Università degli Studi di Ferrara

DOCTORAL COURSE IN
ENGINEERING SCIENCE

Cycle XXXIV

Director: Prof. S. Trillo

Diagnostics and Prognostics of Rotating Machines Through Cyclostationary Methods and Machine Learning

Scientific/Disciplinary field ING/IND 13

Candidate
Dott. Elia Soave

Supervisor
Prof. Giorgio Dalpiaz

Co-Supervisor
Dott. Gianluca D'Elia

Years 2018/2021

.

PREFACE

I graduated *cum laude* in 2018 in Mechanical Engineering at the University of Ferrara discussing the thesis entitled "Tavola vibrante triassiale: validazione di procedure per test dinamici accelerati con vibrazioni random gaussiane" under the supervision of Eng. G. D'Elia. Immediatly after, I joined with a postgraduate fellowship the research group leaded by Prof. G. Dalpiaz at the same university. After eight months I started my Ph.D. in Engineering Science still under the supervision of Prof. G. Dalpiaz.

During the postgraduate fellowship I have been introduced on the research field of the vibration-based diagnostics of rotating machines. In particular, my study has been focused on blind deconvolution methods for the identification of machine faults. In this contest, together with Eng. M. Buzzoni, I re-wrote some existing BD indicators exploiting their cumulative value in order to make them more suitable for industrial applications. This research work led to the publication of one conference papers [1] and one journal paper [2].

In the first year and half of my Ph.D., my research activity has been focused on the development of a new indicator that combines the BD theory with the cyclostationarity for the detection of machine faults exhibiting this behaviour. The new criterion exploits the Fourier-Bessel series expansion in order to reduce the computational time required by the algorithm due to its mathematical nature more compatible with the fault vibration signature. This aspect increases the suitability of the BD for real time analysis, pivotal aspect in the nowadays industrial scenario. The preliminary results of this study have been presented at the Surveillance 2019 conference in Lyon [3] whereas the complete research work, including the extension for systems operating under variable speed regimes, has been published in Mechanical System and Signal Processing journal [4].

The results obtained during the first part have been considered as the starting point for the last part of my Ph.D. The proposed indicator has been exploited as the observation for developing a prognostics model that aims to evaluate the damaging level of a mechanical system during its operating life. My

research activity in this field has been mainly focused on the Hidden Markov Models that divide the component's useful life into a discrete number of state related each other in a probabilistic way. The idea of hidden states regards the probabilistic law that relates the observations, i.e. the considered indicator's values, and the state sequence, i.e. the health condition of the system at a given time instant. These relations are usually defined for a specific data distributions, e.g. Gaussian. Consequently my research in this field has been focused on the definition of a generalized law that enables the consideration of several different laws depending on the value of some parameters. The resulting model has been validated through several run-to failure dataset regarding both laboratory test and real industrial case studies.

Apart from my Ph.D. research topic, I also had the chance to perform several experimental activities; in particular I studied a novel laser triangulation sensor in order to demonstrate that it can be considered as a contactless alternative to accelerometers for the main vibration-based analyses [5]. At the same time I had the opportunity to collaborate with some companies on the resolution of industrial problems both directly with the measurement performed or through numerical models validated through the experimental test conducted [6].

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AI	Artificial Intelligence
ANN	Artificial Neural Network
AR	Auto Regressive
BD	Blind Deconvolution
BIC	Bayesian Information Criterion
BPFI	Ball Pass Frequency Inner race
BPFO	Ball Pass Frequency Outer race
BSF	Ball Spin Frequency
CBM	Condition Based Maintenance
CMS	Cyclic Modulation Spectrum
CS_1	First-order cyclostationary
CS_2	Second-order cyclostationary
CVS	Continuous Vibration Separation
CYCBD	Cyclostationary Maximization Blind Deconvolution
EM	Expectation-Maximization
EMA	Experimental Modal Analysis
FBBD	Fourier-Bessel Blind Deconvolution
FBSE	Fourier-Bessel Series Expansion
FIR	Finite Impulse Response
FSE	Fourier Series Expansion
FRF	Frequency Response Function
GGD	Generalized Gaussian Distribution
GGHMM	Generalized Gaussian Hidden Markov Model
HMM	Hidden Markov Model
ICS	Indicators of cyclostationarity
IMS	Center for Intelligent Maintenance Systems
IQR	Interquartile Range

IRF	Impulse Response Function
KLD	Kullback-Leibler Divergence
LR	Likelihood Ratio
MAPER	Mean Absolute Percentage Error
MCKD	Maximum Correlated Kurtosis Deconvolution
MED	Minimum Entropy Deconvolution
MOMEDA	MultiPoint Optimal Minimum Entropy Deconvolution Adjusted
OMED	Optimal Minimum Entropy Deconvolution
PDF	Probability Density Function
PHM	Proportional Hazards Model
RMSE	Root Mean Square Error
RUL	Remaining Useful Life
SDOF	Single Degree of Freedom
SES	Squared Envelope Spectrum
SISO	Single Input Single Output
SVM	Support Vector Machine
TSA	Time Synchronous Average

1

INTRODUCTION

1.1 DIAGNOSTICS AND PROGNOSTICS IN THE INDUSTRIAL SCENARIO

The topic of this thesis is the diagnostics and prognostics of rotating machines through the application of signal processing techniques directly on the raw vibration signal measured on the system. Diagnostics and prognostics are fields related each other but with significant different meanings. The term diagnostics refers to the detection and identification of defects occurring in a mechanical system during a certain working phase. On the other hand, the prognostics aims to estimate the Remaining Useful Life (RUL), i.e. the residual time to failure, through the analysis of the trend of some diagnostic indicators in order to avoid unnecessary system's shutdowns for maintenance operations. The application of vibration-based analysis is founded on the idea that any modifications inside the considered system directly reflects on the vibration signature, e.g. in form of modification of signal amplitude or frequency content. Consequently, the basis concept behind the vibrational diagnostics and prognostics is the extraction of any information regarding the health state of the system exploiting some signal processing techniques directly on the vibrations generated by a working machine .

In the last years, this kind of analysis attracted the industrial scenario due to some points of strength: the capability of capture the high dynamics of the most common rotating machines and the non-intrusive nature of vibration measurements, pivotal aspect for the end of line quality control operations. Moreover, taking into account complex systems, e.g. vehicle transmissions, composed by several different elements, e.g. shafts, gears or bearings, it is

possible to extract information about them just from a single accelerometer, even though their different natures [7].

On the other hand, the exploitation of vibrational analysis in industrial environment may represent a challenging task. In fact, the information to be extracted from the raw vibration signal are strongly masked by several interferences like background noise or other contributions related to the motion of the system. Another limitation of this analysis is related to the "physical" filter intrinsic to each structure. In fact, the analysis aims to identify the fundamental components related to the excitation given by the rotating elements that compose the system. Unfortunately, the sensor measures the response of the system to the excitation instead of the excitation itself. Each structure acts as a filter that amplifies the excitation only in some frequency ranges, i.e. the natural frequencies, and consequently the measured response is strongly affected by this filter. Directly related to this aspect there is the problem of the transducer position: due to the filter given by the structure, the sensor has to be placed as close to the excitation as possible, e.g. near the bearing cage where the forces are discharged, in order to minimize the effect of the structure.

The aforementioned issue can be overcome (or at least mitigated) by manipulating the raw data in order to extract the machine related information minimizing the effect of the structure and the other interferences. In this direction, in the last decades the researches worked in order to make the classical signal processing techniques suitable for the study of mechanical systems. Several techniques have been developed in the last years, each one fitted for a particular type of signal (stationary, non-stationary, periodic, cyclostationary, etc.). The increasing number of industrial cases considered for the validation of the proposed methods clearly demonstrates the greater attention posed by the academic world to the main industrial needs, e.g. the detection of incipient faults and the prediction of the RUL for increasing the effectiveness of predictive maintenance. Nowadays, the vibration signal processing covers an extended spectrum of applications and consequently a brief overview of the most common approaches and their applications should be explained.

1.2 ROTATING MACHINES DIAGNOSTICS AND PROGNOSTICS THROUGH VIBRATION SIGNALS

In the rotating machines field, gears and rolling element bearings represent fundamental components. In fact, faults occurred on these components lead to dangerous operating conditions, e.g. overheating or high vibrations, or to catastrophic failures in the worst case. For this reason, the identification of incipient fault or, even better, the prediction of possible fault appearance have become pivotal in the diagnostic and prognostic fields.

Gear and bearing faults can be divided into two different families: distributed, i.e. faults that affect the entire component and thus distributed over one shaft revolution, and localized, i.e. faults concentrated in one point like cracks, spalls and pits. This thesis focuses on the latter one. From the gear standpoint, typical localized faults are represented by spalls, pitting on the tooth surface, cracks on the tooth root or manufacturing errors [8]. On the other side, bearing faults may affect both rolling elements and races in form of pitting, brinelling or spalls [9].

Localized faults reflect on significantly changes into the vibration signals measured on gears and bearings due to the appearance of non-stationary components in form of a series of transients, e.g. the train of impulses related to the impact between a race fault and the rollers in a damaged bearing. Clearly, these transients are generated in a dissimilar way for bearings and gears, due to the different natures of the physical phenomena that lead to the localized faults. Speaking about gears, the vibration signature related to two mating gears should contain only the tooth meshing frequency and its higher harmonics [10, 11]. However, usually the signal presents other contributions due to possible faults or not perfect gear involute profiles related to a low quality manufacturing. The target of the vibration-based diagnostics is the identification of these undesired components even if strongly masked by other contributions such as the gear mesh, speed fluctuation effects or background noise. In this context, a suitable methodology for modelling the system response to a localized gear fault is represented by the amplitude/phase modulation with an impulsive component [12, 13].

Moving to bearings, the fault detection and identification is even more challenging than in the gear case due to their very weak vibration levels

with respect to gear mesh related one. The classic deterministic approach for local bearing faults modelling takes into account a Dirac comb, i.e. the ideal impulsive excitation related to the impacts between rolling elements and the fault, convolved with the transfer function of the structure [14]. However, in real cases the impulsive pattern is hidden by modulation effects and variable transfer path [15].

Moreover, it has been demonstrated that the fault related impulse train is not a purely periodic signal but presents random slip effects that reveal the need for a new perspective in the bearing diagnostics field. In this direction, in the last decade the theory of cyclostationarity demonstrated its suitability for describing and analysing bearing related vibration signatures (but it can be also applied on gears and many other mechanical systems). The randomness of the occurence rate and the pulses amplitude have been considered by Antoni and Randall [16, 17] as the base of a new stochastic model of bearing vibration signals. In particular, it has been demonstrated that bearing fault signatures exhibit random hidden phenomena with cyclic behaviour that can be described as a second-order cyclostationary (CS_2) process. In the same way, the cyclostationarity demonstrated to be a suitable tool for modelling localized gear faults. In fact, Capdessus et al. [18] proved that localized gear faults can be described as a first-order cyclostationary (CS_1) process as well as a CS_2 one.

