Mechanical Systems and Signal Processing 114 (2019) 290–327

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

A statistical methodology for the design of condition indicators

Jérôme Antoni [a,*], Pietro Borghesani [b,c]

[a] Laboratoire Vibrations Acoustique, Univ Lyon, INSA-Lyon, LVA EA677, F-69621 Villeurbanne, France
[b] Science and Engineering Faculty, Queensland University of Technology, 2 George St., Brisbane, QLD 4000, Australia
[c] School of Mechanical and Manufacturing Engineering, UNSW Sydney, Kensington, NSW 2052, Australia

ARTICLE INFO

Article history:
Received 16 January 2018
Received in revised form 2 April 2018
Accepted 1 May 2018
Available online 25 May 2018

ABSTRACT

Recent studies in the field of diagnostics and prognostics of machines have highlighted the key role played by non-stationarity – often in the form of cyclostationarity – or non-Gaussianity – often in the form of impulsiveness as characteristic symptoms of abnormality. Traditional diagnostic and prognostic indicators (e.g. the kurtosis) are however sensitive to both types of symptoms without being able to differentiate them. In an effort to investigate how the signal characteristics evolve with the different phases of machine components degradation, this paper proposes a new family of condition indicators able to track cyclostationary or non-Gaussian symptoms independently. A statistical methodology based on the maximum likelihood ratio is introduced as a general framework to design condition indicators. It arrives with the possibility of setting up statistical thresholds, as needed for a reliable diagnosis. The methodology is validated with numerically generated signals and applied to the dataset made available by NSF I/UCR Center for Intelligent Maintenance Systems (IMS). This particular application shows high potential for bearing prognostics by providing condition indicators able to describe different phases of the bearing degradation process.

1. Introduction

Vibration and acoustic-based condition monitoring of machines is still the dominant technology for the diagnostics of faults in a wide variety of machine components. Many diagnostic and prognostic systems are rooted on simple scalar indicators, so-called health or condition indicators (CM), which can be used to detect the presence of a fault and, possibly, to track its evolution in time. Typical examples are provided by the root-mean-square value, the crest factor, the kurtosis, the form factor, the indices of cyclostationarity, etc. [1–5]. Interestingly, most CMs can be categorized in two groups: one dedicated to characterizing a non-Gaussian behaviour of the machine signals – typically in the form of impulsiveness – and another one dedicated to characterizing a non-stationary behaviour – typically in the form of cyclostationarity. This originates from the observation that numerous faults in rotating and alternating machines result in alterations of the signal characteristics which can be described as (*i*) an increase in the impulsiveness of the signal, i.e. a deviation from Gaussianity, and/or (*ii*) a shift of the statistical behaviour of the signal from stationarity to non-stationarity. These two phenomena, traditionally observed in faulty rolling element bearing, fluid-dynamic instabilities [6] and cavitating pumps, damaged gears, have often been described either separately (ignoring one) or, qualitatively, as a single "deviation" from "normality" (to be read as a general term not strictly related to the Gaussian distribution).

* Corresponding author.
E-mail address: antoni@utc.fr (J. Antoni).

https://doi.org/10.1016/j.ymssp.2018.05.012
0888-3270/

List of acronyms

BGCS	Bernoulli-Gauss cyclostationary
BGS	Bernoulli-Gauss stationary
CDF	cumulative density function
CK	correlated kurtosis
CM	condition indicators
DOF	degrees of freedom
EM	expected maximization algorithm
GCS	Gaussian cyclostationary
GG	generalized Gaussian
GGCS	generalized Gaussian cyclostationary
GGS	generalized Gaussian stationary
GLR	generalized likelihood ratio
GLRT	generalized likelihood ratio test
GS	Gaussian stationary
MLE	maximum likelihood function
PDF	probability density function

So far, the most traditional and probably also the most widely employed CM is the kurtosis.[1] Originally defined as the standardized fourth-order central moment (or fourth-order cumulant), the kurtosis is a simple measure of the departure of a probability density function (PDF) from the Gaussian prototype [7]. High positive values of the kurtosis indicate a PDF with heavier tails than the Gaussian, which is symptomatic of the presence of peaks or transients in the signal, as typically produced by incipient faults (another controversial interpretation is a measure of the peakedness of the PDF at the centre [8]). A nice technical discussion about the use of the kurtosis in machine condition monitoring is given in Ref. [9]. The interpretation of kurtosis as a non-Gaussianity measure is confirmed in statistics by its use for the establishment of two typical Normality tests: the D'Agostino K-squared test [10] and the Jarque-Bera test [11]. Less known is the sensitivity of the kurtosis (defined in its time-average version) to non-stationarity [12]. This has recently been revealed in Ref. [13] which draws a strong relationship between the kurtosis and cyclostationary indicators such as the squared envelope spectrum (SES) and the cyclic modulation spectrum. This second nature of the kurtosis is however always combined with its sensitivity to non-Gaussianity and it has been shown that high levels of non-Gaussianity (i.e. strongly impulsive signals) can mask weak non-stationarity symptoms [14].

Other popular health indicators sensitive to non-Gaussian signatures are the crest factor [27], the peak-to-peak, the skewness (third-order central moment), the form factor (ratio of the RMS value to the mean absolute value), the entropy (e.g. of Shannon, Renyi, and approximate types, etc. [15–18]). In principle, any higher-order moment [19] or cumulant could be used (e.g. such as the M6A and M8A which involve 6th and 8th order standardized central moments, respectively [20]), yet orders higher than four (i.e. the kurtosis) are rarely used because of their lack of robustness and poor statistical properties. Incidentally, any measure of sparsity – as recently investigated in signal processing [21] – also happens to be the inverse of a measure of non-Gaussianity [22]. The formerly proposed "smoothness index" also falls in this category [23]. Other non-Gaussian indicators based on power means have been proposed in Ref. [24].

The catalogue of CMs dedicated to measure non-stationary behaviours happens to be much smaller as it mainly contains quantities derived from the theory of cyclostationary processes [25]. One such indicator, which has been rediscovered independently from different approaches, is the "degree of cyclostationarity" [26]. As demonstrated in Ref. [13], it boils down in its most compact form to the sum of peaks of the SES (at multiples of a given cyclic frequency) normalized by the signal power. Reference [28] introduces several indices of higher-order cyclostationarity rooted on cyclic higher-order cumulants, which are theoretically sensitive to both non-stationarity and non-Gaussianity. A recent extension of these indices to cyclo-non-stationary processes – a generalization of cyclostationary processes to speed-varying regimes – is given in Ref. [32].

Although non-exhaustive, this brief overview is enough to evidence the paucity of attempts to design new CMs. This is despite the shortcomings of existing indicators – such as the kurtosis in particular – that have been reported in many instances. One reason is probably historical since the aforementioned indicators are based on simple calculation formulas which are difficult to beat (this was a central motivation of Karl Pearson when he introduced the kurtosis at the beginning of the 20th century), yet this argument no longer applies with modern computational capabilities. Another more profound

[1] In this paper, the kurtosis is defined according to its time-averaged version: given a zero-mean discrete-time signal $x(n)$, it reads $\kappa_4 \stackrel{\text{def}}{=} \lim_{L\to\infty} \frac{\sum_{n=0}^{L-1}|x(n)|^4/L}{\left(\sum_{n=0}^{L-1}|x(n)|^2/L\right)^2}$. This is the form universally used in condition monitoring, although it strictly applies to stationary and ergodic processes, only. The excess kurtosis is defined by $\kappa_4 - 3$.

reason is the lack of principled methods (other than the method based on statistical moments and cumulants for instance) to derive new indicators dedicated to monitor a specific symptom in a signal. As a matter of fact, the prevailing approaches found in the literature is to generate new indicators from old ones, either by applying them on various pre-processed versions of the signal (for instance the FM4, NA4 and NB4 are kurtosis of the difference signal, the residual signal, and envelope signal, respectively [1,29,30]), or by combining them. Typical examples of the latter approach are differences or ratios of RMS values in specific frequency bands and sidebands [1,34]. Nonlinear functions of the RMS value, the kurtosis and the crest-factor is another example used for trending the degradation of rolling element bearing in prognostics [35]. A current widespread strategy is to obtain new indicators from combinations of old ones by running feature reduction algorithms developed in the field of machine learning (e.g. Principal Component Analysis, nonlinear dimensionality reduction) [31,33]. An obvious shortcoming of this approach is that it cannot supply more diagnostic information than is present at the onset in the input features.

This paper tries to fill this gap by introducing a systematic methodology to design CMs tailored to a specific symptom. The principle is based on differentiating the healthy and abnormal states of the machine by two hypotheses, the null and the alternative ones, each characterized by a different PDF. Next, an "optimal" indicator is designed as the logarithm of the ratio of the PDFs, where all unknown parameters are replaced by their maximum likelihood estimates derived under their respective hypotheses. The so-defined indicator is optimal in the sense that it coincides with the statistics of the generalized likelihood ratio test (GLRT) which maximizes the probability of true detection given a fixed probability of false alarm (erroneous rejection of the null hypothesis) [36]. The advantages of this formulation are multiple. First, the introduction of PDFs to characterize the healthy and abnormal states provides an unlimited number of possibilities to monitor a specific behaviour in the development of a fault, such as the emergence of a certain type of non-Gaussianity and/or non-stationarity. In particular, it makes explicit the distinction between non-Gaussian and non-stationary symptoms which can thus be monitored independently. Second, as a by-product of the GLRT, the indicators come together with statistical thresholds against which they can be compared to decide when the null should be rejected for a given risk of false alarm.

It is readily verified that most of the aforementioned classical CMs can be recovered as particular instances of the proposed methodology. However, rather than trying to recover existing indicators, the objective of the paper is rather to demonstrate how new ones can be derived which are specifically tailored to a certain type of symptom. A special effort is made to explore simple scalar indicators which are able to separate impulsiveness (i.e. non-Gaussianity) from non-stationarity, thus allowing the identification of the type of deviation from normality. This forms the basis for a more informed diagnostic and prognostic procedure, thus enabling the user to identify the main phenomena affecting the different stages of component degradation. Optimal scalar indicators will be obtained for different forms on non-stationarity and under different conditions of non-Gaussianity. This methodology will also allow the setup of statistical tests associated with the indicators for a robust and automatic detection of incipient faults.

The rest of the paper is organised as follows. Section 2 will review the concept of generalized likelihood ratio (GLR) and the associated Wilk's theorem. Then sections 3 to 8 will explore how this approach can be used to define optimal indicators of non-stationarity and non-Gaussianity. The proposed indicators will be analysed in detail, exploring their link with similar traditional indicators and will be validated in each section with numerical tests. Section 9 will discuss the interconnection between non-stationary and non-Gaussian indicators and whether they can be interchanged. After a brief Section 10 summarising the theoretical results, the last section before the conclusion will validate the proposed methodology on a typical vibration signals. This will also provide an occasion to show the phenomenological insights that a combined non-Gaussian and non-stationary diagnostic tool can offer on the degradation trend of rolling element bearings.

Notational conventions are as follows. Signal are considered in the discrete time setting. The signal amplitude at the n-th time instant is noted $x(n)$; $\mathbf{x}$ stands for the vector that stacks all the values $x(n)$ for $n = 0, \ldots, L-1$. The PDF of $x(n)$ is noted $p_X(x(n))$. The PDF of vector $\mathbf{x}$ is a multivariate density, taken jointly for all variables $x(n)$, noted $p_X(\mathbf{x})$. $x \sim p_X$ means that random variable x follows the probability law with density $p_X(x)$. $p_X(x|z)$ stands for the PDF of a random variable x conditioned on the value of another variable z. Vector $\boldsymbol{\theta}$ collects all parameters required to express a specific PDF; thus the latter is noted $p_X(\mathbf{x}|\boldsymbol{\theta})$. $\mathcal{N}(x;\mu,\sigma^2)$ stands for the Gaussian PDF, as a function of x, with mean value μ and variance σ^2; $\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the multivariate counterpart with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathcal{GN}(x;\mu,\eta,\beta)$ stands for the generalized Gaussian PDF, as a function of x, with mean value μ, scale parameter η and shape parameter β. χ^2_ν denotes the chi-2 random variable with ν degrees of freedom. Operator $\langle f(n)\rangle \overset{\text{def}}{=} L^{-1}\sum_{n=0}^{L-1} f(n)$ stands for the time average of function $f(n)$.

2. The generalized likelihood ratio as a principle for designing of optimal indicators

2.1. Objectives

The aim of this section is to propose a general principle to design CM indicators. Contrary to many other approaches that verify *a posteriori* the detection capability of a given indicator, the proposed methodology is deductive in the sense that it *designs* the CM to detect a specific symptom in the signal of interest. It also intends to be *optimal*, in the sense that it maximizes the probability of detection among all other candidates, given a probability of false alarm. Several new CMs can thus be designed for diagnostics and prognostics. In some cases, they actually happen to boil down to known indicators of the literature, either directly or after proceeding to some approximations.

Let's assume that the symptom to be detected in the signal is embodied in a statistical property – e.g. a specific form of non-Gaussianity or non-stationarity. The aim is to define the indicator as a transformation of the measured signal, in the form of a functional (e.g. the result is a scalar), which returns high values when the property of interest is present and small values otherwise. Such an indicator therefore measures the "strength" of the property of interest in the signal and can be used for detection by comparison with a threshold. Optimal of the indicator (in a sense to be specified soon) is important in order to end up with a unique definition for a given property.

2.2. The generalized likelihood ratio

The aforementioned objectives are achieved by using the likelihood principle. This is perhaps most simply introduced by using the framework of a statistical test of hypotheses. Let's denote by H_1 and H_0 two alternative hypotheses, which correspond to two different states of the signal: typically H_1 will reflect the presence of the statistical property (i.e. the symptom) to be detected in the signal whereas H_0 will reflect its absence. Let $\mathbf{x} = [x(0), x(1), \ldots, x(L-1)]^T$ denotes the vector of samples of the measured (discrete time) signal, hereafter simply referred to as the "signal". The association of the signal to one or the other hypotheses is next described by different probability density functions (PDFs), say $p_X(\mathbf{x}|H_1, \boldsymbol{\theta}_1)$ and $p_X(\mathbf{x}|H_0, \boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_0$ stand for the sets of parameters assigned to hypotheses H_1 and H_0, respectively:

$$\begin{cases} H_0 : x \sim p_X(\mathbf{x}|H_0, \boldsymbol{\theta}_0) \\ H_1 : x \sim p_X(\mathbf{x}|H_1, \boldsymbol{\theta}_1) \end{cases}. \tag{1}$$

The statistical test then consists in inferring which of the two candidate PDFs best describes the measured signal $\mathbf{x}$, that is whether H_0 should be accepted or rejected. Although there are many possible ways to address this question, the generalized likelihood ratio (GLR) test enjoys several favourable properties as well as optimality. The GLR is defined as [37,38]

$$\Lambda(\mathbf{x}) = \frac{\sup\limits_{\boldsymbol{\theta}_1} p_X(\mathbf{x}|H_1, \boldsymbol{\theta}_1)}{\sup\limits_{\boldsymbol{\theta}_0} p_X(\mathbf{x}|H_0, \boldsymbol{\theta}_0)} = \frac{p_X\left(\mathbf{x}|H_1, \hat{\boldsymbol{\theta}}_1\right)}{p_X\left(\mathbf{x}|H_0, \hat{\boldsymbol{\theta}}_0\right)} \tag{2}$$

where as $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}_1$ are the maximum-likelihood estimates (MLEs) of the parameters of the two models. It is noteworthy that the GLR is eventually a functional of the signal $\mathbf{x}$ through the dependence of both p_X and the MLEs $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}_1$ on $\mathbf{x}$. Intuitively, if the measured signal $\mathbf{x}$ actually satisfies hypothesis H_1, then $p_X\left(\mathbf{x}|H_1, \hat{\boldsymbol{\theta}}_1\right)$ is expected higher than $p_X\left(\mathbf{x}|H_0, \hat{\boldsymbol{\theta}}_0\right)$ and the GLR will be significantly greater than one. On the opposite, if H_0 actually holds, then the GLR will be close to one. Therefore, the decision rule is to reject H_0 is the GLR $\Lambda(\mathbf{x})$ (the so-called "test statistics") is greater than a given threshold $\lambda \geqslant 1$.

It is noteworthy that the GLR – or its logarithm – has been used for long in various fields of science with different terminology (in particular for comparing models), such as the "weight of evidence" [39], the odds-ratio, the LOD (for "log of odds"), the "support" [37].

The GLR enjoys several strong properties. First, it is always greater than or equal to one. Next, according to the Neyman-Pearson criterion, the GLR is an optimal in the sense that it maximizes the probability of detection of state H_1 among all possible detectors given the same rate of false alarm – i.e. the probability of wrongly rejecting H_0 whereas it is true [40]. Third, the GLR theory also provides a useful tool in Wilk's theorem [41]. In particular, as the number of samples of the signal grows towards infinity, the test statistics $-2\ln\Lambda$ is asymptotically distributed as a chi-two (χ^2) with a number of degrees of freedom (DOF) equal to the difference in dimensionality between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. This property allows not only the identification of the most-likely model, but also the provision of a measure of statistical significance of the test.

2.3. Design of indicators

In the following, the logarithm of the GLR is considered since it will generally untangle calculations. All signals are considered white (i.e. delta-correlated) and zero-mean. The assumption of delta-correlation of the signals is only an apparent limitation, since a series of pre-whitening techniques have been demonstrated to be effective in bringing real diagnostic signals into this class (see e.g. [42]). The assumption of zero-mean is also trivially satisfied by first centring the signal (note that in the case of a cyclostationary signal, "centring" means removing the periodic component of the signal, which may be achieved by using the synchronous average or a combfilter [43,44,52]). Given a signal $\mathbf{x}$ and two alternative hypotheses H_1 and H_0, the log-GLR can thus be rewritten into

$$\ln\Lambda(\mathbf{x}) = \sum_{n=0}^{L-1} \ln p_X(x(n)|H_1, \hat{\boldsymbol{\theta}}_1) - \sum_{n=0}^{L-1} \ln p_X(x(n)|H_0, \hat{\boldsymbol{\theta}}_0) \stackrel{def}{=} \mathcal{L}_{H_1}(\mathbf{x}) - \mathcal{L}_{H_0}(\mathbf{x}) \tag{3}$$

where $\mathcal{L}_{H_i}(\mathbf{x})$ stands for the log-likelihood of signal $\mathbf{x}$ (it is, strictly speaking, a functional of the signal $\mathbf{x}$ defined within an arbitrary constant). It is then proposed to define an CM as

$$I_{H_1/H_0}(\mathbf{x}) \stackrel{def}{=} c \cdot \frac{\mathcal{L}_{H_1}(\mathbf{x}) - \mathcal{L}_{H_0}(\mathbf{x})}{L} \geqslant 0 \tag{4}$$

where the normalization by the signal length L is used to make the indicator asymptotically constant (when L grows towards infinity), and $c > 0$ is a calibration factor whose sole role is to possibly simplify the final expression. The fact the proposed indicator is non-negative results from the aforementioned property of the GLR (see previous subsection) and is also proved from the point of view of Kullback-Leibler divergence in Section 2.5. Note that a quantity proportional to $I_{H_1/H_0}(\mathbf{x})$ and known as the "deviance" is sometimes used in statistics to compare competing models [45].

2.4. Wilk's theorem and calibration of the indicators

Being computed on a finite-length signal, the proposed CM is itself an estimate that will statistically fluctuate according to the realization of the underlying stochastic process. It is therefore important to know its statistical distribution, in particular for setting the statistical threshold λ to which it will be compared for deciding whether to reject or not H_0. Quite remarkably, the GLR comes with a useful theorem that provides its asymptotic PDF. Wilk's theorem applies when the two hypotheses to be compared are nested, meaning that H_1 contains the parameters in H_0 and ν additional parameters. This will be the case for all the scenarios investigated in this paper. Therefore, by denoting ν the difference in dimensionality between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_0$, Wilk's theorem [41] establishes that twice the GLR is asymptotically follows a chi-2 distribution with ν DOFs, viz

$$2\big(\mathcal{L}_{H_1}(\mathbf{x}) - \mathcal{L}_{H_0}(\mathbf{x})\big) \underset{L\to\infty}{\sim} \chi^2_\nu. \tag{5}$$

This result is invoked to set $2\ln\lambda = \chi^2_{\nu,1-\alpha}$ in Eq. (2), with significance level α. It also follows that the proposed likelihood indicator is asymptotically distributed as a $(c/2L)\cdot\chi^2_\nu$ under H_0, thus with first two moments

$$\begin{aligned} &\mathbb{E}\{I_{H_1/H_0}(\mathbf{x})\} \approx \frac{c}{2L} \underset{L\to\infty}{\to} 0 \\ &\mathrm{Var}\{I_{H_1/H_0}(\mathbf{x})\} \approx \frac{1}{2}\left(\frac{c}{L}\right)^2 \underset{L\to\infty}{\to} 0. \end{aligned} \tag{6}$$

These equations demonstrate that the indicator converges towards zero under the null hypothesis.

2.5. Link with measures of information

It happens that the proposed likelihood indicators have a close connection with measures of information, such as the "negentropy" and the Kullback-Leibler divergence, as defined in information theory [46]. The link provides a different, yet not less interesting interpretation. It may be elucidated as follows.

(a) *Hypothesis H_1 holds true*

Let us first consider the case where the signal satisfies hypothesis H_1. Then, considering the large sample limit, one has

$$\lim_{L\to\infty}\frac{\mathcal{L}_{H_1}(\mathbf{x})}{L} = \int p_X(\mathbf{x}|H_1,\boldsymbol{\theta}_1)\ln p_X(\mathbf{x}|H_1,\boldsymbol{\theta}_1)d\mathbf{x} \overset{def}{=} -H(p_1) \tag{7}$$

where $H(p_1)$ is the "differential entropy" of the PDF $p_1 \equiv p_X(\mathbf{x}|H_1,\boldsymbol{\theta}_1)$ [47]. The opposite of the differential entropy which appears in the last equality of Eq. (7) is known as the "negentropy" [48], a measure the average amount of information contained in the signal. Similarly, the average amount of information under the alternative hypothesis H_0 is measured by

$$\lim_{L\to\infty}\frac{\mathcal{L}_{H_0}(\mathbf{x})}{L} = \int p_X(\mathbf{x}|H_0,\boldsymbol{\theta}_0)\ln p_X(\mathbf{x}|H_1,\boldsymbol{\theta}_1)d\mathbf{x} \tag{8}$$

A quantity recognized as the opposite of the "cross-entropy". Therefore, the proposed likelihood indicator I_{H_1/H_0} estimates the gain in information when hypothesis H_1 is actually true as opposed to incorrectly modelling the signal with hypothesis H_0. Stated differently, it is the gain in information that one can expect from adopting the correct model for the measured signal.

Interestingly, in the large sample limit, the likelihood indicator tends to be proportional to the Kullback-Leibler divergence [49] under H_1, i.e.

$$\lim_{L\to\infty} I_{H_1/H_0}(\mathbf{x}) \underset{H_1}{=} c\int p_X(\mathbf{x}|H_1,\boldsymbol{\theta}_1)\ln\frac{p_X(\mathbf{x}|H_1,\boldsymbol{\theta}_1)}{p_X(\mathbf{x}|H_0,\boldsymbol{\theta}_0)}d\mathbf{x} \overset{def}{=} c\cdot D_{KL}(p_1||p_0), \tag{9}$$

A quantity which measures the distance[2] between the two PDFs $p_1 \equiv p_X(\mathbf{x}|H_1,\boldsymbol{\theta}_1)$ and $p_0 \equiv p(\mathbf{x}|H_0,\boldsymbol{\theta}_0)$. Among several nice properties, the Kullback-Leibler divergence is non-negative.

(b) *Hypothesis H_0 holds true*

[2] The Kullback-Leibler is not strictly speaking a distance in the mathematical meaning since it is not symmetric and it does not satisfy the triangular inequality.

