benefit of Eq. (37) in order to deals with fluctuating speeds. Moreover, CYCBD can be adapted to maximize different orders of cyclostationarity. Conceptually, this is a key point since the rotating machines exhibit cyclostationary behaviors of different order depending on the type of fault and of the systems [20]. Thus, CYCBD can be more versatile for the rotating machine diagnosis than the MCKD.

To sum up, it has been demonstrated that BD problem expressed as a generalized Rayleigh quotient represents a versatile approach being easily adapted to arbitrary criteria, such as the maximization of κ_p (see Eq. (21)) or ICS_p (see Eq. (31)), just by selecting proper weighting matrices.

4. Comparison considering synthesized signals

The validation of the criterion based on the maximization of ICS_2 using simulated signals is provided in this section. The CYCBD is compared among other techniques already published as: MED (Eq. (17)), OMEDA [2, 3], MCKD [11] and MOMEDA [3].

4.1. Signal description

The simulations have been carried out in Matlab environment exploiting also the Matlab scripts available in Ref. [3]. A Matlab implementation of the proposed cyclostationary BD method is provided online[1] as well. Let us consider different types of cyclostationary signals:

1. periodic impulses with Gaussian distributed amplitudes and additive Gaussian background noise (SNR = -19 dB);
2. periodic impulses having Gaussian distributed amplitudes with jitter effect (following a Gaussian distribution) and additive Gaussian background noise (SNR = -19 dB);
3. a couple of trains of impulses (with different cyclic frequency sets) having Gaussian distributed amplitude and additive Gaussian background noise (SNR = -19 dB);
4. periodic impulses with Gaussian distributed amplitudes and additive Gaussian background noise (SNR = -19 dB) with the addition of a single dominant impulse;
5. train of impulses with Gaussian distributed amplitudes having fluctuating cycle and additive Gaussian background noise (SNR = -19 dB);

[1] `putthelinkhere`

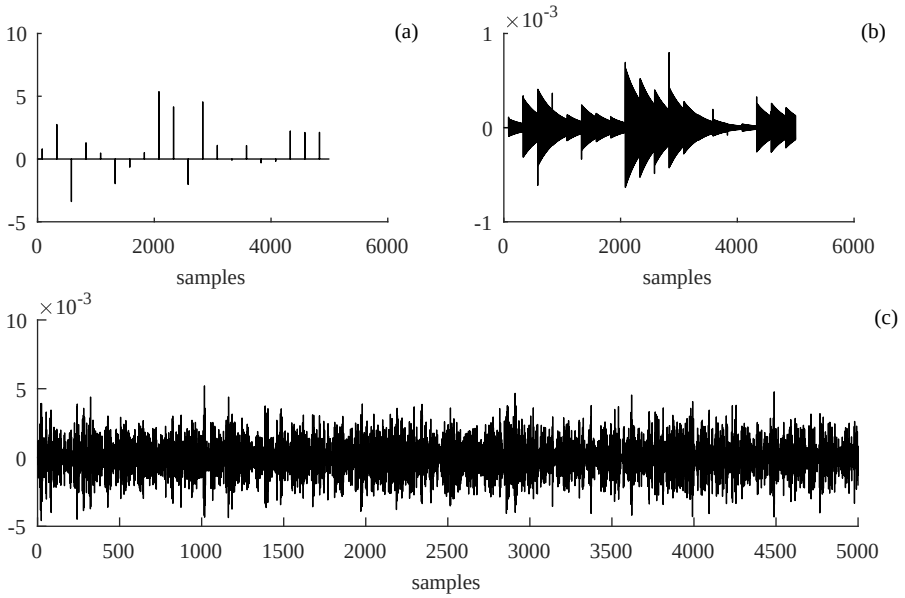

Fig. 2: Simulated signal $\mathbf{x}_{sim,1}$: (a) train of equispaced impulses $\mathbf{s}_{0,1}$ having random (Gaussian) amplitudes with cycle $1/T_{s,1}$; (b) $\mathbf{s}_{0,1}$ convolved with the IRF $\mathbf{g}_{s,1}$; (c) overall signal $\mathbf{x}_{sim,1}$ with SNR = -19 dB.

which, for simplicity, are called $\mathbf{x}_{sim,1}$, $\mathbf{x}_{sim,2}$, $\mathbf{x}_{sim,3}$, $\mathbf{x}_{sim,4}$ and $\mathbf{x}_{sim,5}$, respectively. These simulated signals are cyclostationary in different ways. In fact, $\mathbf{x}_{sim,1}$, $\mathbf{x}_{sim,3}$ and $\mathbf{x}_{sim,4}$ have periodic impulses with (Gaussian) random amplitude while $\mathbf{x}_{sim,2}$ and $\mathbf{x}_{sim,5}$ own also a (Gaussian) random fluctuation of the impulse occurrence rate and a sinusoidal fluctuation of the impulse occurrence rate, respectively.

These signals are expressed in agreement with the previous general formulation (see Eq. (3)), neglecting the periodic component $\mathbf{p}$, viz:

$$\mathbf{x} = \sum_{i=1}^{I} \mathbf{s}_{0,i} * \mathbf{g}_{s,i} + \mathbf{n} * \mathbf{g}_n. \tag{38}$$

where index I indicates the number of the impulsive patterns ($I = 1$ for $\mathbf{x}_{sim,1}$, $\mathbf{x}_{sim,2}$ and $\mathbf{x}_{sim,5}$; $I = 2$ for $\mathbf{x}_{sim,3}$ and $\mathbf{x}_{sim,4}$). The parameters used for the synthesized signals are resumed in Tab. 1 where: F_s is the sampling frequency, T is the time length of the signal, T_s is the impulse period, $\mathbf{g}$ is the IRF, σ_{imp} is the standard deviation of the impulse amplitudes and σ_j is the standard deviation of the jitter. In these simulations, the IRFs $\mathbf{g}_{s,i}$ and $\mathbf{g}_n$ have been modeled as responses of a damped single degree of freedom (SDOF) system to the time domain unit impulse [22]. The damping is assumed viscous and sub-critical, as occurs in many real systems. In terms of displacements, the SDOF IRF for a continuous signal is defined as:

$$\mathbf{x}_{SDOF} = Ae^{-\zeta\omega_n t} \sin(\omega_d t) \tag{39}$$

where A is the response magnitude, ζ the damping coefficient, ω_n the resonance (angular) frequency and $\omega_d = \omega_n \sqrt{1-\zeta^2}$. The IRFs $\mathbf{g}_{s,i}$ and $\mathbf{g}_n$, expressed in terms of accelerations, are merely obtained by taking the second derivative with respect to time of Eq. (39). Tab. 2 recaps the IRFs parameters while Fig. 2 to Fig. 6 display the simulated signals. Each simulated signal owns (at least) a pure cyclostationary source and has been designed with the specific purpose of highlighting the limitations of the considered BD algorithms from different standpoints.

Before examining the results, let us first discuss the general settings used for the different BD techniques in this comparison. As pointed out previously, care should be taken to select the filter length in order to achieve good quality results using BD algorithms: depending on the criterion, different strategies should be adopted. This aspect is argued in detail in Refs. [3, 11]. The filter lengths used for the following comparisons are reported int Tab. 3.

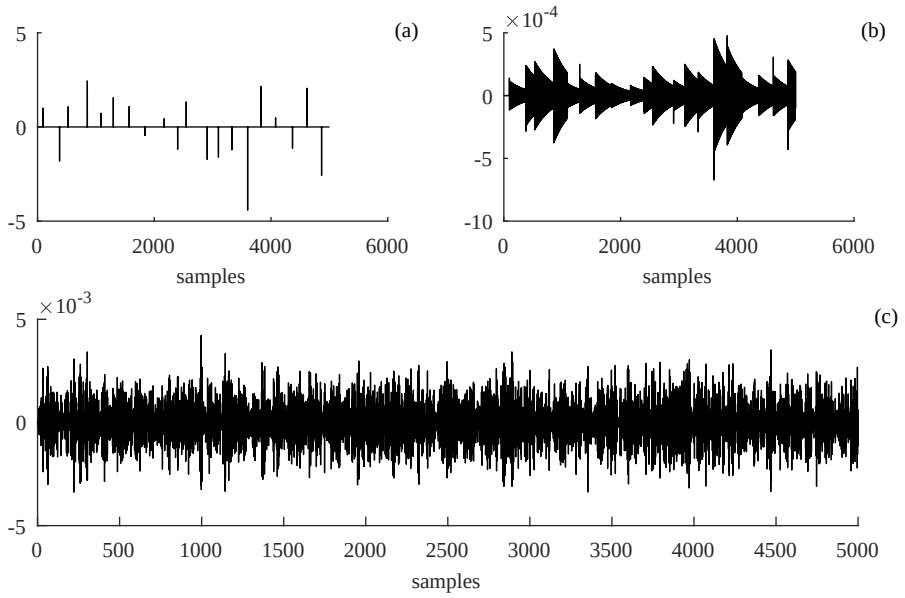

Fig. 3: Simulated signal $x_{sim,2}$: (a) train of equispaced impulses $\mathbf{s}_{0,1}$ having random (Gaussian) amplitudes with cycle $1/T_{s,1}$ and jitter (Gaussian distribution); (b) $\mathbf{s}_{0,1}$ convolved with the IRF $\mathbf{g}_{s,1}$; (c) overall signal $\mathbf{x}_{sim,2}$ with SNR = -19 dB.

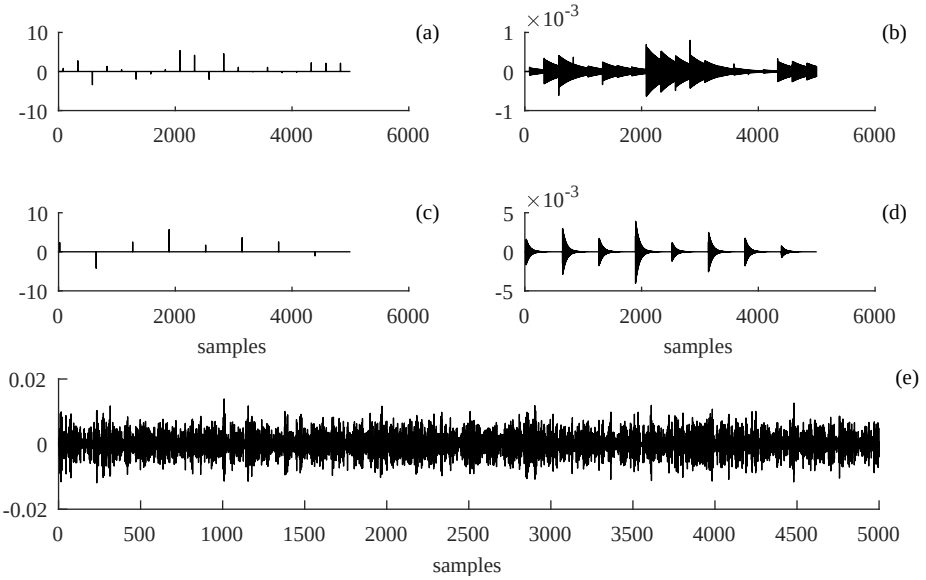

Fig. 4: Simulated signal $\mathbf{x}_{sim,3}$: (a) train of equispaced impulses $\mathbf{s}_{0,1}$ having random (Gaussian) amplitudes with cycle $1/T_{s,1}$; (b) $\mathbf{s}_{0,1}$ convolved with IRF $\mathbf{g}_{s,1}$; (c) train of equispaced impulses $\mathbf{s}_{0,2}$ having random (Gaussian) amplitude with cycle $1/T_{s,2}$; (d) $\mathbf{s}_{0,2}$ convolved with IRF $\mathbf{g}_{s,2}$; (e) overall signal $\mathbf{x}_{sim,3}$ with SNR = -19 dB.