Over the years, several approaches for the detection of localized faults on both gears and rolling element bearings have been presented in the state of art regarding diagnostics and prognostics of mechanical systems. The revolution represented in the diagnostics field by the cyclostationary theory allows the classification of these signal processing techniques into two different families: the cyclostationary strategies and the others. Regarding the gear diagnostics, the following classical strategies have to be mentioned: scalar indicators [19], e.g. kurtosis, order tracking approaches [20], Time Synchronous Average (TSA) [21, 20], demodulation analysis [22], time-frequency techniques [23], cepstrum [24], Blind Deconvolution (BD) methods [25], Auto Regressive (AR) models [26]. From the bearing diagnostics point of view, the envelope analysis (better known as the high-frequency resonance technique) [27] can be still considered as the most common and exploited processing approach for the rolling element bearing fault detection. An exhaustive review about bearing fault diagnosis literature can be found in [28].

Moving to the cyclostationary family, several techniques are worth mentioning: the signal decomposition into CS_1 part and CS_2 part [29], spectral correlation [30], spectral coherence [28], Cyclic Modulation Spectrum (CMS) [31], Squared Envelope Spectrum (SES) [32] and scalar indicators such as the indicators of cyclostationarity (ICS) [33]. Despite the deep changes introduced by the theory of cyclostationarity, it is possible to find several connections with the classic approaches. Between them it has to be remembered the connection between the integrated spectral correlation and the envelope analysis [34], the relation between spectral correlation and spectral kurtosis [35] and the comparison between TSA and first order cyclostationarity [18].

This brief overview aims to present the most common and exploited signal processing techniques for vibration based fault diagnosis. A more general but exhaustive dissertation about this topic can be found in Ref [7].

1.3 RESEARCH OBJECTIVES

This research investigates two different aspects: the fault identification directly from noisy raw vibration signals and its possible application for the definition of a prognostic model for the assessment of the damaging level and the prediction of the RUL of a mechanical system.

Regarding the first field, among the diagnostic approaches proposed in Section 1.2, the BD enables the extraction of the fault related signature from a noisy observation through an iterative or direct algorithm. This method permits to overcome the issue related to the structure's transfer function under the hypothesis of Single Input Single Output (SISO) system (typical situation in the mechanical field). Only recently, the cyclostationary theory has been combined to the BD for improving its diagnostic effectiveness when applied to rotating machines. However, there are some open questions regarding this particular approach:

- Is it possible to reduce the computational cost required by the BD without losing the diagnostic capability of the cyclostationarity?
- Is it possible to exploit a non stationary series expansion for defining a novel criterion that better fits the fault related vibration signature?

- What is the improvement given by a modulated series expansion with respect to the classic Fourier transform?

In this direction, the target of this work is the definition of a BD criterion that rewrites the existing cyclostationary BD indicator through a modulated series expansion. As a result, the proposed indicator better fits the impulsive nature of the fault related vibration signals. Consequently, the computational time is reduced in order to enable the diagnostics as real time as possible, meeting the nowadays industrial needs.

Moving to the prognostic field, the Hidden Markov Models (HMMs) allows the estimation of the health condition of a system through the probabilistic analysis of one or more sets of physical observations, i.e. trends of some diagnostic indicators that describe the damaging level. These observations present a distribution that can be represented by a mixture distribution composed by a number of distributions equal to the number of model states. On this ground some aspects may be detailed studied:

- Do the observation distributions among the different model states belong to the same family?
- Is it possible to consider a generalized distribution in order to take into account the distribution modifications during the working life?
- What is the improvement given by a generalized distribution with respect to a single distribution, e.g. Gaussian or Bernoulli?

In this contest, this thesis proposes a novel HMM based on a generalized Gaussian distribution in order to consider the modifications of the observation distribution through the values of the distribution parameters. The better fitting quality of the model distribution leads to a more accurate estimation of the actual health state and to the prediction of the residual life until the final system failure.

All the aforementioned questions have been carefully examined in this thesis. The possible answers are discussed and demonstrated through the analysis of both simulated and real vibration signals.

1.4 ORGANIZATION OF THE THESIS

This thesis is organized into 5 chapters, where one diagnostic approach for rolling element bearing fault identification and its application as observation for a new prognostic model are described and discussed. For each methodology, after the mathematical explanation, an exhaustive experimental validation is illustrated, taking into account both simulated signals reproducing the classic fault related signature and real case studies.

Before starting with the definition of the proposed diagnostic and prognostic methods, in Chapter 2 the bearing test bench developed at the University of Ferrara is described. Any diagnostic technique requires an experimental validation through tests as similar as possible to the real operating conditions of the system under observation. In this direction, the development of an internal test bench is a pivotal step for the research quality improvement. In this chapter, the test bench design process is illustrated, pointing the attention on the structural and vibrational issues encountered and their resolution through the classic vibration based analyses, e.g. operational runup and Experimental Modal Analysis (EMA). Later in Chapter 2, the experimental tests performed on the previous described test bench and exploited for the validation of all the methodologies presented in this thesis are described, highlighting the target and the typology of dataset obtained.

In Chapter 3, a novel BD indicator defined by combining the theory of the cyclostationarity with the Fourier-Bessel Series Expansion (FBSE) is proposed. The proposed indicator tries to improve the effectiveness of the Generalized Rayleigh Quotient based iterative algorithm known as cyclostationary maximization blind deconvolution (CYCBD) through the exploitation of the FBSE due to its mathematical nature being more suitable for impulsive signals. The comparison between the two methods is carried out considering both synthesized and real signals. The simulated signals reproduce the typical fault related impulsive pattern of a damaged bearings under both constant and variable speed regimes. The target of this validation is to prove the lower computation time required by the proposed algorithm derived from the lower number of frequency harmonics needed by the FBSE for the signal reconstruction. The real signal driven validation concerns both data acquired on the University of Ferrara test bench and datasets provided online by the NASA Prognostic

Data Repository. This part aims to illustrate the sensitivity of the proposed criterion on the damage severity, in order to highlight its suitability also from the prognostic point of view, enabling its exploitation as observation for the prognostic model described in the following chapter.

The second part of this thesis, named Chapter 4, starts from the BD indicator described in the previous chapter in order to define a new prognostic model. The presented model belongs to the family of the Hidden Markov Models (HMMs), where the operating life of a mechanical system is divided into a discrete set of states representing different damage severity levels. The transition of the system between the states has a probabilistic nature. The belonging to one particular state depends on one or more indicators describing the health conditions, i.e. the observations, through a statistic relation. Usually this relation follows a particular statistic distribution, e.g. Gaussian, Bernoulli, etc., for each state. However, the observation distribution may change with respect to different state. The idea of the proposed model is to consider a generic distribution, i.e. the Generalized Gaussian Distribution (GGD), that enables the distribution changing for different states only depending on variable distribution parameters. The proposed model for the 1-D case is validated through two run to failure tests performed on both bearings (University of Ferrara test bench) and planetary gearboxes (real industrial application). The model is then extended to multivariable distributions taking into account other classic diagnostic indicators and validated through the aforementioned datasets.

Finally, Chapter 5 outlines the concluding remarks, summarizing results and pivotal aspects come to light in the thesis. Particular attention is pointed out on the practical implication related to the proposed results, highlighting their possible applications in the 4.0 industry environment.

2

BEARING TEST BENCH AT THE ENGINEERING DEPARTMENT OF THE UNIVERSITY OF FERRARA

2.1 INTRODUCTION

A possible measure of the quality of the academic research, in particular in the engineering field, is given by the possible implementation of its results for the resolution of real issues run into in a particular environment. For this reason, the entire scientific production in the industrial field, e.g. numerical models, diagnostic methods and algorithms, prognostic models, etc., requires an experimental validation carried out, if possible, directly on the physical system object of the study. Unfortunately, the experimental campaigns can not be usually performed directly on the real system during its operating cycle due to the machine downtime needed for mounting the sensors and running the planned tests (clearly incompatible with the idle time and costs reduction at the base of the nowadays industrial scenario). In this direction, even more research laboratories are moving to the construction of internal test benches and the design of dedicated tests that try to reproduce the real system's operating conditions.

Over the years, several datasets have been proposed by both academic and private laboratories. From the rolling element bearings point of view, the most famous and heavily exploited by the scientific community is the one proposed by the center for Intelligent Maintenance Systems (IMS) of the University of Cincinnati [36]. The test bench is composed by four double row bearings Rexnord ZA-2115 mounted on the same shaft driven by an AC motor coupled via rub belts. A radial load of 27.7kN is applied on the shaft (and consequently

on bearings) in the mid point by a spring mechanism. The run-to-failure tests are performed at constant speed of 2000rpm and the vibration signals are measured through mono-axial accelerometers installed on the bearing housing. Another bearing dataset that is worth mentioning is the Case Western Reserve University one [37]. The test rig consists of an electric motor coupled with a dynamometer. The SKF and NTN bearings under test support the motor shaft and single point faults were introduced into them through an electro-discharge machine. Several defects size are tested and the vibration signals are measured with piezoelectric accelerometers placed on the motor housing. An interesting overview about bearing test benches can be found in Ref [38].

Moving to the gearboxes field the dataset supplied by the Southern Methodist University of Dallas [39] should be remembered. The tested gearbox is driven by an electric motor and coupled to a disk brake system in order to increase the contact forces between teeth. The dataset contains vibration signals for three different conditions: healthy, chipped tooth and distributed wear. The test is performed at 1420rpm and the vibrations are measured on the gearbox case through piezoelectric accelerometers. Other examples of interesting gearbox test rigs can be found in Ref [40, 41]. In the same direction, at the Engineering Department of the University of Ferrara, a gearbox test bench has been designed [42]. The investigated gearbox is composed of two stages of helical gears driven by an electric motor and coupled with a load motor. The first stage gear has been milled in order to create four different size of gear tooth spall. The vibration signal is acquired through piezoelectric accelerometers mounted on the gearbox case in proximity of the support bearings of the input shaft.

The diagnostic and prognostic methods proposed in this thesis are mainly developed for the bearing fault identification (they can be easily extended for application on gears and other mechanical systems). For this reason during the data required for the experimental validation had to be measured on a dedicated test rig. This particular need led to the design of a new bearing test bench that is described in detail in this chapter.