Let us now consider the case where the signal actually satisfies hypothesis H_0. If the models are nested – which means that assumption H_1 contains H_0 as a particular case – the estimate $\hat{\boldsymbol{\theta}}_1$ will tend to $\boldsymbol{\theta}_0$ and therefore $p_X\left(\mathbf{x}|H_1,\hat{\boldsymbol{\theta}}_1\right)$ to $p_X(\mathbf{x}|H_0,\boldsymbol{\theta}_0)$. Therefore, indicator I_{H_1/H_0} will tend to zero, viz

$$\lim_{L\to\infty} I_{H_1/H_0}(\mathbf{x}) \underset{H_0}{=} 0. \tag{10}$$

In light of these results, the following proposition holds:

Proposition 1. The scalar indicator $I_{H_1/H_0}(\mathbf{x})$ measures the extra information conveyed by the signal when it follows the model specified by hypothesis H_1 as opposed to H_0.

Interestingly, it will turn out in what follows that the information conveyed by the signal is transmitted by its "envelope". The latter will take different forms depending on the choice of the null and alternatives hypotheses H_0 and H_1. However, even if a link with entropy is made explicit in Eq. (7), the corresponding indicators differ from the other entropic definitions found in the literature [15–18].

In the following sections, CMs will be derived for testing different forms of non-stationarity and non-Gaussianity in a signal. Each of these sections will include (i) the analytical derivation of the indicator and its (ii) statistical threshold derived from Wilk's theorem, (iii) the discussion of potential approximate forms and their relationship to traditional or recently proposed indicators and (iv) a numerical validation of the procedure.

3. Testing for cyclostationarity under the Gaussian hypothesis

This section discusses the case of a signal $\mathbf{x}$ distributed according to a Gaussian law with null mean.

3.1. The GLR test

In this case, the null hypothesis H_0 describes a stationary signal and it is tested against the alternative hypothesis, H_1, describing a cyclostationary signal with a known period of N samples (a whole number is assumed hereafter for the sake of simplicity, while the general case with a non-integer period is discussed in Appendix B).

The null and alternative hypotheses read

$$\begin{cases} H_0 : \mathbf{x} \sim \mathcal{N}(\mathbf{x};\mathbf{0},\sigma^2\mathbf{I}) \\ H_1 : \mathbf{x} \sim \mathcal{N}(\mathbf{x};\mathbf{0},\mathrm{Diag}(\sigma^2(n))) \end{cases} \tag{11}$$

where symbol $\mathcal{N}$ stands for the (possibly complex-valued) Gaussian PDF (see Eq. (84) of Appendix A), $\mathrm{Diag}(a_n)$ stands for the diagonal matrix with diagonal elements a_n and $\sigma^2(n)$ is N-periodic variance – i.e. $\sigma^2(n)=\sigma^2(n+N), N\in\mathbb{N}$. Based on the log-likelihood worked out in Appendix A, the following result is found:

Proposition 2. The scalar indicator testing the Gaussian cyclostationary hypothesis (GCS) against the Gaussian stationary (GS) one is

$$I_{GCS/GS}(\mathbf{x}) = c\frac{\mathcal{L}_{GCS}(\mathbf{x})-\mathcal{L}_{GS}(\mathbf{x})}{L} = \ln\langle s^2(n)\rangle - \langle \ln s^2(n)\rangle \tag{12}$$

where c = 2 when $\mathbf{x}$ is real and c = 1 when it is complex and where $s^2(n)$ is the N-periodic component of the squared envelope of the signal, viz,

$$s^2(n)\frac{1}{K} \overset{\text{def}}{=} \sum_{k=0}^{K-1}|x(n+kN)|^2, \tag{13}$$

with $K=\lfloor L/N\rfloor$ is the number of N-sample periods of in the L-sample long signal.

(See Eqs. (86) and (87) for the expressions of $\mathcal{L}_{GCS}(\mathbf{x})$ and $\mathcal{L}_{GCS}(\mathbf{x})$, respectively.)

The squared envelope $s^2(n)$ in Eq. (13) is the MLE of the N-periodic variance obtained by the synchronous average, a quantity central importance in condition monitoring [43]. Note that $\langle s^2(n)\rangle$ stands for the time average value of $s^2(n)$ over the period $n=0,\ldots,N-1$, and thus represents the total power of the signal; it is also the MLE for the variance σ^2 under the stationary assumption H_0.

By virtue of the concavity of the logarithm, the so-defined indicator is again checked to be always non-negative: $I_{GCS/GS}\geqslant 0$. Another expression of the same indicator is

$$I_{GCS/GS}(\mathbf{x}) = -\left\langle \ln\left(\frac{s^2(n)}{\langle s^2(n)\rangle}\right)\right\rangle \tag{14}$$

which is to be interpreted as the time average of the logarithm of the N-periodic component of the squared envelope normalized by the signal power. Referring to Section 2.5, it is also construed as the information conveyed by the envelope $s^2(n)$ of the signal – i.e. by a periodic modulation as opposed to no modulation at all.

The indicator is scale invariant (it does not depend on the units of the original signal), which provides advantages for diagnostics in variable operating conditions.

3.2. Approximations of the indicator: Link with the kurtosis and with the sum of the peaks of the SES

(a) *Time-domain derivation*

Assuming that the time fluctuations in $s^2(n)$ are small as compared to the mean value the quantity $\langle s^2(n)\rangle$, it is possible to approximate Eq. (12) by a truncated Taylors' series

$$I_{GCS/GS}(\mathbf{x}) \approx \frac{1}{2}\frac{\langle\left(s^2(n)-\langle s^2(n)\rangle\right)^2\rangle}{\langle s^2(n)\rangle^2} = \frac{1}{2}\left(\underbrace{\frac{\langle s^2(n)^2\rangle}{\langle s^2(n)\rangle^2}}_{\kappa_4} - 1\right) \stackrel{\text{def}}{=} I^*_{GCS/GS}(\mathbf{x}). \tag{15}$$

Interestingly, the quantity

$$\kappa_4 \stackrel{\text{def}}{=} \frac{\langle s^2(n)^2\rangle}{s^2(n)^2} \tag{16}$$

appearing Eq. (15) is recognized as the kurtosis of $s^2(n)^{\frac{1}{2}}$ or, equivalently, as the relative variance of the squared envelope $s^2(n)$ (the link between the kurtosis and the variance of the squared envelope is also discussed in [18]). A connection to the SES is given hereafter.

(b) *Frequency-domain derivation*

Being the synchronous average of the envelope, $s^2(n)$ is also easily expressed as a function of the signal's squared envelope spectrum (SES):

$$s^2(n) = \sum_{h=-N/2}^{N/2} SES_{hK}\cdot e^{j2\pi\frac{hn}{N}}. \tag{17}$$

where the index hK represents the h-th harmonics of the fundamental cyclic frequency K in the squared-envelope spectrum SES, which is defined in this study as:

$$SES_k \stackrel{\text{def}}{=} \frac{1}{L}\sum_{n=0}^{L}|x(n)|^2\cdot e^{-j2\pi\frac{kn}{N}}. \tag{18}$$

Using the relationship with the SES and Parseval's theorem (remember that $s^2(n)$ is N-periodic under H_1), the approximate indicator (16) can be expressed as:

$$I^*_{GCS/GS}(\mathbf{x}) = \frac{\sum_{h=1}^{N/2}|SES_{hK}|^2}{|SES_0|^2} \tag{19}$$

where K is the index of the component with cyclic frequency corresponding to the period N. Although derived in a different way, $I^*_{GCS/GS}(\mathbf{x})$ is recognized equivalent to "degree of cyclostationarity" of Ref. [26]. As demonstrated in Ref. [13], $I^*_{GCS/GS}(\mathbf{x})$ is also strictly linked to the kurtosis κ_4, which has the expression

$$\kappa_4 = 1 + 2\frac{\sum_{k=1}^{L/2}|SES_k|^2}{|SES_0|^2} \geqslant 1 \tag{20}$$

with the identity $\kappa_4 = 1 + 2I^*_{GCS/GS}$ in case of observing only one cycle of the signal (i.e. when $L = N$ and $K = 1$). Akin to the indicators proposed in Ref. [13], $I^*_{GCS/GS}$ can therefore be interpreted as a cycle-specific version of the kurtosis.

3.3. Comparison with the correlated kurtosis

At first glance, indicator $I_{GCS/GS}$ shows similarity with the recently proposed "correlated kurtosis", CK, that aims at measuring higher-order degrees of cyclostationarity [50,51]. Yet this is only apparent. The CK is defined as the ratio of the arithmetic average of the synchronous geometric average, i.e.

$$CK(\mathbf{x}) \overset{\text{def}}{=} \frac{\frac{1}{N}\sum_{n=0}^{N-1}\left(\prod_{k=0}^{K-1}|x(n+kN)|^2\right)}{\langle x^2(n)\rangle^K}, \tag{21}$$

whereas $I_{GCS/GS}$ involves the geometric average of the synchronous arithmetic average:

$$e^{-N\cdot I_{GCS/GS}(\mathbf{x})} = \frac{\prod_{n=0}^{N-1}\left(\frac{1}{K}\sum_{k=0}^{K-1}|x(n+kN)|^2\right)}{\langle x^2(n)\rangle^N}. \tag{22}$$

The difference between these two formulas is illustrated in Fig. 1. One drawback of the correlated kurtosis as compared to $I_{GCS/GS}$ is that it collapses as soon as one of the values $x(n+kN)$ entering in the geometric average approaches zero. This will hardly happen with $I_{GCS/GS}$ where the geometric average is calculated on values whose squares have been synchronously averaged, therefore with vanishing probability of being nil.

3.4. Statistical threshold

A statistical threshold can be obtained using Wilks' theorem. In this case $(2\ln\Lambda)$ is asymptotically χ^2-distributed under H_0, with degrees of freedom equal to N–1 (the difference in dimensionality of the two estimators $\langle s^2(n)\rangle$ and $s^2(n)$). Assuming that $L = KN$ (true asymptotically), the CM is therefore distributed as

$$I_{GCS/GS}(\mathbf{x}) \sim \frac{c}{2L}\chi^2_{N-1} \tag{23}$$

with $c = 2$ when $\mathbf{x}$ is real and $c = 1$ when it is complex. This implies in particular that, under H_0,

$$\begin{aligned}\mathbb{E}\{I_{GCS/GS}(\mathbf{x})\} &\underset{N\to\infty}{\to} \frac{c(N-1)}{2L} \approx \frac{c}{2K}\\ \mathrm{Var}\{I_{GCS/GS}(\mathbf{x})\} &\underset{N\to\infty}{\to} 2c\frac{N-1}{2L^2} \approx \frac{c}{LK}.\end{aligned} \tag{24}$$

The statistical test is thus to reject H_0 if $I_{GCS/GS}(\mathbf{x}) \geqslant (c/2L)\chi^2_{N-1,1-\alpha}$ with significance level α.

3.5. Numerical validation

The indicators $I_{GCS/GS}$ and $I^*_{GCS/GS}$ (and the associated statistical threshold) are tested on two sets of numerically generated signals. The first set consists of multiple generations of identically distributed GS signals,

$$\mathbf{x}^{(r)} \sim \mathcal{N}(\mathbf{x}; 0, \sigma^2\mathbf{I}),\ r = 0,\ldots,R-1 \tag{25}$$

while the second set is a series of GCS signals,

$$\mathbf{x}^{(r)} \sim \mathcal{N}(\mathbf{x}; 0, \mathrm{Diag}\{\sigma_r^2(n)\}),\ r = 0,\ldots,R-1 \tag{26}$$

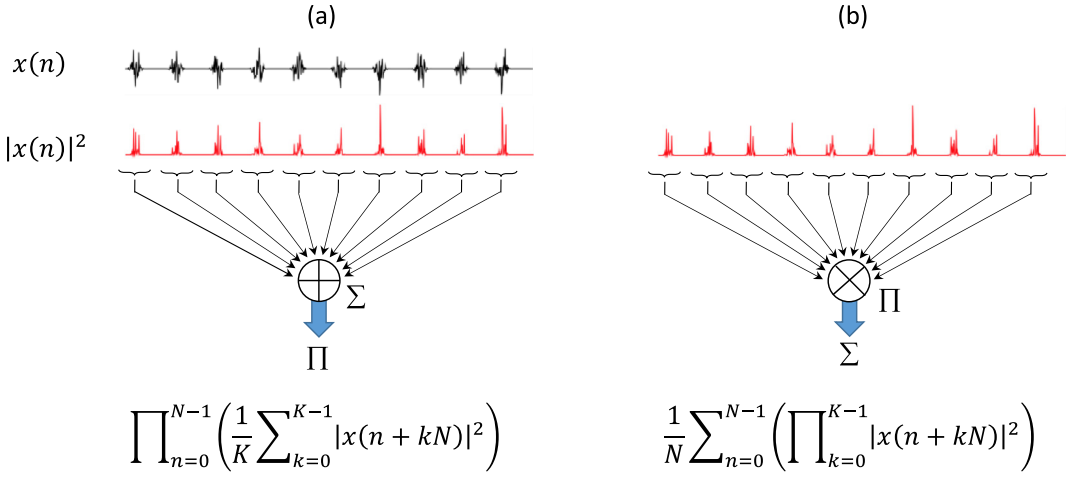

Fig. 1. Schematic comparison of (a) indicator $I_{GCS/GS}$ with (b) the correlated kurtosis. While the former involves the geometric average of the synchronous arithmetic average, the later involves the arithmetic average of the synchronous geometric average.

characterized by an increasing modulation depth m_r:

$$\sigma_r^2(n) = \frac{\sigma^2\left(1 + m_r \sin\left(\frac{2\pi n}{N}\right)\right)}{1 + m_r}, \; r = 0, \ldots, R-1. \tag{27}$$

The numerical tests of Fig. 2, performed with $L = 10,000$, $N = 100$ and $\sigma^2 = 3$, confirm the analytical results, with a similar performance of the two indicators $I_{GCS/GS}$ and $I^*_{GCS/GS}$. In particular, they are almost indistinguishable for $m_r < 0.3$, they are both sensitive to the cyclostationary component from $m_r > 0.05 - 0.06$ and they clearly exceed the threshold for $m_r > 0.1$. The behaviour of the kurtosis is similar, but shows a slightly later separation of the behaviours of the two signals: there is no sensitivity to the cyclostationary signal for $m_r < 0.1$.

A similar test, performed increasing the length of the signals to $L = 100,000$, confirms the convergence of the indicators and the superiority of the newly developed indicators versus the kurtosis (see Fig. 3).

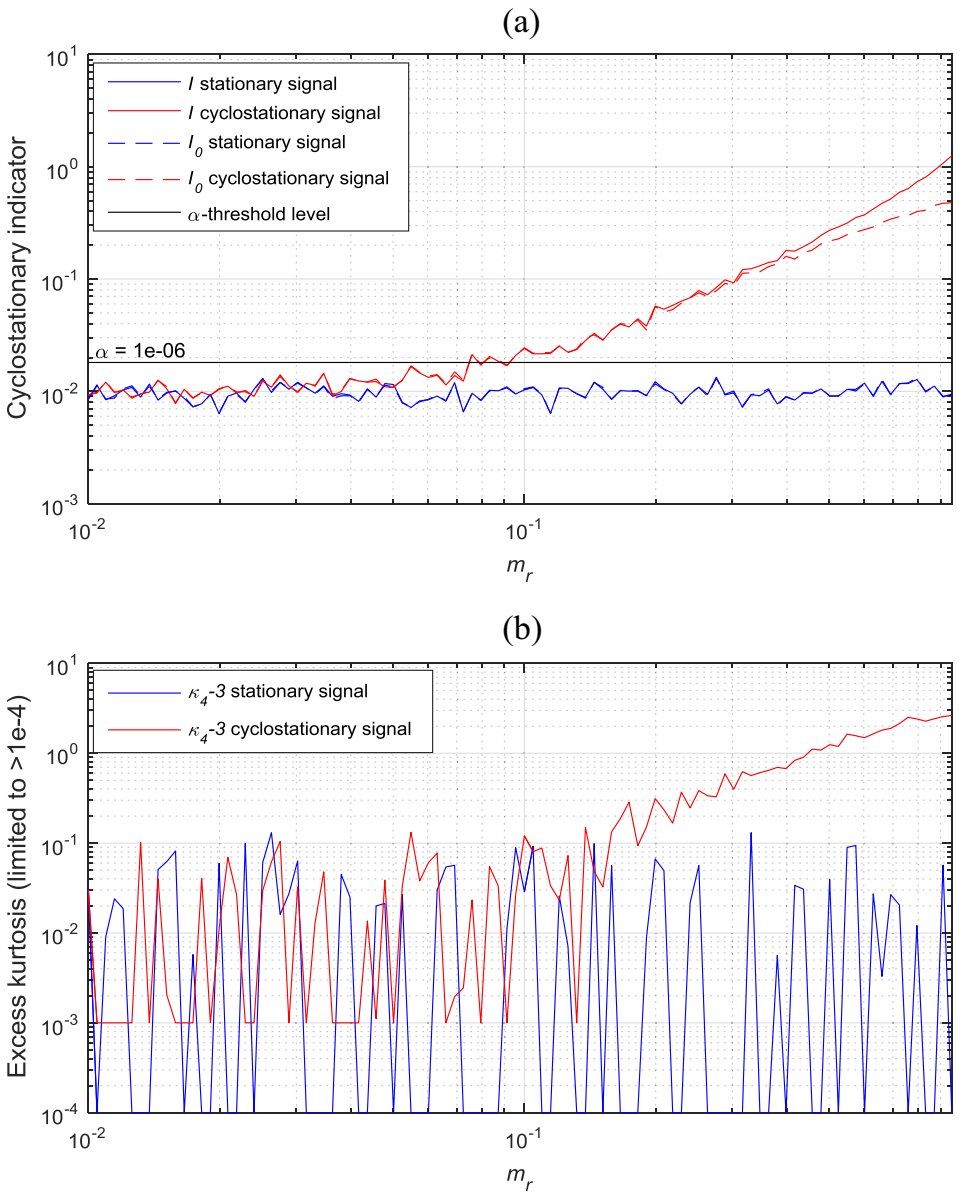

Fig. 2. Performance of (a) condition indicators $I_{GCS/GS}$ (exact) and $I^*_{GCS/GS}$ (approximate) and (b) excess kurtosis as a function of the modulation depth m_r on a series of numerically generated signals ($L = 10,000$, $N = 100$ and $\sigma^2 = 3$). A significance level $\alpha = 1 \cdot 10^{-6}$ has been used for the statistical threshold.

4. Testing cyclostationary under non-Gaussian conditions

In the previous section, cyclostationarity was tested under the commonly adopted assumption of a Gaussian signal. However, as described in many experimental studies, the insurgence of a fault in a mechanical component often results also in a deviation from Gaussianity, usually in the leptokurtic direction (i.e. impulsive signals). Moreover, non-damage symptomatic impulsive components might be present in the signal even in healthy machines due to harsh operating conditions (e.g. combustion engines, mining and coal milling machinery). As shown in Ref. [14], exogenous impulsive components can completely compromise the effectiveness of tools based on the squared envelope spectrum, thus rendering inapplicable the use of the kurtosis, $I_{GCS/GS}$ and $I^*_{GCS/GS}$. This is due to the hybrid sensitivity of the squared envelope to both non-stationarity and non-Gaussianity.

Therefore, this section will release the hypothesis of Gaussianity and pursue optimal indicators for testing for the presence of generalized Gaussian cyclostationarity (GGCS) against generalized Gaussian stationarity (GGS). It is noteworthy that GGCS include processes which are higher-order cyclostationary as opposed to simply second-order cyclostationary (i.e. GCS) [28].

A zero-mean GG PDF has the form:

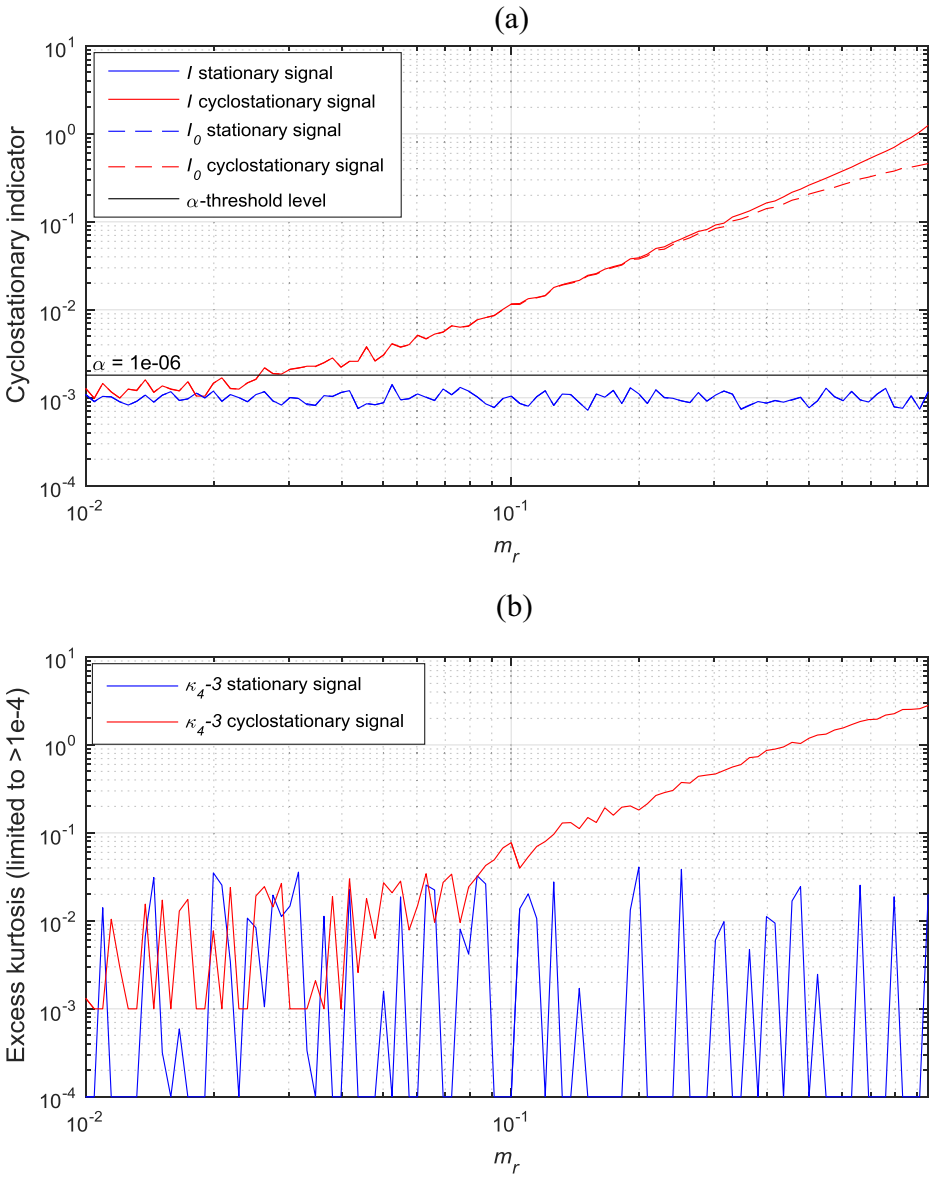

Fig. 3. Same experiment as in Fig. 2 with $L = 100,000$.

$$p_X(x;\eta,\beta)=\frac{\beta}{2\pi^{(c-1)}\Gamma(c/\beta)\eta}e^{-\left(\frac{|x|}{\eta}\right)^{\beta}}\overset{\text{def}}{=}\mathcal{GN}(x;0,\eta,\beta),\quad \beta>0,\eta>0 \tag{28}$$

with $c=1$ when x is real and $c=2$ when it is complex [53]. This PDF allows the description of highly impulsive signals, and has the moments listed in Table 1. In particular the parameter $\beta>0$ defines the shape of the distribution: $\beta=2$ corresponds to the special case of the normal distribution, smaller β values result in increasingly leptokurtic distributions, $\beta=1$ corresponds to the Laplacian distribution and, as $\beta\to\infty$, the PDF converges to an uniform distribution. The scale parameter. Governs the variance of the distribution and, in case of $\beta=2$ (Gaussian), $\eta=\sigma/\sqrt{2}$. By means of an example, Fig. 4 illustrates three realizations of random processes with $\beta=1$, $\beta=0.5$ and $\beta=0.1$.

The GG PDF is therefore used in this chapter to separate the detection of non-stationarity from non-Gaussianity, with the first revealed in a time-dependency of η and the second dependent only on the parameter β.