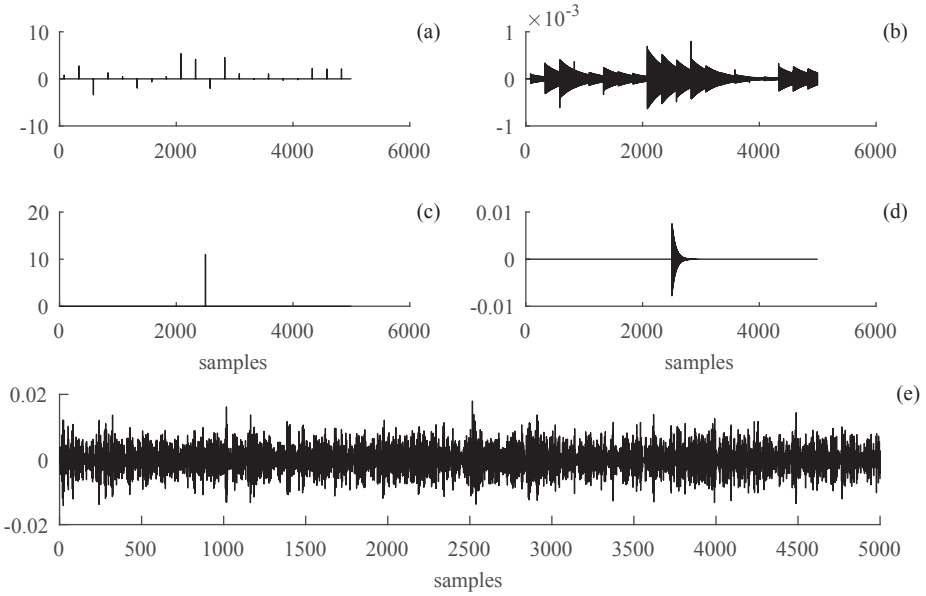

Fig. 5: Simulated signal $\mathbf{x}_{sim,4}$: (a) train of equispaced impulses $\mathbf{s}_{0,1}$ having random (Gaussian) amplitudes with cyclic frequency $1/T_{s,1}$; (b) $\mathbf{s}_{0,1}$ convolved with IRF $\mathbf{g}_{s,1}$; (c) the single dominant peak $\mathbf{s}_{0,2}$; (d) $\mathbf{s}_{0,2}$ convolved with IRF $\mathbf{g}_{s,2}$; (e) overall signal $\mathbf{x}_{sim,4}$ with SNR = -19 dB.

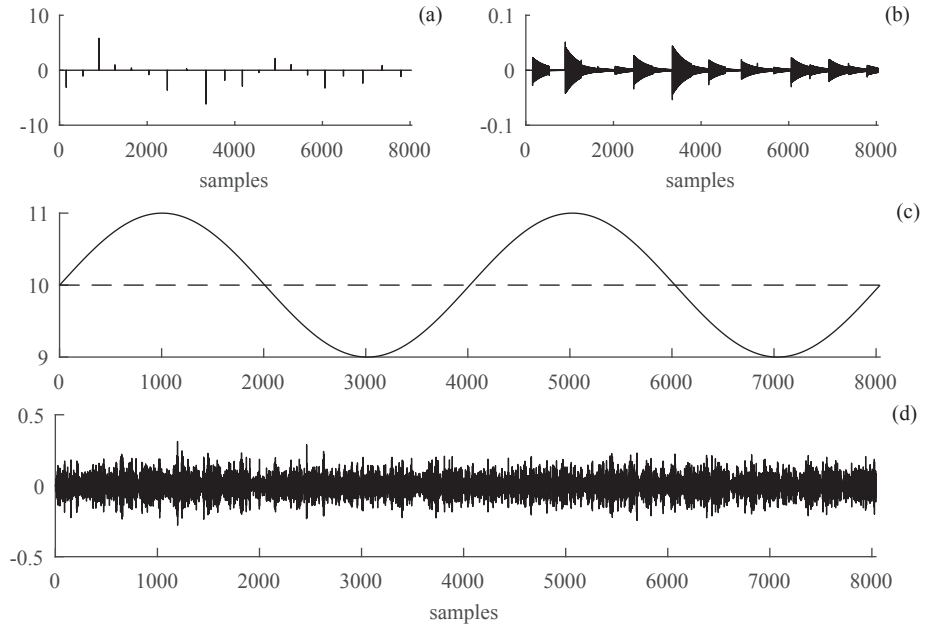

Fig. 6: Simulated signal $\mathbf{x}_{sim,5}$: (a) train of equispaced impulses $\mathbf{s}_{0,1}$ having random (Gaussian) amplitudes with variable cycle $1/T_{s,1}$; (b) $\mathbf{s}_{0,1}$ convolved with IRF $\mathbf{g}_{s,1}$; (c) cyclic frequency values fluctuating around the mean value (dotted line); (d) Gaussian noise $\mathbf{n}$ convolved with IRF $\mathbf{g}_n$; (e) overall signal $\mathbf{x}_{sim,5}$ with SNR = -19 dB.

Tab. 1: Parameters used for the synthesized signals.

	$\mathbf{x}_{sim,1}$	$\mathbf{x}_{sim,2}$	$\mathbf{x}_{sim,3}$	$\mathbf{x}_{sim,4}$	$\mathbf{x}_{sim,5}$
$L\ (samples)$	5000	5000	5000	5000	5000
$T_{s,1}\ (samples)$	250	250	250	250	250
$T_{s,2}\ (samples)$	-	-	625	-	–
$\sigma_{s,1}$	1.500	1.500	1.500	1.500	1.500
$\sigma_{s,2}$	-	-	3.500	-	-
σ_j	-	0.025	-	-	-
$SNR\ (dB)$	-19	-19	-19	-19	-19

Tab. 2: Parameters used for the computation of the IRFs.

	$\mathbf{g}_{s,1}$	$\mathbf{g}_{s,2}$	$\mathbf{g}_n$
A	$1.963\,10^{-10}$	$3.591\,10^{-10}$	$1.258\,10^{-10}$
ζ	0.004	0.01	0.05
$\omega_n\ (rad/s)$	19.894	39.788	63.662

4.2. Results and discussion

Fig. 7 summarizes the outputs obtained for different BD algorithms considering $\mathbf{x}_{sim,1}$. Note that source signal s_0 is buried under strong background noise ($SNR = -19\ dB$) and all the output signals are normalized by their respective maximum value. This normalization is allowed since BD is unable to recover the actual source amplitude. Furthermore, it should be remarked that the estimated sources are delayed with respect to the output signal due to the filtering. In this specific application, the FIR filters introduce a constant delay with respect to the frequency domain which can be easily corrected. From now, the filter delay has been kept since it does not affect the result discussion. From the qualitative standpoint, the best result is achieved by MED, MCKD and CYCBD that provide a satisfying representation of the source signal. The impulses are extracted with their correct rate of repetition as well as their relative magnitudes. The MCKD performance is expected since it is based on a cyclostationary criterion. However, the correlated kurtosis is very sensitive to slight changes of the fundamental fault cycle. Hence, the jitter effect or even very slight changes on the impulse repetition rate should affect the MCKD results, as demonstrated by further examples. Unfortunately, the smallest impulses are difficult, even impossible, to be detected since they are overwhelmed by the background noise that is still present in the recovered signal. MOMEDA and OMEDA are able to properly extract just the prominent peaks, since they present a remarkable background noise kept by the recovered signal.

From these results, it is clear that the outputs of MED and CYCBD appear very similar. This outcome is not surprising since it has been demonstrated in Section 3.2 that there is a strong mathematical similarity between these two criteria. However, CYCBD owns the capability to recover signals characterized by certain cyclic frequencies and this feature may be exploited for diagnostics purpose as investigated in the following section.

In the same fashion, the results regarding $\mathbf{x}_{sim,2}$ are summarized in Fig. 8. In this case, the random part of the signal is higher than in the previous example because of the jitter, which follows a Gaussian distribution. The presence of the jitter strongly worsens the MCKD output since the repetition of the impulses is no longer at constant rate. Again, MOMEDA returns to be not effective for the recover of cyclostationary sources since it assumes a periodic source to extract.

So far, no remarkable differences have been found between MED and CYCBD considering $\mathbf{x}_{sim,1}$ and $\mathbf{x}_{sim,2}$. The examples represented by $\mathbf{x}_{sim,3}$ and $\mathbf{x}_{sim,4}$ emphasize the differences between these algorithms. The simulated signal

Tab. 3: Filter lengths (in samples) used for benchmarking BD algorithms with simulated signals.

x_{sim}	MED	OMEDA	MCKD	MOMEDA	CYCBD	CYCBDang
1-4	300	300	300	300	300	-
5	-	-	100	400	-	250

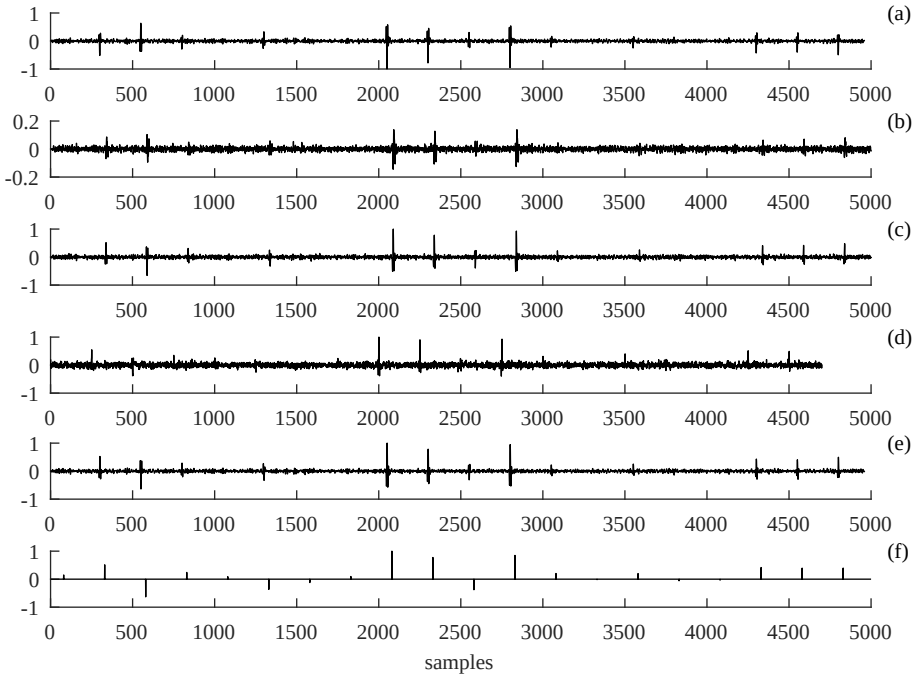

Fig. 7: Comparison of the BD results regarding $\mathbf{x}_{sim,1}$: (a) MED, (b) OMEDA, (c) MCKD, (d) MOMEDA, (e) CYCBD and (f) the target source.

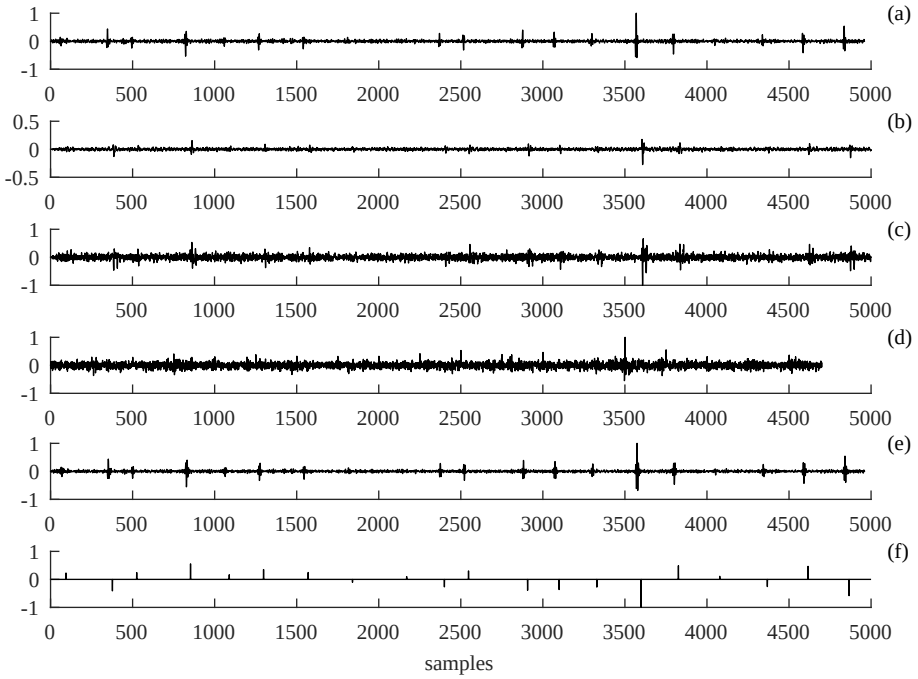

Fig. 8: Comparison of the BD results regarding $\mathbf{x}_{sim,2}$: (a) MED, (b) OMEDA, (c) MCKD, (d) MOMEDA, (e) CYCBD and (f) the target source.