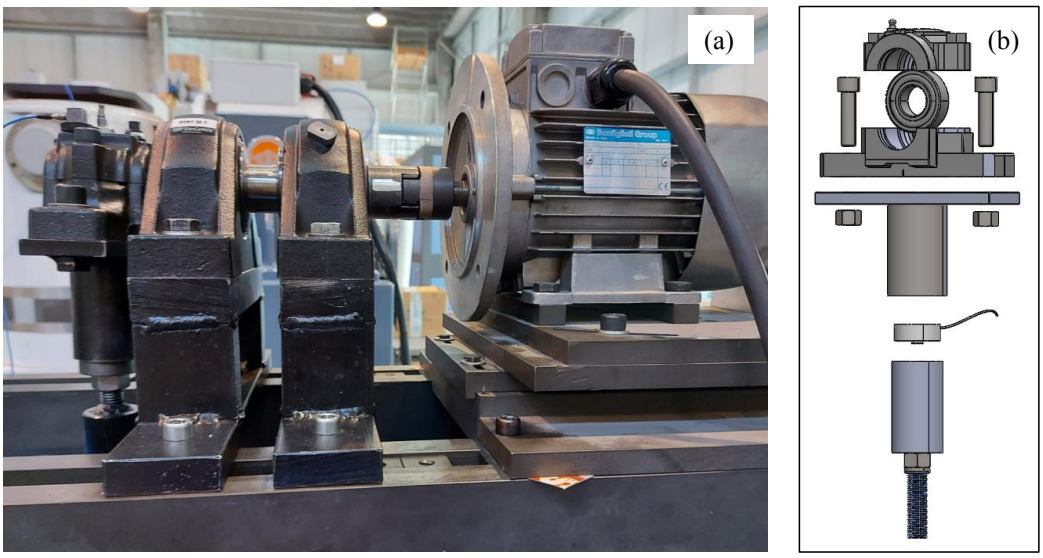

Figure 1: Bearing test bench at the University of Ferrara: (a) Overall system and (b) Exploded view of the tested bearing house

2.2 BEARING TEST BENCH DESIGN

Starting from the model of the aforementioned test rigs, the test bench at the University of Ferrara has been designed in order to reproduce as better as possible the real bearing operating conditions. As shown in Fig.1(a) the tested bearing is cantilever mounted on a shaft supported by other two bearings and driven by an electric motor. The load is regulated by a spring system and applied on the tested bearing through a lever. The instantaneous applied load is measured with a cell load insert between the piston (installed at the end of the lever) and the bearing housing (see Fig.1(b)).

One of the fundamental aspects to be controlled during bearing experimental tests is the consistency of the load for the entire test length due to its direct influence on the fault detection effectiveness [15]. Taking into account real system, even with high manufacturing quality, the shaft unbalance combined

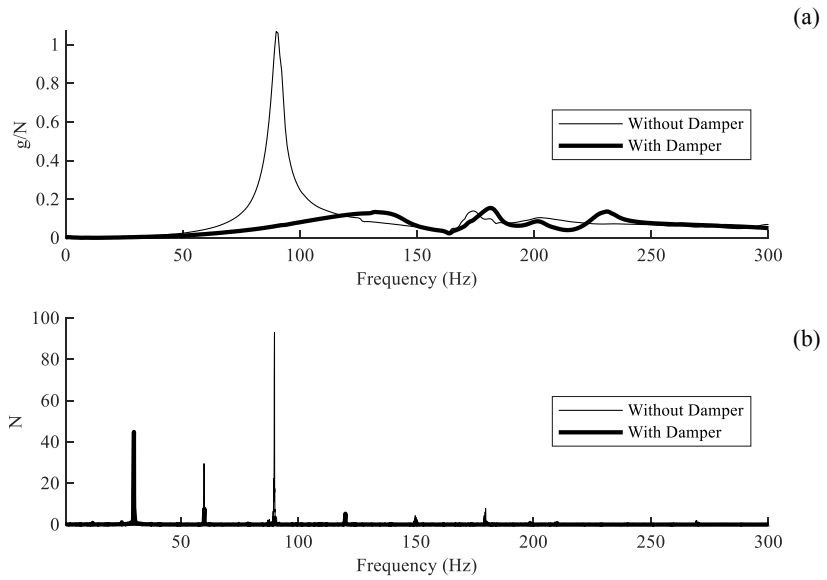

Figure 2: Effect of damper applied on lever system: (a) Point FRF on the beam extremity and (b) Spectrum of load applied on the tested bearing

with the support bearing clearances and possible shaft misalignment lead to variable induced forces with the periodicity of the shaft rotation [43, 44]. The main drawback related to the load application through a lever system is the possible presence of beam natural frequencies that may be excited by the aforementioned variable forces in the low frequency range, i.e. lower than 100Hz. This excitation is consequently increased by the frequency dependent structure filter nearby the resonances leading to a raising fluctuation of the load applied on the tested bearing. This matter is clearly explained in Fig.2: the point Frequency Response Function (FRF), i.e. the ratio between the spectra of the response and the excitation measured in the same point during an EMA, of the beam presents a natural frequency at 90Hz (thin line in Fig.2(a). Consequently the main frequency content of the load signal is related to the third harmonic of the rotational frequency (thin line in Fig.2(b)), taking into account a fundamental one of 30Hz. After the EMA, this resonance has been found as related to the first bending mode of the beam in the direction of application of the load. In order to mitigate this phenomenon, a damper has been added in this direction at the extremity of the lever reducing the load fluctuation during

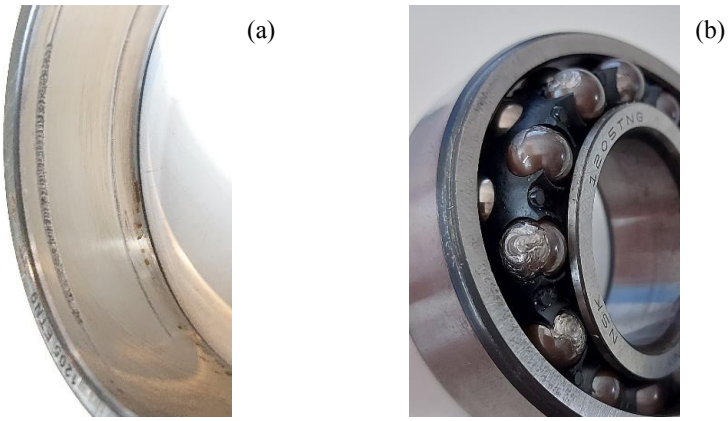

Figure 3: Run-to-failure test related defects: (a) Campaign 1 and (b) Campaign 2

the shaft revolution. The effect of the damper is illustrated in Fig.2 with the bold lines: the natural frequency is translated at 140Hz with lower amplitude and wider peak due to the increased damping. The resulting load spectrum presents only significant amplitude for the fundamental rotational frequency and the overall load variation, i.e. the sum of all the harmonics in Fig.2(b), is strongly reduced.

The above presented test bench allows to perform several bearing test typologies, e.g. run-to failure, stationary test, i.e. under constant speed, and variable speed test. The datasets exploited for the validation of the algorithm and the model proposed in this work are described in detail in the following section.

2.3 DATASETS EXPLOITED IN THIS THESIS

The aforementioned test rig has been exploited for carrying out several experimental tests that can be divided into two main groups: run-to-failure tests, i.e. system under constant operating condition leading to the appearance of natural defects, and stationary test with artificial defects of different sizes.

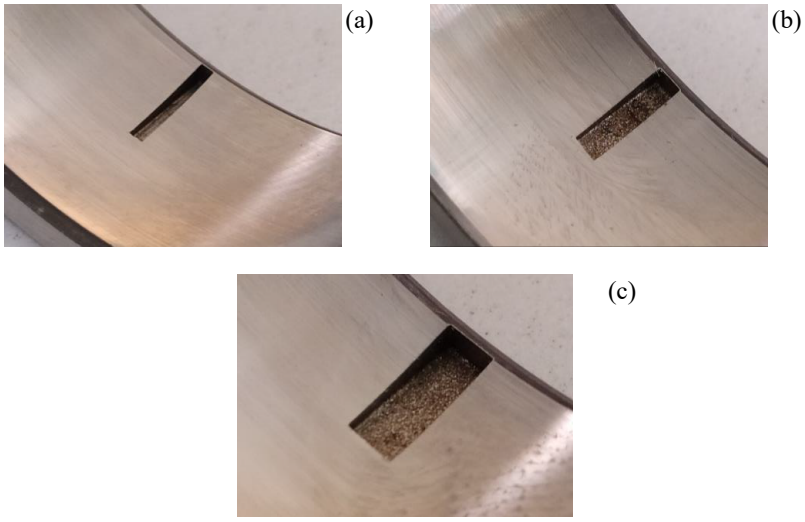

Figure 4: Artificial outer race defects: (a) 0.8mm, (b) 1.6mm, (c) 2.4mm

Both data typologies have been taken into account for the validation of the methodologies proposed in this thesis.

Between the first family, two different experimental campaigns have been considered. Both datasets have been measured for system operating at 2400rpm with a load of 3000N applied on the tested bearing type NSK 1205ETN9. The vibration signal has been measured through a mono-axial piezoelectric accelerometer mounted in radial direction on the bearing housing and acquired continuously by means of a NI cRio system with a sample frequency of 51.2kHz. The results of the two tests have been significantly different:

- Campaign 1 has been stopped after 21 days and an outer race fault has been detected as shown in Fig 3(a).
- Campaign 2 ended after 13 days due to a deep rolling element defect appeared on several rollers (see Fig 3(b)).

These datasets, hereafter called Dataset 1.1 and Dataset 1.2, respectively, will be exploited in Chapter3 in order to demonstrate the suitability of the proposed indicator for the assessment of the bearing damage level and in Chapter4 for the training process and the validation of the proposed prognostic model.

The sensitivity of the novel BD indicator to the fault dimension will be demonstrated through another dataset regarding three artificially damaged bearings type NSK 1205ETN9. Rectangular shaped single point faults of different dimension, i.e. 0.8mm, 1.6mm, 2.4mm, have been introduced on the outer race through an electro-discharge machine as shown in Fig 4. The operating parameters are the same already described for the first two datasets (rotational speed of 2400rpm and load of 3000N applied of the tested bearing) and the vibration signals have been acquired with a sample frequency of 51.2kHz for 30s with a LMS SCADAS acquisition system. This dataset will be referred as Dataset 2 in the following chapters.

3

FAULT IDENTIFICATION THROUGH FOURIER-BESSEL BASED BLIND DECONVOLUTION

3.1 INTRODUCTION

Nowadays, one of the most challenging tasks in the diagnostic field is represented by the identification of impulsive fault related patterns due to their weakness, in particular in the early stages of the damaging process. For this reason these signatures are often strongly masked by background noise, mechanical interferences, e.g. meshing components, and the effect of the dynamic response of the structure that dominate the raw acquired signals. In fact, as already mentioned in Section1.1 each excitation given by a mechanical component is filtered by the Impulse Response Function (IRF) of the system. Unfortunately, usually this physical filter is unknown and consequently it is impossible to isolate the fault related excitation from the measured signal. In this contest, BD techniques allow the extraction of impulsive patterns directly from noisy observations under the hypothesis of linear time-invariant system.

The basis idea of the BD consists in modelling a general phenomenon as the sum of several sources convolved with different filters and an additive background noise. This concept may be applied to several fields such as telecommunications, image processing and rotating machinery and for this reason BD has been widely exploited in the last decades. The first documented BD method has been proposed by Wiggins [45] in 1976 in the seismic signal processing field. This method, known as Minimum Entropy Deconvolution (MED), aims to recover an impulsive source through an iterative algorithm based on the maximization of the kurtosis (also named Varimax norm) related

to the deconvolved signature. In the same direction, Cabrelli [46] proposed an alternative algorithm, known as Optimal Minimum Entropy Deconvolution (OMED), that enables the direct resolution of the BD problem exploiting a novel indicator, called D-Norm. The main issue of the application of MED in the rotating machine diagnostics field is directly related to the mathematical nature of the kurtosis. In fact, both the iterative and the direct algorithm tend to deconvolve a single peak instead of a train of impulses, typical signal waveform of faulty rotating systems.