4.1. The GLR test

Under this extended modelling hypothesis, the hypothesis test is transformed in the following:

$$\begin{cases} H_0 : x\sim G\mathcal{N}(\mathbf{x};0,\eta,\beta)\\ H_1 : x\sim G\mathcal{N}(\mathbf{x};0,\eta(n),\beta)\end{cases} \tag{29}$$

where $\eta(n)$ is periodic with known period N. The two models are nested (necessary hypothesis for the application of Wilk's theorem), since β is included in the set of model parameters to be estimated for both H_0 and H_1, with conditional probability densities of the observed signal $\mathbf{x}$:

$$p_X(\mathbf{x}|H_0,\eta_0,\beta_0)=\prod_{n=0}^{L-1}\frac{\beta_0}{2\pi^{(c-1)}\Gamma(c/\beta_0)\eta_0}e^{-\left(\frac{|x(n)|}{\eta_0}\right)^{\beta_0}} \tag{30}$$

and

$$p_X(\mathbf{x}|H_1,\eta_1(n),\beta_1)=\prod_{k=0}^{K-1}\prod_{n=0}^{N-1}\frac{\beta_1}{2\pi^{(c-1)}\Gamma(c/\beta_1)\eta_1(n)}e^{-\left(\frac{|x(kN+n)|}{\eta_1(n)}\right)^{\beta_1}}. \tag{31}$$

In this case, the following CM is established based on the results of Appendix A.

Proposition 3. The scalar indicator testing the generalized Gaussian cyclostationary (GGCS) hypothesis against the generalized Gaussian stationary (GGS) one is

$$\begin{aligned} I_{GGCS/GGS}(\mathbf{x})&=2\frac{\mathcal{L}_{GGCS}(\mathbf{x})-\mathcal{L}_{GGS}(\mathbf{x})}{L}\\ &=2\hat{\beta}_0^{-1}\ln\langle s^{\hat{\beta}_0}(n)\rangle-2\hat{\beta}_1^{-1}\langle\ln s^{\hat{\beta}_1}(n)\rangle+2C(c,\hat{\beta}_0,\hat{\beta}_1)\end{aligned} \tag{32}$$

where $s^{\beta}(n)$, $\beta=\hat{\beta}_0,\hat{\beta}_1$, is the N-periodic component of the β-power envelope of the signal,

$$s^{\beta}(n)\frac{1}{K}\overset{\text{def}}{=}\sum_{k=0}^{K-1}|x(n+kN)|^{\beta} \tag{33}$$

with $K=\lfloor L/N\rfloor$ is the number of N-sample periods of in the L-sample long signal,$\hat{\beta}_0$ and $\hat{\beta}_1$ are the MLEs of the shape parameters β_0 and β_1 (obtained as explained in Appendix A), and

$$C\left(c,\hat{\beta}_0,\hat{\beta}_1\right)=\ln\left(\frac{\left(\hat{\beta}_0\right)^{\frac{1}{\hat{\beta}_0}-1}\Gamma\left(c/\hat{\beta}_0\right)}{\left(\hat{\beta}_1\right)^{\frac{1}{\hat{\beta}_1}-1}\Gamma\left(c/\hat{\beta}_1\right)}\right)+\left(\frac{1}{\hat{\beta}_0}-\frac{1}{\hat{\beta}_1}\right) \tag{34}$$

with c = 2 when $\mathbf{x}$ is real and c = 1 when it is complex.

Table 1
Moments of the generalized Gaussian distribution ($c=1$ or 2 for real or complex variables, respectively).

Moment	Moment order	Expression
Mean	1	0
Variance	2	$\sigma^2=\eta^2\frac{\Gamma(3/\beta)}{\pi^{(c-1)}\Gamma(c/\beta)}$
Skewness	3	0
Kurtosis	4	$\kappa=\pi^{(c-1)}\frac{\Gamma(5/\beta)\Gamma(c/\beta)}{\Gamma(3/\beta)^2}$

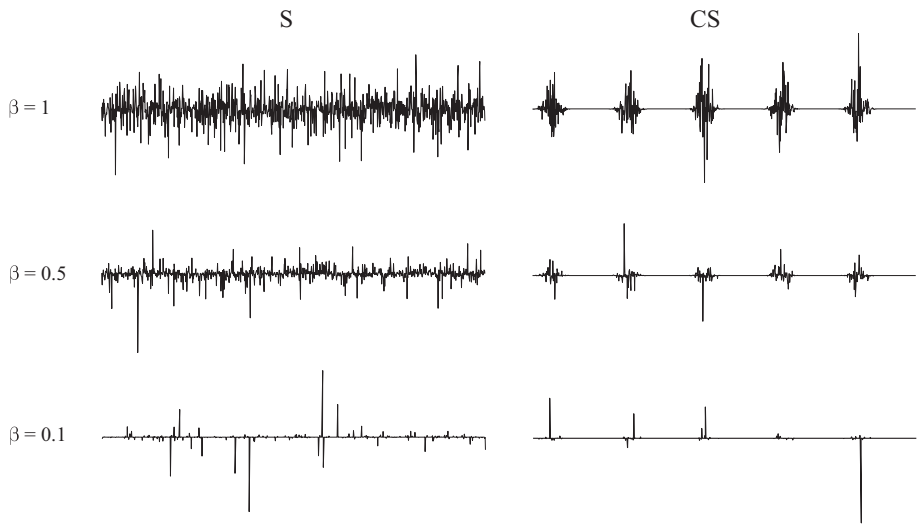

Fig. 4. Realizations of generalized Gaussian processes with $\beta = 1$, $\beta = 0.5$ and $\beta = 0.1$ in the stationary (left column) and cyclostationary (right column) cases.

Note that the factor 2 is used in the definition in order to recover the indicator of the Gaussian case, $I_{GCS/GS}(\mathbf{x})$, as a particular case when $\hat{\beta}_0 = \hat{\beta}_1 = 2$.

(See Eqs. (92) and (93) for the expressions of $\mathcal{L}_{GGCS}(\mathbf{x})$ and $\mathcal{L}_{GGS}(\mathbf{x})$, respectively.)

Therefore the quantity

$$I_{GGCS/GGS}(\mathbf{x}) = 2\left\langle \ln\left(\frac{\langle s^{\hat{\beta}_0}(n)\rangle^{1/\hat{\beta}_0}}{\left(s^{\hat{\beta}_1}(n)\right)^{1/\hat{\beta}_1}}\right)\right\rangle + 2C\left(c,\hat{\beta}_0,\hat{\beta}_1\right) \tag{35}$$

is an optimal indicator of cyclostationarity under the assumption of GG signals. The indicator is seen to essentially consist in comparing the power means taken at distinct exponents $\hat{\beta}_0$ and $\hat{\beta}_1$. Referring to Section 2.5, it is also interpreted as the extra information conveyed by the power-envelope $s^{\hat{\beta}_1}(n)$ of the signal as compared to the power-envelope $s^{\hat{\beta}_0}(n)$. The scale-invariance of the indicator is surely a useful property in diagnostic applications, for instance for comparing signals under variable conditions.

Compared to the Gaussian case discussed in the previous section, there is a clear similarity in the logarithmic term, which consists of a ratio of the synchronous averaged $\hat{\beta}_1$-power envelope with the average $\hat{\beta}_0$-power of the signal. Note that under the special case of a Gaussian assumption for both H_0 and H_1, the additional term $C\left(c,\hat{\beta}_0,\hat{\beta}_1\right)$ is null and the indicator boils down to $I_{GCS/GS}(\mathbf{x})$ in Eq. (14).

It is noteworthy that the constant $C\left(c,\hat{\beta}_0,\hat{\beta}_1\right)$ appearing in definition of the indicator must be kept in order to allow quantitative comparisons between signals characterized by different values of the shape parameters $\hat{\beta}_0$and$\hat{\beta}_1$. This is particularly important if the indicator is to be used for trending (e.g. in condition monitoring or prognostics) when the shape parameters are allowed to evolve with the state of the device under study. However, in the special situations where the shape parameters *i*) are known or *ii*) forced to be constant for all states of signals to be compared or *iii*) identical ($\hat{\beta}_0 = \hat{\beta}_1$), then the additive constant C can be removed from the above indicator, which then simplifies in

$$I'_{GGCS/GGS}(\mathbf{x}) = 2\left\langle \ln\left(\frac{\langle s^{\hat{\beta}_0}(n)\rangle^{1/\hat{\beta}_0}}{\left(s^{\hat{\beta}_1}(n)\right)^{1/\hat{\beta}_1}}\right)\right\rangle. \tag{36}$$

4.2. Approximations of the indicator

Using the approximation $\hat{\beta}_0 \approx \hat{\beta}_1$ a first approximation of the indicator is found as

$$I'_{GGCS/GGS}(\mathbf{x}) \approx \frac{2}{\hat{\beta}_1}\left\{\ln\langle s^{\hat{\beta}_1}(n)\rangle - \langle \ln s^{\hat{\beta}_1}(n)\rangle\right\}. \tag{37}$$

This already shows a clear similarity with the Gaussian counterpart of Eq. (14). Then, applying a truncated Taylor series expansion (akin to the Gaussian case), an even simpler formulation of the indicator is obtained as

$$I_{GGCS/GGS}(\mathbf{x}) = \frac{2}{\hat{\beta}_1}\left\langle \ln\left(\frac{\langle s^{\hat{\beta}_1}(n)\rangle}{s^{\hat{\beta}_1}(n)}\right)\right\rangle \approx \frac{1}{\hat{\beta}_1}\left(\underbrace{\frac{\langle s^{\hat{\beta}_1}(n)^2\rangle}{\langle s^{\hat{\beta}_1}(n)\rangle^2}}_{\kappa_{2\hat{\beta}_1}} - 1\right) \stackrel{\text{def}}{=} I^*_{GGCS/GGS}(\mathbf{x}) \tag{38}$$

where

$$\kappa_{2\beta} \stackrel{\text{def}}{=} \frac{\langle s^{\beta}(n)^2\rangle}{\langle s^{\beta}(n)\rangle^2} \tag{39}$$

is recognized as the kurtosis of $s^{\beta}(n)^{\frac{1}{2}}$ or, equivalently, as the relative variance of the power-envelope $s^{\beta}(n)$.

The above definition of $I^*_{GGCS/GGS}$ provides also a meaningful interpretation of the way in which the approximate indicator detects the cyclostationary behaviour: akin to its Gaussian counterpart of Eq. (16), the indicator focusses on the deviation of the synchronous averaged power-envelope $s^{\hat{\beta}_1}(n)$ from its average value.

The value of $I^*_{GGCS/GGS}$ can also be expressed as a function of the envelope spectrum, in this case calculated on the $\hat{\beta}_1$-power envelope (PES), rather than the squared envelope, i.e.

$$I^*_{GGCS/GGS}(\mathbf{x}) = \frac{2}{\hat{\beta}_1}\frac{\sum_{h=1}^{N/2}|PES_{hK}|^2}{|PES_0|^2} \tag{40}$$

with

$$PES_k \frac{1}{L} \stackrel{\text{def}}{=} \sum_{h=1}^{N/2}|x(n)|^{\hat{\beta}_1} e^{-j2\pi\frac{nk}{L}}. \tag{41}$$

It is again checked that $I^*_{GGCS/GGS}(\mathbf{x})$ boils down to $I^*_{GCS/GS}(\mathbf{x})$ when $\hat{\beta}_1 = 2$ (i.e. for Gaussian signals).

Of particular importance is the limit case $\beta_1 \to 0$ which corresponds to highly impulsive environments. Using the fact that $X^{\beta} = e^{\beta \ln X} \to \beta \ln X$ for $X > 0$ when $\beta \to 0$, the power envelope in Eq. (41) can then be replaced by the log-envelope:

$$LES_k = \frac{1}{L}\sum_{h=1}^{N/2}\ln|x(n)|e^{-j2\pi\frac{nk}{L}}. \tag{42}$$

This defines the spectrum of the log-envelop (LES) which has been introduced in Ref. [14] to provide robustness in the detection of cyclostationary signals in the presence of strong impulsive noise. Substituting into Eq. (40), the corresponding CM (simplified after removing the scale factor $2/\beta_1$) reads

$$I^*_{GGCS/GGS}(\mathbf{x}) = \frac{\sum_{h=1}^{N/2}|LES_{hK}|^2}{|LES_0|^2}. \tag{43}$$

4.3. Statistical threshold

Exploiting Wilk's theorem, it is possible to establish a threshold for $I_{GGCS/GGS}$. Given that the H_0model has only 2 DOFs (η, β) and the H_1 model has $N+1$ DOFs ($\eta(n)$, β):

$$2\ln\Lambda \sim \chi^2_{N-1}. \tag{44}$$

Therefore the optimal indicator $I_{GGCS/GGS}$ is, as in the Gaussian case, asymptotically distributed according to a $N-1$ DOF chi-square:

$$I_{GGCS/GGS}(\mathbf{x}) \sim \frac{1}{L}\chi^2_{N-1} \tag{45}$$

with the same properties as its Gaussian counterpart.

4.4. Some numerical results

The indicators $I_{GGCS/GGS}$ and $I^*_{GGCS/GGS}$ (and the corresponding statistical threshold) are tested on two sets of numerically generated signals. The first set consists of multiple generations, $r = 0, \ldots, R-1$, of identically distributed GGS signals

$$x^{(r)}(n) \sim \mathcal{GN}(x; 0, \eta, \beta) \tag{46}$$

with $n = 0, \ldots, L-1$, while the second set is a series of GGCS signals component,

$$x^{(r)}(n) \sim \mathcal{GN}(x; 0, \eta_r(n), \beta), r = 0, \ldots, R \tag{47}$$

with an increasing modulation depth m_r:

$$\eta_r(n) = \frac{\eta(1 + m_r \sin(2\pi n/N))}{1 + m_r}, r = 0, \ldots, R. \tag{48}$$

The scale parameter for each GGS signals is set at $\eta = 3$ and the shape parameter at $\beta = 1$. The same shape parameter $\beta = 1$ is chosen for the GGCS signals, while the time-varying shape parameters $\eta_r(n)$ has a common cycle length $N = 100$ and modulation depth m_r in the range 10^{-2} to $10^{-0.02}$. All signals are generated with length $L = 10,000$ (see Figs. 6–8).

The validation of the asymptotic chi-square behaviour is obtained by applying the analytically calculated threshold to a series of 1000 identically distributed GGS signals, with the same statistical properties as in the previous numerical example (scale parameter $\eta = 3$ and the shape parameter $\beta = 1$). The threshold is calculated for a series of significance levels and for each the actual number of exceedances (out of 1000 signals) is obtained. The ideal asymptotic behaviour would result in a straight 45° line in the plane test significance level versus exceedances. This procedure is repeated varying the length of the signals from $L = 2,000$ to $L = 10,000$, to show further improvement of the behaviour.

5. Testing generalized Gaussianity against Gaussianity under the cyclostationary assumption

One may wonder whether the price to pay for the flexibility of the generalized Gaussian cyclostationary model is really worthwhile as compared to the simplicity of the Gaussian cyclostationary model (compare Eqs. (12) and (32) of Propositions 1 and 2, respectively). The answer to this question is rigorously answered by another test which explicitly takes into account the "cost" of each model.

Proposition 4. The scalar indicator testing the generalized Gaussian cyclostationary (GGCS) hypothesis against the Gaussian cyclostationary (GCS) one is

$$I_{GGCS/GCS}(\mathbf{x}) = 2\frac{\mathcal{L}_{GGCS}(\mathbf{x}) - \mathcal{L}_{GCS}(\mathbf{x})}{L} = I_{GGCS/GS}(\mathbf{x}) - \frac{2}{c} \cdot I_{GCS/GS}(\mathbf{x}) \tag{49}$$

where $I_{GGCS/GS}(\mathbf{x})$ is as defined in Eq. (32) with the MLE $\hat{\beta}_0$ replaced by the known value 2, and where c = 2 when **x** is real and c = 1 when it is complex.

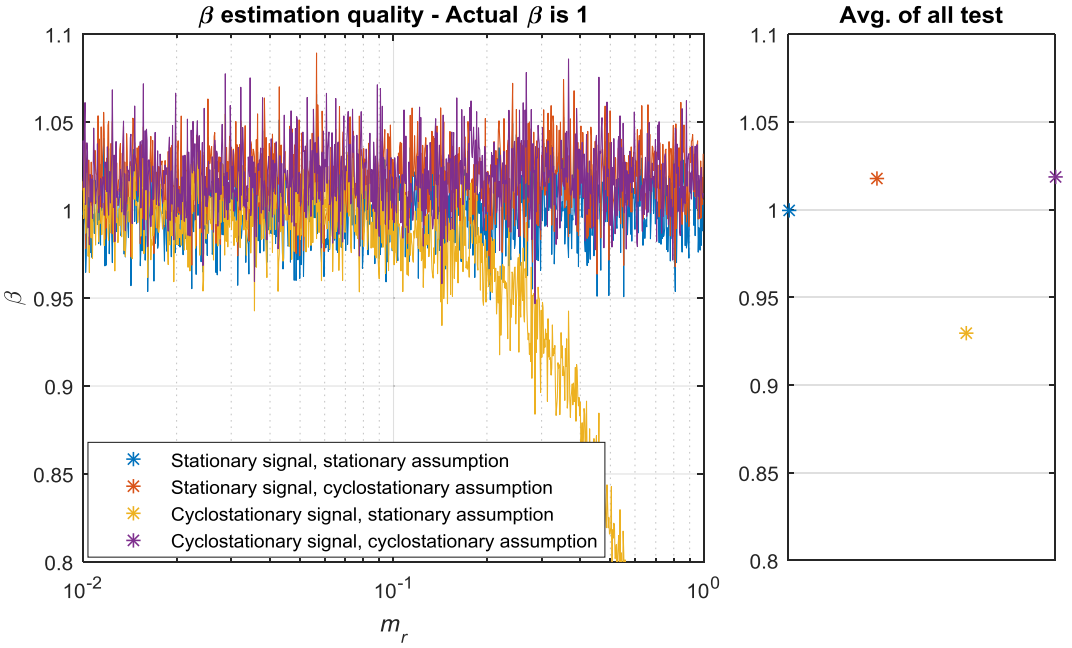

Fig. 5. Estimation of the shape parameter β for generalized Gaussian stationary and cyclostationary signals. The stationary signals are sampled from a generalized Gaussian distributions with constant parameters $\eta = 3$ and $\beta = 1$, while the cyclostationary signals are generated with a periodically-time varying scale parameter η with increasing modulation depth m_r.

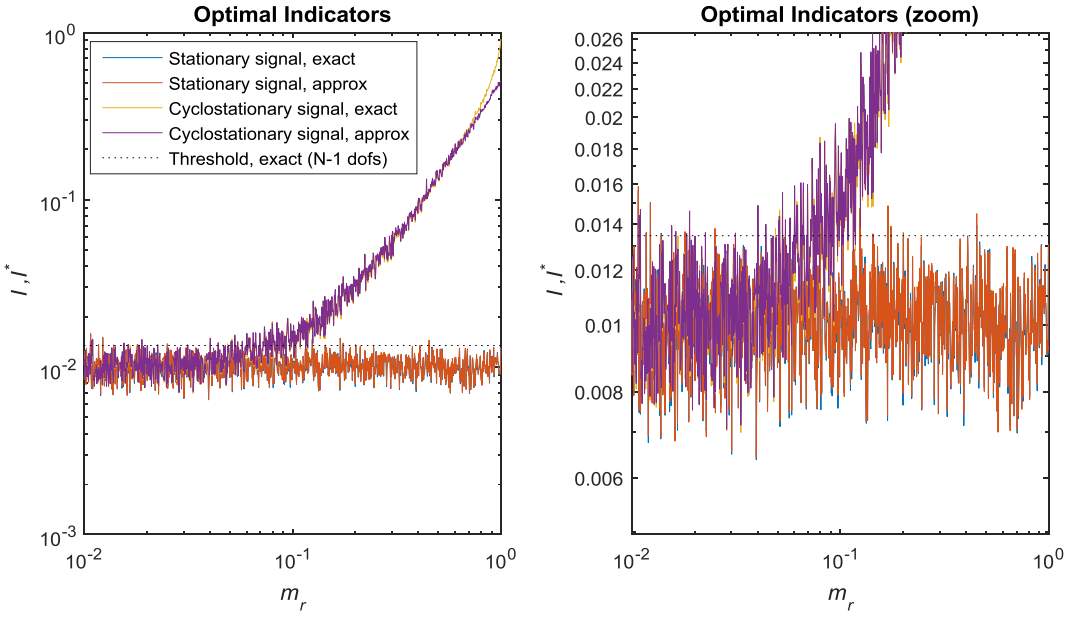

Fig. 6. Condition indicator $I_{GGCS/GGS}$ (exact) and $I^*_{GGCS/GGS}$ (approximate) for generalized Gaussian cyclostationary and stationary signals, compared to 1% significance threshold levels. The signals are generated as in Fig. 5.

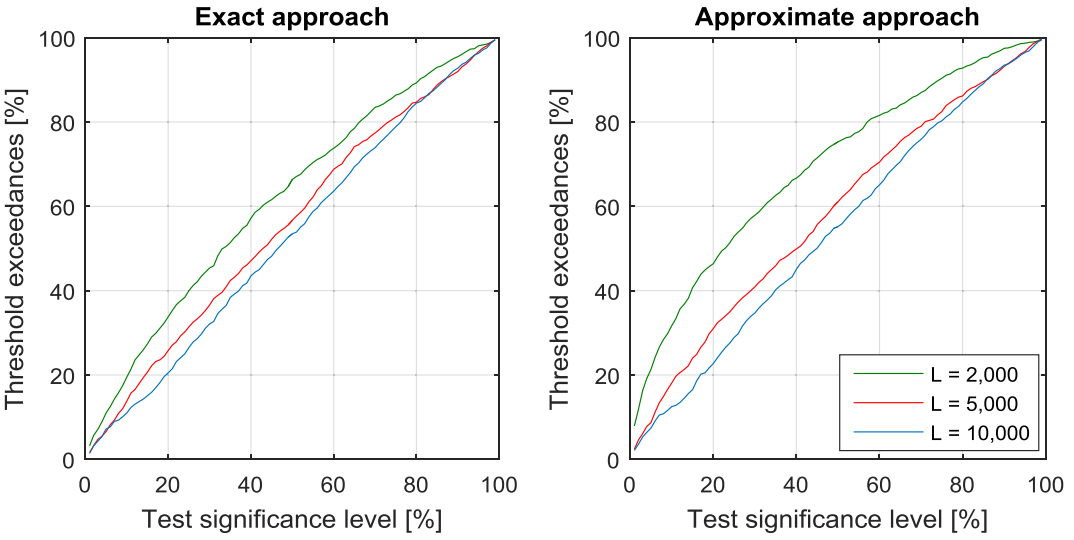

Fig. 7. Validation of the asymptotic behaviour of Wilk's threshold for the stationary generalized Gaussian signal and effect of signal length: (left) Exact; (right) approximate.

Using Wilk's theorem, the likelihood indicator $I_{GGCS/GCS}(\mathbf{x})$ is asymptotically distributed according to a 1-DOF chi-square under H_0 (H_0 has N DOFs and H_1 has $N+1$ DOFs),

$$I_{GGCS/GCS}(\mathbf{x}) \sim \frac{1}{L}\chi^2_1. \tag{50}$$

How the CMs $I_{GCS/GS}(\mathbf{x})$, $I_{GGCS/GGS}(\mathbf{x})$ and $I_{GGCS/GCS}(\mathbf{x})$ can be used conjointly in order to identify the most likely statistical property of a symptom – i.e. GS, GGS, GCS or GGCS – will be addressed in details in Section 10.