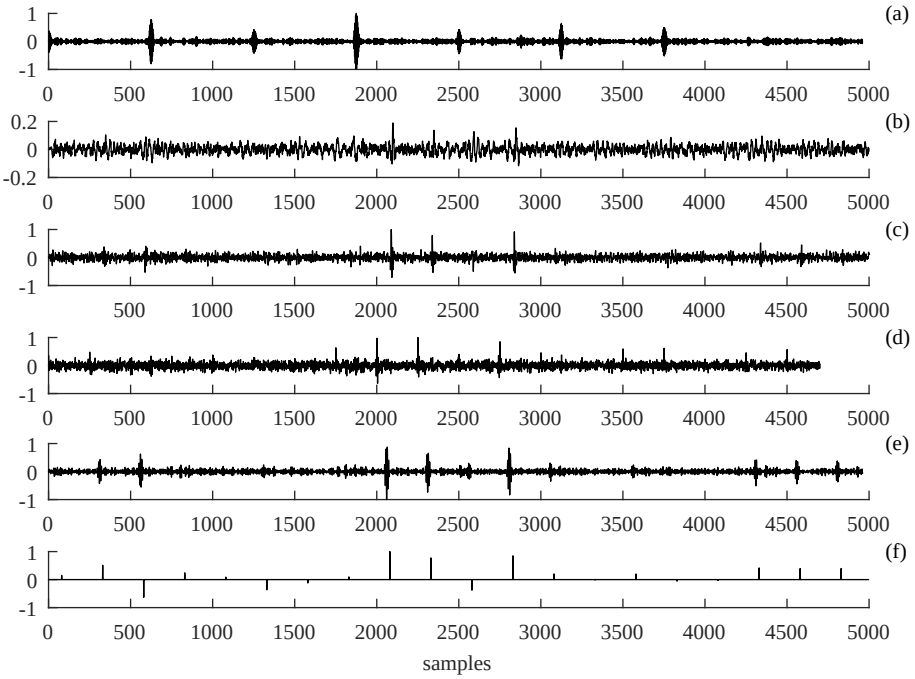

Fig. 9: Comparison of the BD results regarding $\mathbf{x}_{sim,3}$ focusing on fault period $T_{s,1}$: (a) MED, (b) OMEDA, (c) MCKD, (d) MOMEDA, (e) CYCBD and (f) the target source.

$\mathbf{x}_{sim,3}$ has been obtained by adding a second pattern of impulses with a different cyclic frequency. This signal may be useful in order to inspect the capability of CYCBD to discriminate different types of fault, depending on their periodicity. The expectation is that MED will recover the source exhibiting the lower number of impulses, i.e. the impulses with period $T_{s,2}$, since the lower the number of impulses, the higher the value of the kurtosis. This behavior should be followed also by OMEDA. Assuming that the source of interest is the one having period $T_{s,1}$, the results shown in Fig. 9 agrees with the previous prediction. Indeed, MED (diagram (a) in Fig. 9) recovers the source having period $T_{s,2}$ whereas CYCBD is the only one that recovers correctly the desired impulsive source. Moreover, note that MOMEDA is not biased by the presence of another impulsive pattern, but the provided estimation appears very noisy and the recovered peaks are barely observable; analogously, MCKD returns similar results to MOMEDA. OMEDA completely fail the source recovering. Fig. 10 collects the results obtained with the same data of Fig. 9, but focusing on the target source having period $T_{s,2}$. As expected, CYCBD and the other BD techniques that depend on the selected fault occurrence rate, i.e. MCKD and MOMEDA, can properly extract the target signal (diagram (f) in Fig. 10). Note that, in this case, all the recovered signals are globally better than those collected in Fig. 9. This behavior can be explained by the fact that, since the two mixed impulsive patterns have similar amplitudes (see diagram (a) and (c) in Fig. 4), the maximized criteria tends to be more effective for the pattern, exhibiting the minor amount of impulses in the full time span. The simulated signal $\mathbf{x}_{sim,4}$ (see Fig. 5) addresses the case of a train of impulses with Gaussian distributed amplitude with a single dominant impulse. Likewise to $\mathbf{x}_{sim,3}$, MED as well as OMEDA deconvolve the single dominant peak instead of the train of impulses, as reported in Fig. 11. Moreover, also MCKD fails to provide the desired results.

Care should be paid to signals that exhibit more than one impulsive source. In fact, all the BD techniques addressed in this research are valid under the hypothesis of the presence of a single impulsive pattern to deconvolve; in other words, they basically refer to SISO systems. Hence, the proposed attempt to extract multiple sources only demonstrates that, under certain conditions, CYCBD - and to a lesser extent MCKD and MOMEDA - can overcome to this intrinsic limit. But the results shown in Figs. 9 and 10 can not be considered as a definitive proof of the

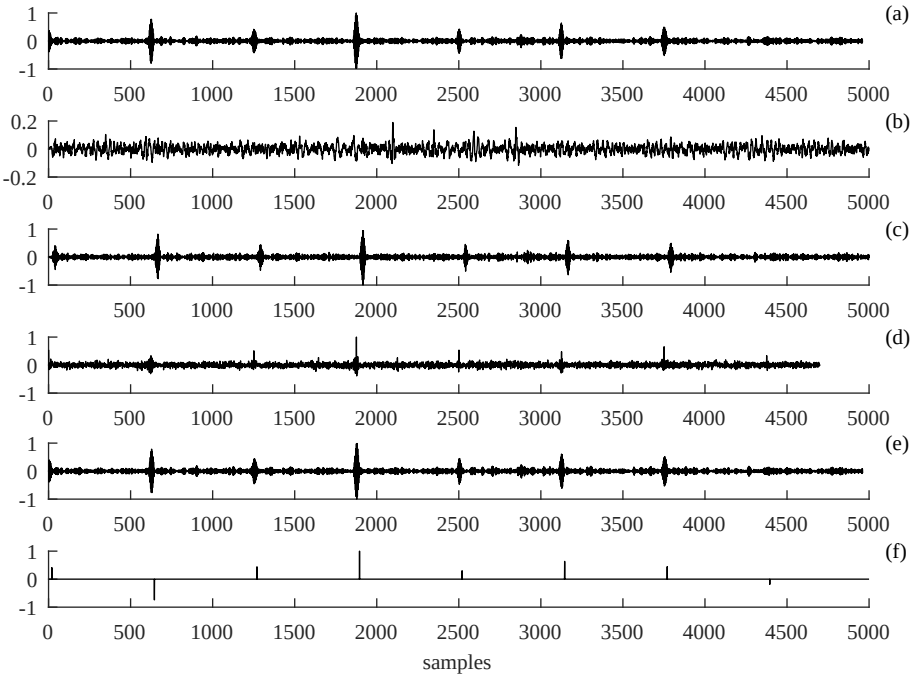

Fig. 10: Comparison of the BD results regarding $\mathbf{x}_{sim,3}$ focusing on fault period $T_{s,2}$: (a) MED, (b) OMEDA, (c) MCKD, (d) MOMEDA, (e) CYCBD and (f) the target source.

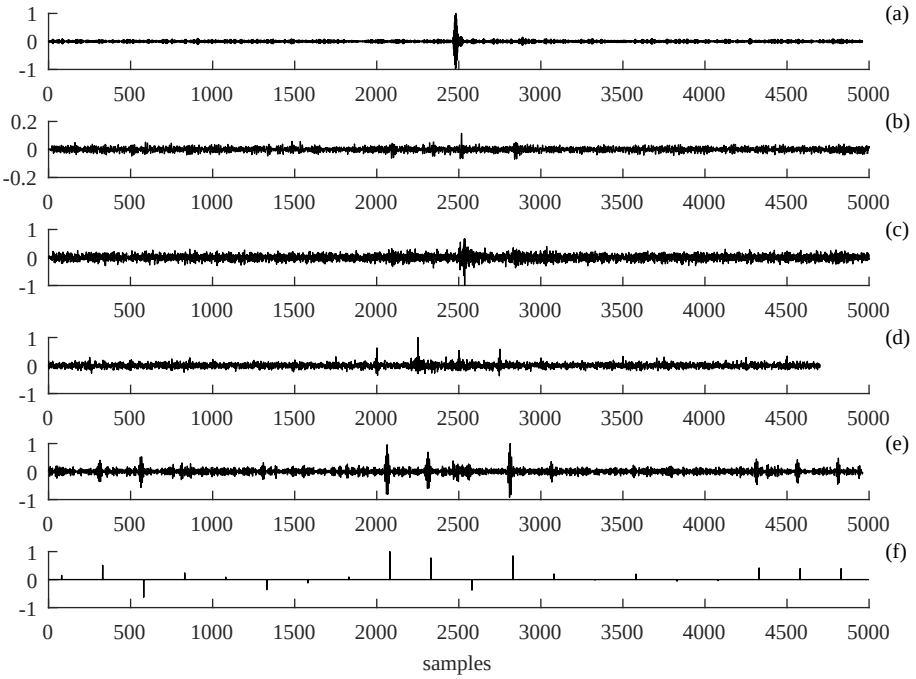

Fig. 11: Overall comparison of the BD results regarding $\mathbf{x}_{sim,4}$: (a) MED, (b) OMEDA, (c) MCKD, (d) MOMEDA, (e) CYCBD and (f) the target source.

effectiveness of CYCBD on signal separation considering multiple sources.

Finally, $\mathbf{x}_{sim,5}$ concerns the case of a train of impulses that are equispaced in the angle domain but they have a cyclic frequency varying in the time domain. In practice, this could be the case of bearing faults or gear faults under variable operating conditions. As reported in Fig. 6(c), the cyclic frequency oscillates sinusoidally with a mean value of 10 Hz and an amplitude of ± 1 Hz. Fig. 12 summarizes the results of the last numerical example and does not display the results of MED and OMEDA since the goal is to underline the limitations of MCKD and MOMEDA regarding non-equispaced impulses. In this case, the CYCBDang is driven by the tacho reference while MCKD and MOMEDA are performed by using the average impulse period, which is 10 Hz. It can be easily noticed that CYCBDang recovers the original source with good accuracy in terms of relative amplitude as well as impulse occurrence rate. However, both MCKD and MOMEDA are not capable to recover the original source providing a noisy output.

This section has remarked the advantages of CYCBD for the analysis of cyclostationary signals in comparison with BD algorithms already published in the literature by using dedicated examples. In general, CYCBD returns better results with respect to the other BD algorithms when the source is purely cyclostationary. The simulated results highlight the superiority of CYCBD with respect to MOMEDA, which is expected since MOMEDA fits with periodic sources. Moreover, MCKD is overcome too even if correlated kurtosis is a cyclostationary criterion. This is likely due to the definition of correlated kurtosis. Indeed, as remarked in Section 2.2.2, correlated kurtosis can lose its effectiveness for long noisy responses [11] as well as impulses not perfectly equispaced. However, this preliminary benchmark has been performed considering ad hoc simulated signals with marked cyclostationary behavior. Thus, further verifications are carried out in the next section considering real signals.

5. Application to real vibration signals

This section is devoted to the investigation of the CYCBD method considering two different applications. The first one addresses a gear tooth spall identification on a two-stage gearbox; the second one regards a run-to-failure bearing

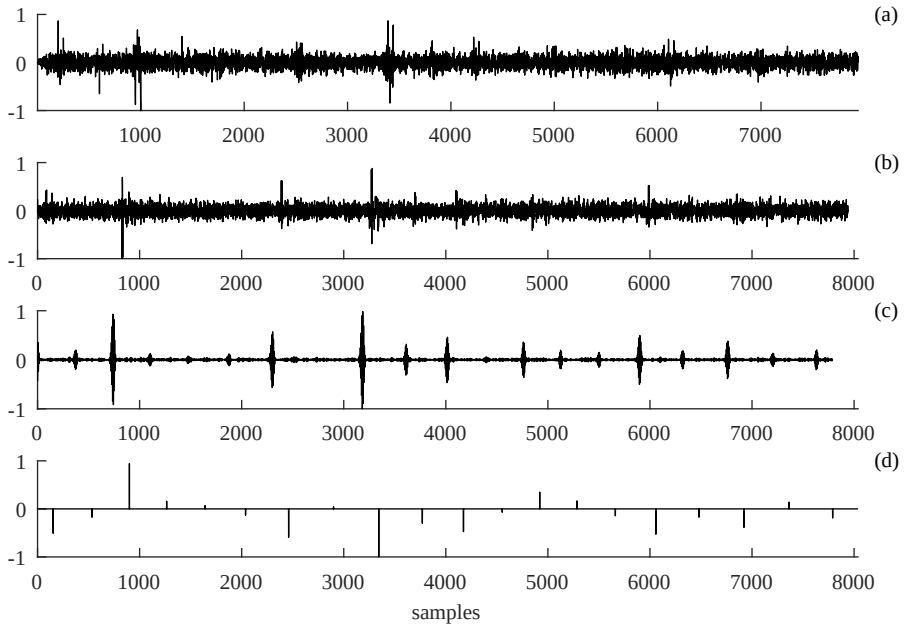

Fig. 12: Overall comparison of the results regarding $\mathbf{x}_{sim,5}$: (a) MCKD, (b) MOMEDA, (c) CYCBDang and (d) the target source.

test.

5.1. Diagnosis of gear tooth spall

The first application deals with the detection of a gear tooth spall. First, a qualitative comparison of the results of several BD algorithm is presented; second, a procedure for gearbox diagnostics for the identification and quantification of seeded gear spalls based on BD technique is proposed and discussed.

5.1.1. Experimental setup

The first experimental verification of the proposed algorithm has been assessed by means of a dedicated test rig, shown in Fig. 13, located at the Engineering Department of the University of Ferrara. Detailed information about this test rig can be found in Ref. [23]. The vibration signals in the radial direction have been collected by means of B&K piezoelectric accelerometer placed on the bearing support of the first stage pinion with sampling frequency 10.2 kHz for a total time duration of 4 s. The measurement campaign has been conducted using LMS SCADAS 310 controlled by LMS Test.Lab.