In order to overcome this drawback and increase its effectiveness for diagnostic purposes, over the years MED has been combined with other signal processing techniques. From the gear diagnostics standpoint, Endo and Randall [25] proposed the application of MED directly on the AR residual in order to enhance the impulsive pattern inside a vibration signal. This method has been later exploited for discriminating gear tooth spalls from tooth root cracks by Endo et al. [47]. In the same field, Zhang et al. [48] proposed the combination of MED and Continuous Vibration Separation (CVS) for planetary gearboxes analysis. Moving to the bearing diagnostics, Sawalhi et al. [49] presented a particular application of MED driven by the maximization of the spectral kurtosis on the envelope spectrum. The spectral kurtosis has been also considered as BD indicator by He et al. [50] for the identification of multiple bearing faults.

Despite the MED improvements above described, the need for a method fitted for the rotating machine diagnosis led to the definition of novel BD criteria based on kurtosis and D-Norm but taking into account the fault period related to the system revolution. McDonald et al. [51] defined a new BD algorithm, known as Maximum Correlated Kurtosis Deconvolution (MCKD), based on the so called correlated kurtosis, i.e. a redefinition of the kurtosis that takes into account an a priori known period between the peaks. Based on the same idea, the Multi Point D-Norm has been proposed by McDonald and Zhao [52] as the indicator for a new BD method called MultiPoint Optimal Minimum Entropy Deconvolution Adjusted (MOMEDA). Both these methods are driven by the idea to deconvolve a periodic train of impulses in order to better fit the characteristic waveform of faulty rotating machines. In this direction, Miao et al. [53] improved the MCKD for bearing fault detection through the automatic pulse period estimation by using the auto-correlation function of the envelope signal.

All these indicators share the mathematical idea of identifying the fault appearance through the measure of the data distribution departure from the ideal Gaussian condition that characterizes vibration signals. However, as mentioned in Sec.1.2, in the last decades the relevance of cyclostationarity for rotating machinery diagnostics has been widely recognized [29]. For this reason, the coexistence of impulsiveness, i.e. the deviation from the ideal Gaussianity, and cyclostationarity, i.e. the deviation from the ideal stationary condition, inside the fault related signature needs to be considered in order to enhance the diagnostic capability of BD. This fundamental relation has been exhaustively illustrated by Antoni and Borghesani [54] from a statistical point of view.

Coming back to BD, although the correlated kurtosis can be interpreted as a cyclostationary criterion, it has been defined empirically, without an explicit investigation about its statistical nature [2]. In a different way, MOMEDA can not be seen as a cyclostationary BD method being it based on a purely periodic indicator that leads to the extraction of a train of equispaced impulses, without considering the random slip effects typical of cyclostationary signatures. In order to fill the existing gap between BD and cyclostationarity, recently Buzzoni et al. [55] proposed a novel BD iterative algorithm, called CYCBD. This method aims to extract the source that maximizes the ICS, i.e. the pattern that exhibits the maximum cyclostationary behaviour, through the iterative resolution of an eigenvalue problem based on the generalized Rayleigh quotient. The powerful of this method is represented by its adaptability to other criteria as well as the enhanced effectiveness given by the direct exploitation of cyclostationarity.

The impulse like nature of a machine fault reflects on waveforms composed by a train of fast amplitude modulations, i.e. transients with variations of the signal's energy localized in an extremely narrow time span. The mathematical nature of the Fourier Series Expansion (FSE) on which the ICS is based, i.e. the sum of constant amplitude sinusoidal functions, may no longer represent the best fitted transform for describing localized faults, due to the high number of series term required [56]. This research tries to enhance the diagnostic capability of the ICS through the exploitation of the FBSE, being it based on amplitude modulated sinusoidal functions that decay in time according to a specific law [57, 58]. The mathematical base of the novel BD criterion allows the reduction of the computational time due to the lower number of series

terms, i.e. the number of fault frequency harmonics required for the source estimation, without losing the effectiveness of the ICS.

Firstly, a brief overview about some BD generalities with a particular focus on the cyclostationarity based method is given. Then the proposed BD indicator is described, highlighting the possibility of writing it in form of a generalized Rayleigh quotient through a specific weighting matrix. The performances of the novel method, named Fourier-Bessel Blind Deconvolution (FBBD), are compared with the CYCBD taking into account both simulated signals and real measured signals. The analysis of synthesized signals in the time domain, i.e. simulating constant speed regimes, as well as in the angular domain, i.e. representing variable regimes, aims to demonstrate the effect of the different transform natures on the number of harmonics required for the reconstruction of the target source. The robustness of the method is studied through the analysis of the filter length effect and the release from the a priori knowledge of the fault period. The analysis of real signals from both academic test benches and real industrial examples illustrates the diagnostic capability of the proposed indicator, pointing the attention on the sensitivity to the damage severity.

3.2 PROBLEM STATEMENT

In general, the vibration signal acquired from a SISO linear time-invariant system, hereafter called $\mathbf{x}$, can be seen as the linear sum of three different contributions as shown in Fig. 5: an impulse-like pattern $\mathbf{s}_0$ related to the excitation given by a localized fault, a pure periodic component $\mathbf{p}$ (e.g. the gear mesh) and a Gaussian background noise $\mathbf{n}$, all convolved with the respective IRFs, such as:

$$\mathbf{x} = \mathbf{s}_0 * \mathbf{g}_s + \mathbf{p} * \mathbf{g}_p + \mathbf{n} * \mathbf{g}_n \tag{1}$$

where, $\mathbf{g}_s$, $\mathbf{g}_p$ and $\mathbf{g}_n$ are the IRFs related to $\mathbf{s}_0$, $\mathbf{p}$ and $\mathbf{n}$, respectively, and * refers to the convolution operation. For the sake of clarity, from now bold capital letters refer to matrices and bold lowercase letters refer to vectors. The target of BD is the estimation, i.e. the deconvolution, of the fault related input source $\mathbf{s}_0$ directly from the noisy observation $\mathbf{x}$, viz:

$$\mathbf{s} = \mathbf{x} * \mathbf{h} = (\mathbf{s}_0 * \mathbf{g}_s + \mathbf{p} * \mathbf{g}_p + \mathbf{n} * \mathbf{g}_n) * \mathbf{h} \approx \mathbf{s}_0 \tag{2}$$

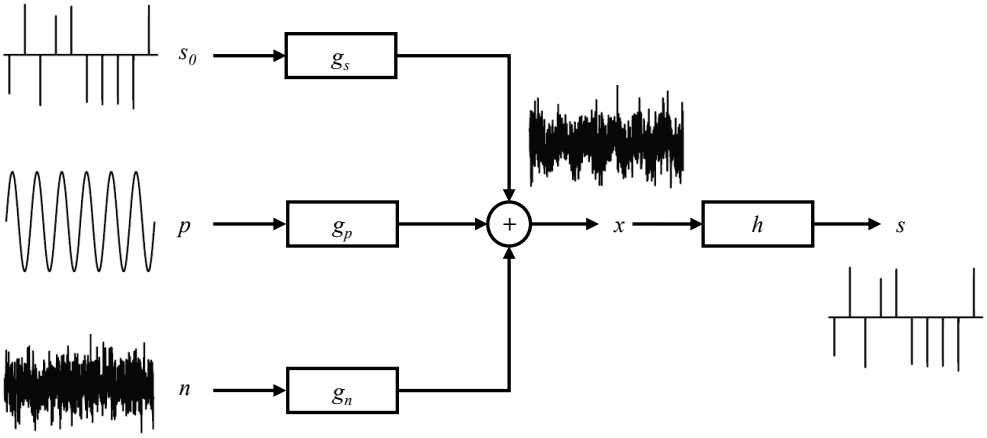

Figure 5: General scheme of BD for SISO mechanical system

where, $\mathbf{s}$ is the deconvolved source and $\mathbf{h}$ is the inverse filter, assumed to be a Finite Impulse Response (FIR) filter, that highlights the impulsive pattern $\mathbf{s}_0$ minimizing the other contributions, i.e. $\mathbf{p}$ and $\mathbf{n}$. A convenient way to express the convolution operation for discrete signals in matrix form is the following:

$$\mathbf{s} = \mathbf{X}\mathbf{h} \tag{3a}$$

$$\begin{bmatrix} s[0] \\ \vdots \\ s[L-1] \end{bmatrix} = \begin{bmatrix} x[N-1] & \dots & x[0] \\ \vdots & \ddots & \vdots \\ x[L-1] & \dots & x[L-N-2] \end{bmatrix} \begin{bmatrix} h[0] \\ \vdots \\ h[N-1] \end{bmatrix} \tag{3b}$$

where L and N are the number of samples of $\mathbf{s}$ and $\mathbf{h}$, respectively.

The mathematical problem described in Eq.2 is based on some simple assumptions: the samples of $\mathbf{s}_0$ are independent identically distributed random variables exhibiting a specific statistical property (e.g. impulsiveness or cyclostationarity), the system is linear time-invariant and all the other contributions are additive and do not share the same characteristics of the target source to be deconvolved. The weak point of the problem is related to the fact that the IRFs are not available. This issue can be overcome by considering an arbitrary criterion that leads to a solution based on an a priori assumption, e.g. assuming that a certain statistical property is maximized by the desired source. If the frequency contents of the different IRFs are not overlapped, i.e. they exhibit

different statistical properties, the estimation of the inverse filter **h** leads to the maximization of the researched signature. It is important to underline that the main drawback of BD methods is their inability to recover the real source magnitude, i.e. they are amplitude invariant methods. On the other hand, BD allows the extraction of informations regarding the waveform, in particular taking into account the fault related characteristics, pivotal aspects for diagnostic purposes.

3.3 OVERVIEW ABOUT CYCLOSTATIONARITY BASED BLIND DECONVOLUTION

As stated in Sec.3.1, the source related statistical property maximized by the classic BD methods, i.e. MED, OMED, MCKD and MOMEDA, is related to the data deviation from the ideal Gaussian conditions, i.e. impulsiveness, without taking into account the deviation from stationarity, i.e. the cyclostationarity, typical of rotating machines fault related signatures. In order to fill this gap, Buzzoni et al. [55] proposed a new BD methods, named CYCBD, that tries to enhance the BD diagnostic capability through the exploitation of the cyclostationary behaviour that describes localized faults on rotating systems.