6. The case of unknown cyclic period

In practical cases, often the exact cyclic frequency of a fault is unknown. For instance it has been shown that the exact fault frequency of rolling element bearings might be few percentages different from the theoretical value at low bearing

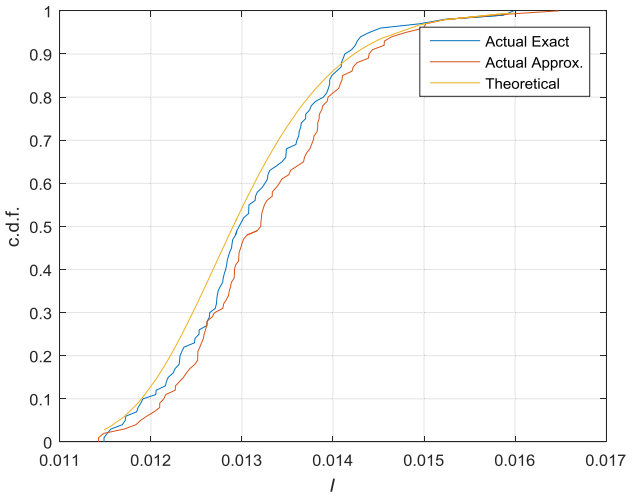

Fig. 8. Numerical validation of the analytical CDF expression.

loads [54]. The methodology proposed in the previous Sections 3-5 must therefore be modified to apply over a range of possible periods N. This is done by including N in the set of parameters to be estimated in Eqs. (1)–(3). The principle is illustrated hereafter in the case of $I_{GGCS/GGS}$, yet it applies similarly to $I_{GCS/GS}$ and $I_{GGCS/GCS}$.

Let's consider the test of hypotheses

$$\begin{cases} H_0 : x \sim \mathcal{GN}(\mathbf{x}; 0, \eta, \beta) \\ H_1 : x \sim \mathcal{GN}(\mathbf{x}; 0, \eta(n), \beta, N) \end{cases} \tag{51}$$

where the cyclic period N is now included in the set of unknown parameters. From Eq. (2), the associated GLR is

$$\Lambda(\mathbf{x}) = \frac{\max\limits_{N_{\min} \leqslant N \leqslant N_{\max}} p\left(\mathbf{x}|H_1, \hat{\eta}^{(N)}(n), \hat{\beta}_1^{(N)}\right)}{p\left(\mathbf{x}|H_0, \hat{\eta}, \hat{\beta}_0\right)}, \tag{52}$$

which is maximized over the range of possible values, $[N_{\min}; N_{\max}]$ of N. Therefore, the CM is obtained as

$$I_{GGCS/GGS}(\mathbf{x}) = \max_{N_{\min} \leqslant N \leqslant N_{\max}} I_{GGCS/GGS}(\mathbf{x}, N) \tag{53}$$

where $I_{GGCS/GGS}(\mathbf{x}, N)$ denotes the indicator of Eq. (30) conditioned to a given period N. This is implemented by estimating $\hat{\eta}^{(N)}(n)$, $\hat{\beta}_1^{(N)}$ in an outer line-search over N within the specified range $N_{\min} \leqslant N \leqslant N_{\max}$ and finally the values of $\hat{\eta}^{(\hat{N})}(n)$ and $\hat{\beta}_1^{(\hat{N})}$ are selected as the MLEs.

The thresholds previously established when N was knew unfortunately do not apply any more when N is replaced by its MLE. They should ne replaced by new expressions obtained as follows. The cumulative density function (CDF) of $I_{GGCS/GGS}(\mathbf{x}, N)$ is a product of the CDFs of $I_{GGCS/GGS}(\mathbf{x}, N)$, each following a chi-square distribution with $N-1$ DOFs:

$$\text{Prob}\{I_{GGCS/GGS}(\mathbf{x}) < z\} = \prod_{N=N_{\min}}^{N_{\max}} \text{Prob}\{I_{GGCS/GGS}(\mathbf{x}, N) < z\} = \prod_{N=N_{\min}}^{N_{\max}} F_{\chi^2_{N-1}}(z). \tag{54}$$

An analytic expression of a threshold for this test is intractable, however a numerical evaluation can be done by calculating the p-value simply using the CDF product of the single $I_{GGCS/GGS}(\mathbf{x}, N)$. The p-value (analogous to the test-significance value specified for the thresholding of previous sections) will indicate the relative likelihood of H_0 versus H_1, with low p-values indicating a propension to reject H_0 (signal not stationary). The same procedure can be applied to the approximate version $I^*_{GGCS/GGS}$, calculated on the basis of the single-period indicators of Eq. (38).

A numerical test, analogous to the ones presented in Section 4, has been executed to validate this derivation. A series of 100 GGS signals with length $L = 10,000$, $\eta = 3$ and $\beta = 1$ were generated and the corresponding $I_{GGCS/GGS}$ and $I^*_{GGCS/GGS}$ indicators estimated with $90 \leqslant N \leqslant 110$ (see Fig. 9).

The same numerical test has then been extended to 100 GGCS, with the same properties as the ones described in Section 4, with actual period $N = 100$. The statistical test was executed for potential cyclostationary behaviour with cyclic period in the range $90 \leqslant N \leqslant 110$.

As discussed before, in this case obtaining a threshold for a specified significance level is more complex than just calculating the p-value of the indicator. This is done for the values of Fig. 9 and reported in Fig. 10. In order to allow visualisation

on a logarithmic scale (where zero-values are not representable), the p-values were saturated at 10^{-16}. These numerical examples show the high sensitivity of the indicators of cyclostationarity, even in the case of unknown periods, and their useful statistical properties.

7. Testing for non-Gaussianity against Gaussianity

The previous Sections 3-6 have focused on detecting the presence of cyclostationarity versus stationarity as the manifestation of a fault signature. A different route often followed in the literature is to model a fault signature as a departure from Gaussianity, yet under the implicitly stationary assumption. The problem can still be solved in light of the proposed likelihood indicators. Several connexions are drawn with other indicators used in the literature.

7.1. The GLR test and statistical threshold

The problem is to test a non-Gaussian stationary hypothesis against the Gaussian stationary one. Resorting to the GG PDF as a rather general way to capture non-Gaussian behaviours, the following proposition arrives as a direct consequence of the material introduced in Section 4.

Proposition 5. The likelihood indicator testing the generalized Gaussian stationary (GGS) hypothesis against the Gaussian stationary (GS) one is

$$I_{GGS/GS}(\mathbf{x}) = 2\frac{\mathcal{L}_{GGS}(\mathbf{x}) - \mathcal{L}_{GS}(\mathbf{x})}{L} = \ln\langle|x(n)|^2\rangle - \frac{2}{\hat{\beta}}\ln\langle|x(n)|^{\hat{\beta}}\rangle + 2C(c,2,\hat{\beta}) \tag{55}$$

with $\hat{\beta}$ the MLE of the shape parameter (obtained as explained in Appendix A) and $C(c,2,\hat{\beta})$ given in Eq. (34).

(See Eqs. (93) and (87) for the expressions of $\mathcal{L}_{GGS}(\mathbf{x})$ and $\mathcal{L}_{GS}(\mathbf{x})$, respectively.)

Under the Gaussian stationary assumption, the likelihood indicator $I_{GGS/GS}(\mathbf{x})$ is asymptotically distributed according to a 1-DOF chi-square (H_0 has 1 DOF (η) and H_1 has 2 DOFs (η,β)),

$$I_{GGCS/GGS}(N) \sim \frac{1}{L}\chi_1^2. \tag{56}$$

In the special situation where the shape parameter is known or happens to be the same for all signals to be compared, the additive constant C can be removed from the above indicator, which then simplifies in

$$I'_{GGS/GS}(\mathbf{x}) = 2\ln\left(\frac{\langle|x(n)|^2\rangle^{\frac{1}{2}}}{\langle|x(n)|^{\hat{\beta}}\rangle^{\frac{1}{\hat{\beta}}}}\right). \tag{57}$$

7.2. Connection with other indicators

Without surprise, the likelihood indicator based on the GG PDF allows the recovery of several other indicators as particular cases. By setting $\beta = 1$ in Eq. (57), $\exp\left(I'_{GGS/GS}(\mathbf{x})/2\right)$ is found proportional to the "sparsity measure" of Ref. [22], which is actually similar to the form factor used in electronics [59]. By using the fact that $X^{\beta} = e^{\beta \ln X} \to \beta \ln X$ for $X > 0$ and $\beta \to 0$, $\exp\left(I'_{GGS/GS}(\mathbf{x})/2\right)$ is found equal to the smoothness index of Ref. [23]. The kurtosis and the crest factor are similarly recovered by setting $\beta = 4$ and $\beta = \infty$, respectively, yet $\exp\left(-I'_{GGS/GS}(\mathbf{x})/2\right)$ is then considered instead of $\exp\left(I'_{GGS/GS}(\mathbf{x})/2\right)$. Obviously, many other CMs can thus be defined by changing the value of β in Eq. (57). An even greater flexibility is allowed by considering ratios of power means, say P_{β_1}/P_{β_0} with $P_{\beta} \doteq \langle|x(n)|^{\beta}\rangle^{1/\beta}$, by following Eq. (36). However the advantage of designing an CM from the proposed likelihood methodology should not be overlooked; in, particular, it is reminded here that:

(i) ratios of power means usually contain pre-defined values of β_1 and β_0 whereas in the proposed indicators there is the possibility to estimate the shape parameters from the data,
(ii) ratios of power means do not generally come with statistical thresholds,
(iii) no account of non-stationarity is taken in ratios of power means, contrary to the more general situation considered in Proposition 3 paper.

8. Testing for impulses with the Bernoulli-Gauss (cyclo)stationary model

Of particular interest is the case where machine faults manifest themselves as a series transients, typically as a result of brief impacts. Although this is often monitored by means of the kurtosis – which is known to be very sensitive to a few impulses – it is not necessarily the optimal approach. As a matter of fact, the kurtosis has been derived in Section 7.2 as

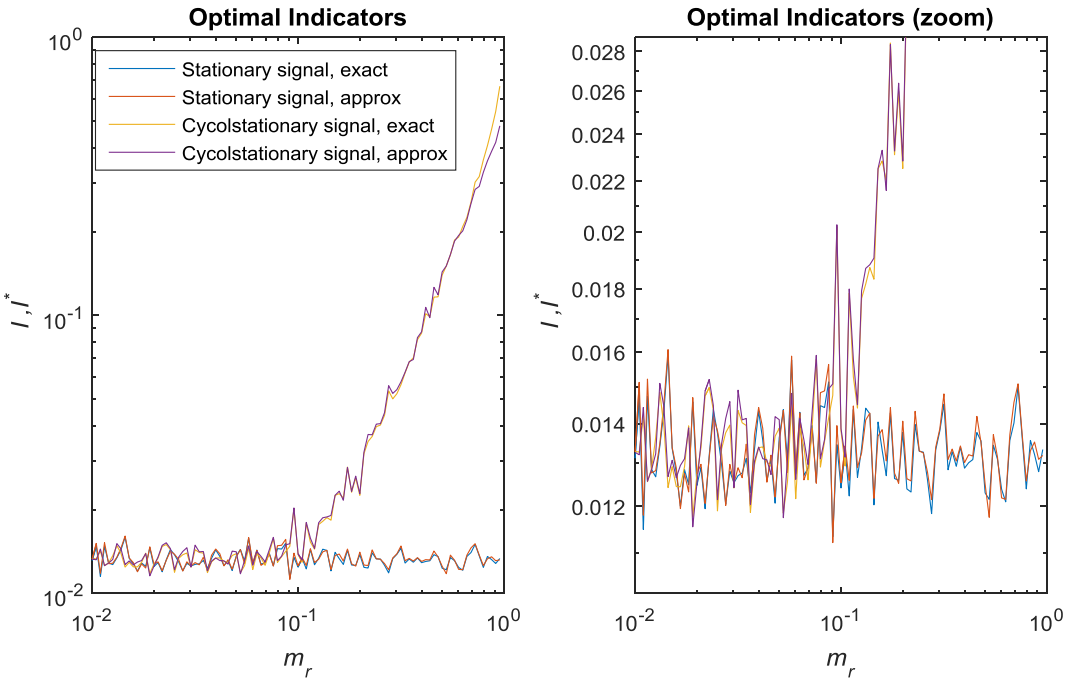

Fig. 9. Condition indicator $I_{GGCS/GGS}$ (exact) and $I^*_{GGCS/GGS}$ (approximate) for stationary and cyclostationary signals under the assumption of unknown cyclic period in the range $90 \leqslant N \leqslant 110$. The stationary signals are sampled from a generalized Gaussian distributions with $\eta = 3$ and $\beta = 1$, while the cyclostationary signals are generated with different levels of the non-stationarity modulation depth m_r.

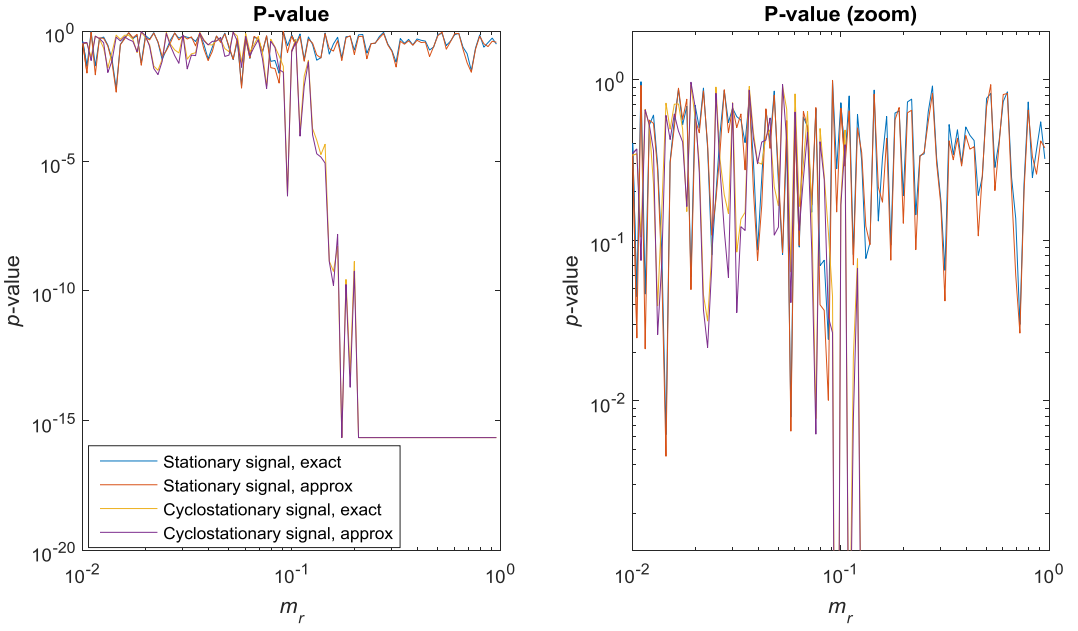

Fig. 10. P-value of the condition indicator $I_{GGCS/GGS}$ (exact) and $I^*_{GGCS/GGS}$ (approximate) for stationary and cyclostationary signals under the assumption of unknown cyclic period in the range $90 \leqslant N \leqslant 110$.

an approximation under the generalized Gaussian case, which is a good but not the best model for impulses. An interesting alternative is the Bernoulli-Gauss model in which the signal switches between two states: a "normal" state where it takes values according to a stationary Gaussian PDF and an "abnormal" state which corresponds to the occurrence of impulses. The model is first described in the stationary situation where impulses arrive without synchronization and next in the cyclostationary situation where they are repetitive with a given cycle.

8.1. Bernoulli-Gauss stationary model (BGS)

The Bernoulli-Gauss stationary model at time instant n reads

$$x(n) = x_0(n)(1-\zeta(n)) + x_1(n)\zeta(n) \tag{58}$$

where $x_0(n)$ and $x_1(n)$ are two centred stationary random variables, distributed with Gaussian PDFs with variances σ_0^2 and σ_1^2, respectively, with $\sigma_0^2 < \sigma_1^2$, and $\zeta(n)$ is a latent variable that takes values 0 or 1 according to a Bernoulli distribution with probability p. In other words, an impulse $x_1(n)$ will occur with probability $p \overset{\text{def}}{=} prob(\zeta(n)=1)$ and variance σ_1^2 at time n. This model is now tested against the assumption of a stationary Gaussian distribution. Since σ_0^2, σ_1^2, p and $\zeta(n)$ are all unknown, they have to be estimated by maximum likelihood in the GLRT. This is efficiently accomplished with the Expected Maximization (EM) algorithm, which is briefly summarised hereafter (see Appendix E for its derivation).

EM algorithm

Step 0 (q = 0): initialize $z^{[0]}(n)$ and $p^{[0]}$ (e.g. with random values)

Step q ($q \leftarrow q+1$):

i) estimate the variance components as

$$\sigma_0^{2[q]} = \frac{\sum_{n=0}^{L-1}(1-z^{[q]}(n))|x(n)|^2}{\sum_{n=0}^{L-1}(1-z^{[q]}(n))} \text{ and } \sigma_1^{2[q]} = \frac{\sum_{n=0}^{L-1}z^{[q]}(n)|x(n)|^2}{\sum_{n=0}^{L-1}z^{[q]}(n)}, \tag{59}$$

ii) estimate the posterior mean of the latent variable as

$$z^{[q]}(n) = \frac{1}{1+\frac{1-p^{[q]}}{p^{[q]}}\frac{\sigma_1^{[q]}}{\sigma_0^{[q]}}\exp\left(\frac{|x(n)|^2}{c}\left(\frac{1}{\sigma_1^{2[q]}}-\frac{1}{\sigma_0^{2[q]}}\right)\right)}, \tag{60}$$

with c = 2 if x is real c = 1 if it is complex,

iii) estimate the probability as

$$p^{[q]} = \frac{1}{L}\sum_{n=0}^{L-1} z^{[q]}(n). \tag{61}$$

Stop at convergence.

Proposition 6. The likelihood indicator testing the Bernoulli-Gauss stationary hypothesis against the Gaussian stationary one is

$$\begin{aligned} I_{BGS/GS}(\mathbf{x}) &= c\frac{\mathcal{L}_{BGS}(\mathbf{x}) - \mathcal{L}_{GS}(\mathbf{x})}{L} \\ &= \ln\left((1-\hat{p})\hat{\sigma}_0^2 + \hat{p}\hat{\sigma}_1^2\right) - \left((1-\hat{p})\ln\hat{\sigma}_0^2 + \hat{p}\ln\hat{\sigma}_1^2\right) + c.C_S(\hat{p},\mathbf{z}) \end{aligned} \tag{62}$$

where $\hat{p}$, $\hat{\sigma}_0^2$ an, $\hat{\sigma}_1^2$ stand for the MLEs of the model parameters, as returned at convergence of the EM algorithm (i.e. $\hat{p} = p^{[\infty]}$, $\hat{\sigma}_0^2 = \sigma_0^{2[\infty]}$ *and* $\hat{\sigma}_1^2 = \sigma_1^{2[\infty]}$), k = 2 when x is real and c = 1 when it is complex, $\mathbf{z} = \left[z^{[\infty]}(0),\ldots,z^{[\infty]}(L-1)\right]^T$, and

$$C_S(p,\mathbf{z}) = \langle H(z(n))\rangle - H(p), \tag{63}$$

$H(p) = -p\ln p - (1-p)\ln(1-p)$ the entropy associated with the Bernoulli law of probability p and $\langle\cdots\rangle$ the time averaging operator.

(See Eqs. (110) and (87) for the expressions of $\mathcal{L}_{BGS}(\mathbf{x})$ and $\mathcal{L}_{GS}(\mathbf{x})$, respectively.)

Referring to Section 2.5, indicator $I_{BGS/GS}$ is interpreted as the information conveyed by the envelope $z(n)$ of the signal.

Under the Gaussian stationary assumption, the likelihood indicator $I_{BGS/GS}(\mathbf{x})$ is asymptotically distributed according to a 2-DOF chi-square (H_0 has 1 DOF (σ_0^2) and H_1 has 3 DOFs (σ_0^2, σ_1^2 and p),

$$I_{BGS/GS}(\mathbf{x}) \sim \frac{c}{2L}\chi_2^2. \tag{64}$$

Several remarks are in order at this juncture. First, the same formulas apply to real and complex signals except for the estimation of the latent variable in Eq. (60). Second, it is noteworthy that the structure of indicator $I_{BGS/GS}$ is very similar to that of $I_{GCS/GS}$ (compare Eqs. (62) and (12)) where the time average operator $\langle\cdots\rangle$ is replaced by the weighted average $(1-\hat{p})X_0 + \hat{p}X_1$ with X_i, $i=1,2$ standing either for $\ln\hat{\sigma}_i^2$ or $\hat{\sigma}_i^2$. In particular, it is readily checked from Eq. (59) that $\ln\left((1-\hat{p})\hat{\sigma}_0^2 + \hat{p}\hat{\sigma}_1^2\right) = \ln\langle s^2(n)\rangle$ with $s^2(n)$ as defined in Eq. (13). Third, the posterior mean of the latent variable, $z(n)$, reflects the same information as the classical envelope of the signal. Hence it may be helpful to use it as a diagnostic tool *per se*.

8.2. Bernoulli-Gauss cyclostationary model (BGCS)

The Bernoulli-Gauss stationary model is easily modified to take into account the fact that impulses arrives repetitively with a given cycle. The only modification is to make p a periodic function of time, $p(n) = p(n+T)$. Therefore, Eq. (59) is unchanged, $p^{[q]}(n)$ is substituted for $p^{[q]}$ in Eq. (60), and Eq. (61) is replaced by

$$p^{[q]}(n) = \frac{1}{M}\sum_{k=0}^{M-1} z^{[q]}(n+kT). \tag{65}$$

Proposition 7. The likelihood indicator testing the Bernoulli-Gauss cyclostationary hypothesis against the Gaussian stationary one is

$$\begin{aligned} I_{BGCS/GS}(\mathbf{x}) &= c\frac{\mathcal{L}_{BGCS}(\mathbf{x}) - \mathcal{L}_{GS}(\mathbf{x})}{L} \\ &= \ln\left((1-\bar{p})\hat{\sigma}_0^2 + \bar{p}\hat{\sigma}_1^2\right) - \left((1-\bar{p})\ln\hat{\sigma}_0^2 + \bar{p}\ln\hat{\sigma}_1^2\right) + c.C_{CS}(\bar{\mathbf{p}}, \mathbf{z}) \end{aligned} \tag{66}$$

with

$$C_{CS}(\mathbf{p}, \mathbf{z}) = \langle H(z(n))\rangle - \langle H(p(n))\rangle \tag{67}$$

and $\bar{p} = \frac{1}{N}\sum_{n=0}^{N-1}\hat{p}(n)$.

(See Eqs. (111) and (87) for the expressions of $\mathcal{L}_{BGCS}(\mathbf{x})$ and $\mathcal{L}_{GS}(\mathbf{x})$, respectively.)

Under the Gaussian stationary assumption, the likelihood indicator $I_{BGCS/GS}(\mathbf{x})$ is asymptotically distributed according to a $N+1$ DOF chi-square (H_0 has 1 DOF (σ_0^2) and H_1 has $N+2$ DOFs (σ_0^2, σ_1^2 and $p(n)$, $n = 0,\ldots,N-1$)),

$$I_{BGS/GS}(\mathbf{x}) \sim \frac{c}{2L}\chi^2_{N+1}. \tag{68}$$

8.3. Testing cyclostationarity against stationarity under the Bernoulli-Gauss hypothesis

In practice, the Bernoulli-Gauss stationary model is found to capture very well the statistical behaviour of repetitive impulses, even if they are cyclostationary. The question then arises whether the refinement brought by the cyclostationary model is really necessary. Since the two models are nested, (the Bernoulli-Gauss cyclostationary model accepts the Bernoulli-Gauss stationary one as a particular case), they can be easily tested.

Proposition 8. The likelihood indicator testing the Bernoulli-Gauss cyclostationary hypothesis against the Bernoulli-Gauss stationary one is

$$I_{BGCS/BGS}(\mathbf{x}) = I_{BGCS/GS}(\mathbf{x}) - I_{BGS/GS}(\mathbf{x}) \tag{69}$$

with $I_{BGCS/GS}(\mathbf{x})$ and $I_{BGS/GS}(\mathbf{x})$ as given in Eqs. (66) and (62), respectively.