The investigated gearbox is composed of two stages of helical gears: the first one having 18 and 71 teeth and the second one 12 and 55 teeth. Four different sizes of gear tooth spall have been realized via milling process in the 71 teeth wheel (first stage) in order to verify the sensitivity of the proposed criterion. Table 4 collects the details of the artificial defects sorted according to the percentage ratio between the spall size and the whole tooth face. More information about the reproduction of the gear tooth spall by milling process can be found in Ref. [24]. All the steady condition tests have been carried out considering the following nominal conditions: input shaft speed of 3600 rpm and nominal load of 48.8 Nm.

It should be noted that the test condition just described is particularly unfavorable for the spall detection. Firstly, gear tooth spall is harder to detect in helical gears than in spur gears since the contact in helical gear is smoother. This feature, in general, favors the reduction of impulsive components due to the contact among teeth. Secondly, considering that the higher the load, the better the contact among gear teeth, the test load is significantly lower than the nominal load of the gearbox in actual working condition.

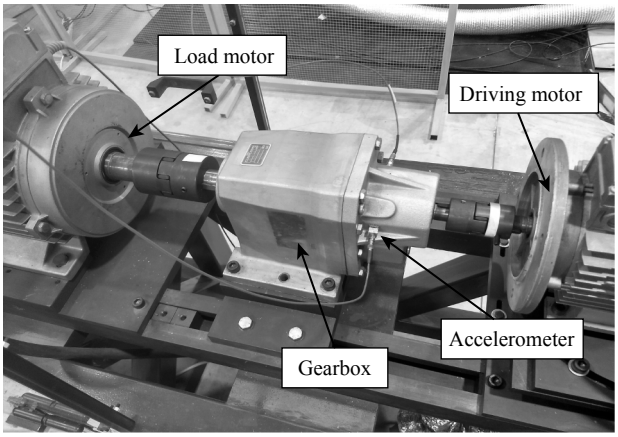

Fig. 13: Experimental setup.

Tab. 4: Description of the gear tooth spalls considered in the experimental campaign [25].

ID	Fault description
Sp12.5	2 *mm* along the tooth profile, 0.6 *mm* depth, 2 *mm* across the tooth face (12.5 % of the tooth face width)
Sp25	2 *mm* along the tooth profile, 0.6 *mm* depth, 4 *mm* across the tooth face (25 % of the tooth face width)
Sp50	2*mm* along the tooth profile, 0.6 *mm* depth, 7.8 *mm* across the tooth face (50 % of the tooth face width)
Sp100	2 *mm* along the tooth profile, 0.6 *mm* depth, 15.5 *mm* across the tooth face (100 % of the tooth face width)

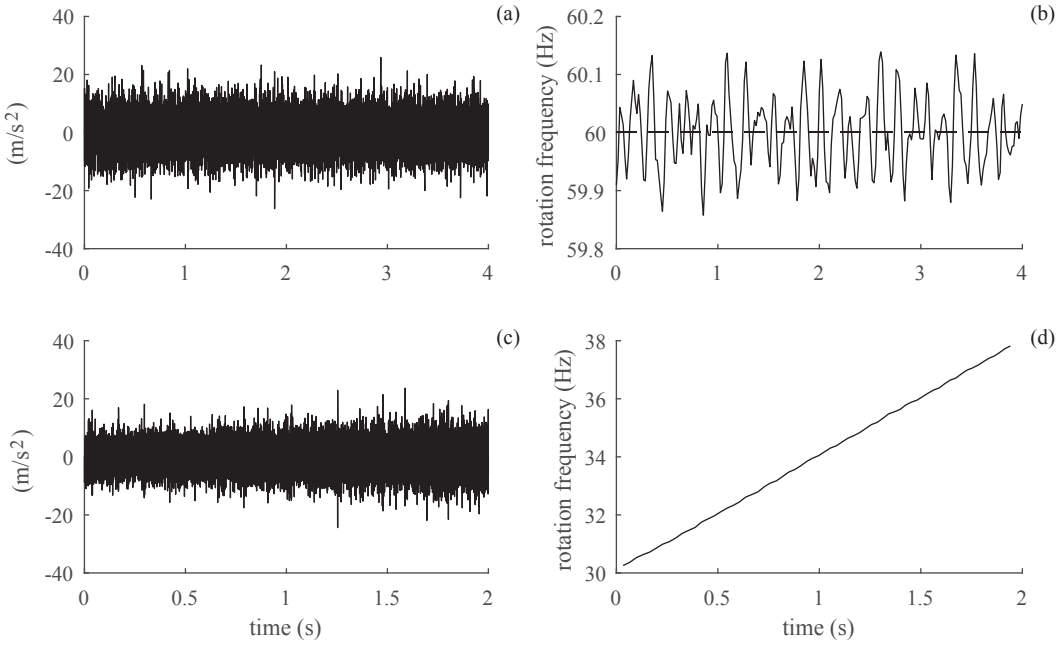

Fig. 14: Measured vibration signals in the sp100 case at (a) constant speed and (b) variable speed with (b,d) their speed profiles.

A variable speed test has been also carried out in order to validate experimentally the CYCBDang algorithm. In this case, the test has been performed in run-up condition between $30\,Hz$ and $34\,Hz$ referenced to the input shaft rotation frequency, with a load of $45.8\,Nm$. An example of the acquired vibration signals and their speed profiles is reported in Fig. 14.

5.1.2. Result discussion: constant regime tests

The results presented in this section have been carried out considering the following filter lengths: 50 samples for MED and MOMEDA, 40 samples for MCKD (with 5 shifts), 800 samples for MOMEDA and 700 samples for CYCBD. The CYCBDang has not been considered since the benefits of using this algorithms should be negligible in a constant speed case. All the recovered signals have been normalized by their respective absolute maximum value in order to facilitate their comparison. Furthermore, the MCKD has been performed using the whole signal in order to highlight its intrinsic limitation to deal with long signal and with non-constant speed [11].

Figs. 15, 16, 17 and 18 summarize the results related to cases Sp12.5, Sp25, Sp50 and Sp100, respectively. The first observation is that MED correctly recover the fault train only in case Sp100 (see Fig. 18), i.e. where the defect should be more evident, whereas OMEDA appears ineffective in all the considered cases. The tendency of OMEDA to be recover a single large impulse is highlighted in Fig. 15(b) and to a lesser extent in Fig. 16(b). As expected, the challenging test conditions highlighted the significant limitations of MED and OMEDA when applied to mechanical vibration signals.

Now, let us discuss the results related to BD algorithms designed for rotating machine diagnostics (MCKD and MOMEDA) compared to the proposed ones. The MCKD returns satisfactory results just in cases Sp100 and Sp50 but in the remaining cases the algorithm does not recover correctly the train of impulses. These results can be explained by the fact that, as reported in Fig. 14, the instantaneous speed is not perfectly constant. Considering that the definition of correlated kurtosis (see Eq. (7)) is based on the correlation of signal segments having a fixed lag mT_s, which correspond to a constant fault period, the MCKD could fail on the extraction of very slight impulsive sources buried in background noise and mechanical interferences. The other BD algorithms examined in this comparison – namely MOMEDA and CYCBD – are able to properly deconvolve the sequence of impulses due to the gear tooth spall. De facto, this preliminary result is not surprising since the gear tooth spall can be modeled as a composition of CS1 signal

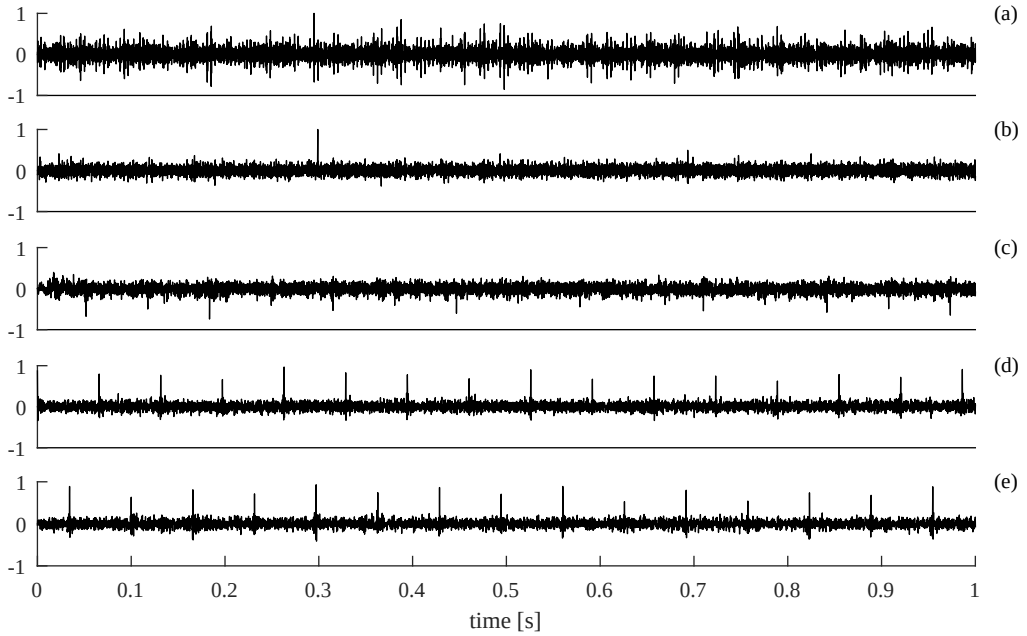

Fig. 15: Comparison of the results with 12.5 % spalling (Sp12.5) considering: (a) MED, (b) OMEDA, (c) MCKD, (d) MOMEDA and (e) CYCBD.

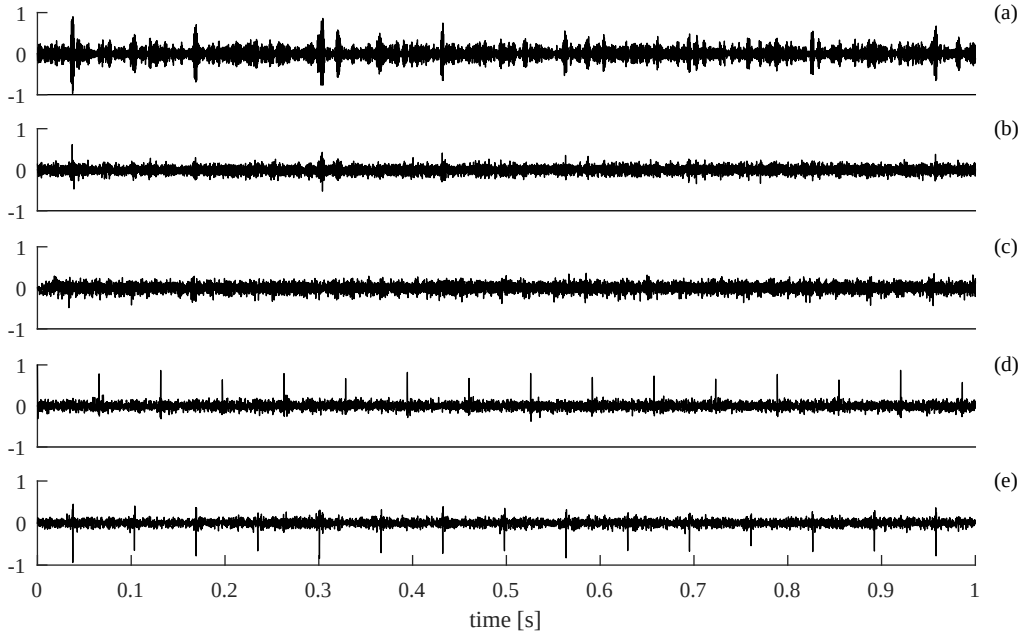

Fig. 16: Comparison of the results with 25 % spalling (Sp25) considering: (a) MED, (b) OMEDA, (c) MCKD, (d) MOMEDA and (e) CYCBD.