In general, a process can be defined as cyclostationary if some statistical properties exhibit a periodic behaviour. Several authors [29] demonstrated that the vibration signal from faulty rotating machines is well modelled as a cyclostationary process. In particular, real mechanical signals are composed by a mixture of CS_1 and CS_2 processes. The CS_1 part is related to the periodic contribution and describes the deterministic part of the signal. The CS_2 part is the random part, i.e. the residual signal, described by a periodic autocorrelation function. In this contest, it is possible to define the cyclic frequency as the carrier frequency of an hidden periodicity inside the signal energy, usually resulting from physical phenomena such as faulty gears or bearings. For a generic discrete signal the cyclic frequency can be defined as:

$$\alpha = \frac{k}{T} \tag{4}$$

where k is the sample index and T is the cyclic fault period. The CYCBD method is based on the maximization of the cyclostationary behaviour through

the estimation of the source with the maximum ICS. This indicator has been proposed by Raad et al. [33] and its effectiveness for diagnostic purposes has been exhaustively studied and applied to several mechanical components [33, 59]. The second-order ICS is defined as:

$$ICS_2 = \frac{\sum_{k>0} |c_s^k|^2}{|c_s^0|^2} \tag{5}$$

with

$$c_s^k = \frac{1}{L-N+1} \sum_{n=N-1}^{L-1} |s[n]|^2 e^{-j2\pi \frac{k}{T} n} \tag{6a}$$

$$c_s^0 = \frac{\|\mathbf{s}\|^2}{L-N+1} \tag{6b}$$

Note that the first considered sample is $N-1$ instead of 0 in order to reduce possible numerical artifacts on the reconstructed source related to the convolution operation [52]. Buzzoni demonstrated how the ICS_2 can be written in form of Generalized Rayleigh Quotient as:

$$ICS_2 = \frac{\mathbf{h}^H \mathbf{X}^H \mathbf{W} \mathbf{X} \mathbf{h}}{\mathbf{h}^H \mathbf{X}^H \mathbf{X} \mathbf{h}} = \frac{\mathbf{h}^H \mathbf{R}_{XWX} \mathbf{h}}{\mathbf{h}^H \mathbf{R}_{XX} \mathbf{h}} \tag{7}$$

where the weighting matrix $\mathbf{W}$ contains all the cyclic frequencies α of interest. It can be noticed that the maximization of the ICS_2 with respect to the filter coefficient $\mathbf{h}$ corresponds to the eigenvector associated to the maximum eigenvalue of the following eigenvalue problem [60]:

$$\mathbf{R}_{XWX} \mathbf{h} = \mathbf{R}_{XX} \mathbf{h} \lambda \tag{8}$$

where, λ is equivalent to the maximum ICS_2. The CYCBD is based on the iterative resolution of the eigenvalue problem described in Eq.8 finalized to the maximization of λ, i.e. the ICS_2.

3.4 PROPOSED INDICATOR

In this section, after a brief overview about the mathematical nature of the FBSE and its main characteristics, the novel BD criterion is proposed and discussed.

3.4.1 *Fourier-Bessel series expansion*

The mathematical complexity of the CYCBD directly reflects on an higher computational time required for the analysis with respect to the classic BD methods, based on the maximization of simple statistics like the kurtosis or the D-Norm. This disadvantage is completely balanced by the higher diagnostic capability of this method given by the direct exploitation of the cyclostationarity for the reconstruction of the fault related excitation. The idea at the base of this research activity is to preserve the diagnostic effectiveness of the CYCBD improving it through the reduction of the computational time given by the redefinition of the basis indicator. This aspect is pivotal in order to make the BD more suitable for real time condition monitoring, going towards the nowadays industrial needs.

Considering the impulse-like nature of a localized fault related vibration signature and the mathematical definition of the second-order ICS given in Eq.5, the latter may not represent the best fitted indicator for the fault detection on rotating machinery. In fact, it should be noticed from Eq.6a that the numerator of the ICS_2 is nothing but the Fourier transform of the instantaneous power of the signal $\mathbf{s}$. As a matter of fact, the FSE is perfectly fitted for periodic signals, being it based on constant amplitude sinusoidal functions, and this characteristics may contrast with the fast transient that describes the fault appearance inside a vibration signal. Applying this mathematical consideration to the CYCBD, it entails a number of cyclic harmonics of α required for the signal description that tends to increase inversely proportional to the transient extension in time (or in angle, speaking about rotating machines). As a result, the higher the number of considered harmonics, the higher is the number of operations comprised into the algorithm, consequently leading to an increasing computational cost.

According to this consideration, the use of a generalized series expansion, i.e. based on modulated functions, may represent a suitable tool for speeding up the analysis without losing the diagnostic capability reached by the cyclostationarity based BD. In this direction, the FBSE may represent a valid solution

for better describing the fault related transients due to its non stationary nature. For a generic time discrete signal x(n) the FBSE is defined as:

$$x(n) = \sum_{i=1}^{L} C_i J_0\left(\frac{\beta_i n}{L}\right), \quad n = 0, 1, ..., L-1 \tag{9}$$

where L is the length of $\mathbf{x}$, J_0 is the zero-order Bessel function and C_i are the Fourier-Bessel coefficients, defined as follows:

$$C_i = \frac{2}{L^2[J_1(\beta_i)]^2} \sum_{n=0}^{L-1} n x(n) J_0\left(\frac{\beta_i n}{L}\right) \tag{10}$$

where J_1 is the first order Bessel function and β_i are the positive roots of $J_0 = 0$. The application of FBSE moves the analysis to the β domain that is direct related to the frequency domain, i.e. the resulting domain for the FSE, according to Schroeder [57]:

$$\beta_i \approx \frac{2\pi f_i L}{f_s} \tag{11}$$

where f_s is the sampling frequency of $\mathbf{x}$. The direct correlation between frequency and β domain described in Eq.11 explains the suitability of FBSE for vibration signal analysis (usually based on the study of the frequency content of the signal) and justifies its application for diagnostic purposes in several fields such as gears [56] and electric motors [61].

Eq.9 highlights the non-stationary nature of the FBSE given by the Bessel function decay within the signal range L. According to the aforementioned relation between transient duration and number of series terms, the FBSE may need a lower number of harmonics for the estimation of the excitation pattern. All these aspects allow the basis idea of the proposed BD indicator for enhancing the BD suitability making the algorithm as real time as possible.

3.4.2 *Fourier-Bessel based blind deconvolution criterion*

As previously stated, the numerator of the ICS_2 given in Eq.6a can be seen as the Fourier transform of the instantaneous power of the signal $\mathbf{s}$. Combining

this aspect with the improvement achievable through the exploitation of modulated nature of the FBSE, the definition of the novel BD criterion starts from the re-writing of Eq.6a through the FBSE, viz:

$$c_F^k = \frac{2}{(L-N+1)^2[J_1(\beta_k)]^2} \sum_{n=N-1}^{L-1} n|s(n)|^2 J_0\left(\frac{\beta_k n}{L-N+1}\right) \tag{12}$$

According to the definition of cyclic frequency given in Eq.4, the roots of the zero-order Bessel function can be expressed as:

$$\beta_k = \frac{2\pi \frac{k}{T}(L-N+1)}{f_s} \tag{13}$$

It has to be underlined how the fault related cyclic frequency is directly considered inside the re-estimated numerator of the ICS_2, named c_F^k, through the roots β_k enabling the fault detection through the FBSE. Now it is possible to substitute Eq.12 and Eq.6b into Eq.5. Moving to the matrix form, the revised form of the ICS_2, hereafter called ICS_{2FB}, is defined as follows:

$$ICS_{2FB} = \frac{4}{(L-N+1)^2|\mathbf{J}_1\mathbf{J}_1^H|^2} \frac{|\mathbf{s}|^{2H}\mathbf{J}_{0_n}\mathbf{J}_{0_n}^H|\mathbf{s}|^2}{|\mathbf{s}^H\mathbf{s}|^2} \tag{14}$$

with

$$\mathbf{J}_{0_n} = \begin{bmatrix} J_0\left(\frac{\beta_1(N-1)}{L-N+1}\right)(N-1) & \dots & J_0\left(\frac{\beta_K(N-1)}{L-N+1}\right)(N-1) \\ \vdots & \ddots & \vdots \\ J_0\left(\frac{\beta_1(L-1)}{L-N+1}\right)(L-1) & \dots & J_0\left(\frac{\beta_K(L-1)}{L-N+1}\right)(L-1) \end{bmatrix} \tag{15a}$$

$$\mathbf{J}_1 = [J_1(\beta_1)\dots J_1(\beta_k)\dots J_1(\beta_K)] \tag{15b}$$

where K is the number of considered cyclic frequency harmonics.

At this juncture, substituting Eq.3a in Eq.14 after a simple manipulation it is possible to write the ICS_{2FB} in form of generalized Rayleigh quotient as:

$$ICS_{2FB} = \frac{\mathbf{h}^H\mathbf{X}^H\mathbf{W}\mathbf{X}\mathbf{h}}{\mathbf{h}^H\mathbf{X}^H\mathbf{X}\mathbf{h}} = \frac{\mathbf{h}^H\mathbf{R}_{XWX}\mathbf{h}}{\mathbf{h}^H\mathbf{R}_{XX}\mathbf{h}} \tag{16}$$

where the weighting matrix $\mathbf{W}$ is given by:

$$\mathbf{W} = \begin{bmatrix} \ddots & & 0 \\ & \frac{4\mathbf{J}_{0_n}\mathbf{J}_{0_n}^{H}|\mathbf{s}|^2}{(L-N+1)^2|\mathbf{J}_1\mathbf{J}_1^{H}|^2} & \\ 0 & & \ddots \end{bmatrix} \frac{1}{\sum_{n=N-1}^{L-1} s[n]^2} \tag{17}$$

It should be noticed that the term inside the diagonal matrix comprises all the periodic components of $|\mathbf{s}|^2$, i.e. the cyclic frequencies to be investigated.

As demonstrated by Buzzoni for the ICS_2, the maximization of the ICS_{2FB} with respect to the filter coefficients $\mathbf{h}$ is equivalent to the maximum eigenvalue of the generalized eigenvalue problem described in Eq.8. The proposed method, hereinafter referred as FBBD, analogously to the CYCBD estimates the fault related excitation source through an iterative algorithm summarized as follows:

STEP 1: Assume a guess of $\mathbf{h}$. The target source is an independent distributed random variable and consequently a whitening filter is a good solution for the initialization of $\mathbf{h}$. At the same time, the convergence of the method is strongly influenced by the presence of deterministic sources (e.g. shaft unbalance, misalignment or gear mesh harmonics) that may lead to an increasing number of required iterations. Thus, a suitable solution for the computation of the whitening filter can be represented by an AR model filter, e.g. estimating the filter coefficients by using the Yule-Walker equations, in order to attenuate the predictable components, return a flat spectral density and consequently simplify the convergence of the method [26];

STEP 2: Estimate the weighting matrix $\mathbf{W}$ through Eq.17 after the computation of $\mathbf{s}$ through Eq.3a;

STEP 3: Solve Eq.8 obtaining the maximum ICS_{2FB} and the related filter coefficients $\mathbf{h}$;

STEP 4: Repeat from Step 2 with the filter $\mathbf{h}$ computed in Step 3 until convergence.

3.4.3 *Extension to the angular domain*

It is a matter of fact that the angular speed of rotating machines in real operating conditions often presents fluctuations around the constant working target. Consequently, the periodicities hidden into the vibration signatures should be investigated in the angular domain rather than in the time domain in order to release the analysis from the rotational speed [29, 62]. In this direction, the FBBD may be further enhanced for the rotating system diagnostics through the extension of the proposed indicator into the angular domain.