Using Wilk's theorem, the likelihood indicator $I_{BGCS/BGS}(\mathbf{x})$ is asymptotically distributed according to a $N-1$ DOF chi-square under H_0 (H_0 has 3 DOFs and H_1 has $N+2$ DOFs),

$$I_{BGCS/BGS}(\mathbf{x}) \sim \frac{c}{2L}\chi^2_{N-1}. \tag{70}$$

The behaviour of the CMs $I_{BGS/GS}(\mathbf{x})$, $I_{BGCS/GS}(\mathbf{x})$ and $I_{BGCS/BGS}(\mathbf{x})$ is illustrated in Figs. 11 and 12 when applied on Bernoulli-Gauss stationary and cyclostationary processes with length $L = 10^5$, $p = 0.01$, $\sigma_0^2 = 1$, and σ_1^2 ranging from 1 to 100. It is seen that indicator $I_{BGS/GS}(\mathbf{x})$ detects the Bernoulli-Gauss property as soon as $\sigma_1^2/\sigma_0^2 > 2$ (a lower bound that obviously depends on L) irrespectively of whether the signal is actually stationary or cyclostationary. Indicator $I_{BGCS/GS}(\mathbf{x})$ correctly detects a Bernoulli-Gauss cyclostationary property from about the same ratio of variances, yet it wrongly detects a Bernoulli-Gauss stationary property for $\sigma_1^2/\sigma_0^2 > 3$. This is fixed by inspecting indicator $I_{BGCS/BGS}(\mathbf{x})$ which always stays below the statistical threshold in the stationary case and quickly exceeds it in the cyclostationary case.

9. Discussion on non-Gaussian versus non-stationary indicators

It is largely admitted that the manifestation of faults in rotating machines modify the vibration signals through an "abnormal" increase of non-stationarity or of non-Gaussianity. A typical example is that of local transients (typically due to repeated contacts with a damaged element), which are treated in the literature as either non-stationary or non-Gaussian (sometimes called "nonlinear") signatures. Interestingly, both points of view reflect the fact that the information about an abnormal event is evidenced by a departure from stationarity and/or from Gaussianity in the signals.

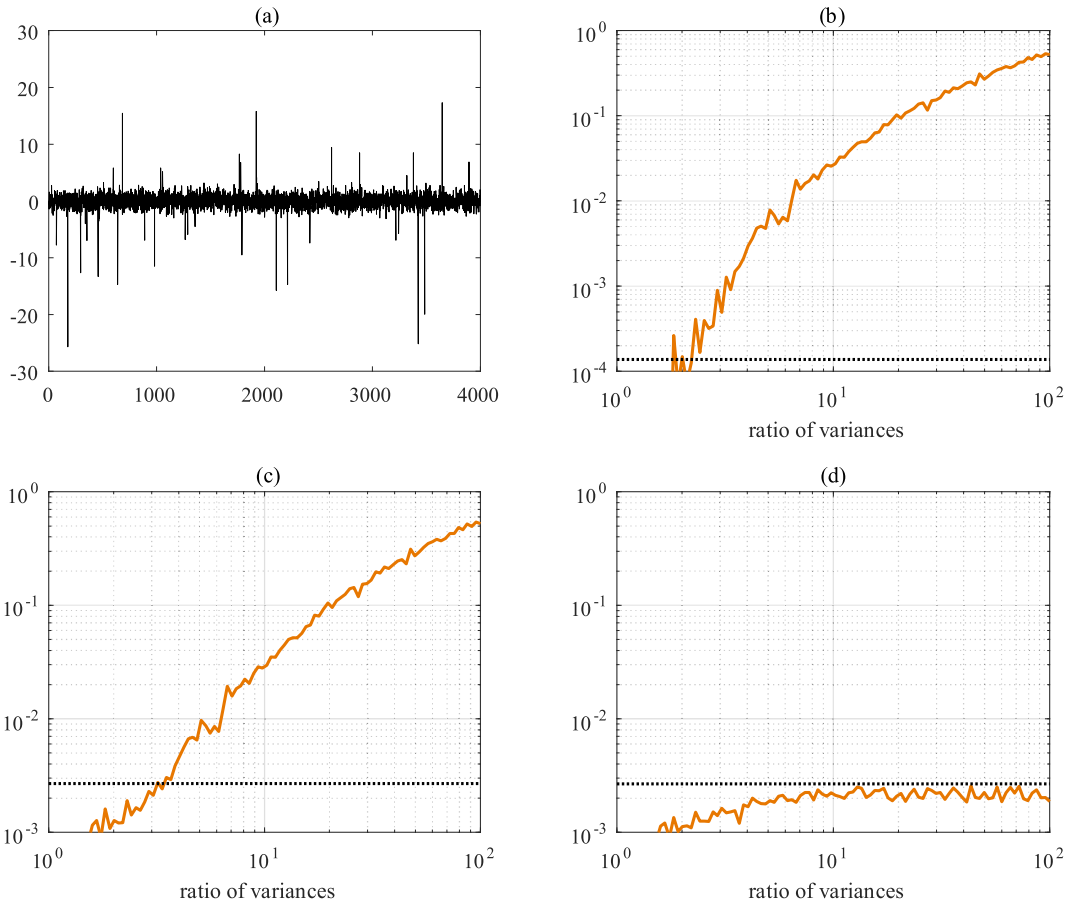

Fig. 11. Simulations of Bernoulli-Gauss stationary process with $L = 10^5$, $p = 0.01$, $\sigma_0^2 = 1$, and σ_1^2 ranging from 1 to 100. (a) One realization of the process shown over 4000 samples. Evolutions of indicators (b) $I_{BGS/GS}(\mathbf{x})$, (c) $I_{BGCS/GS}(\mathbf{x})$ and (d) $I_{BGCS/BGS}(\mathbf{x})$ as functions of ratio σ_1^2/σ_0^2 together with their statistical thresholds at significance level at 0.001.

Since non-stationarity and non-Gaussianity are well defined and non-interchangeable concepts in signal processing, it might seem surprising that some indicators dedicated to non-Gaussianity also detect non-stationary signatures, as typically discussed for the kurtosis in Section 3.2. This raises two questions:

(1) in which sense are the two visions – non-stationary and non-Gaussianity – equivalent?
(2) are indicators designed for non-Gaussianity detection also optimal to detect non-stationarity?[3]

The formalism proposed in this paper happens to provide accurate answers to these two questions.

9.1. In which sense are non-stationarity and non-Gaussianity equivalent?

It is first shown in this subsection that a non-stationary Gaussian stochastic process is equivalent to a non-Gaussian leptokurtic (i.e. with positive excess kurtosis) stochastic process after averaging its PDF in time. The discussion is first limited to the real non-stationary Gaussian case, because of its practical importance, and then extended to a general non-stationary PDF. Without loss of generality, all processes are assumed centred – e.g. as obtained after removal of the mean value.

(a) *The Gaussian case*

[3] Note that the inverse question is also legitimate: "are indicators designed for non-stationarity detection also optimal to detect non-Gaussianity?" However, since the former usually have a higher complexity than the latter, the substitution has a limited practical advantage. In addition, the answer is obviously negative in the special case of cyclostationary signals since the notion of a cyclic period is irrelevant after stationarization.

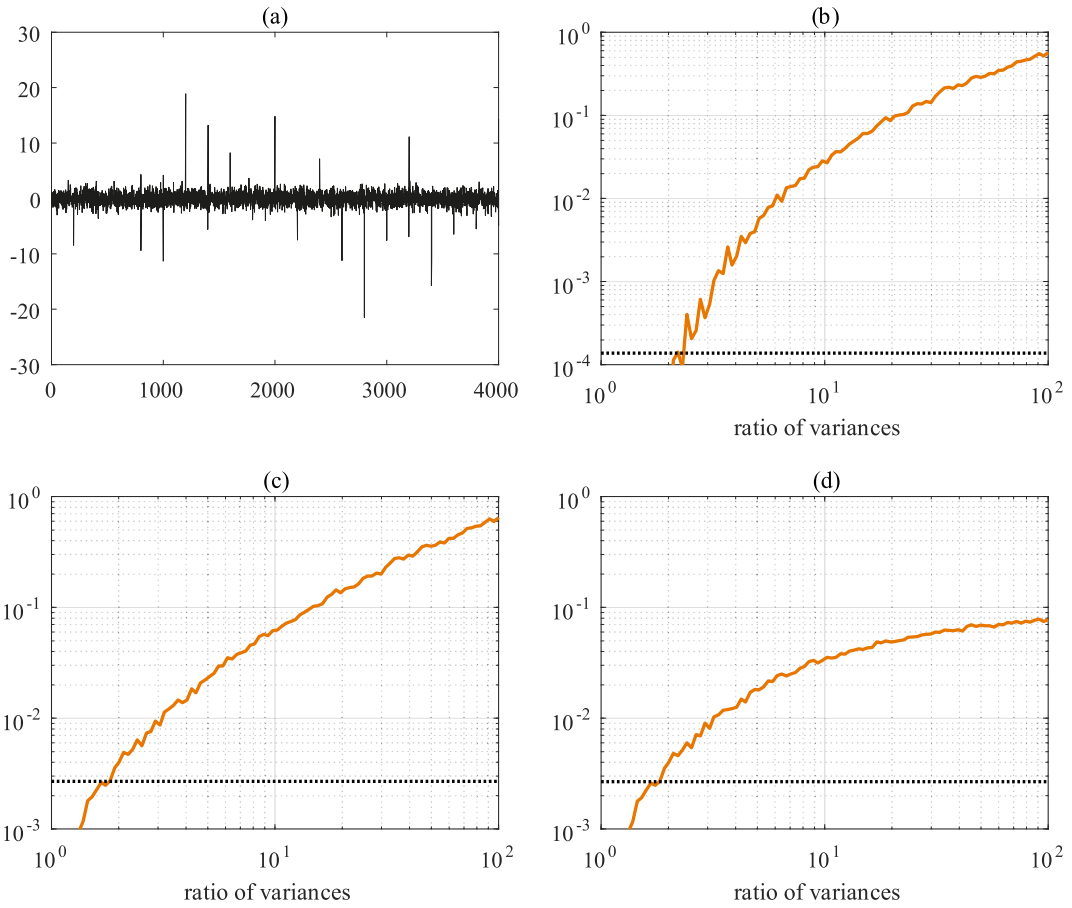

Fig. 12. Simulations of Bernoulli-Gauss cyclostationary process with cyclic period $N = 200$, $L = 10^5$, $p = 0.01$, $\sigma_0^2 = 1$, and σ_1^2 ranging from 1 to 100. (a) One realization of the process shown over 4000 samples. Evolutions of indicators (b) $I_{BGS/GS}(\mathbf{x})$, (c) $I_{BGCS/GS}(\mathbf{x})$ and (d) $I_{BGCS/BGS}(\mathbf{x})$ as functions of ratio σ_1^2/σ_0^2 together with their statistical thresholds at significance level at 0.001.

The formal way to define a non-stationary Gaussian process is by introducing a time-varying variance, as done in Section 3 (see Eq. (11)). The corresponding time-varying PDF is to be understood with respect to an ensemble of realizations of the stochastic process: each new experience outputs a new random carrier, yet under the same (deterministic) variance modulation. At a given (discrete) time instant n, the set of all generated signals follows a Gaussian law with variance $\sigma^2(n)$, i.e.

$$x(n) \sim p_X(x;n) = \frac{1}{\sqrt{2\pi}\sigma(n)} e^{-\frac{|x|^2}{2\sigma^2(n)}}, \ x \in \mathbb{R} \tag{71}$$

which is a function of time n. Fig. 13(a) illustrates the meaning of this PDF and, in particular, that it represents the "vertical" distribution of values at a fixed time instant.

The usual engineering practice is however to consider the distribution of values of one realization of the process along the time axis, of which an estimator is returned by the usual time histogram, as illustrated in Fig. 13(b). Strictly speaking, this is fundamentally different from the former definition of the PDF and equality of the two distributions only applies in the special case of stationary and ergodic processes. The time histogram actually corresponds to the time-average of $p_X(x;n)$,

$$\bar{p}_X(x) = \langle p_X(x,n) \rangle \stackrel{def}{=} \lim_{L\to\infty} \frac{1}{L} \sum_{n=0}^{L-1} p_X(x,n). \tag{72}$$

This is the PDF that would follow the non-stationary process after shuffling randomly all its samples so as to destroy non-stationary patterns, a practice known as "stationarization" – see Fig. 13(c)–(d). After stationarization, each sample of the signal is distributed according the mixture of Gaussian (72), that its PDF is drawn randomly in the family of Gaussians (71) with variances $\sigma^2(n)$, $n = 0,\ldots,L-1$, with an uniform probability $1/L$.

Another insightful interpretation of the PDF $\bar{p}_X(x)$ is that of a stationary Gaussian process whose variance is itself a random variable with PDF $p(\sigma^2)$. It is proved in the appendix that

$$\bar{p}_X(x) = \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|x|^2}{2\sigma^2}} p(\sigma^2) d\sigma^2 \tag{73}$$

with

$$p(\sigma^2) = \sum_i |\dot{\sigma}_i^2|^{-1} \tag{74}$$

where $\dot{\sigma}_i^2$ stands for the time derivative of the variance $\sigma^2(t)$ – re-expressed as a continuous function of time – evaluated at all time instants t_i such that $\sigma^2(t_i)$ equals a given value σ^2. This is recognized as a scaled-mixture of Gaussians model, i.e. a Gaussian distribution whose variance is itself a random stationary variable. Fig. 14 displays the shape of $p_X(\sigma^2)$ for a sinusoidal modulation $\sigma^2(t) = 1 + B\sin(2\pi t/T)$ (continuous time), which corresponds to the well-known "inverse sinusoidal" PDF,

$$p(\sigma^2) = \frac{T}{\pi\sqrt{B^2 - (\sigma^2 - 1)^2}}, \ |B| < 1, \quad 1 - B \leqslant \sigma^2 \leqslant 1 + B. \tag{75}$$

It is further proved in the appendix that Eq. (73) defines a leptokurtic distribution, i.e. one with a kurtosis greater than of a Gaussian (also including the cases where the kurtosis is infinite).

Proposition 9. To each non-stationary Gaussian process corresponds an equivalent stationary non-Gaussian leptokurtic process with identical time-averaged moments whose PDF is given by Eqs. (73) and (74).

Proposition 9 establishes the equivalence between two different manifestations of the same information, as a departure from the canonical stationary Gaussian case: one with a deterministic time-varying variance and the other one with a stochastic stationary variance. However, care should be taken not to conclude that the stochastic processes are identical: this is not true in general and equivalence holds only for their time-averaged moments. It should also be noted that the proposition assigns an equivalent stationary non-Gaussian leptokurtic process to a given non-stationary Gaussian process, but not the inverse. It is easily checked that to a given stationary non-Gaussian leptokurtic process may corresponds several non-stationary Gaussian processes (e.g. the PDF in Fig. 13(d) may have been produced by many other time-varying PDFs different from that in Fig. 13(a)).

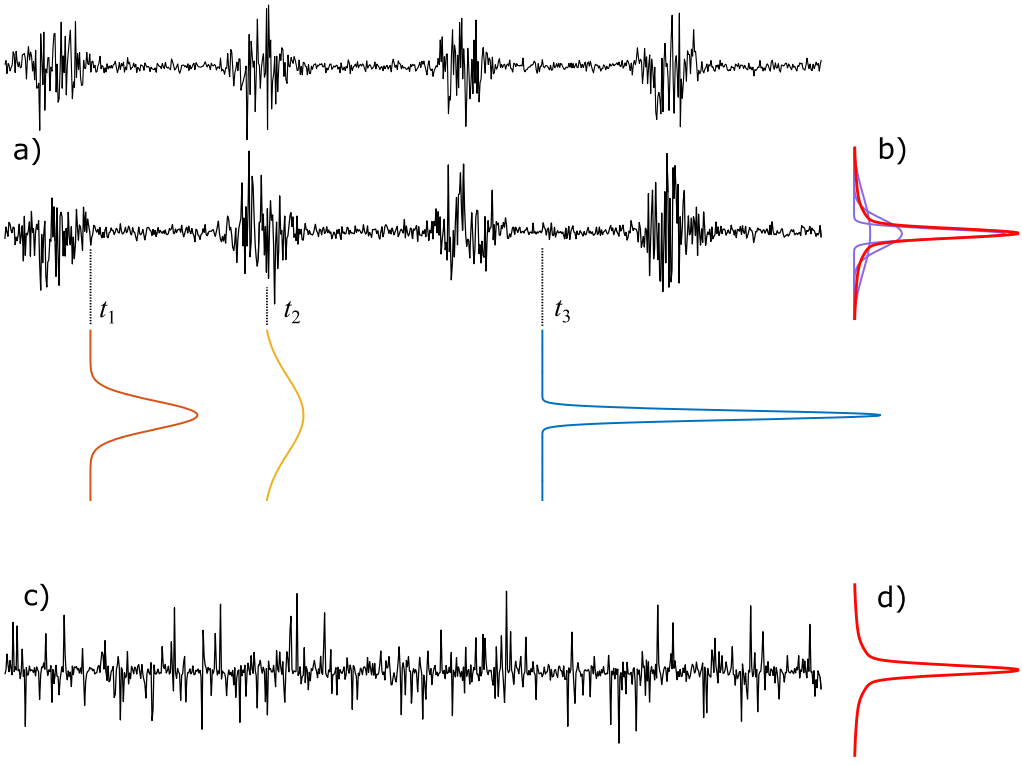

Fig. 13. (a) Example of two realizations of a non-stationary process generated by a time-varying Gaussian PDF whose shape is displayed at time instants t_1, t_2, and t_3. (b) The corresponding histogram obtained by averaging the time-varying Gaussian PDF. (c) The equivalent stationary process generated by the non-Gaussian PDF (c).

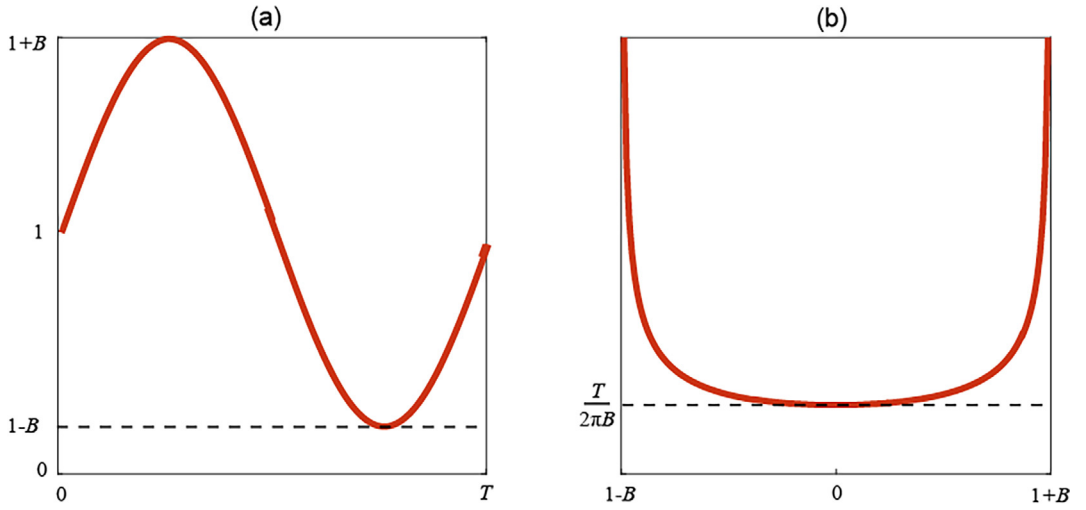

Fig. 14. (a) Sinusoidally varying variance $\sigma^2(t)$ over one period T and (b) the corresponding PDF $p(\sigma^2)$.

(b) *The general (non-Gaussian) case*

Eventually, the above results are easily extended to a general non-stationary PDF.

Proposition 10. To each non-stationary process with PDF $p_X(x;\theta(t))$, with continuous time-dependent parameter $\theta(t)$, $t \in \mathbb{R}$, corresponds an equivalent stationary process with identical time-averaged moments whose PDF is given by

$$\bar{p}_X(x) = \int p_X(x;\theta)p(\theta)d\theta \tag{76}$$

where

$$p(\theta) = \sum_i |\dot{\theta}_i|^{-1} \tag{77}$$

such that $\theta(t_i) = \theta$, with $\dot{\theta}_i$ the time-derivative of $\theta(t)$. If $C_4(t)$ denotes the time-dependent kurtosis of $p_X(x;\theta(t))$, then $\bar{p}_X(x)$ has a kurtosis $\kappa_4 \geqslant \langle C_4(t)\rangle$.

9.2. Are indicators designed for detecting non-Gaussianity also optimal to detect non-stationarity?

In light of the conclusion of the previous subsection, one may wonder whether the equivalence also holds for indicators and, after all, if non-Gaussian indicators are versatile. The question is particularly relevant since the later are probably easier to handle and are, de facto, found more ubiquitous in the literature. It is proved in this subsection that the answer is unfortunately negative.

The problem is formulated by means of the likelihood ratio introduced in Section 2.2. If a non-stationary indicator is better than a non-Gaussian one, then its assigned probability of detection must be the largest under the non-stationary assumption. Let's denote H_{NS} the non-stationary assumption and H_S the stationary one. The problem then boils down to checking the sign of the log likelihood ratio (taking the limit when the signal length tends to infinity),

$$\ln\frac{p_X(\mathbf{x}|H_{NS})}{p_X(\mathbf{x}|H_S)} \tag{78}$$

on which the definition of indicators is rooted. In order to account for the actual PDF of the signal samples $\mathbf{x} = [x(0), x(1), \cdot sx(L-1)]^T$, the expected value of the log-likelihood ratio with respect to $p_X(\mathbf{x}|H_{NS})$ is taken, i.e.

$$\mathbb{E}\left\{\ln\frac{p_X(\mathbf{x}|H_{NS})}{p_X(\mathbf{x}|H_S)}\right\} = \int \ln\frac{p_X(\mathbf{x}|H_{NS})}{p_X(\mathbf{x}|H_S)} p_X(\mathbf{x}|H_{NS})d\mathbf{x}. \tag{79}$$

This is recognized as the Kullback-Leibler divergence between the two PDFs $p_X(\mathbf{x}|H_{NS})$ and $p_X(\mathbf{x}|H_S)$, a quantity which is always nonnegative as mentioned in Section 2.5 [49]. Therefore, if one designs a scalar indicator $I_{NS} = \langle \ln p_X(\mathbf{x}|H_{NS})/p_X(\mathbf{x}|H_0)\rangle$ to detect non-stationarity and another indicator $I_S = \langle \ln p_X(\mathbf{x}|H_S)/p_X(\mathbf{x}|H_0)\rangle$ based on the stationarized PDF $p_X(\mathbf{x}|H_S)$ to detect non-Gaussianity, both referenced to an arbitrary assumption H_0 (e.g. stationary Gaussian), it will hold that

$$\mathbb{E}\{I_{NS}\} - \mathbb{E}\{I_S\} = \left\langle \mathbb{E}\left\{ \ln \frac{p_X(\mathbf{x}|H_{NS})}{p_X(\mathbf{x}|H_S)} \right\} \right\rangle \geqslant 0 \tag{80}$$

where the ensemble average $\mathbb{E}$ is with respect to all realizations of the non-stationary stochastic process $\mathbf{x}$. Therefore, the following result holds:

Proposition 11. When the signal is actually non-stationary, non-Gaussian indicators cannot be better than non-stationary ones.

It is conjectured that Proposition 11 also holds for the GLRT – where parameters are replaced by their maximum likelihood estimates – at least asymptotically.

10. Summary and discussion

10.1. Summary tables

A summary of the CMs found in the previous sections is given in Tables 2 and 3. Table 2 collects indicators linked to all combinations of stationary, cyclostationary, Gaussian and generalized Gaussian hypotheses, whereas Table 3 is about the combinations of stationary, cyclostationary, Gaussian and Bernoulli-Gauss hypotheses. The last row in each Table shows the existence of a "closure relationship" between the competing hypotheses. This is also illustrated in Fig. 15 and it will be used in the next subsection to define a strategy for testing nested models.