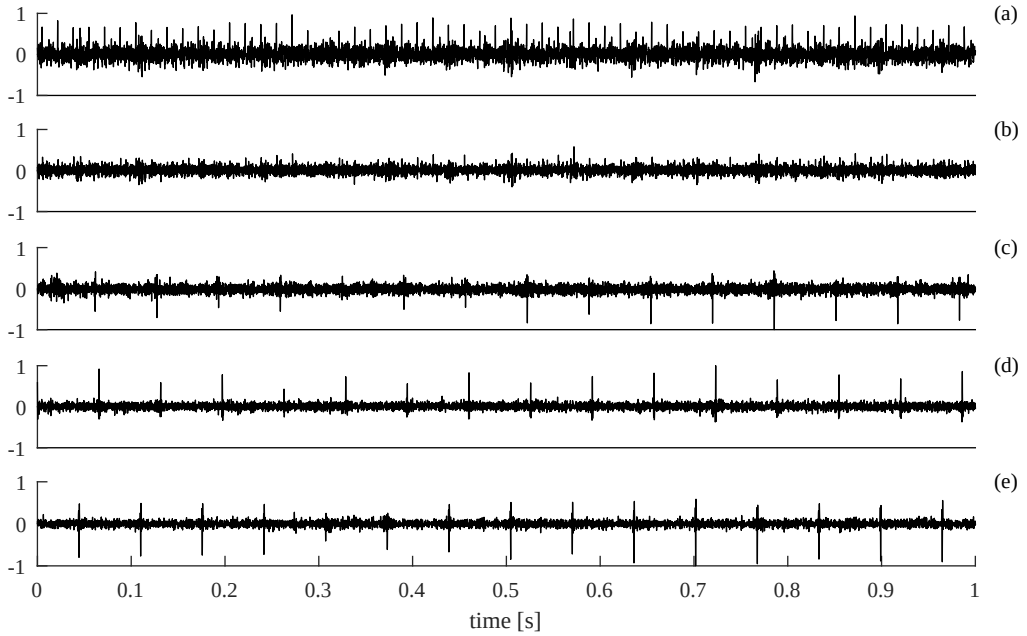

Fig. 17: Comparison of the results with 50 % spalling (Sp50) considering: (a) MED, (b) OMEDA, (c) MCKD, (d) MOMEDA and (e) CYCBD.

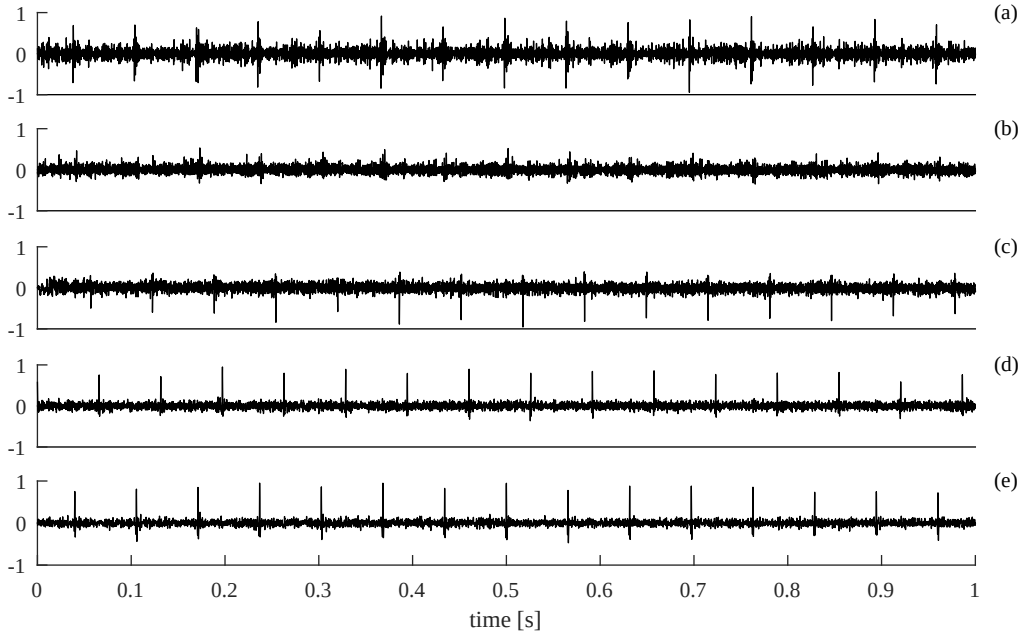

Fig. 18: Comparison of the results with 100 % spalling (Sp100) considering: (a) MED, (b) OMEDA, (c) MCKD, (d) MOMEDA and (e) CYCBD.

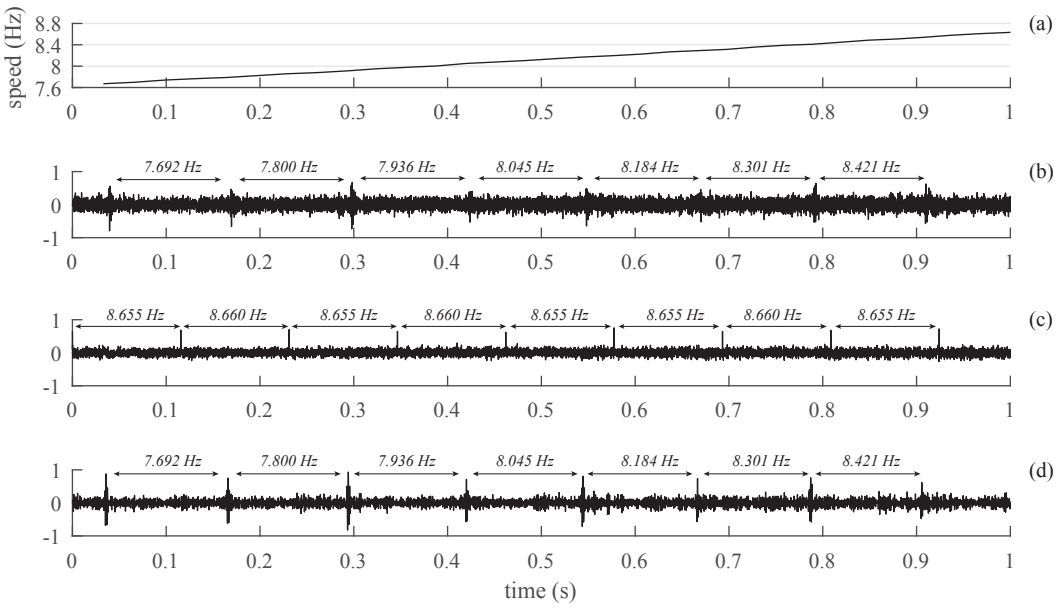

Fig. 19: BD results of the run-up test: (a) instantaneous speed of the intermediate shaft, (b) MCKD output, (c) MOMEDA output and (d) CYCBDang output.

and CS2 signal [17, 26]. On these grounds, even if the criterion of MOMEDA is based on the extraction of a periodic signal rather than a cyclostationary one, the MOMEDA results are as good as the CYCBD results. Therefore, this first comparison highlights the superiority of the proposed CYCBD algorithm with respect to MED, OMEDA and MCKD for the gear tooth spall identification.

Finally, this experimental results have shown that:

- MED and MOMEDA cannot deal with small gear tooth spalls with small loaded gears;
- MCKD returns better results than MED and MOMEDA but only for Sp50 and Sp100 cases;
- MOMEDA and CYCBD have comparable outputs.

This investigation has shown also that the maximized criterion of the CYCBD is sensitive to the spall size. This feature will be exploited later for the design of a diagnostic procedure for gear tooth spalls as well as for pointing out the differences between CYCBD and MOMEDA.

5.1.3. Results discussion: run-up test

The run-up experiment has been performed in order to confirm the observation made on the simulated case with variable speed, with specific reference to the simulated signal $\mathbf{x}_{sim,5}$ and Fig. 12. The goal is to verify the limits of MCKD and MOMEDA to deal with train of impulses having a variable period with respect to CYCBDang. In this case, just the Sp100 have been considered. As done before, MCKD and MOMEDA are referenced to a mean impulse period while CYCBDang is driven by the tacho signal.

The experimental results of the run-up are shown in Fig. 19 and discussed hereafter. For the sake of clarity, only 1 *s* is shown. In this case MCKD as well as OMEDA have been neglected since they are not sensitive to the periodicity of the impulse train. The results have been carried out by using a filter length of 50 samples for all the BD methods. It is clear that the best results is returned by CYCBDang and MOMEDA. In fact, the impulses extracted by using the MCKD are buried under strong background noise. For this reason, some impulses are barely visible. It should be noted that the time intervals, which are actually expressed in frequency in Fig. 19, are almost constant

for the MOMEDA results. The spacing of the impulses is in agreement with the prior periodicity selected (i.e. the average rotation frequency of the intermediate shaft), which is 8.656 Hz. However, the rotation frequency is linearly increasing, that is in contrast with the constant spacing of the extracted impulses.

Finally, this experiment has shown that MCKD returns a very noisy result even if the spacing between the impulses is correct and MOMEDA extracts a train of impulses without a physical meaning but reflect the periodicity of the target vector $\mathbf{t}$ (see Eq. (15)). Instead, CYCBDang returns a clear extraction of the impulses with the proper occurrence rate.

5.1.4. A diagnostic procedure for the gear tooth spall identification

The strong points of CYCBD previously highlighted may be exploited in order to design a diagnostic procedure for the identification (and the quantification) of gear tooth spall for gearboxes operating at constant speed. Also if the results presented in the previous subsection are pretty easy to be interpreted, it's a matter of fact that the diagnostics of machines by means of vibration analysis is directing to be less dependent on the user interpretation by using of simple indicators that objectively quantify the the machine condition. For this purpose, a methodology can be advanced that gives a simpler interpretation of the data returning information about both the presence and the severity of the gear tooth spall.

In a similar manner to Refs. [3, 11], the final value of the maximized criterion may be exploited considering the percentage difference between the healthy case and the faulty one. Thus, calling such a percentage difference as F, it reads:

$$F = \frac{C - C_{ref}}{C_{ref}} 100 \tag{40}$$

where C and C_{ref} are the maximized criterion values (see Eqs (7), (13) and (23)) for the faulty case and the healthy one, respectively. This indicator can be used in order to estimate both the size and the position of the gear tooth spall. From the physical standpoint, a positive deviation of F indicates the possible presence of a defect; analogously, a negative deviation as well as very low positive F may be interpreted as the absence of a defect. Furthermore, the value of F can used to determine the severity of the fault since the greater F, the greater the spall size.

Let us apply the proposed indicator for the diagnostics of the gearbox under investigation. This gearbox is composed of two stages that correspond to 3 rotating axes having different rotation frequencies. Performing BD algorithms using the incorrect fault periodicity should lead to negative value of F for all the considered spall sizes. On the contrary, when the proper fault period is taken into account, F should exhibit positive values, increasing with the spall size. Before discussing the results, it should be noted that the cyclic frequency sets used in CYCBD must be computed avoiding overlapped frequencies among the sets in order to guarantee the uniqueness of the information carried by each frequency set.

The final results on the use of indicator F are summarized in Fig. 20, considering MCKD (a), MOMEDA (b) and CYCBD (c). MED and OMEDA have been left out on purpose since they are not able to distinguish repetitive impulses having different periods. Moreover, the CYCBD results has been obtained by using a filter length of 200 samples. As demonstrated later, the filter length has a limited influence on the CYCBD results. As mentioned previously, three fault periods have been taken into account, namely the rotation frequencies of the input shaft (first column of Fig. 20), of the intermediate shaft (second column of Fig. 20) and output shaft (third column of Fig. 20). Remember that the seeded fault in the 71 teeth gear is synchronized with the period of the intermediate shaft. Hence, negative values of F should be expected regarding to the periods of the input and the output shafts whereas F must be positive for the intermediate shaft just because it corresponds to the fault period.

The data related to the intermediate shaft period (Fig. 20) highlight that MCKD and MOMEDA fail on the full detection of the gear tooth spall. In fact, negative F values occur despite the presence of the defect. Furthermore, unsatisfying results are also achieved considering the period of the input shaft and of the intermediate one, since non-negligible positive values of F are present. However, this experimental evidence is not an absolute proof of the ineffectiveness of such methods. Indeed, it should be remarked that the considered test condition is particularly hard to detect due to the low load and the gear type (helical). The results perhaps could be improved finding a more convenient filter length - and M shifts for MCKD - by trial-and-error. However, as demonstrated in the following subsection, CYCBD results are robust with respect to the selected filter length.