As suggested by Borghesani et al. [63], it is possible to move from time to angular domain avoiding resampling through a change of variable starting from the measure of the instantaneous speed or the instantaneous angular position. According to this methodology, under the assumption of time/angle dependent signal the c_F^k can be rewritten as:

$$c_F^k = \frac{2}{\Theta\dot{\Theta}J_1^2(\beta_k)} \sum_{n=N-1}^{L-1} \theta(n)s(n)J_0\Big(\frac{\beta_k\theta(n)}{\Theta}\Big)\dot{\theta}(n) \tag{18a}$$

$$\Theta = \sum_{n=N-1}^{L-1} \Delta\theta_n \tag{18b}$$

$$\dot{\Theta} = \sum_{n=N-1}^{L-1} \dot{\theta}(n) = \frac{\Theta}{\Delta t} \tag{18c}$$

The positive roots of the zero-order Bessel functions may be expressed in the angular domain by dividing both numerator and denominator of Eq.13 by the rotational frequency, viz:

$$\beta_k = \frac{2\pi\Omega_k(L-N+1)}{M} \tag{19}$$

where Ω_k is the cyclic order and M is the maximum investigable order according to the sampling frequency f_s.

The extension of the FBBD to the angular domain through Eq.18 is hereafter referred as AngleFBBD. The overall proof for moving from Eq.12 to Eq.18 can be found in Appendix A.1.

3.5 APPLICATION ON SYNTHESIZED SIGNALS

This section provides the validation of the proposed BD method using simulated signals that reproduce faulty signatures under both stationary and non stationary operating conditions. The results are compared with those obtained through the application of CYCBD in order to illustrate how the different mathematical nature of the two criteria reflects on the number of considered harmonics and thus on the computational time.

3.5.1 *Description of simulated signals*

The experimental validation through synthesized signals has been carried out in Matlab environment exploiting also the code provided in Ref. [55]. The simulation regards two signals reproducing typical fault related cyclostationary patterns under both stationary conditions, i.e. system operating at constant speed, and non-stationary conditions, i.e. angular speed fluctuating according to a specific law:

1. train of cyclic impulses with Gaussian distributed amplitudes and additive Gaussian background noise (SNR = -18dB);

2. impulse train with cyclic frequency fluctuating according to a sinusoidal law and additive Gaussian background noise (SNR = -18dB);

which, for simplicity, are called $\mathbf{x}_{\mathrm{time}}$ and $\mathbf{x}_{\mathrm{angle}}$, respectively. The parameters used for the generation of the simulated signals are summarized in Tab.1 where f_s is the sampling frequency, T is the impulse period and σ is the standard deviation of the amplitude of the impulses.

These signals are generated according the general formulation of BD given in Eq.1, neglecting the periodic pattern $\mathbf{p}$, as follows:

$$\mathbf{x} = \mathbf{s}_0 * \mathbf{g}_s + \mathbf{n} * \mathbf{g}_n \tag{20}$$

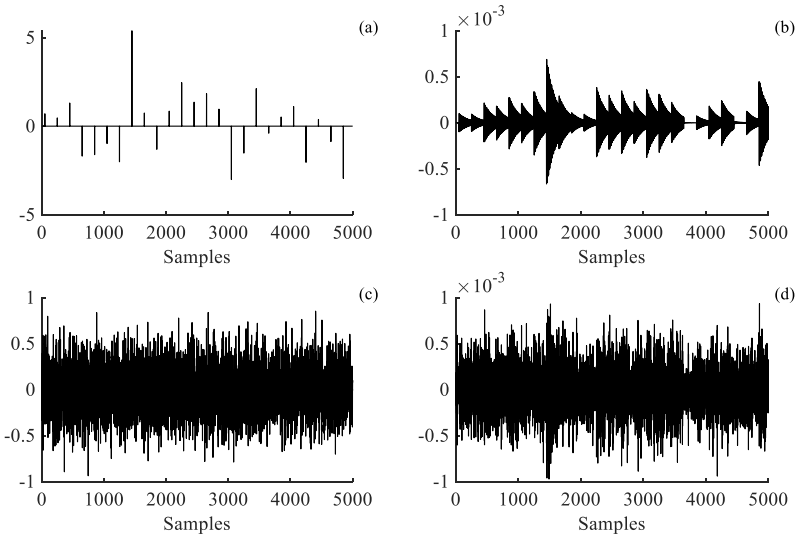

Figure 6: Synthesized signal $\mathbf{x}_{time}$: (a) impulsive pattern $\mathbf{s}_0$ with cyclic period T and Gaussian distributed amplitudes, (b) $\mathbf{s}_0$ convolved with its IRF $\mathbf{g}_s$, (c) additive Gaussian background noise with SNR = -18dB, (d) overall signal

For these simulations the IRFs $\mathbf{g}_s$ and $\mathbf{g}_n$ have been modelled according to the response of a damped Single Degree of Freedom (SDOF) system to a unitary impulse in the time domain [64], viz:

$$\mathbf{x} = Ae^{-\zeta\omega_n t}\sin(\omega_d t) \tag{21a}$$

$$\omega_d = \omega_n\sqrt{1-\zeta^2} \tag{21b}$$

where A is the response amplitude, ζ is the damping coefficient and ω_n is the resonance frequency. Starting from Eq.21, under the hypothesis of viscous and sub-critical damping, $\mathbf{g}_s$ and $\mathbf{g}_n$ can be obtained in terms of acceleration by taking the second derivative with respect to time. Tab.2 reports all the parameters used for the computation of the IRFs. Fig.6 and Fig.7 show the simulated signals $\mathbf{x}_{time}$ and $\mathbf{x}_{angle}$ implemented in Matlab environment.

3.5.2 *Results and discussion: stationary conditions*

Before analysing the comparison between the results obtained with both FBBD and CYCBD on $\mathbf{x}_{time}$, it is necessary to discuss the input parameters chosen for

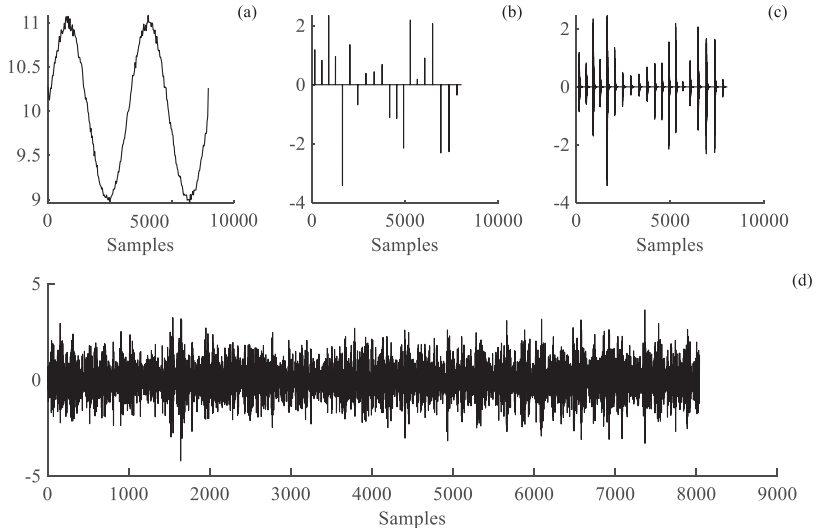

Figure 7: Synthesized signal $\mathbf{x}_{angle}$: (a) fluctuating sinusoidal cyclic frequency, (b) impulsive pattern $\mathbf{s}_0$ with cyclic period T and Gaussian distributed amplitudes, (c) $\mathbf{s}_0$ convolved with its IRF $\mathbf{g}_s$, (d) overall signal

Table 1: Parameters used for the synthesized signals

	f_s (Hz)	L (samples)	T (samples)	σ (/)	SNR dB
$\mathbf{x}_{time}$	1000	5000	200	2.0	-18
$\mathbf{x}_{angle}$	4000	8000	364 ÷ 440	2.0	-18

Table 2: Parameters used for the IRFs

	A		ζ		ω_s (rad/s)	
	$\mathbf{g}_s$	$\mathbf{g}_n$	$\mathbf{g}_s$	$\mathbf{g}_n$	$\mathbf{g}_s$	$\mathbf{g}_n$
$\mathbf{x}_{time}$	1.67210^{-10}	1.14810^{-10}	0.007	0.06	942	2638
$\mathbf{x}_{angle}$	1.67210^{-10}	1.14810^{-10}	0.003	0.06	1570	2638

the application of both methods. In particular, it has been demonstrated [51]

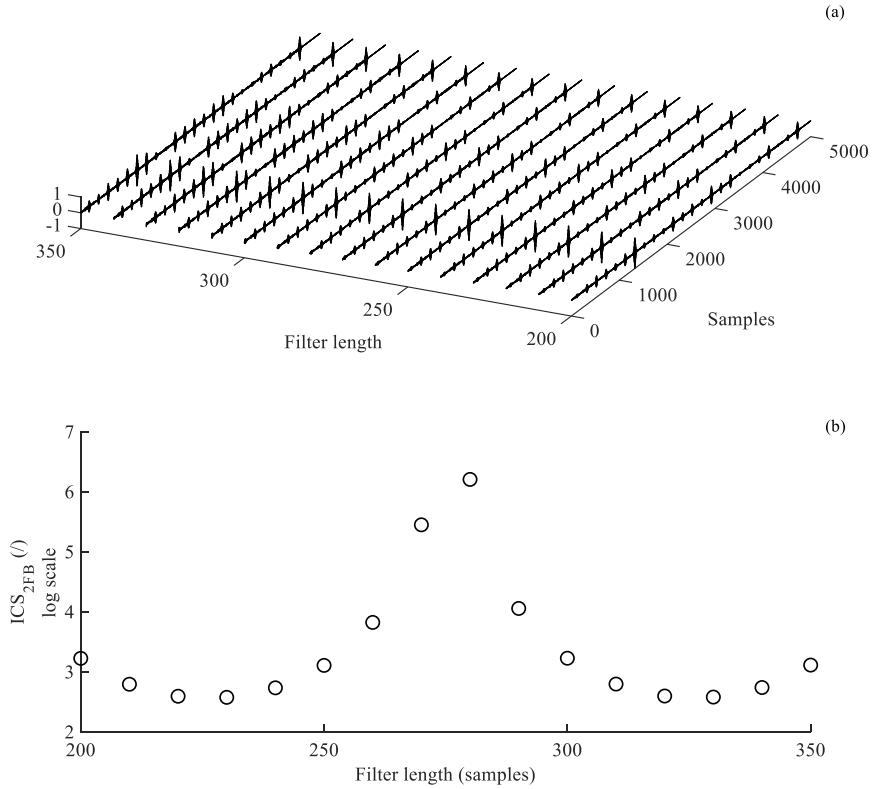

Figure 8: Effect of filter length N on the FBBD analysis: (a) estimated sources, (b) ICS_{2FB} values

how the filter length N plays a pivotal role in the final quality of the estimated source and consequently it has to be carefully set. In general, the basic rule for the correct extraction of the hidden periodicity requires the use of a filter length higher than the fault periodicity in order to deconvolve an entire period and identify the fault related pulses. Furthermore, it has to be reminded the amplitude invariant nature of the BD that allows the normalization of the estimated sources with respect to their maximum value in order to make them comparable each other. Fig.8 explains the effect of the filter length on the FBBD analysis in terms of estimated sources and maximized ICS_{2FB} obtained by applying the method taking into account filter lengths from 200 to 350 samples. The behaviour of the algorithm from the two points of view is quite contrasting.