The tables show analogies among the different indicators. It is seen that they all involve the difference between the logarithm of the time average and the time average of the logarithm. In the case of the indicators of Table 2, this difference is systematically applied on the "envelope" of the signal, whether it is $s^2(n)$ for the Gaussian case or the equivalent power envelope $\langle s_n^\beta \rangle^{2/\beta}$ for the generalized Gaussian case. These envelopes all have a squared-magnitude unit, but under a non-Gaussian hypothesis a "β-weighting" of the samples is included. The latter has the effect of reducing the impact of the extreme values of the signal in highly impulsive cases (low β). As discussed in Section 2.5, the concept of envelope is directly linked to the information conveyed by the signal. The notion of an envelope is less explicit for the Bernoulli-Gauss model, although it is contained in the posterior mean of the latent variable as mentioned in Section 8.1. More prevalent is the information conveyed by switches of the variances between two possible values.

10.2. Testing multiple nested models

The CMs in Tables 2 and 3 may be used to test a set of nested models. As illustrated in Fig. 15(a), the GGCS assumption corresponds to the most general model which contains GCS, GS and GGS as particular cases. In turn, GCS and GGS both contains GS. The first diagonal in the diamond separates cyclostationary models from stationary ones. The second diagonal separates generalized Gaussian models from Gaussian ones. This suggests that, given an unknown signal, several alternative models can be tested. For instance, by following the left branch, one first tests the GS model against the GCS one by calculating $I_{GCS/GS}$ and comparing it to its threshold. If the GS hypothesis is rejected, one next tests the GCS model against the GGCS one. Alternatively, by following the right branch, one first tests the GS model against the GGS one and, if successful, next tests the GGS model against the GGCS one. A last possibility is to follow the middle branch and directly test GS against GGCS. There are overall five different possible tests, but only three are independent as evidenced by the two closure relationships in the bottom of Fig. 15(a). Obviously, only one out of the five models is true. As exemplified in Section 11.1, it might happen that more than one model are accepted after running the statistical tests (this would happen when two wrong hypotheses are compared); in this case the simplest model should be retained.

Table 2
Condition indicators under (cyclo)stationary (generalized) Gaussian hypotheses.

Hypothesis test:	Indicator: (C stands for an additive constant)
Gaussian Cyclostationary (GCS) versus Gaussian Stationary (GS)	$I_{GCS/GS} = \ln\langle s^2(n)\rangle - \langle \ln s^2(n)\rangle$ with $s^2(n) = \frac{1}{K}\sum_{k=0}^{K-1} \lvert x(n+kN)\rvert^2$
Generalised Gaussian Cyclostationary (GGCS) versus Generalised Gaussian Stationary (GGS)	$I_{GGCS/GGS} = \ln\left(\langle s_n^{\hat{\beta}_0}\rangle^{2/\hat{\beta}_0}\right) - \langle \ln\left(s_n^{\hat{\beta}_1}\right)^{2/\hat{\beta}_1}\rangle + C$ with $s^\beta(n) = \frac{1}{K}\sum_{k=0}^{K-1} \lvert x(n+kN)\rvert^\beta$
Generalised Gaussian Cyclostationary (GGCS) versus Gaussian Cyclostationary (GCS)	$I_{GGCS/GCS} = \langle \ln s^2(n)\rangle - \langle \ln\left(s_n^{\hat{\beta}}\right)^{2/\hat{\beta}}\rangle + C$
Generalised Gaussian Stationary (GGS) versus Gaussian Stationary (GS)	$I_{GGS/GS} = \langle \ln \lvert x(n)\rvert^2\rangle - \ln\langle \lvert x(n)\rvert^{\hat{\beta}}\rangle^{2/\hat{\beta}} + C$
Generalised Gaussian Cyclostationary (GGCS) versus Gaussian Stationary (GS)	$I_{GGCS/GS} = I_{GGCS/GCS} + I_{GCS/GS}$ $= I_{GGCS/GGS} + I_{GGS/GS}$

Table 3
Condition indicators under (cyclo)stationary (Bernoulli) Gauss hypotheses.

Hypothesis test:	Indicator: (C stands for an additive constant)
Bernoulli-Gauss Stationary (BGS) versus Gaussian Stationary (GS)	$I_{BGS/GS} = \ln\left((1-\hat{p})\hat{\sigma}_0^2 + \hat{p}\hat{\sigma}_1^2\right) - \left((1-\hat{p})\ln\hat{\sigma}_0^2 + \hat{p}\ln\hat{\sigma}_1^2\right) + C$
Bernoulli-Gauss Cyclostationary (BGCS) versus Gaussian Stationary (GS)	$I_{BGCS/GS} = \ln\left((1-\bar{p})\bar{\sigma}_0^2 + \bar{p}\bar{\sigma}_1^2\right) - \left((1-\bar{p})\ln\bar{\sigma}_0^2 + \bar{p}\ln\bar{\sigma}_1^2\right) + C$
Bernoulli-Gauss Cyclostationary (BGCS) versus Bernoulli-Gauss Stationary (BGS)	$I_{BGCS/BGS} = I_{BGCS/GS} - I_{BGS/GS}$
Bernoulli-Gauss Cyclostationary (BGCS) versus Gaussian Cyclostationary (GCS)	$I_{BGCS/GCS} = I_{BGCS/GS} - I_{GCS/GS}$

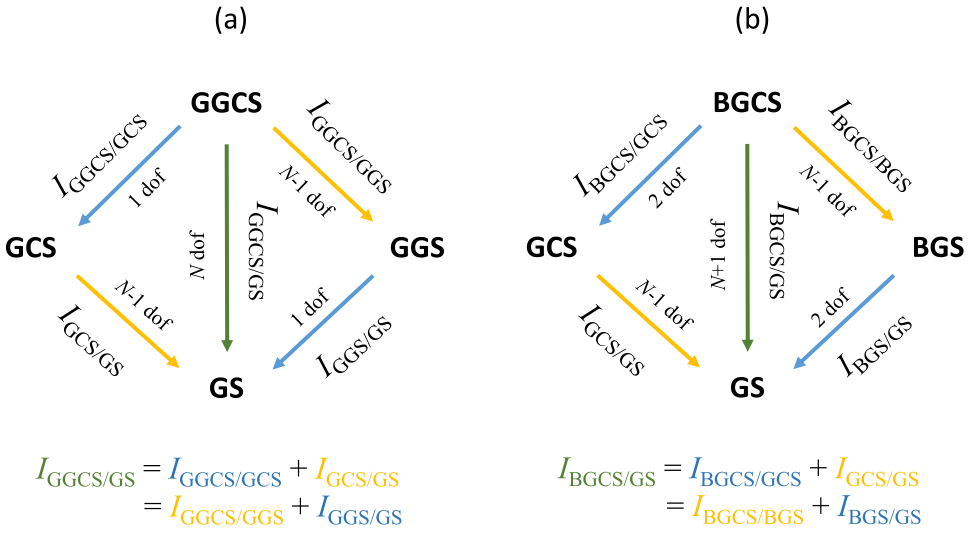

Fig. 15. Principle for testing nested models. Each arrow shows an existing inclusion between two competing models: the model at the arrow tip is contained in the model at the arrow tail. Orange arrows separate cyclostationary models from stationary models. Blue arrows separate non-Gaussian models from Gaussian models. a) (Cyclo)stationary (generalized) Gaussian models. b) (Cyclo)stationary (Bernoulli) Gauss models. Also indicated are the numbers of degrees of freedom (DOF) separating each model and used in the statistical tests. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Similarly, the Bernoulli-Gauss PDF can be used in place of the generalized Gaussian to model non-Gaussianity, thus providing similar imbricated tests. This is illustrated in Fig. 15(b). Although Fig. 15(a) and (b) share the lower left branch GCS/GS, it is emphasized that the two families of models based on BG and GG PDFs are not nested and it is therefore not possible to test one against the other.

11. Experimental tests

11.1. Analysis of a vibration signal with presence of a bearing fault

This subsection illustrates the use of the proposed CMs on a vibration signal measured on the casing of a gearbox in the presence of a ball fault in one of the bearings. The sampling frequency is 48 kHz and the signal duration is 1.4 s. The signal was first pre-processed in order to remove the effect of masking vibrations due to gears and to meet the assumption of whiteness required for the computation of the CMs (see Section 2.3). Although this can be done in various ways [56], these two objectives were achieved at once by applying a maximum-kurtosis blind deconvolution algorithm inspired from Ref. [57] (another plausible approach is to simply whiten the data by spectral equalization or cepstral editing [42]). Fig. 16(a) and (b) compares the signal before and after blind deconvolution, the sample kurtosis then increasing from 2.9 to 42.3. Next, the exact fault period was retrieved by using the procedure described in Section 7. Since the theoretical fault period is about 930 samples (corresponding to twice the ball spin frequency), the cyclostationary indicators $I_{GCS/GS}$, $I_{GGCS/GGS}$, $I_{BGCS/GS}$ and $I_{BGCS/BGS}$ were computed for a range of values between $N_{\min} = 870$ and $N_{\max} = 1000$. The results displayed in Fig. 17 show a maximum reached by the indicators at about $N = 931.2$. Based on this estimate, all indicators of Tables 2 and 3 were eventually computed. Their values are listed in Fig. 18 together with their statistical thresholds at significance level $\alpha = 0.001$. The CDF in Eq. (54) was simply approximated by the CFD of a chi-squared with probability α/K, with $K = N_{\max} - N_{\min} = 130$ the number of tested periods in the range $[N_{\min}; N_{\max}]$. It is seen that all null hypotheses are rejected and that finally the most complex models, GGCS or BGCS, are selected. As mentioned before, it is not possible to further test the GGCS model against the BGCS since they are not nested. They are therefore two plausible alternative models. The

estimated exponent in the GGCS model, $\hat{\beta} = 1.3$, indicates a slightly leptokurtic PDF. The ratio of variances in the BGCS model is estimated as $\hat{\sigma}_1^2/\hat{\sigma}_0^2 = 62.8$. The estimated posterior mean of the latent variable, $z(n)$, displayed in Fig. 16(c), clearly indicates a non-stationary behaviour. It provides useful diagnostic information by capturing all fault transients, even in time intervals where they are hardly seen after blind deconvolution.

The periodically time-varying standard deviations $\hat{\sigma}_{GCS}(n)$ and $\hat{\sigma}_{GGCS}(n)$ of the GCS and GGCS models are compared in Fig. 19(a) and (b). The former seems overestimated, probably due to some extreme values that are not well accounted for in the Gaussian model. This is also compared with the periodically time-varying standard deviation of the BGCS model, $\hat{\sigma}_{BGCS}(n)$ in Fig. 19(d), obtained by the square root of

$$\hat{\sigma}^2(n) = \hat{p}(n)\hat{\sigma}_1^2 + (1 - \hat{p}(n))\hat{\sigma}_0^2 \tag{81}$$

with $\hat{p}(n)$ the periodically-time varying probability shown in Fig. 19(c). A moderate difference in shape and in magnitude is seen between $\hat{\sigma}_{BGCS}(n)$ and $\hat{\sigma}_{GGCS}(n)$. One advantage of the BGCS model over the GGCS model is to provide a decomposition of the variance into a stationary part and a transient part.

Finally, the same experiment was repeated, yet with an arbitrary period of $N = 978$ samples. This is to illustrate how the CMs behave when the tested cyclic frequency does not match a potential fault frequency. The estimated standard deviations of the GCS, GGCS and BGCS models are displayed in Fig. 20. It is seen that the last two are nearly constant, as expected in this case. However, the standard deviation of the GCS model shows high estimation errors due to the fact that the signal hardly

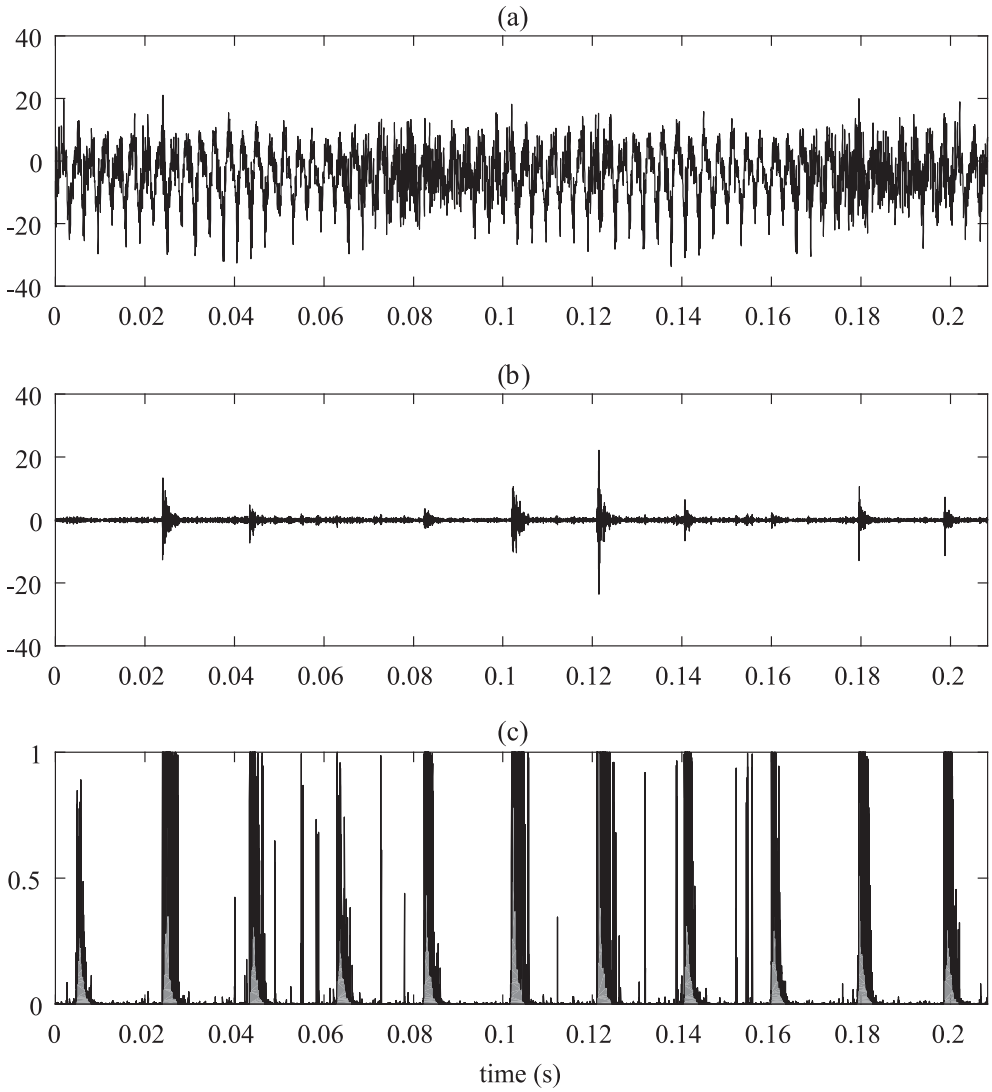

Fig. 16. (a) Measured vibration signal ($\kappa_4 = 2.9$). (b) Signal after blind deconvolution ($\kappa_4 = 42.3$). (c) Posterior mean of latent variable, $z(n)$.

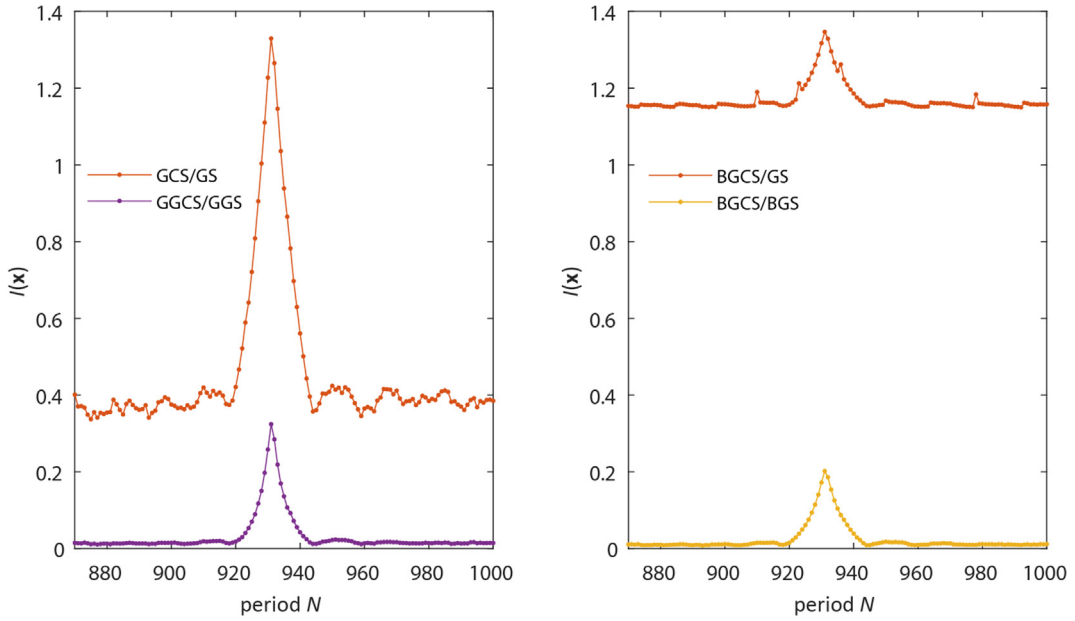

Fig. 17. Cycle retrieval by searching for the maximum values of the cyclostationary indicators (a) $I_{GCS/GS}(\mathbf{x},N)$, $I_{GGCS/GGS}(\mathbf{x},N)$ and (b) $I_{BGCS/GS}(\mathbf{x},N)$, $I_{BGCS/BGS}(\mathbf{x},N)$ computed for a range of values of the period N.

agrees with the underlying model. The values taken by the CMs are displayed in Fig. 21 together with the statistical thresholds at significance level $\alpha = 0.001$. It is seen that the two models, GGCS and GGS, are accepted by the statistical tests in Fig. 21(a). However, since GGCS is rejected when compared to GGS, the latter only should be ultimately retained. The fact that the GGCS succeeded the statistical tests against GCS (and GCS against GS) is because these two hypotheses were actually wrong. The same observation holds for the statistical tests in Fig. 21(b): the BGS model should ultimately be accepted, although BGCS succeeded against GS because these two hypotheses were wrong. As illustrated by this example, the conclusion is to retain the simplest of all models when more than one model are accepted. Here, they correctly correspond to the GGS (with $\hat{\sigma}_1^2/\hat{\sigma}_0^2 = 73.2$) and BGS (with $\hat{\beta} = 0.7$) models (two possible alternatives that cannot be tested against each other) which try to explain the presence of the fault transients in the signal when the tested cyclic frequency is not the good one.

11.2. Illustration of the condition indicators for trending

In this subsection the proposed technique is applied to the IMS bearing prognostic dataset [58]. The dataset related to an outer race fault is investigated (Campaign 2, Channel 1). The BPFO is expected for this dataset to appear at 236 Hz, corresponding to a normalized period of $N = 84.6$ samples. The evolution of the kurtosis and RMS values are displayed in Fig. 22 as functions of the record number in the experiment; these indicators have been used in several research works dealing with this dataset, for instance to set up Remaining Useful Life models. The values of all indicators in Tables 2 and 3 have been calculated using the procedure proposed in this paper. The data were first pre-processed as in the previous subsection

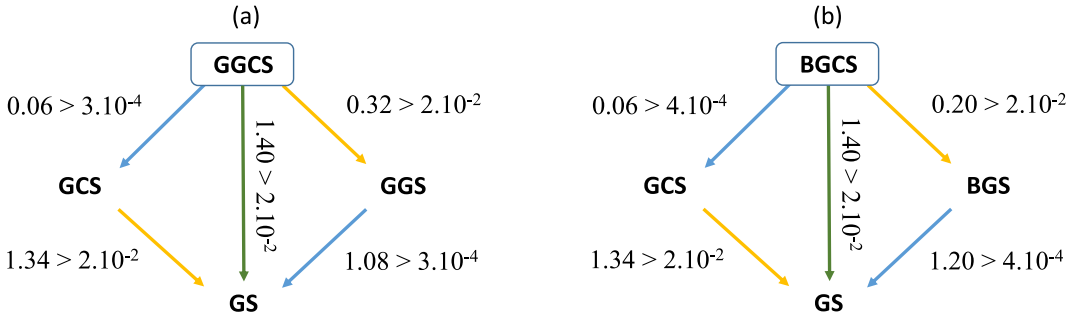

Fig. 18. Values of the indicators for testing between competing models together with their statistical thresholds at significance level 0.001, assuming the cyclic period of a ball fault, $N = 931.2$. (a) (Cyclo)stationary (generalized) Gaussian models. (b) (Cyclo)stationary (Bernoulli) Gauss models. The retained models are shown in boxes.

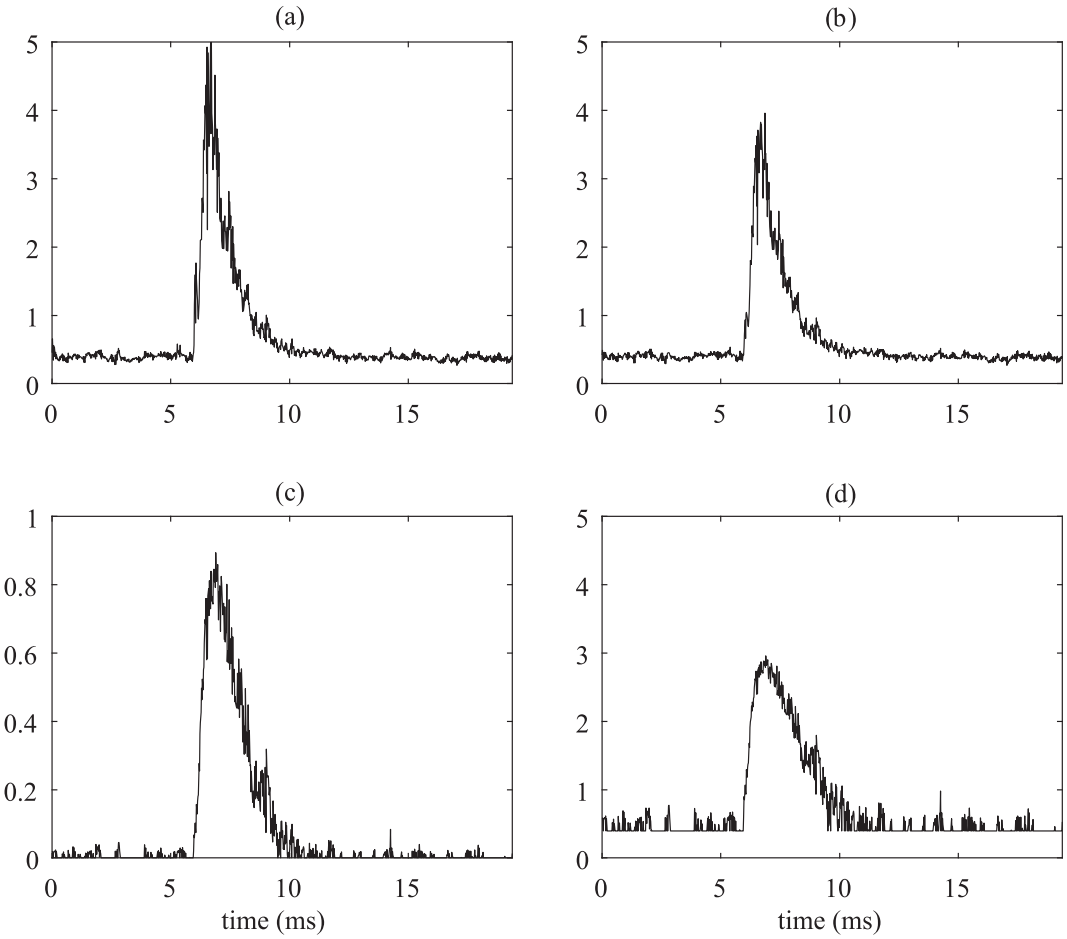

Fig. 19. Cyclic period of a ball fault, $N = 931.2$. Estimated periodic standard deviation $\hat{\sigma}(n)$ in (a) the cyclostationary Gaussian model, (b) the cyclostationary generalized Gaussian model and (d) the cyclostationary Bernoulli-Gauss model. (c) Estimated periodic probability $\hat{p}(n)$ in the cyclostationary Bernoulli-Gauss model.

with maximum-kurtosis blind deconvolution. The exact cyclic period being unknown, it had to be retrieved by using a range of possible periods from $N_{\min} = 80$ to $N_{\max} = 90$ samples centred on the expected fault period. The evolutions of the indicators based on the generalized Gaussian model, $I_{GGS/GS}(\mathbf{x})$, $I_{GGCS/GGS}(\mathbf{x})$, $I_{GCS/GS}(\mathbf{x})$ and $I_{GGCS/GCS}(\mathbf{x})$, are displayed in Fig. 23(a) and (b). The indicator $I_{GGCS/GS}(\mathbf{x}) = I_{GGS/GS}(\mathbf{x}) + I_{GGCS/GGS}(\mathbf{x})$ is shown in Fig. 24(b). Indicators based on the Bernoulli-Gauss model were found almost identical to those of the Generalized Gaussian model and are therefore not shown. It is seen that $I_{GGS/GS}(\mathbf{x})$ (Fig. 23(d)) stands well above the statistical threshold at significance level $\alpha = 0.001$ from the beginning of the test, whereas $I_{GGCS/GGS}(\mathbf{x})$ (Fig. 23(b)) remains (on the average) below the statistical threshold until record number 534. In addition, $I_{GGCS/GCS}(\mathbf{x})$ (Fig. 23(a)) passes below its statistical threshold at the end of the experiment, after record number 969. This separates three different phases:

- A first phase from record 1 to 533 where the vibration signal is stationary non-Gaussian (GGS), with an exponent $\hat{\beta}_0$ of about 1.5 as seen in Fig. 22(b). Although non-Gaussian, this phase corresponds to a healthy state of the tested bearing.
- A second phase from record 534 to 969 where the vibration signal is cyclostationary non-Gaussian (GGCS), with an exponent $\hat{\beta}_1$ slightly fluctuating between 1.5 and 1.8 as seen in Fig. 22(b). This phase is marked by a sudden jump of cyclostationary and then stabilizes.
- A last phase from 970 to the last record where the vibration signal suddenly becomes Gaussian, while still remaining cyclostationary, with an exponent $\hat{\beta}_1$ close to 2 as seen in Fig. 22(b). This phase corresponds to advanced damage.