On the other hand, interesting results are achieved by CYCBD. The gear tooth spall is correctly identified in all the considered cases, since positive values of F are achieved in all the considered cases. In addition, the severity of the

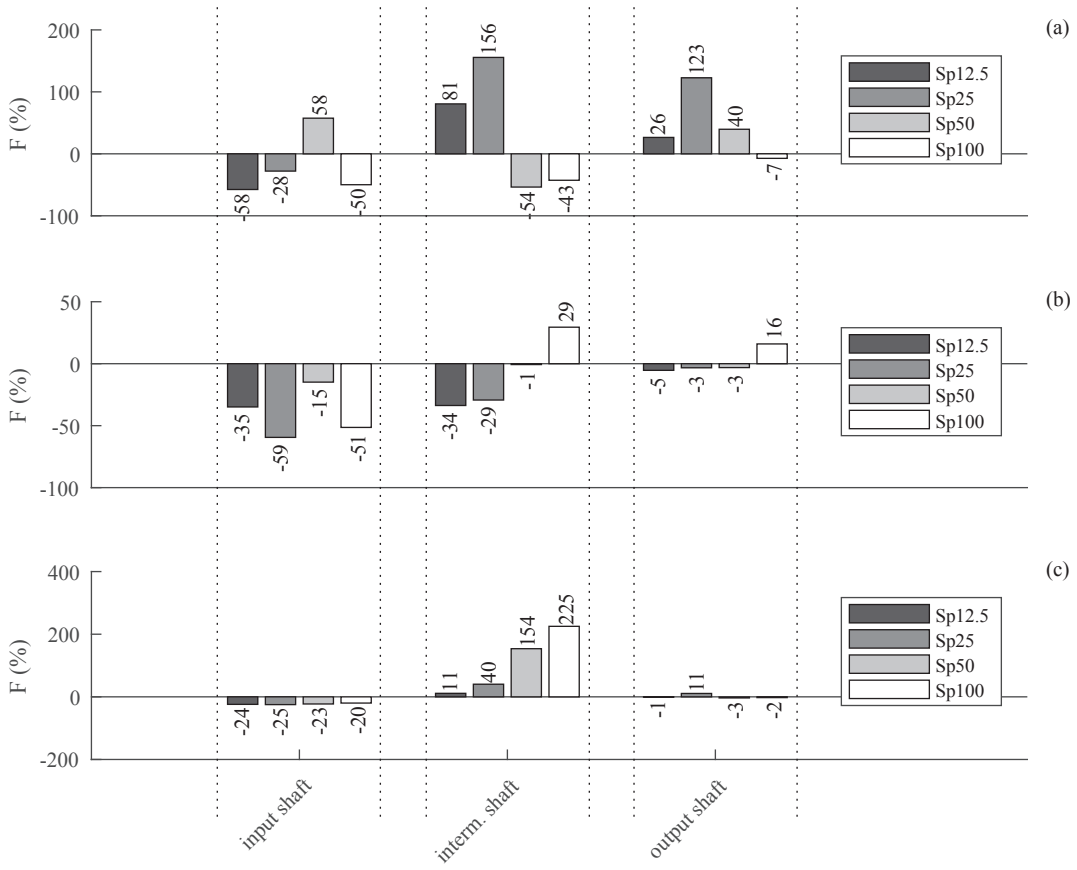

Fig. 20: Chart collecting the values of F for (a) MCKD, (b) MOMEDA, (c) CYCBD and (d) CYCBDang. The percentage values are displayed on the top of each bar.

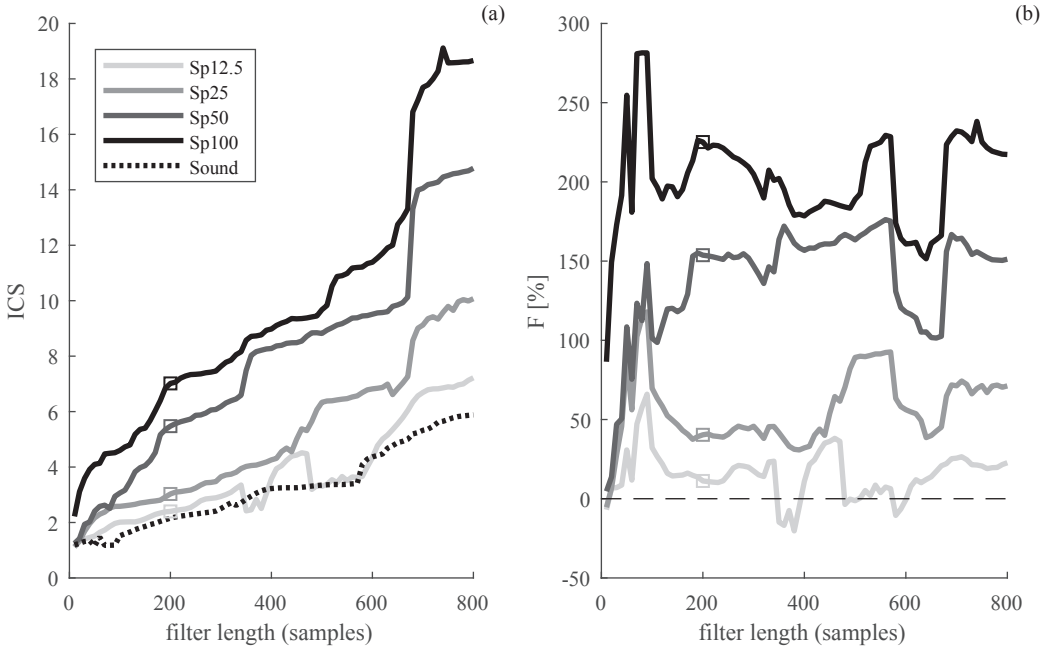

Fig. 21: Sensitivity analysis of CYCBD referenced to (a) the ICS_2 values and (b) the F values. The square marks indicate the number of samples used in Fig. 20.

fault is also well represented because F increases according to the size of the seeded fault. This results is particularly relevant because it indicates that CYCBD, de facto, is sensitive to the dimension of the fault, which can be of great value in order to monitor the evolution of the defects.

The results of CYCBD for the input shaft are satisfying since they are all negative, indicating that the 18 teeth gear is healthy. Regarding the output shaft, low positive values similar to the Sp12.5 case are present. This result can be interpreted as the presence of a small defect in the 55 teeth gear. However, the proposed method globally gives adequate results considering the slight size of the fault in case Sp12.5 and the unfavorable test conditions due to the small load. Indeed, in more favorable cases (spur gears, higher load..) this method should be even more effective.

5.1.5. Sensitivity analysis

Despite the promising results obtained in the previous subsection, a sensitivity analysis must be carried out in order to assess the effect of different filter lengths on the final results. It is a matter of fact that blind deconvolution techniques achieve different results depending on the considered filter length. Thus, this aspect is investigated in more detail below. In this sensitivity analysis, CYCBD has been performed with FIR filter length, N, varying from 10 samples to 800 samples taking into account the cyclic frequency set related to the fault period. Greater values of N are not considered because of the high computational efforts which is unbearable for real time application and, in general, for industrial purposes. The effects of various values of N have been tested in terms of maximized ICS_2, values of F and number of iterations.

Fig. 21(a) collects the values of ICS_2 whereas F values are displayed in Fig. 21(b). ICS_2 values estimated with both the proposed algorithms are in agreement with the related fault size, i.e. the greater the fault, the greater the ICS_2. Furthermore, the value of the final estimated ICS_2 seems directly related to the filter length. However, longer filter lengths should be taken into account in order to clarify this relationship. This investigation is not performed in this work because of the extreme computational effort involved. A sudden increment is observed between 670 samples and 680 samples. This effect occurs very likely because in that region of the diagram the filter length approaches the length of the fault period (approximately 673 samples). Interesting remarks can be made observing the values of F collected in diagrams (a) of Fig. 21. This diagram clearly show that the proposed methodology is globally consistent

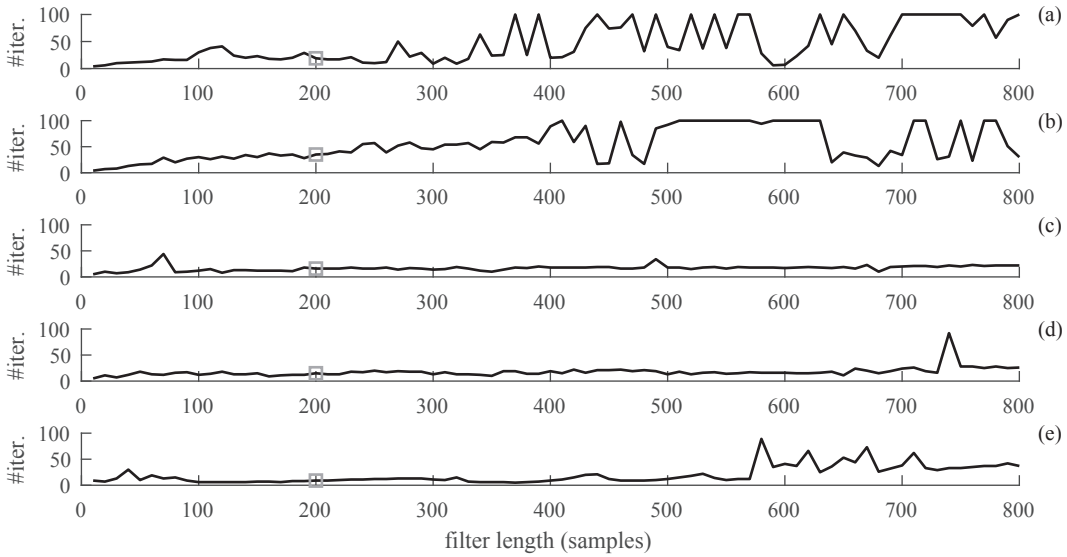

Fig. 22: Sensitivity analysis of CYCBD concerning the number of iterations in the cases: (a) Sp12.5, (b) Sp25, (c) Sp50, (d) Sp100 and (e) healthy. The gray square marks indicate the number of samples used in Fig. 20.

considering a wide range of N. The effect of a unlucky choice of N seems limited to a slight range of values and concerns only the identification of case Sp12.5. Hence, the detection of the smallest defect (Sp12.5) is not always guaranteed.

The number of iterations as a function of the filter length shown in Fig. 22 can be considered as an indicator of the quality of the deconvolved signal. In fact, a high number of iterations implies a slow convergence rate and consequently a slow algorithm. Two stoppage criteria have been adopted in the proposed algorithm: the first one regards the percentage difference of two consecutive final values of the maximized criterion; the second one regards the maximum number of iterations, which is activated if the condition of the first one is not met. Hence, the observation of how the number of iterations changes according to the variation of N may lead to pivotal consideration about the proper selection of the filter length. Fig. 22 indicates that a large number of iterations is globally needed for the smaller spall sizes, i.e. Sp12.5 and Sp25. The other cases do not highlight particular convergence issues. Hence, taking into account the convergence rate and the computational effort, Fig. 22 suggest to limit N to 200 samples.

Thus, according to both Figs. 21 and 22, in the considered experimental case N should not be greater than 200 samples. In fact, in this range, satisfying results can be achieved avoiding slow convergence rate associated with strong computational effort. Furthermore, Fig. 21 is a consistent proof of the robustness of the method considering different size faults.

5.2. Early diagnosis of an outer-race bearing fault

The second application addresses the detection of a bearing fault considering a run-to-failure experiment. The capability of the CYCBD to monitor the fault development is investigated and discussed.

5.2.1. Experimental setup

The vibration signals used in this example are part of the data set provided by the Center for Intelligent Maintenance Systems of the University of Cincinnati [27]. According to the scheme shown in Fig. 23, the test rig is composed by four bearings type Rexnord ZA-2115 installed in the same shaft. For this test, the shaft speed has been fixed at 2000 rpm and a load of 27.7 kN has been applied to bearings 2 and 3. Each bearing is monitored by to accelerometers PCB 253B33 mounted in radial direction. Further detail about the experimental apparatus as well as the

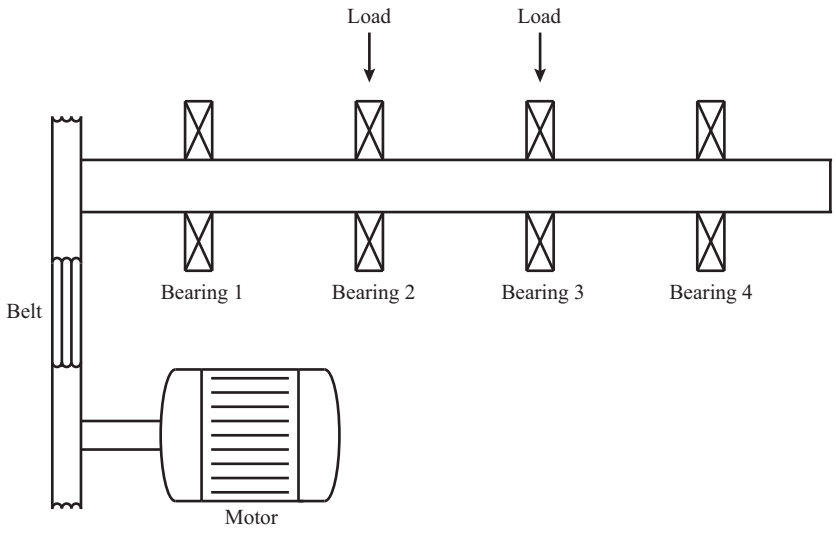

Fig. 23: Schematic of the IMS test rig.

whole dataset is available online[2]. The vibration signals have been continuously acquired with a sampling frequency of 20.48 kHz, collecting 1 second of samples each 10 minutes. The test has been stopped after 7 days (corresponding to 16.4 minutes of actual acquisition) revealing an outer race fault occurred in the first bearing.

5.2.2. Result discussion

On the basis of the previous discussion in Section 5.1.4, data can be analyzed by observing the variations of maximized BD criteria with respect to time. To this purpose, CYCBD output (i.e. the ICS_2 final value) can be used as an indicator capable to identify both the kind and the severity of the fault. In this respect, two different frequencies of interest, the BPFO and the ball pass frequency inner race (BPFI), are considered. BPFO and BPFI represents the characteristic frequencies of an outer race bearing fault and an inner race bearing fault, respectively. As done previously in Section 5.1.4, the effectiveness of CYCBD, using both the SISO and SIMO formulations this time, is verified and compared with MCKD and MOMEDA.