Since all the tested lengths are higher than the cyclic period, different values of N do not seem to affect the reconstruction quality of the deconvolved cyclostationary source (see Fig.8(a)). On the other hand, the BD criterion is strongly dependent on the filter length (Fig.8(b)). This aspect can be explained considering the definition of the weighting matrix **W** given in Eq.17: all the periodic components of **s** are comprised inside **W** and consequently, different values of N lead to different estimations of the eigenvalue, i.e. the maximized criterion. Therefore, according to these considerations, it is possible to define a procedure for the optimal filter length evaluation based on the maximization of the ICS_{2FB} over a range of retrieved values of N, viz:

$$ICS_{2FB} = \max_{N \in [N_{min}, N_{max}]} ICS_{2FB}(N) \tag{22}$$

where $ICS_{2FB}(N)$ denotes the proposed indicator conditioned to the period N. As already stated, the lower limit N_{min} has to be set higher than the fault period. The choice of the higher limit N_{max} is driven by the need of limiting the computational time, exponentially increasing with the filter length due to the convolution operation. According to the aforementioned protocol, the following analysis has been carried out with $N = 280$ samples. The same filter length has been considered for the CYCBD analysis.

The target waveform and the resulting patterns deconvolved through both FBBD (left side) and CYCBD (right side) taking into account an increasing number of considered cyclic harmonics, i.e. 10, 30, 50, 100 (from top to bottom), are compared in Fig.9. It has to be noticed that the filtering operations required by the BD methods leads to the deconvolution of sources delayed with respect to the target waveform. Nevertheless, the choice of FIR filters allows to ignore this aspect for this specific application due the constant delay [65] applied to all the patterns that consequently does not affect the comparison. Fig.9 shows how both FBBD and CYCBD enable the detection of the correct pulse periodicity with the actual relative pulse amplitude. However, it can be seen that FBBD estimates zero-mean patterns, i.e. pulses symmetrical with respect to the zero, regardless the number of considered terms (Fig.9(b-d-f-h)). The same behaviour can be noticed for CYCBD for low harmonic number (Fig.9(c)). Nevertheless, in this case an increasing number of series terms (Fig.9(e-g-i)) enables the estimation of the correct pulse sign. However, usually in the rotating machines diagnostics field the main attention has to be pointed out on the fault frequency

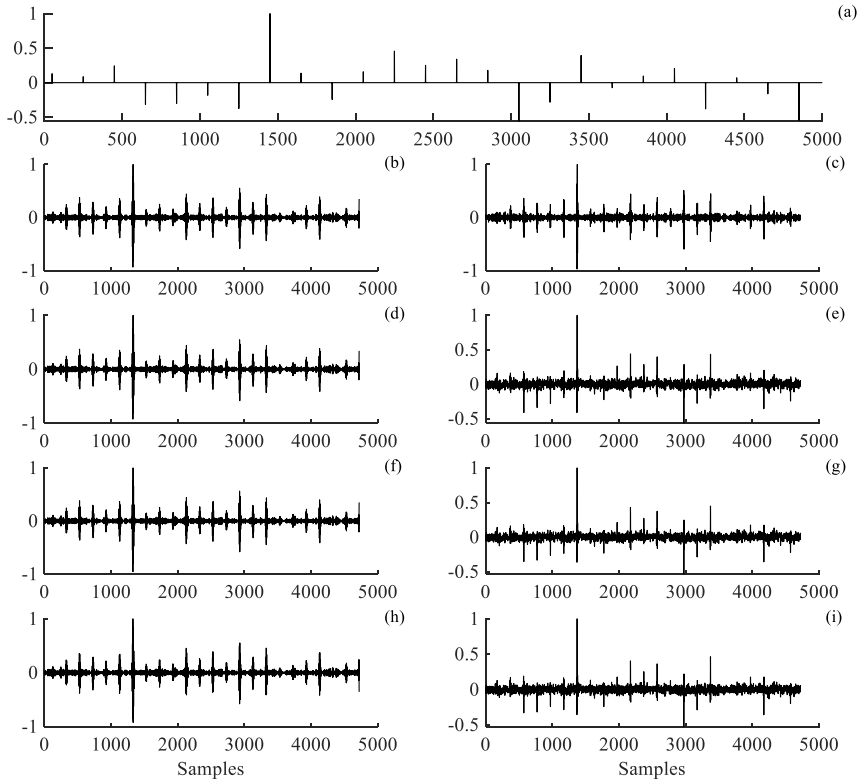

Figure 9: Estimated sources for the analysis of $\mathbf{x}_{time}$ for increasing number of considered cyclic harmonics (from top to bottom): (a) target impulsive pattern, (b-d-f-h) results for FBBD method, (c-e-g-i) results for CYCBD method

and on the relative amplitude between different damage severities instead of the absolute amplitude of a vibration signature. Therefore, taking also into account the BD inability of recovering the real excitation amplitude, the issue related to the peak sign estimation can be considered irrelevant for diagnostic purposes.

The effect of the different mathematical natures of FBSE and FSE can be explained through the comparison of the Euclidean distances, i.e. the squared sum of the distance between each point of two signals [66], between the sources estimated with both FBBD and CYCBD and the target pattern, depicted in Fig.10. From the CYCBD standpoint (dotted line), there is a strong variation

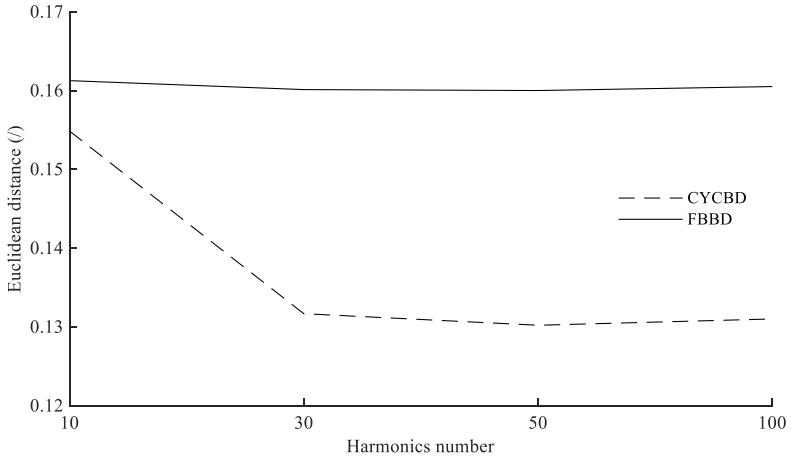

Figure 10: Euclidean distance between the deconvolved source and the target excitation pattern

on the reconstruction quality, i.e. the measure of how the estimated source reproduces the target impulsive pattern, moving from 10 to 30 considered harmonics. This can be easily understood looking to Fig.9(c-e) where the deconvolved waveform deeply changes in terms of peaks amplitude. A further increment of harmonics taken into account in the algorithm does not seem to modify the obtained signature as confirmed by the constant Euclidean distance in Fig.10 for 50 and 100 harmonics. On the contrary, considering the FBBD results (continuous line), the Euclidean distance remains quite constant independently from the number of cyclic harmonics. As a consequence, the extracted pattern reaches the maximum reconstruction quality even with 10 harmonics and does not change with an increasing number of considered terms in Fig.9(b-d-f-h). This results clearly demonstrates the consideration previously explained in Sec.3.4.1: the modulated nature of the FBSE better fits for the description of transient signals due to the lower number of series expansion terms required for the modelling of amplitude modulation localized in short time spans. The effects on the BD method efficiency directly depends on the weighting matrix $\mathbf{W}$ definition. In fact, its computation requires the analysis of each cyclic harmonic one by one at each algorithm iteration, thus the higher the number of required considered harmonics, the higher the number of algorithm operations leading to an increasing computational cost.

Table 3: Computational times for $\mathbf{x}_{time}$

Harmonics	10 (s)	30 (s)	50 (s)	100 (s)
CYCBD	2.81	4.09	5.64	8.92
FBBD	3.13	4.42	5.97	9.06

Quantitatively speaking, Tab.3 reports the computational times required by FBBD and CYCBD for the analysis of $\mathbf{x}_{time}$. These times refer to a desktop computer Dell XPS 8700 equipped with an Intel CoreTm i7-4790 @ 3.6GHz processor. Although the higher mathematical complexity of the FBSE reflects on a higher computational time with the same number of harmonics, the effect of the lower number of series terms required for the pulses description can be seen by comparing the values referred to the signals that reach the maximum reconstruction quality for each method. In fact, taking into account the FBBD analysis with 10 harmonics and the CYCBD analysis with 30 harmonics the difference is not negligible, 0.9s lower for the proposed method. Therefore, the ICS_{2FB} may represent a more suitable indicator for real time condition monitoring, pivotal aspect in the nowadays industrial scenario ruled by the need of time (and consequently costs) reduction.

Moreover, Fig.9 highlights another advantaging aspect of the proposed criterion. One of the critical points regarding the BD results is related to the lower magnitude peaks, challenging (even impossible) to be detected due to the residual background noise still present in the estimated signal. It can be noticed that the FBBD whitened sources comprise background noise with lower specific weight in the overall waveform. Therefore, the proposed indicator seems to enable the extraction of lower amplitude peaks, usually covered by the residual background noise. An example can be provided by the comparison of the final waveform, i.e. the sources reaching the maximum reconstruction quality, from both two methods in the range between 3500 and 4000 samples (Fig.9(b-e)). Due to the lower specific weight of the background noise, the FBBD enables the recovering of these low amplitude peaks, even masked in the CYCBD estimated source.

FBBD, such as CYCBD and other BD methods, requires the a priori knowledge of the carrier cyclic frequency of the fault related pattern. However, the

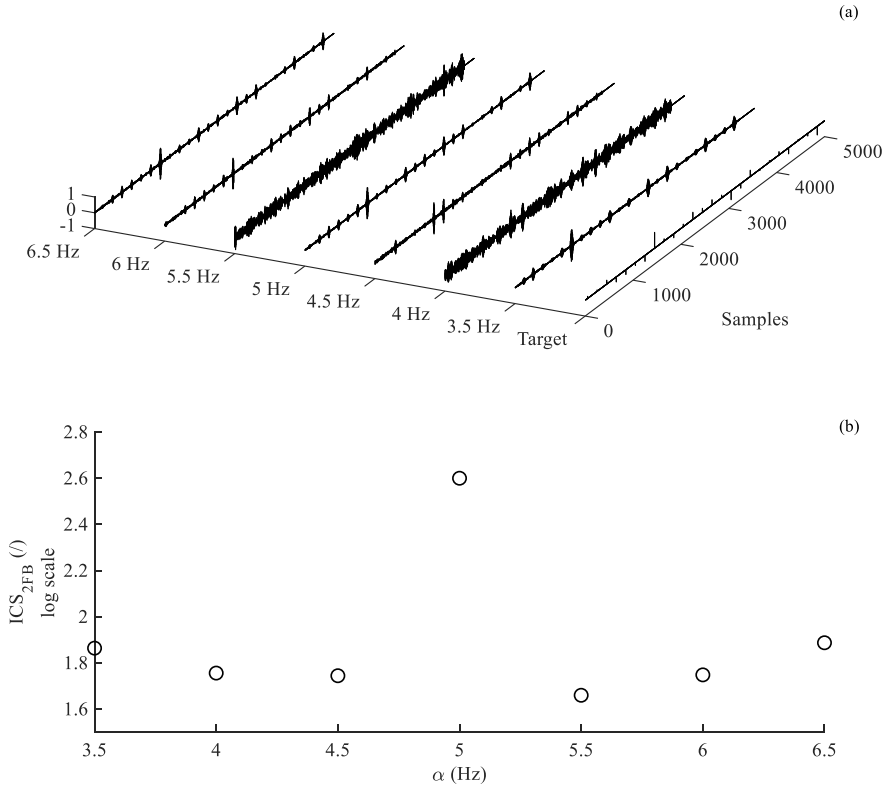

Figure 11: Effect of the a priori knowledge of the fault related cyclic frequency on the FBBD analysis: (a) estimated sources, (b) ICS_{2FB} values