The above observations demonstrate that the degradation of the bearing is here evidenced by a marked emergence of cyclostationarity, and not non-Gaussianity. Indeed, the signals are non-Gaussian from the beginning of the experiment, before the occurrence of the fault, and they tend to become more and more Gaussian with the increase of degradation. This

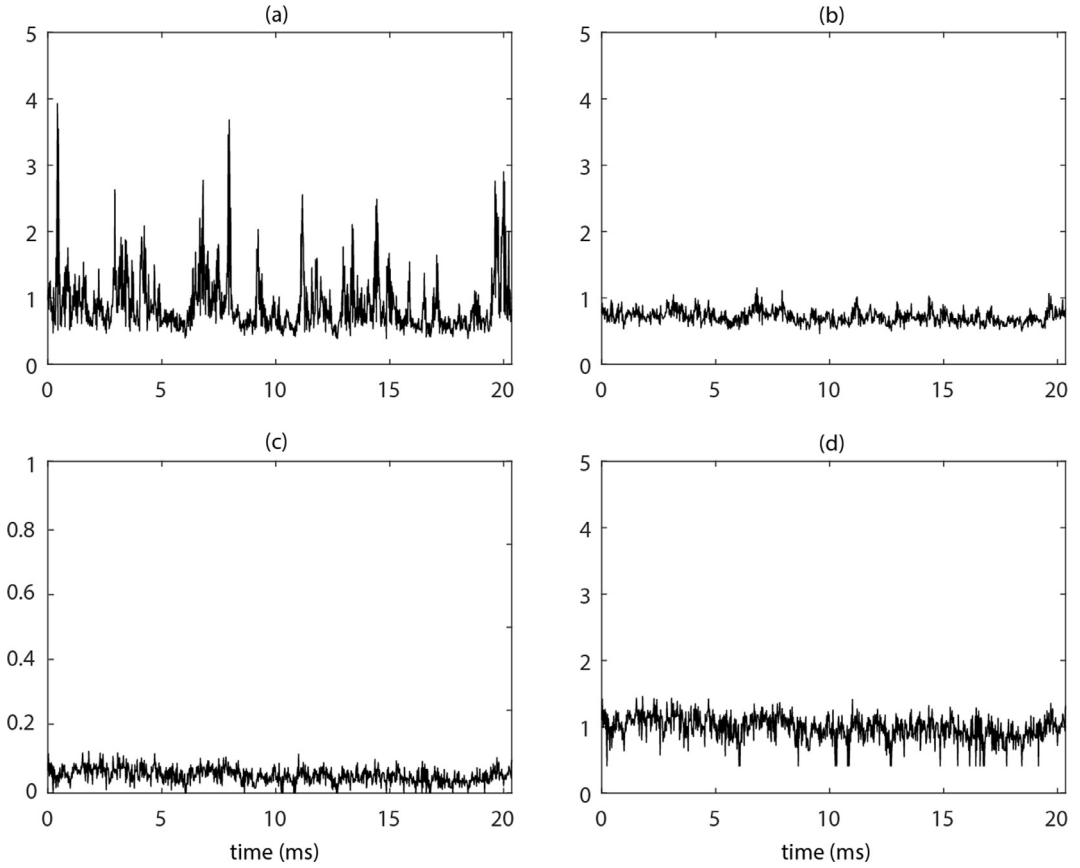

Fig. 20. Wrong cyclic period $N = 978$. Estimated periodic standard deviation $\hat{\sigma}(n)$ in (a) the cyclostationary Gaussian model, (b) the cyclostationary generalized Gaussian model and (d) the cyclostationary Bernoulli-Gauss model. (c) Estimated periodic probability $\hat{p}(n)$ in the cyclostationary Bernoulli-Gauss model.

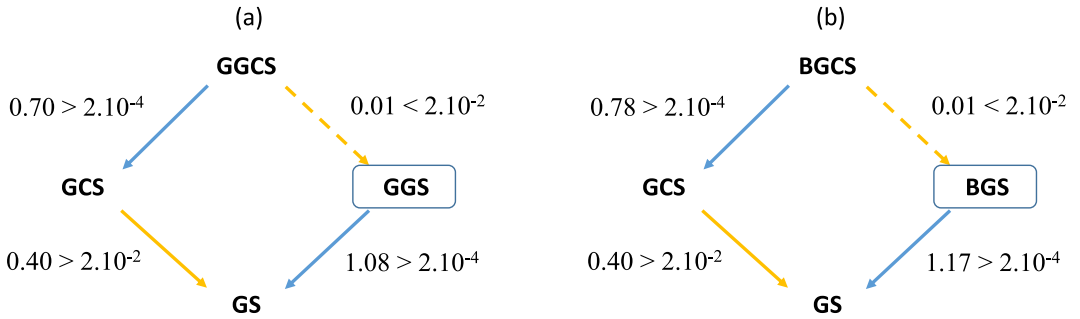

Fig. 21. Values of the indicators for testing between competing models together with their statistical thresholds at significance level 0.001, assuming a wrong cyclic period $N = 978$. (a) (Cyclo)stationary (generalized) Gaussian models. (b) (Cyclo)stationary (Bernoulli) Gauss models. The retained models are shown in boxes. The tests which failed are marked with dotted arrows.

aspect seems to have been ignored by most publications dealing with this dataset. It also indicates that the kurtosis (see Fig. 22(a)) is a misleading indicator in this situation as it wrongly suggests an increase of non-Gaussianity. The best indicator to follow in this situation is $I_{GGCS/GGS}(\mathbf{x})$ (or equivalently $I_{BGCS/BGS}(\mathbf{x})$ for the Bernoulli-Gauss model) which tracks the evolution of cyclostationarity only; this is because it exceeds the threshold only when degradation appears. Indicator $I_{GGCS/GS}(\mathbf{x}) = I_{GGCS/GGS}(\mathbf{x}) + I_{GGS/GS}(\mathbf{x})$ in Fig. 24(a) – which may have been thought more relevant than $I_{GGCS/GGS}(\mathbf{x})$ because it measures a distance to both non-Gaussianity and non-stationarity – rejects the null hypothesis even in the healthy stage and it suffers from more fluctuation in the degradation phase (just as the kurtosis and RMS value in Fig. 22(a)).

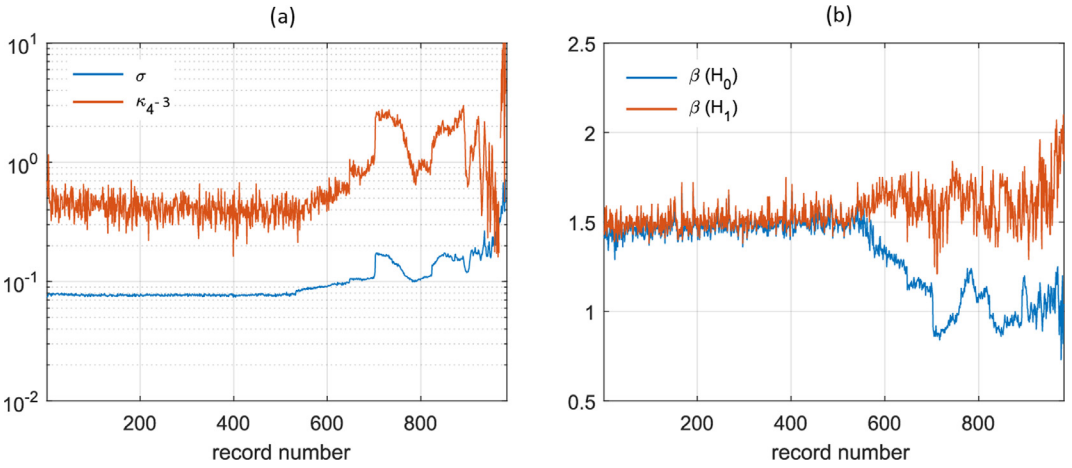

Fig. 22. IMS dataset, Campaign 2, Channel 1 (Outer race failure). (a) The traditional indicators, RMS value σ and excess kurtosis $\kappa_4 - 3$, displayed as functions of the record number. (b) Estimated exponents $\hat{\beta}_0$ and $\hat{\beta}_1$ in the generalized Gaussian model as functions of the record number.

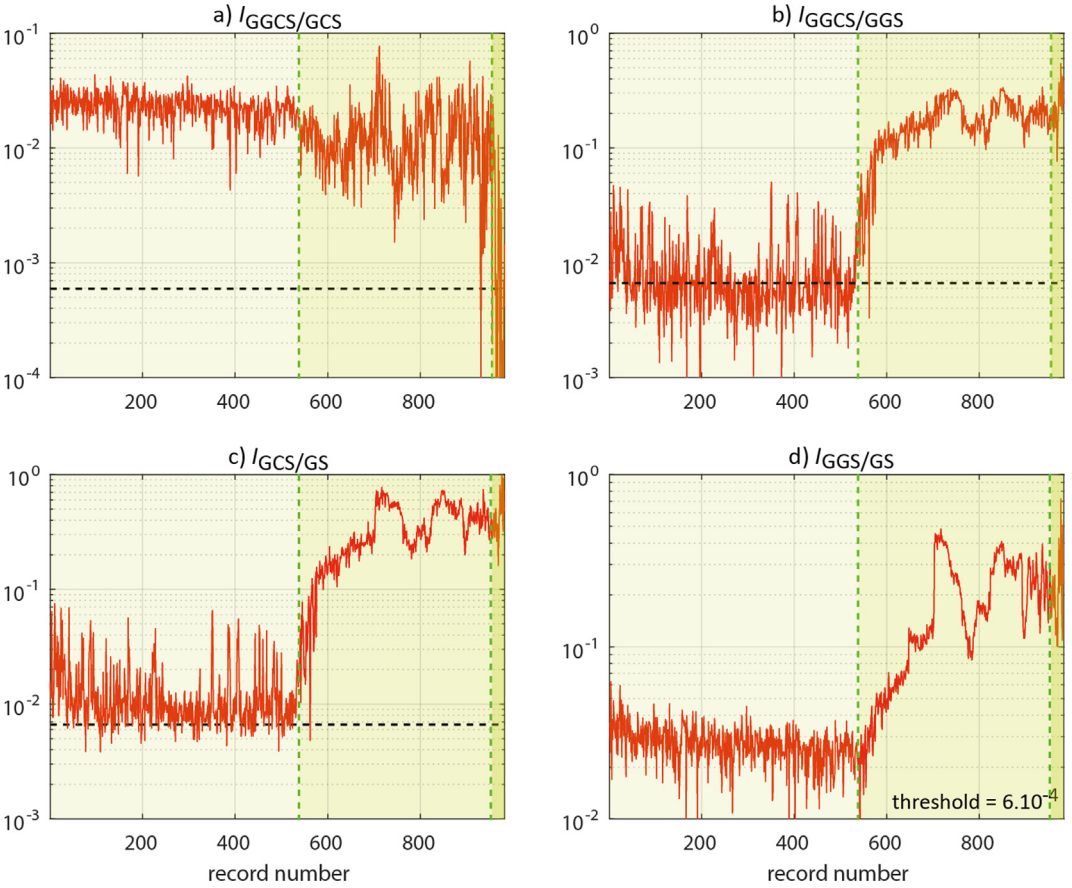

Fig. 23. Condition indicators (a) $I_{GGCS/GCS}(\mathbf{x})$, (b) $I_{GGCS/GGS}(\mathbf{x})$, (c) $I_{GCS/GS}(\mathbf{x})$ and (d) $I_{GGS/GS}(\mathbf{x})$ together with their statistical thresholds (black dotted lines) at significance level $\alpha = 0.001$, as functions of the record number.

A last quantity of interest, returned by the Bernoulli-Gauss model, is the ratio of the energy delivered by the fault to the total energy in the signal (see Fig. 24(b)), as calculated by

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}_0^2 + \hat{\sigma}_1^2}. \tag{82}$$

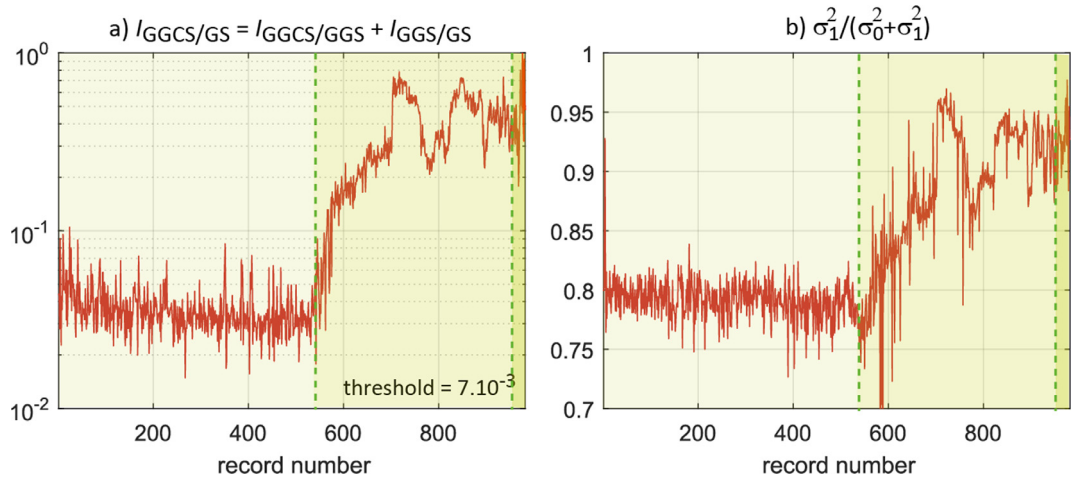

Fig. 24. Condition indicators (a) $I_{GGCS/GS}(\mathbf{x})$ and (b) $\hat{\sigma}_1^2/(\hat{\sigma}_0^2+\hat{\sigma}_1^2)$ from the Bernoulli-Gauss model, as functions of the record number.

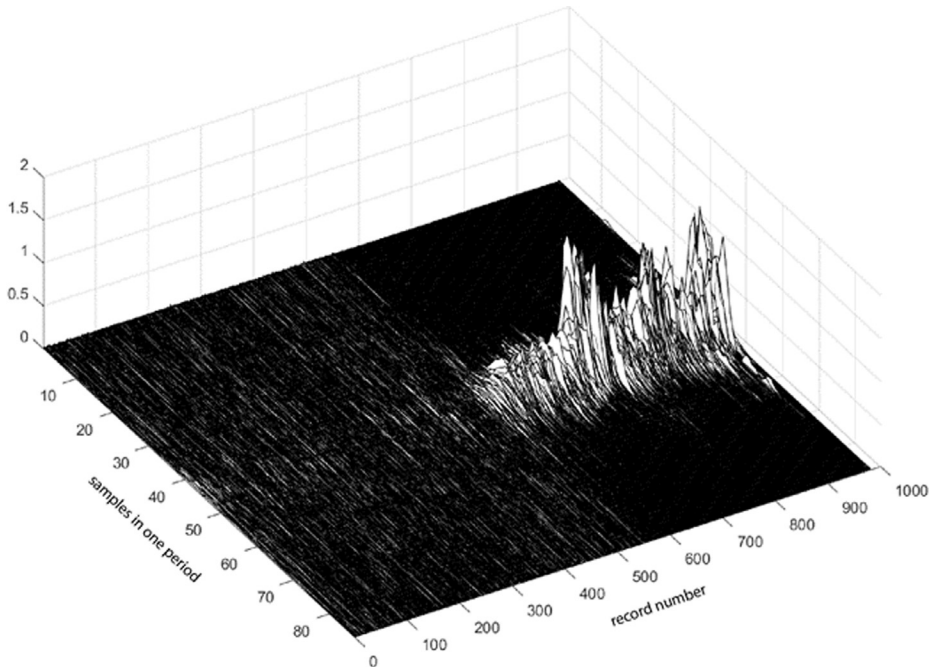

Fig. 25. Periodically time-varying variance of the fault signals with respect to record number (signals have been resynchronized by attaching the centre of gravity of their envelopes to the middle of the period).

Contrary to the RMS value which may increase for any reason, an increase of this quantity necessarily correlates with an intensification of the contact forces on the fault surface. As such, it may serve as a useful indicator in prognostics, just as the periodically time-varying variance of the fault signal, calculated as

$$\hat{\sigma}^2(n) = \hat{p}(n)\hat{\sigma}_1^2, \tag{83}$$

and shown in Fig. 25.

12. Conclusions

Condition indicators are scalar metrics which aim at resuming the information contained in a signal about the state of the machine. They are commonly used for monitoring, diagnostic, trending, and prognostic purposes. This paper has introduced

a rigorous methodology, based on the generalized likelihood ratio, to design condition indicators that are optimal for detecting a given symptom in the signal. Optimality is to be understood with respect to the maximization of the probability of detection for a given probability of false alarm (false positive). An important by-product of the proposed methodology is to provide a statistical threshold, which is crucial in condition monitoring systems. By following these principles, several new condition indicators have been introduced for the first time, which account for a large variety of possible fault symptoms. They are all embodied into two comprehensive models: the cyclostationary generalized Gaussian on the one hand and the cyclostationary Bernoulli-Gauss model on the other hand. Besides, several known indicators have been recovered as particular cases or as approximations under these assumptions. One particularity of the proposed methodology is to make a clear separation between non-Gaussianity and non-stationarity. This turns out especially relevant to detect fault symptoms that are both impulsive and cyclostationary, each of these properties carrying information on the condition of the machine. In addition, it has been proved that although non-stationarity is often detected with non-Gaussian indicators, this approach is actually suboptimal.

This work gathers converging efforts aimed at rationalising the increasingly dispersive literature on machine diagnostic techniques. This, in turn, allows a clarification of the relationship between competitive alternatives, and therefore more informed choices in the design of digital signal processing algorithms for diagnostic and prognostic applications. An important conclusion is that there is no such thing as a universal "best" indicator: each indicator is – often implicitly – related to the comparison of two statistical hypotheses and is therefore optimal only in this perspective.

Acknowledgements

This work was performed within the framework of the Labex CeLyA of Université de Lyon, operated by the French National Research Agency (ANR-10-LABX-0060/ANR-11-IDEX-0007). The authors acknowledge both the assistance received by using the NASA Ames Prognostics Data Repository http://ti.arc.nasa.gov/project/prognostic-data-repository and the donators of the data: Center for Intelligent Maintenance Systems (IMS), University of Cincinnati [58].

Appendices

A. Expressions of log-likelihoods under various hypotheses

(a) *Gaussian cyclostationary (GCS)*

Under the GS hypothesis, the log-likelihood is

$$\mathcal{L}_{GCS}(\mathbf{x}) = \ln \prod_{n=0}^{L-1} \frac{e^{-\frac{|x(n)|^2}{c\hat{\sigma}^2(n)}}}{(c\pi\hat{\sigma}^2(n))^{\frac{1}{c}}} \tag{84}$$

where $c = 2$ when $x(n)$ is real and $c = 1$ when it is complex and where the MLE of the N-periodic variance $\hat{\sigma}^2(n)$ is

$$s^2(n) = \frac{1}{K}\sum_{k=0}^{K-1}|x(n+kN)|^2 \tag{85}$$

with $K = \lfloor L/N \rfloor$ the number of N-sample periods in the L-sample long signal. After substitution and assuming that $KN = L$ for simplicity (the signal can always be truncated to meet this condition), it comes

$$\mathcal{L}_{GCS}(\mathbf{x}) = -\frac{L}{c}\left(\ln(c\pi e) + \ln(s^2(n))\right) \tag{86}$$

(b) *Gaussian stationary (GS)*

The GS case arrives as a particular instance of the GCS case with a time-invariant variance $\sigma^2(n) \equiv \sigma^2$; thus, the log-likelihood is obtained as

$$\mathcal{L}_{GS}(\mathbf{x}) = -\frac{L}{c}\left(\ln(k\pi e) + \ln\langle s^2(n)\rangle\right) \tag{87}$$

where the MLE of the variance is $\hat{\sigma}^2 = L^{-1}\sum_{n=0}^{L-1}|x(n)|^2 = K^{-1}\sum_{n=0}^{K-1}s^2(n) = \langle s^2(n)\rangle$.