Considering this kind of data representation, a threshold can be designed in order to establish both fault type and first fault manifestations. In a general context, a threshold based on the presence of outliers, i.e. observations far from a given distribution pattern, is a reasonable choice. Frequently, outliers indicate the presence of some kind of anomaly due, for instance, to a bearing failure. Hence, the idea is to define a threshold from observations related to a supposed healthy state – e.g. samples acquired during the test beginning – that defines the indicator value beyond which a fault may be occurred.

A consistent rule of thumb for identifying suspected outliers is the Tukey's method that has been already successfully used for pass/fail decision tests by using vibration-based scalar indicators in other research works [19, 28]. The Tukey's method is quiet general and does not require prior distribution knowledge. The only limitation is that Tukey's method could lose effectiveness for data having non-symmetric probability distribution. Moreover, for bearing fault monitoring applications, another necessary condition is that the system operates in stationary conditions because all the vibration signature changes should be associated to faults rather than regime variations. According to Ref. [29], this method exploits the interquartile range (IQR), namely the distance between first and third quartile, in order to define two classes of outliers: "outside" and "far out". The first class refers to mild outliers (i.e. data not so far from the reference distribution) and it is defined as 1.5 times the IQR distance. The second one refers to extreme outliers (i.e. data significantly distant from the reference distribution) and it is defined as 3 times the IQR distance. In this application, the threshold is represented by the latter class just for the upper limit since it is expected that the considered indicators rise when a fault occurs.

[2] `http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository` Last consulted on 11/10/2017

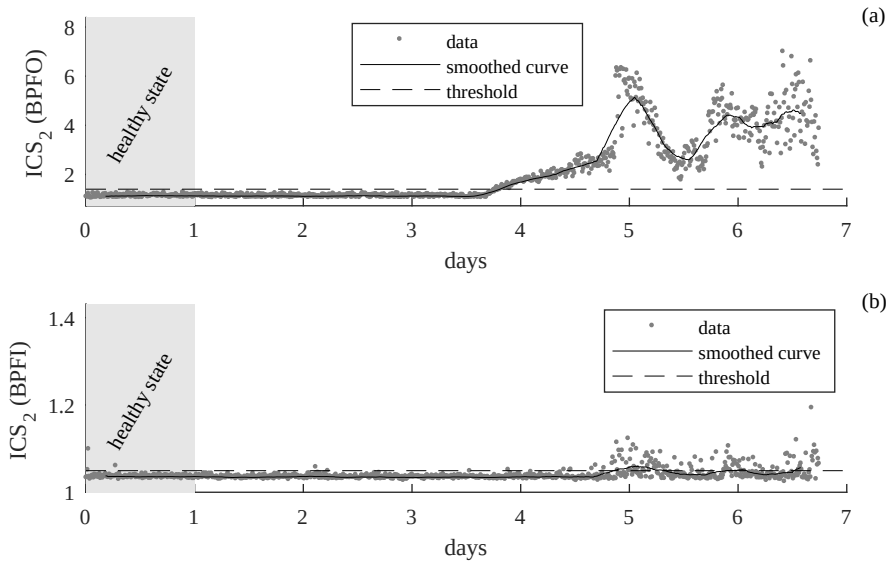

Fig. 24: ICS_2 final values of CYCBD (SISO) for the sensor on Bearing 1 referenced to (a) the BPFO and (b) the BPFI.

The comparison of BD methods is performed in terms of bearing fault identification and early detection dedicated diagrams from Fig. 24 to Fig. 27. In these diagrams, dark gray data points refer to the maximized BD criterion that works as a scalar indicator referring to a specific type of bearing fault (outer race fault or inner race fault). The threshold has been estimated by using the Tukey's method on the data acquired in the first day of recording, under the hypothesis that all the bearings are healthy in that time interval. Note that time window related to the healthy condition has been highlighted in light gray.

Moreover, it is expected a certain dispersion of the indicator even when the bearings are healthy. This is due to unpredictable phenomena such as measurement uncertainty and other events unrelated to bearing faults. Such phenomena could produce also odd data points that can be misinterpreted. Consequently, it is convenient to smooth the data points in order to estimate a consistent trend that simplify the data verification with respect to a given threshold. The data smoothing has been performed by the moving average method, which can be easily implemented by convolving a given time series with a fixed rectangular window of 50 samples.

Fig. 24 collects the ICS_2 final values estimated by SISO CYCBD for the sensor on Bearing 1 with a FIR filter length of 80 samples. Fig. 24 highlights that, considering the BPFO, a clear increasing trend can be detected after 3.8 days, with three consecutive local sudden deviations in correspondence to day 5, 6 and 6.5. This global fluctuations of the indicator trend observed in Fig. 24, are likely due to propagation phenomena of the bearing fault. As reported in Ref. [30], after the defect appearance, the propagation mechanism of bearing faults is composed by consecutive propagation and smoothing effects. The diagram related to the BPFI shows a slighter increasing trend due to the presence of a certain number of values above the threshold. This happen especially when also the ICS_2 values referenced to the BPFO significantly deviate. This behavior is in agreement with the physical interpretation of the fault development, since a marked growing trend of the indicator related to the outer race fault is expected. By the way, the presence of some data points beyond the threshold in the BPFI case could make the interpretation of the data difficult.

The SIMO approach may improve the results because the weighting matrix estimation is enhanced when several responses are considered. Since the quality of the CYCBD output is strictly related to the accuracy of the extracted cyclic components, a good estimation of the weighting matrix leads to a better source estimation. In this particular case, the low frequency resolution of the measurements (1 Hz) is not ideal for bearing diagnostics, thus the SIMO approach can be used to counter-balance this limitation by exploiting the presence of multiple sensors on the test bench. Fig. 25 show the results obtained by performing SIMO CYCBD taking into account all the available accelerometers.

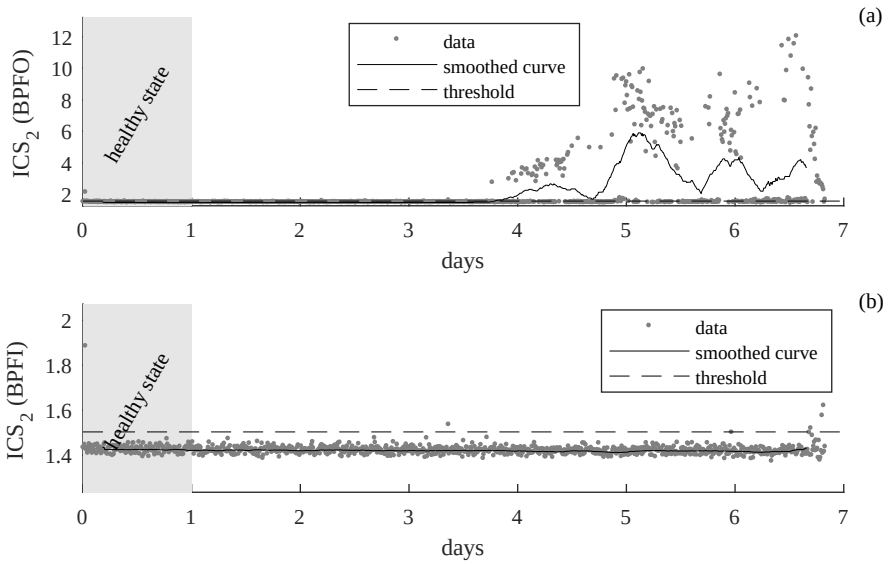

Fig. 25: ICS_2 final values of CYCBD (SIMO) referenced to (a) the BPFO and (b) the BPFI taking into account all the available sensors (on Bearings 1 to 4).

It can be observed that the result quality has been strongly improved by performing the SIMO CYCBD. In this case, the outer race fault results are not significantly improved since becomes visible after 3.8 days, as highlighted also by SISO CYCBD results. However the diagrams related to the inner race fault are generally enhanced. In fact, the ICS_2 values referenced to the inner race fault shows decreasing number of outliers as the number of considered sensors increases. This leads to a reduced chance to detect false positives since the smoothed trend never across the threshold, in particular in Fig. 25. Thus, the SIMO approach not only enhances the early detection of the bearing fault but also improves the interpretation of the data. However, the computational effort the SIMO CYCBD is remarkable compared to the SISO CYCBD. Furthermore, it should be noted that when accelerometers far from the excitation (i.e. the outer-race fault in Bearing 1) are involved, data points below the threshold may appear in the diagrams concerning BPFO. This can be explained because the effectiveness of SIMO CYCBD can decrease considering accelerometers very far from the source implying to a reduction of the SNR, even if the system is very stiff.

As done before, the results obtained by MCKD and MOMEDA are reported in Figs. 26 and 27, respectively. For the sake of clarity, the correlated kurtosis values in Fig. 26 are expressed in logarithmic scale. For the MCKD, the filter length has been set to 120 samples using 7 shifts. Moreover, only 2000 samples of each records has been taken into account, according to the suggestion in Ref. [11]. The results obtained with MCKD referenced to the BPFO are worse than the SIMO CYCBD in terms of early fault detection and of data dispersion. However, with some limitations, the MCKD allows for the bearing fault identification. Indeed, the fault typology is correctly detected, as well as the increasing trend of the BPFO indicator. However, the threshold is crossed after 4.5 days that is approximately one day later with respect to CYCBD results. Comparing Figs. 24 and 26, it is further clear that correlated kurtosis appears less sensitive with respect to ICS_2 to changes in the vibration signature due to the outer-race bearing fault. Thus, after these observations, CYCBD results should be preferred.

Considering now MOMEDA results in Fig. 27, the filter length has been set to 120 samples as in the MCKD. In this case the indicator of MOMEDA (i.e. multipoint kurtosis [3]) gives unsatisfactory results. In fact, some data point from day 5 to the last day are above the threshold but the smoothed curves, which represents the data trend, is always below the threshold. Actually, this is not surprising since the bearing faults are well modeled as second-order cyclostationary signals [20] but MOMEDA is sensitive to periodic signals. Even if the BPFO highlights a weak increasing trend, the data dispersion is too high to be reliable. In fact, the fitted curve remains below the threshold in the whole dataset.

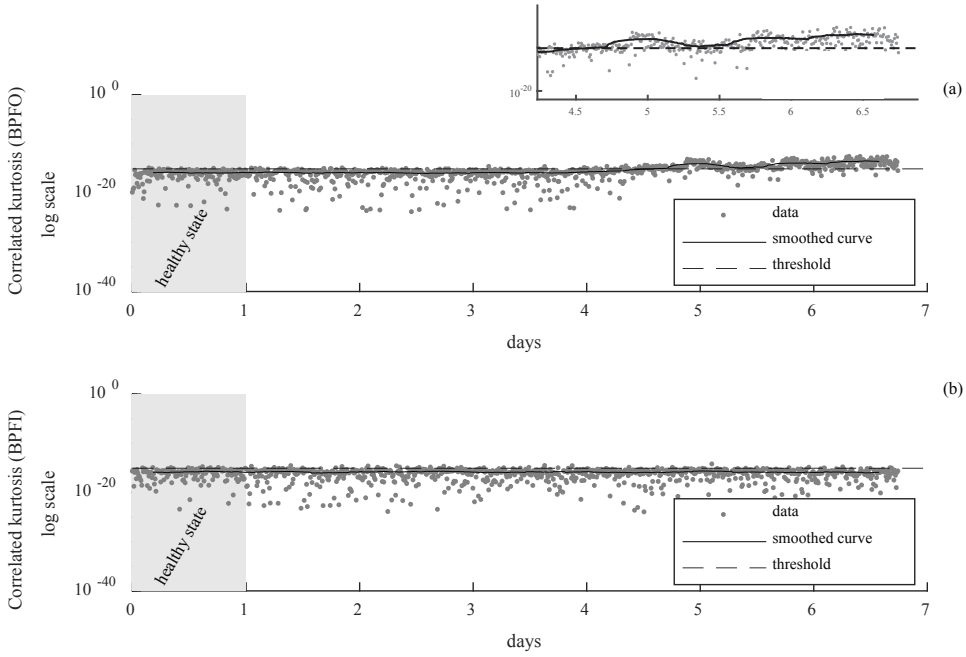

Fig. 26: Correlated kurtosis final values of MCKD referenced to (a) the BPFO and (b) the BPFI.