(c) *Generalized Gaussian cyclostationary (GGCS)*

Under the GGCS hypothesis, the log-likelihood is

$$\begin{aligned}\mathcal{L}_{GGCS}(\mathbf{x}) &= \ln \prod_{n=0}^{L-1} \frac{\hat{\beta}}{2\pi^{(2-c)}\Gamma\left(c/\hat{\beta}\right)\hat{\eta}(n)} e^{-\left(\frac{|x(n)|}{\hat{\eta}(n)}\right)^{\hat{\beta}}} \\ &= L\cdot\left(\ln\hat{\beta} - \ln 2 - (2-c)\ln\pi - \ln\Gamma\left(c/\hat{\beta}\right) - \langle\ln\hat{\eta}(n)\rangle - \left\langle\left(\frac{|x(n)|}{\hat{\eta}(n)}\right)^{\hat{\beta}}\right\rangle\right)\end{aligned} \tag{88}$$

where $c = 2$ when $x(n)$ is real and $c = 1$ when it is complex [53] and where the scale parameter $\hat{\eta}(n)$ is a N-periodic function of time. Its MLE is readily obtained as

$$\hat{\eta}(n) = \left(\hat{\beta} \cdot s^{\hat{\beta}}(n)\right)^{\frac{1}{\hat{\beta}}}, \tag{89}$$

with

$$s^{\beta}(n) = \frac{1}{K}\sum_{k=0}^{K-1}|x(n+kN)|^{\beta} \tag{90}$$

representing the N-sample synchronous average of the β-powered envelope of x (boiling down to $s^2(n)$ of the Gaussian case when $\beta = 2$). The MLE for the shape parameters does not have a closed form expression. The differentiation of the Gamma function with respect to β results in the following equations defining the MLE $\hat{\beta}$,

$$\frac{\hat{\beta} + c \cdot \psi^{(0)}\left(\frac{c}{\hat{\beta}}\right)}{\hat{\beta}^2} - \frac{1}{L}\sum_{n=0}^{N-1}\sum_{k=0}^{K-1}\left(\frac{|x(n+kN)|}{\hat{\eta}\left(n;\hat{\beta}\right)}\right)^{\hat{\beta}} \ln\left(\frac{|x(n+kN)|}{\hat{\eta}\left(n;\hat{\beta}\right)}\right) = 0 \tag{91}$$

with $\psi^{(0)}$ the digamma function and $\hat{\eta}\left(\hat{\beta}\right)$ representing the functional dependency expressed by Eq. (89). Solving Eq. (91) with a zero-finding algorithm, it is possible to find $\hat{\beta}$ and, by substituting, the expression of $\hat{\eta}$. Finally, the log-likelihood reads

$$\mathcal{L}_{GGCS}(\mathbf{x}) = -L \cdot \left(\hat{\beta}^{-1} - \ln\hat{\beta} + \hat{\beta}^{-1}\langle\ln(\hat{\beta} \cdot s^{\hat{\beta}}(n))\rangle + \ln\Gamma\left(\frac{c}{\hat{\beta}}\right) + \ln 2 + (2-c)\ln\pi\right). \tag{92}$$

(d) *Generalized Gaussian stationary (GGS)*

The GGS arrives as a particular instance of the GGCS case with a time-invariant scale parameter $\eta(n) \equiv \eta$; thus, the log-likelihood is obtained as

$$\mathcal{L}_{GGS}(\mathbf{x}) = -L \cdot \left(\hat{\beta}^{-1} - \ln\hat{\beta} + \hat{\beta}^{-1}\ln(\hat{\beta} \cdot \langle s^{\hat{\beta}}(n)\rangle) + \ln\Gamma\left(\frac{c}{\hat{\beta}}\right) + \ln 2 + (2-c)\ln\pi\right) \tag{93}$$

where the MLE of the scale parameter is $\hat{\eta} = \left(\hat{\beta}\langle s^{\hat{\beta}}(n)\rangle\right)^{\frac{1}{\hat{\beta}}}$ with $\langle s^{\beta}(n)\rangle = K^{-1}\sum_{n=0}^{K-1} s^{\beta}(n)$ and the MLE $\hat{\beta}$ is found as the zero of

$$\frac{\hat{\beta} + c \cdot \psi^{(0)}\left(\frac{c}{\hat{\beta}}\right)}{\hat{\beta}^2} - \frac{1}{L}\sum_{n=0}^{N-1}\sum_{k=0}^{K-1}\left(\frac{|x(n+kN)|}{\hat{\eta}\left(\hat{\beta}\right)}\right)^{\hat{\beta}} \ln\left(\frac{|x(n+kN)|}{\hat{\eta}\left(\hat{\beta}\right)}\right) = 0. \tag{94}$$

B. Consideration of non-integer periods

The discussion in the paper has assumed for simplicity that N is a whole number, however it can easily be accommodated to non-integer periods by interpolating the signal: since the squared signal is essentially dominated by low frequencies, a zero-th or first order interpolation will be found good enough in practice. The counterpart of the synchronous average in Eq. (7) in the case of zero-th order interpolation reads

$$\hat{\sigma}^2(n) = \frac{1}{K}\sum_{k=0}^{K-1}|x([n+kN])|^2,\ n = 0, \ldots, \lfloor N\rfloor,\ K = \lfloor L/\lfloor N\rfloor\rfloor \tag{95}$$

where symbols $[]$ and $\lfloor\rfloor$ stand for the nearest and greatest integer functions, respectively (e.g. $[n+kN]$ is the nearest integer to $n+kN$). The same algorithm applies to all other synchronous averages used in the paper.

C. Proof of Proposition 9

The proof follows as a particular case of Proposition 10 where $p_X(x;\theta(t))$ is given by Eq. (71) with $\theta(t) = \sigma^2(t)$ and $C_4(t) \equiv 0$ and where continuous time (variable t) is eventually replaced by discrete time (index n).

D. Proof of Proposition 10

Let

$$M_p(t) \overset{def}{=} \int x^p p_X(x;\theta(t))dx \tag{96}$$

defines the time-dependent moment of order p. Then, its time-averaged version is

$$\bar{M}_p = \langle M_p(t)\rangle = \int x^p \langle p_X(x;\theta(t))\rangle dx, \tag{97}$$

with

$$\langle p_X(x;\theta(t))\rangle = \int_T p_X(x;\theta(t))\frac{dt}{T} = \int_{\cup\Theta_i} p_X(x;\theta)\frac{d\theta}{|\dot{\theta}|}, \tag{98}$$

where the last equality results from the change of variable $t \to \theta$ and $\bigcup\Theta_i$ denotes the union of support sets where $\theta(t)$ is monotonic. This proves Eqs. (76) and (77).

Moreover, the time-varying kurtosis is defined as

$$C_4(t) \overset{def}{=} M_4(t) - 3|M_2(t)|^2. \tag{99}$$

Therefore,

$$\begin{aligned}\langle C_4(t)\rangle &= \langle M_4(t)\rangle - 3\langle |M_2(t)|^2\rangle \\ &= \underbrace{\bar{M}_4 - 3|\bar{M}_2|^2}_{(i)} + 3\underbrace{\left(|\langle M_2(t)\rangle|^2 - \langle |M_2(t)|^2\rangle\right)}_{(ii)}\end{aligned} \tag{100}$$

where term (i) is recognized as the definition of κ_4 and term (ii) is non-positive according to Cauchy-Schwartz inequality. This proves that $\kappa_4 \geqslant \langle C_4(t)\rangle$.

E. Derivation of the EM algorithm

The log-likelihood of the GMM reads

$$\ln p_X(\mathbf{x}|H_1) = \sum_{n=0}^{L-1} \ln\left(\mathcal{N}\left(x(n);0,\sigma_0^2\right)prob(\zeta(n)=0) + \mathcal{N}\left(x(n);0,\sigma_1^2\right)prob(\zeta(n)=1)\right). \tag{101}$$

Finding closed-form expressions for the MLEs of the variances σ_0^2 and σ_1^2 and of the probability $prob(\zeta(n)=1)$ is unfortunately intractable, yet a simple iterative process is offered by the EM algorithm. The principle consists in assuming momentarily, at iteration q, that the value of the latent variable $\zeta(n)$ is known and in taking the conditional expectation of the log-likelihood conditional to the data $\mathbf{x}$ and the estimates $\boldsymbol{\theta}^{[q-1]} \doteq \left[\sigma_0^{2[q-1]}, \sigma_1^{2[q-1]}, \pi^{[q-1]}\right]^T$ from the previous iteration $q-1$. This reads

$$\underset{\boldsymbol{\zeta}|\mathbf{x},\boldsymbol{\theta}^{[q-1]}}{\mathbb{E}}\left\{\ln p_{X,\zeta}(\mathbf{x},\boldsymbol{\zeta}|H_1)\right\} \tag{102}$$

where

$$\begin{aligned}p_{X,\zeta}(\mathbf{x},\boldsymbol{\zeta}|H_1) &= p_{X|\zeta}(\mathbf{x}|\boldsymbol{\zeta},H_1)p(\boldsymbol{\zeta}) \\ &= (1-\zeta(n))\mathcal{N}\left(x(n);0,\sigma_0^2\right)prob(\zeta(n)=0) + \zeta(n)\mathcal{N}\left(x(n);0,\sigma_1^2\right)prob(\zeta(n)=1)\end{aligned} \tag{103}$$

Inserting Eq. (103) into Eq. (102) and using $p \overset{\text{def}}{=} prob(\zeta(n)=1)$, ones arrives at

$$\begin{aligned}&\underset{\boldsymbol{\zeta}|\mathbf{x},\boldsymbol{\theta}^{[q-1]}}{\mathbb{E}}\left\{\ln p_{X,\zeta}(\mathbf{x},\boldsymbol{\zeta}|H_1)\right\} \\ &= \underset{\boldsymbol{\zeta}|\mathbf{x},\boldsymbol{\theta}^{[q-1]}}{\mathbb{E}}\left\{\sum_{n=0}^{L-1}(1-\zeta(n))\left(\ln\mathcal{N}\left(x(n);0,\sigma_0^2\right)+\ln(1-p)\right) + \zeta(n)\left(\ln\mathcal{N}\left(x(n);0,\sigma_1^2\right)+\ln p\right)\right\} \\ &= \sum_{n=0}^{L-1}\left((1-z^{[q]}(n))\left(-\frac{1}{c}\left(\ln\sigma_0^2 + \frac{|x(n)|^2}{\sigma_0^2}\right)+\ln(1-p)\right) + z^{[k]}(n)\left(-\frac{1}{c}\left(\ln\sigma_1^2+\frac{|x(n)|^2}{\sigma_1^2}\right)+\ln p\right) - \frac{1}{c}\ln(c\pi)\right)\end{aligned} \tag{104}$$

with $c = 2$ when $\mathbf{x}$ is real and $c = 1$ when it is complex and where the notation $z^{[q]}(n) \overset{def}{=} \underset{\boldsymbol{\zeta}|\mathbf{x},\boldsymbol{\theta}^{[q-1]}}{\mathbb{E}}\{\zeta(n)\}$ has been used in the last line. Setting to zero the derivative of (102) with respect to σ_0^2, σ_1^2 and p then directly returns the MLEs of Eqs. (59) and (61) (or (65) in the cyclostationary case). The evaluation of $z^{[q]}(n)$ is now carried on by using Bayes' theorem, i.e. $p(\zeta|x) = p(x|\zeta)p(\zeta)/p(x)$. Hence,

$$z^{[q]}(n) \overset{def}{=} \mathbb{E}\{\zeta(n)|x(n),\boldsymbol{\theta}^{[q-1]}\} = \frac{p(x(n)|\zeta(n)=1)prob(\zeta(n)=1)}{p(x(n)|\zeta(n)=1)p(\zeta(n)=1)+p(x(n)|\zeta(n)=0)prob(\zeta(n)=0)}$$
$$= \frac{p^{[q-1]}\mathcal{N}\left(x(n);0,\sigma_1^{2[q-1]}\right)}{(1-p^{[q-1]})\mathcal{N}\left(x(n);0,\sigma_0^{2[q-1]}\right)+p^{[q-1]}\mathcal{N}\left(x(n);0,\sigma_1^{2[q-1]}\right)} \tag{105}$$

which, after division by $p^{[q-1]}\mathcal{N}\left(x(n);0,\sigma_1^{2[q-1]}\right)$, proves Eq. (60).

F. Log-likelihood under the Bernoulli-Gauss hypothesis

(a) *Stationary case*

The log-likelihood is evaluated as follows. First, the identity

$$\mathcal{L}_{BGS}(\mathbf{x}) = \ln p_X(\mathbf{x}|H_1)$$
$$= \sum_{n=0}^{L-1}\sum_{\zeta(n)} q(\zeta(n)) \ln \frac{p_{X,\zeta}(x(n),\zeta(n)|H_1)}{q(\zeta(n))} - \sum_{n=0}^{L-1}\sum_{\zeta(n)} q(\zeta(n)) \ln \frac{p_{\zeta|X}(\zeta(n)|x(n),H_1)}{q(\zeta(n))} \tag{106}$$

is used, which can be shown to hold for any PDF $q(\zeta)$ of the latent variable ζ. It is proved in Ref. [55] that convergence of the EM algorithm is reached when the last term vanishes, that is when

$$q(\zeta(n)) = p_{\zeta|X}(\zeta(n)|x(n),H_1). \tag{107}$$

Thus, after expanding the first term, the log-likelihood reduces to

$$\mathcal{L}_{BGS}(\mathbf{x}) = \sum_{n=0}^{L-1}\sum_{\zeta(n)} q(\zeta(n)) \ln p_{X,\zeta}(x(n),\zeta(n)|H_1) - \sum_{n=0}^{L-1}\sum_{\zeta(n)} q(\zeta(n)) \ln q(\zeta(n)). \tag{108}$$

Now, recognizing that $q(\zeta=1) = p_{\zeta|X}(\zeta=1|x) = z$ and $q(\zeta=0) = p_{\zeta|X}(\zeta=0|x) = 1-z$,

$$\mathcal{L}_{BGS}(\mathbf{x}) = \underset{\boldsymbol{\zeta}|\mathbf{x},\boldsymbol{\theta}^{[q-1]}}{\mathbb{E}}\{\ln p_{X,\zeta}(\mathbf{x},\boldsymbol{\zeta}|H_1)\} - \sum_{n=0}^{L-1}((1-z(n))\ln(1-z(n)) + z(n)\ln z(n)) \tag{109}$$

where the first term was evaluated in Eq. (104). Finally, plugging the MLEs (59) and (61) in the latter yields

$$\mathcal{L}_{BGS}(\mathbf{x}) = L((1-\hat{p})\ln(1-\hat{p}) + \hat{p}\ln\hat{p})$$
$$-\frac{L}{c}\left((1-\hat{p})\ln\hat{\sigma}_0^2 + \hat{p}\ln\hat{\sigma}_1^2 + \ln(c\pi e)\right)$$
$$-\sum_{n=0}^{L-1}((1-z(n))\ln(1-z(n)) + z(n)\ln z(n)) \tag{110}$$

(b) *Cyclostationary case*

Following similar lines as above, the log-likelihood under the Bernoulli-Gauss cyclostationary hypothesis is found as

$$\mathcal{L}_{BGCS}(\mathbf{x}) = K\sum_{n=0}^{M-1}((1-\hat{p}(n))\ln(1-\hat{p}(n)) + \hat{p}(n)\ln\hat{p}(n))$$
$$-\frac{L}{c}\left((1-\bar{p})\ln\hat{\sigma}_0^2 + \bar{p}\ln\hat{\sigma}_1^2 + \ln(c\pi e)\right)$$
$$-\sum_{n=0}^{L-1}((1-z(n))\ln(1-z(n)) + z(n)\ln z(n)) \tag{111}$$

with $\bar{p}$ as defined after Eq. (66).

References

[1] P. Večeř, M. Kreidl, R. Šmíd, Condition indicators for gearbox condition monitoring systems, Acta Polytechnica 45 (6) (2005) 35–43.
[2] J. Zhu, T. Nostrand, C. Spiegel, B. Morton, Survey of condition indicators for condition monitoring systems, Annual Conference of the Prognostics and Health Management Society 5 (2014) 1–13.
[3] G. D'Elia, M. Cocconcelli, R. Rubini, G. Dalpiaz, Evolution of gear condition indicators for diagnostics of planetary gearboxes, in: Proceedings of the 8th International Conference on Surveillance, Roanne Institute of Technology, France, 20–21 October 2015.
[4] A. Rai, S.H. Upadhyay, A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings, Tribology Int. 96 (2016) 289–306.
[5] V. Sharma, A. Parey, A review of gear fault diagnosis using various condition indicators, Procedia Eng. 144 (2016) 253–263.
[6] P. Pennacchi, P. Borghesani, S. Chatterton, A cyclostationary multi-domain analysis of fluid instability in Kaplan turbines, Mech. Syst. Signal Process. 60–61 (2015) 375–390.

[7] A.M. Fiori, M. Zenga, Karl Pearson and the origin of Kurtosis, Int. Stat. Rev. 77 (1) (2009) 40–50.
[8] P.H. Westfall, Kurtosis as peakedness, 1905–2014. R.I.P, The American Statistician 68 (3) (2014) 191–195.
[9] C. Pachaud, R. Salvetat, C. Fray, Crest factor and kurtosis contributions to identify defects inducing periodical impulsive forces, Mech. Syst. Signal Process. 11 (6) (1997) 903–916.
[10] R.B. D'Agostino, A. Belanger, R.B. D'Agostino, A suggestion for using powerful and informative tests of normality, Am. Stat. 44 (4) (1990) 316–321.
[11] C.M. Jarque, A.K. Bera, A test for normality of observations and regression residuals, Int. Stat. Rev./Rev. Int. Stat. 55 (2) (1987) 163–172.
[12] F.A. Zakaria et al., Walking analysis: Empirical relation between kurtosis and degree of cyclostationarity, in: 2013 2nd International Conference on Advances in Biomedical Engineering, Tripoli, 2013, pp. 93–96.
[13] P. Borghesani, P. Pennacchi, S. Chatterton, The relationship between kurtosis- and envelope-based indexes for the diagnostic of rolling element bearings, Mech. Syst. Signal Process. 43 (1–2) (2014) 25–43.
[14] P. Borghesani, J. Antoni, CS2 analysis in presence of non-Gaussian background noise – Effect on traditional estimators and resilience of log-envelope indicators, Mech. Syst. Signal Process. 90 (2017) 378–398.
[15] R. Yan, R.X. Gao, Approximate entropy as a diagnostic tool for machine health monitoring, Mech. Syst. Signal Process. 21 (2007) 824–839.
[16] H. Hong, M. Liang, Fault severity assessment for rolling element bearings using the Lempel-Ziv complexity and continuous wavelet transform, J. Sound Vib. 320 (1) (2009) 452–468.
[17] Y. Li, H. Li, W. Wang, B. Wang, C0 complexity: An alternative complexity measure for degradation status of rolling bearings, 2017 Prognostics and System Health Management Conference (PHM-Harbin), Harbin, 2017, pp. 1–8.
[18] J. Antoni, The infogram: Entropic evidence of the signature of repetitive transients, Mech. Syst. Sig. Process. 74 (2016) 73–94.
[19] M. Cocconcelli, G. Curcurù, R. Rubini, Statistical evidence of central moments as fault indicators in ball bearing diagnostics, in: Proceedings of the 9th International Conference on Surveillance, INSA Euro-Mediterranean, Fes, Morocco, 22–24 May 2017.
[20] H.R. Martin, Statistical moment analysis as a means of surface damage detection, in: Proceedings of the 7th International Modal Analysis Conference, Society for Experimental Mechanics, Schenectady, NY, 1989, pp. 1016–1021.
[21] N. Hurley, S. Rickard, Comparing measures of sparsity, IEEE Trans. Inf. Theory 55 (10) (2009) 4723–4741.
[22] P.W. Tse, D. Wang, The design of a new sparsogram for fast bearing fault diagnosis: Part 1 of the two related manuscripts that have a joint title as "Two automatic vibration-based fault diagnostic methods using the novel sparsity measurement – Parts 1 and 2", Mech. Syst. Sig. Process. 40 (2) (2013) 499–519.
[23] I.S. Bozchalooi, M. Liang, A smoothness index-guided approach to wavelet parameter selection in signal de-noising and fault detection, J. Sound Vib. 308 (2007) 246–267.
[24] D. Wang, K.-L. Tsui, Theoretical investigation of the upper and lower bounds of a generalized dimensionless bearing health indicator, Mech. Syst. Sig. Process. 98 (2018) 890–901.
[25] S. Delvecchio, G. D'Elia, G. Dalpiaz, On the use of cyclostationary indicators in IC engine quality control by cold tests, Mech. Syst. Sig. Process. 60–61 (2015) 208–228.
[26] G. Živanović, W. Gardner, Degrees of cyclostationarity and their application to signal detection and estimation, Signal Process. 22 (3) (1991) 287–297.
[27] J. Igba, K. Alemzadeh, C. Durugbo, E.T. Eiriksson, Analysing RMS and peak values of vibration signals for condition monitoring of wind turbine gearboxes, Renew. Energy 91 (2016) 90–106.
[28] A. Raad, J. Antoni, M. Sidahmed, Indicators of cyclostationarity: Theory and application to gear fault monitoring, Mech. Syst. Sig. Process. 22 (3) (2008) 574–587.
[29] H.J. Decker, R.F. Handschuh, J.J. Zakrajsek, An enhancement to the NA4 gear vibration diagnostic parameter, NASA TM-106553, ARL-TR-389, NASA and the US Army Research Laboratory, 1994.
[30] J. Zakrajsek, D. Townsend, H. Decker, An analysis of gear fault detection methods as applied to pitting fatigue failure data, the systems engineering approach to mechanical failure prevention, 47th Meeting of the MFPG, 1993.
[31] I. Khelf, L. Laouar, A.M. Bouchelaghem, D. Rémond, S. Saad, Adaptive fault diagnosis in rotating machines using indicators selection, Mech. Syst. Sig. Process. 40 (2) (2013) 452–468.
[32] K. Gryllias, H. Andre, Q. Leclere, J. Antoni, Condition monitoring of rotating machinery under varying operating conditions based on Cyclo-Non-Stationary Indicators and a multi-order probabilistic approach for Instantaneous Angular Speed tracking, IFAC-PapersOnLine 50 (1) (2017) 4708–4713.
[33] P. Baraldi, G. Bonfanti, E. Zio, Differential evolution-based multi-objective optimization for the definition of a health indicator for fault diagnostics and prognostics, Mech. Syst. Sig. Process. 102 (2018) 382–400.
[34] C. Hu, W. Smith, R. Randall, Z. Peng, Development of a gear vibration indicator and its application in gear wear monitoring, Mech. Syst. Sig. Process. 76–77 (2016) 319–336.
[35] S. Sassi, B. Badri, M. Thomas, Tracking surface degradation of ball bearings by means of new time domain scalar indicators, Int. J. COMADEM 11 (3) (2008) 36–45.
[36] C. Helstrom, Elements of Signal Detection and Estimation, Prentice Hall, New Jersey, 1995.
[37] A.W.F. Edwards, Likelihood, Cambridge University Press, Cambridge 1972, expanded edition., Johns Hopkins University Press, Baltimore, 1992.
[38] J.A. Rice, Mathematical Statistics and Data Analysis, second ed., Duxbury Press, 1994.
[39] C.-E. Särndal, A class of explicata for "information" and "weight of evidence", Revue de l'Institut International de Statistique / Review of the International Statistical Institute 38 (2) (1970) 223–235.
[40] J. Neyman, E.S. Pearson, On the problem of the most efficient tests of statistical hypotheses, Philos. Trans. Royal Soc. London Ser. A 31 (1933) 289–337.
[41] S.S. Wilks, The large-sample distribution of the likelihood ratio for testing composite hypotheses, Ann. Math. Statist. 9 (1) (1938) 60–62.
[42] P. Borghesani, P. Pennacchi, R.B. Randall, N. Sawalhi, R. Ricci, Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions, Mech. Syst. Sig. Process. 36 (2) (2013) 370–384.
[43] S. Braun, The extraction of periodic waveforms by time domain averaging, Acoustica 32 (1975) 69–77.
[44] E. Bechhoefer, M. Kingsley, A review of time synchronous average algorithms, Annual Conference of the Prognostics and Health Management Society, 2009.
[45] P. McCullagh, J. Nelder, Generalized Linear Models, second ed., Chapman & Hall/CRC, 1989.
[46] T. Cover, J. Thomas, Elements of Information Theory, Wiley, New York, 1991.
[47] C. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423.
[48] L. Brillouin, Science and Information Theory, Academic Press, 1962.
[49] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1) (1951) 79–86.
[50] G.L. McDonald, Q. Zhao, M.J. Zuo, Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection, Mech. Syst. Sig. Process. 33 (2012) 237–255.
[51] Y. Miao, M. Zhao, J. Lin, Y. Lei, Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings, Mech. Syst. Sig. Process. 92 (2017) 173–195.
[52] J. Antoni, F. Bonnardot, A. Raad, M. El Badaoui, Cyclostationary modelling of rotating machine vibration signals, Mech. Syst. Sig. Process. 18 (6) (2004) 1285–1314.
[53] M. Novey, T. Adali, A. Roy, A complex generalized Gaussian distribution: characterization, generation, and estimation, IEEE Trans. Signal Process. Archive 58 (3) (2010) 1427–1433.
[54] P. Borghesani, R. Ricci, S. Chatterton, P. Pennacchi, Fault symptoms of rolling element bearings under variable operating conditions: A multi domain analysis, in: Proceedings of the ASME Design Engineering Technical Conference, 2012.

[55] R.M. Neal, G.E. Hinton, A View of the Em Algorithm that Justifies Incremental, Sparse, and other Variants, Learning in Graphical Models, Volume 89 of the series NATO ASI Series, Kluwer Academic Publishers, 1993, pp. 355–368.
[56] R.B. Randall, J. Antoni, Rolling element bearing diagnostics—a tutorial, Mech. Syst. Sig. Process. 25 (2) (2011) 485–520.
[57] H. Endo, R. Randall, Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter, Mech. Syst. Sig. Process. 21 (2) (2007) 906–919.
[58] J. Lee, H. Qiu, G. Yu, J. Lin, Rexnord Technical Services, Bearing Data Set, NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository), NASA Ames Research Center, Moffett Field, CA, 2007.
[59] R. Pallás-Areny, J.G. Webster, Analog Signal Processing, first ed., Wiley-Blackwell, 1999.