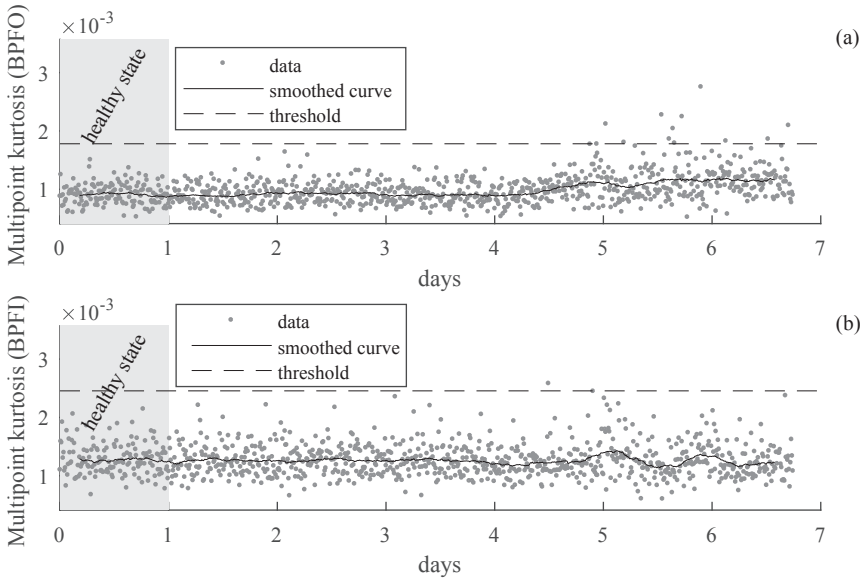

Fig. 27: Multipoint kurtosis final values of MOMEDA referenced to (a) the BPFO and (b) the BPFI.

Finally, this application has shown the effectiveness of CYCBD in order to identify an outer race bearing fault considering the natural advancement of the bearing fault. This is all the more accurate as BD is performed considering several sensors placed on the test rig.

6. Conclusions

In this paper a novel blind deconvolution algorithm based on the iteratively minimizing a Rayleigh quotient has been introduced. The strong point of the proposed method is the flexibility: several tailored criteria can be easily implemented thanks to the presence of a weighting function driving the deconvolution. A BD method based on the indicators of cyclostationarity, called CYCBD, has been deduced. In this scenario, the cyclostationary framework has been explicitly used for the first time in order to design a BD method with application to machine diagnosis.

The advantages of CYCBD have been highlighted in five different synthetic signals enlightening its capability to recover impulsive patterns exhibiting cyclostationary behavior. Particular care has been dedicated to compare CYCBD to MOMEDA and MCKD since represent the most recent BD methods for the diagnosis of rotating machines proposed in the literature.

Considering real signals, the effectiveness of CYCBD has been investigated and discussed, both in qualitative and quantitative terms, in comparison with other BD methods proposed in the literature, first considering the gear tooth spall identification by means of a dedicated experimental campaign. This comparison has led to design a diagnostic procedure based on the proposed cyclostationary criterion for gear tooth spalling in gearboxes operating at constant regimes (or accounting small speed fluctuations). The main advantage of this method is that requires a limited user interaction by exploiting an indicator based on the relative value of the maximized criterion between the healthy condition and the current one. The method robustness has been further demonstrated by means of an extended sensitivity analysis taking into account the effect of the FIR filter length on different critical parameters of the algorithm. Moreover, the proposed BD method has been applied for the bearing fault identification using a run-to-failure dataset achieving satisfactory results. In this case, CYCBD provides excellent diagnostic performances with respect to the other BD algorithms in terms of early fault detection and identification.

All the results highlight that the maximized value of ICS_2 through CYCBD can be considered as a robust fault indicator. Specifically, the experimental results, according to the simulated ones, demonstrated the superiority of the proposed criterion dealing with cyclostationary signals in two different experimental cases. In particular, CYCBD overcomes MOMEDA dealing with cyclostationary sources since MOMEDA is based on a criterion sensitive to periodic sources. Concurrently, the cyclostationary criterion used in CYCBD, i.e. ICS_2, appears more consistent with respect to the criterion used in MCKD, namely the correlated kurtosis. CYCBD proved to be effective also in non-constant regime cases, which represents an open issue until now. CYCBDang has highlighted good performances dealing with vibration signals in variable speed conditions. In such circumstances, MOMEDA and MCKD do not give satisfactory results since they have been designed only for the extraction of equispaced impulse trains.

External Resources

A Matlab implementation of the proposed blind deconvolution method together with an interactive demo can be found at: `put_the_link_here`

Acknowledgments

This work has been realized through the contribution of "5 per mille assegnato all'Università degli Studi di Ferrara - dichiarazione dei redditi dell'anno 2014" and the support of the international mobility fellowship for PhD students provided by the University of Ferrara.

Appendix A. Relationship between differential entropy and kurtosis

The differential entropy $\mathbb{H}$ of a random variable x is defined as:

$$\mathbb{H} = -\int f(x) \ln f(x) dx \tag{A.1}$$

where $f(x)$ is the probability density function (pdf) of x. If x is zero-mean and of unit variance, the truncated version of the Gram-Charlier expansion of $f(x)$ is given by:

$$f(x) \approx \varphi(u) \left[1 + \frac{\kappa_3}{3!\sigma^3} H_3(u) + \frac{\kappa_4}{4!\sigma^4} H_4(u)\right] \tag{A.2}$$

where

$$\varphi(u) = \frac{e^{\frac{-u^2}{2}}}{\sqrt{2\pi}} \tag{A.3}$$

H_m is the Chebyshev-Hermite polynomial of order m and κ_n refers to the cumulant of order n. For the sake of simplicity, the argument of H and φ are neglected hereafter. The expansion reported in Eq. (A.2) is valid under the hypothesis that $f(x)$ is close to a Gaussian pdf [31]. Eq. (A.2) can be further simplified if we assume that $f(x)$ is symmetric, which implies that κ_3 is nil.

On these grounds, considering the Taylor's expansion $\ln(1+\psi) \approx \psi - \frac{\psi^2}{2}$, the substitution of Eq. (A.2) in Eq. (A.1) gives:

$$\begin{aligned} \mathbb{H} &\approx -\int \varphi\left(1 + \frac{\kappa_4 H_4}{4!}\right)\left[\ln\varphi + \frac{\kappa_4 H_4}{4!} - \frac{1}{2}\left(\frac{\kappa_4 H_4}{4!}\right)^2\right] du \\ &\approx -\int \varphi \ln\varphi + \varphi\frac{\kappa_4 H_4}{4!} - \frac{\varphi}{2}\left(\frac{\kappa_4 H_4}{4!}\right)^2 + \varphi\ln\varphi\frac{\kappa_4 H_4}{4!} + \varphi\left(\frac{\kappa_4 H_4}{4!}\right)^2 - \frac{\varphi}{2}\left(\frac{\kappa_4 H_4}{4!}\right)^3 du. \end{aligned} \tag{A.4}$$

Recalling the hypothesis of x approximately Gaussian, this expression can be simplified by noting that:

- κ_4 is small and it's equivalent to the kurtosis, viz:

$$\kappa_4 = \mathbb{E}\left[(u)^4\right] = Kurt[x] \tag{A.5}$$

- the third-order term is infinitely smaller than the second-order terms under the hypothesis;
- the Chebyshev-Hermite polynomials are orthogonal with respect to φ:

$$\int \varphi H_i H_j du = j!\delta_{ij} \tag{A.6}$$

  where δ_{ij} is the Kronecker delta;

- $\int \varphi \ln\varphi dx$ is the differential entropy of a Gaussian distribution with zero-mean and unit variance, such as

$$-\int \varphi \ln\varphi dx \approx \frac{1 + \ln 2\pi}{2}. \tag{A.7}$$

Thus, after some manipulation, Eq. (A.4) can be rewritten as:

$$\mathbb{H} \approx \frac{1 + \ln 2\pi}{2} - \frac{1}{48} Kurt[x]^2 \tag{A.8}$$

giving the link between the kurtosis and the differential entropy.

References

[1] R. A. Wiggins, Minimum entropy deconvolution, Geoexploration 16 (1978) 21–35.

[2] C. A. Cabrelli, Minimum entropy deconvolution and simplicity: A noniterative algorithm, Geophysics 50 (1985) 394–413.

[3] G. L. McDonald, Q. Zhao, Multipoint Optimal Minimum Entropy Deconvolution and Convolution Fix: Application to vibration fault detection, Mechanical Systems and Signal Processing 82 (2017) 461–477.

[4] J.-Y. Lee, A. Nandi, Blind Deconvolution of Impacting Signals Using Higher-Order Statistics, Mechanical systems and signal processing 12 (1998) 357–371.

[5] J.-Y. Lee, A. Nandi, Extraction of Impacting Signals Using Blind Deconvolution, Journal of Sound and Vibration 232 (2000) 945–962.

[6] J. Obuchowski, R. Zimroz, A. Wyłomańska, Blind equalization using combined skewness-kurtosis criterion for gearbox vibration enhancement, Measurement: Journal of the International Measurement Confederation 88 (2016) 34–44.

[7] H. Endo, R. Randall, Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter, Mechanical Systems and Signal Processing 21 (2007) 906–919.

[8] H. Endo, R. B. Randall, C. Gosselin, Differential diagnosis of spall vs. cracks in the gear tooth fillet region: Experimental validation, Mechanical Systems and Signal Processing 23 (2009) 636–651.

[9] N. Sawalhi, R. B. Randall, H. Endo, The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis, Mechanical Systems and Signal Processing 21 (2007) 2616–2633.

[10] D. He, X. Wang, S. Li, J. Lin, M. Zhao, Identification of multiple faults in rotating machinery based on minimum entropy deconvolution combined with spectral kurtosis, Mechanical Systems and Signal Processing 81 (2016) 235–249.

[11] G. L. McDonald, Q. Zhao, M. J. Zuo, Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection, Mechanical Systems and Signal Processing 33 (2012) 237–255.

[12] J. Antoni, Cyclostationarity by examples, Mechanical Systems and Signal Processing 23 (2009) 987–1036.

[13] B. Jelonnek, D. Boss, K.-D. Kammeyer, Generalized eigenvector algorithm for blind equalization, Signal Processing 61 (1997) 237–264.

[14] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[15] J.-L. Lacoume, P.-O. Amblard, P. Comon, Statistiques d'ordre supérieur pour le traitement du signal, MASSON, 1997.

[16] W. Wang, A. K. Wong, Autoregressive Model-Based Gear Fault Diagnosis, Journal of Vibration and Acoustics 124 (2002) 172.

[17] A. Raad, J. Antoni, M. Sidahmed, Indicators of cyclostationarity: Theory and application to gear fault monitoring, Mechanical Systems and Signal Processing 22 (2008) 574–587.

[18] B. Kilundu, X. Chiementin, J. Duez, D. Mba, Cyclostationarity of Acoustic Emissions (AE) for monitoring bearing defects, Mechanical Systems and Signal Processing 25 (2011) 2061–2072.

[19] S. Delvecchio, G. D'Elia, G. Dalpiaz, On the use of cyclostationary indicators in IC engine quality control by cold tests, Mechanical Systems and Signal Processing 60-61 (2015) 208–228.

[20] J. Antoni, F. Bonnardot, A. Raad, M. El Badaoui, Cyclostationary modelling of rotating machine vibration signals, Mechanical Systems and Signal Processing 18 (2004) 1285–1314.

[21] P. Borghesani, P. Pennacchi, S. Chatterton, R. Ricci, The velocity synchronous discrete Fourier transform for order tracking in the field of rotating machinery, Mechanical Systems and Signal Processing 44 (2014) 118–133.

[22] J. P. Den Hartog, Mechanical Vibrations, Dover Pubns, 2003.

[23] G. Dalpiaz, G. D'Elia, S. Delvecchio, Design of a test bench for the vibro-acoustical analysis and diagnostics of rotating machines, in: The Second World Congress on Engineering Asset Management & The Fourth International Conference on Condition Monitoring, Harrogate, UK, 2007, pp. 777–792.

[24] G. D'Elia, Fault detection in rotating machines by vibration signal processing techniques, Phd thesis, Alma Mater Studiorum Università di Bologna, 2008. doi:10.6092/unibo/amsdottorato/952.

[25] G. D'Elia, S. Delvecchio, G. Dalpiaz, Gear spall detection by non-stationary vibration signal analysis, in: ISMA2008 International Conference on Noise and Vibration Engineering, Leuven, Belgium, 2008, pp. 777–792.

[26] C. Capdessus, M. Sidahmed, J. Lacoume, Cyclostationary Processes: Application in Gear Faults Early Diagnosis, Mechanical Systems and Signal Processing 14 (2000) 371–385.

[27] J. Lee, H. Qiu, G. Yu, J. Lin, Rexnord Technical Services, Bearing Data Set, IMS, University of Cincinnati, NASA Ames Prognostics Data Repository, 2007. URL: http://ti.arc.nasa.gov/project/prognostic-data-repository.

[28] M. Malagoó, E. Mucchi, G. Dalpiaz, Fault detection in heavy duty wheels by advanced vibration processing techniques and lumped parameter modeling, Mechanical Systems and Signal Processing 70-71 (2016) 141–160.

[29] J. W. Tukey, Exploratory data analysis, Pearson, 1977.

[30] T. Williams, X. Ribadeneira, S. Billington, T. Kurfess, Rolling Element Bearing Diagnostics in Run-To-Failure Lifetime Testing, Mechanical Systems and Signal Processing 15 (2001) 979–993.

[31] A. Hyvarinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley & Sons, Inc., 2001.