considered cyclic period directly induces the estimation of the weighting matrix **W**.The choice of the correct frequency leads to the estimation of **W** that weights the autocorrelation matrix according to a specific power flow hidden inside the signal. As a consequence, if the measured signal is characterized by a cyclostationary behaviour, this power flow is maximized at a given cyclic frequency [55]. In order to explain this aspect, the robustness of the FBBD with respect to the considered cycle frequency has been analysed taking into account a set of different carrier frequencies. Being the sample frequency of $\mathbf{x}_{\mathrm{time}}$ set at 1000Hz, the pulse frequency is 5Hz. thus, the analysis has been performed oh a set of frequencies from 3.5Hz to 6.5Hz in steps of 0.5Hz. Fig.11 depicts the influence of the fundamental frequency oh the FBBD results, in terms of

estimated impulsive pattern and maximized BD criterion. It is interesting to observe that, unlike the effect of the filter length N, the choice of an unsuitable cyclic frequency strongly affects the results also from the estimated source point of view. In fact, as it can be seen in Fig.11(a), a carrier frequency different from the actual one leads to the recovering of a waveform which differs (even strongly) from the target pattern. Fig.11(b) demonstrates the aforementioned relation between cyclic frequency and weighting matrix $\mathbf{W}$. In fact, the ICS_{2FB} reaches the maximum value when the correct frequency is considered for the analysis and the difference with respect the other values is significant (it has to be noticed the logarithmic scale of the plot).

The real industrial environment represents a very challenging field of application for BD methods and consequently the robustness of the proposed algorithm has to be demonstrated taking also into account different interferences in the measured signal that may affect the diagnostic capability of FBBD.

3.5.2.1 *Effect of strong interferences*

In order to prove the effectiveness of the proposed BD method, the comparison between FBBD and CYCBD has been also performed taking into account two different families of interferences:

- Distributed interference: the Gaussian background noise weight effect is analysed by considering increasing values of SNR in the original signal, i.e. -21dB and -25dB. These signals are displayed in Fig.12.
- Localized interference: a single random pulse convolved with $\mathbf{g}_s$ is added to the original signal. The resulting waveform is described in Fig.13.

Fig.14 describes the sources deconvolved through both FBBD (left side) and CYCBD (right side) for increasing SNR (from top to bottom) taking into account the number of cyclic harmonics that ensures the maximum reconstruction quality, i.e. 10 and 30, respectively. The results clearly highlights one of the aspects emerged in the previous section: the capability of FBBD to detect the weakest peaks due to the lower residual background noise in the final pattern. Fig.14 shows how, for increasing SNR values, the CYCBD extracts a

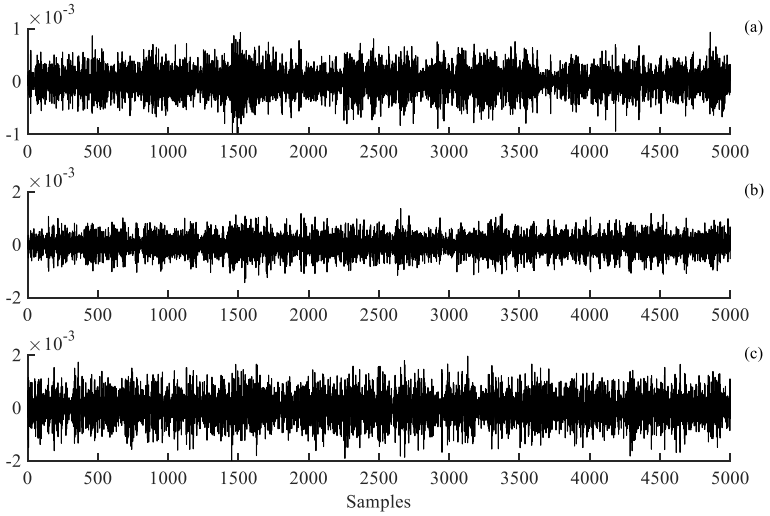

Figure 12: Simulated signals for increasing SNR values: (a) original signal with −18dB, (b) −21dB, (c) −25dB

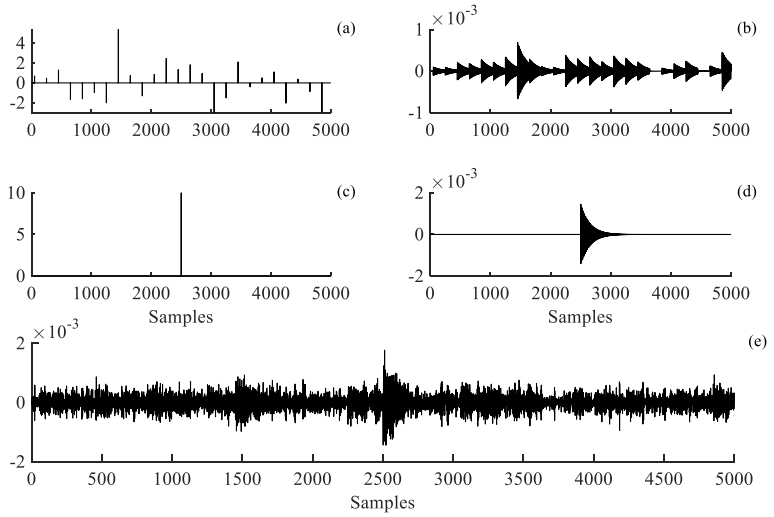

Figure 13: Synthesized signal with strong random pulse: (a) impulsive pattern $\mathbf{s}_0$, (b) $\mathbf{s}_0$ convolved with $\mathbf{g}_s$, (c) strong random pulse, (d) random pulse convolved with $\mathbf{g}_s$, (e) overall signal

waveform with residual background noise with higher specific weight on the

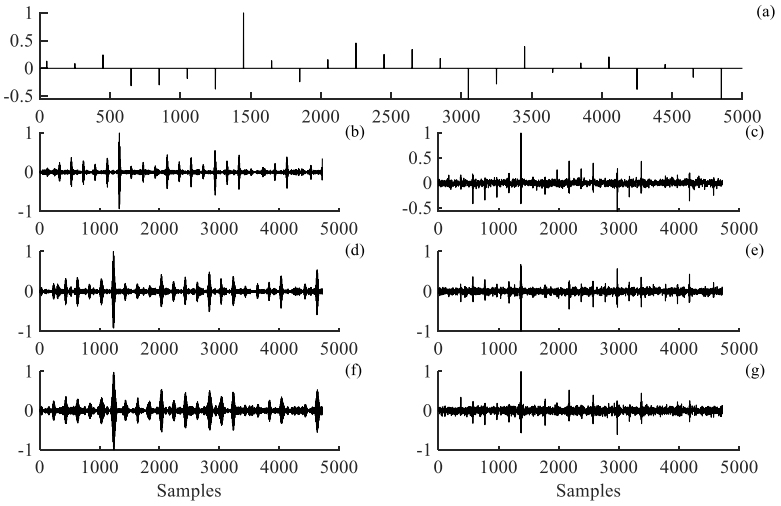

Figure 14: Estimated sources from $\mathbf{x}_{time}$ for increasing SNR (from top to bottom): (a) target pattern, (b-d-f) from FBBD analysis, (c-e-g) from CYCBD analysis

overall signal and consequently even more peaks (each of them with significant amplitude) are difficult to be detected at high SNRs. For example, taking into account the case with -25dB in Fig.14(g), the method is no longer able to extract all the peaks after sample 3500 being them buried under the residual noise. On the contrary, the FBBD allows the identification of the weakest peaks, independently from the considered SNR due to a residual background noise that seems to keep a constant specific weight on the overall estimated source for all cases.

Moving to the localized interference case, Fig.15 describes the results on the signal depicted in Fig.13 obtained through both FBBD and CYCBD in the case of maximum reconstruction quality reached. In both cases, the strong single pulse is still identifying in the estimated sources but the introduced interference does not disable the detection of the fault related cyclostationary pattern. As well as the extraction of the lower amplitude peaks, this analysis reveals another interesting aspect. Comparing Fig.15(b-c) it is possible to note the presence of some artificial peaks around the strong random one between samples 2300 and 2600. However, the source estimated through FBBD seems to reduce this numerical artefact due to the presence of some peaks only before

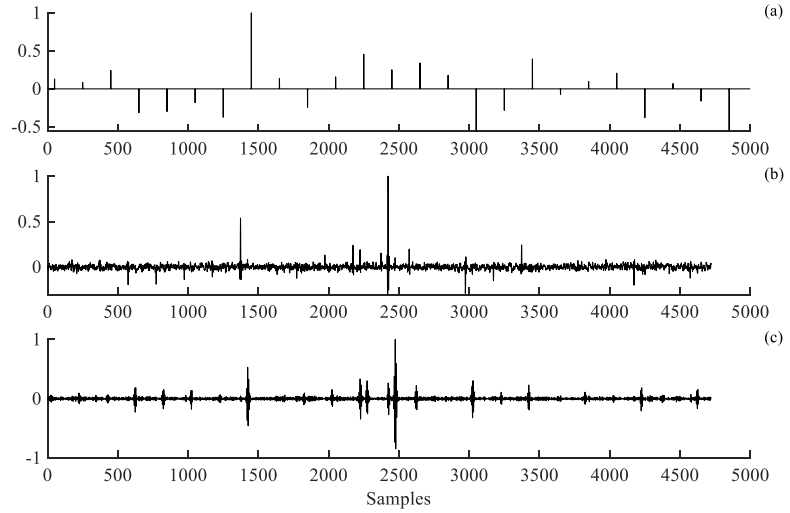

Figure 15: Estimated sources from $\mathbf{x}_{time}$ in case of presence of strong random pulse: (a) target pattern, (b) from CYCBD analysis, (c) from FBBD analysis

the interference instead of the CYCBD case where those peaks can be seen also after the added one.

The above discussed results proved the robustness of the proposed BD algorithm to strong interferences, highlighting the suitability of FBBD for the real industrial environment, main target for of the research in the mechanical field.

3.5.3 *Results and discussion: non-stationary conditions*

Before analysing the results obtained through the application of FBBD and CYCBD on $\mathbf{x}_{angle}$, some considerations regarding the selection of the optimal FIR filter length are mandatory also in this contest, even more due to the non-stationarity given by the fluctuating cyclic period hidden inside the signal. In order to respect the main rule related to the choice of filter lengths higher than the fault periods, the consideration stated in Sec.3.5.2 may be extended to variable cyclic periods by taking into account the highest value into the signal. As a result, the period range used for the ideal filter identification in Eq.22 must be set higher than the highest period to be recovered. For the

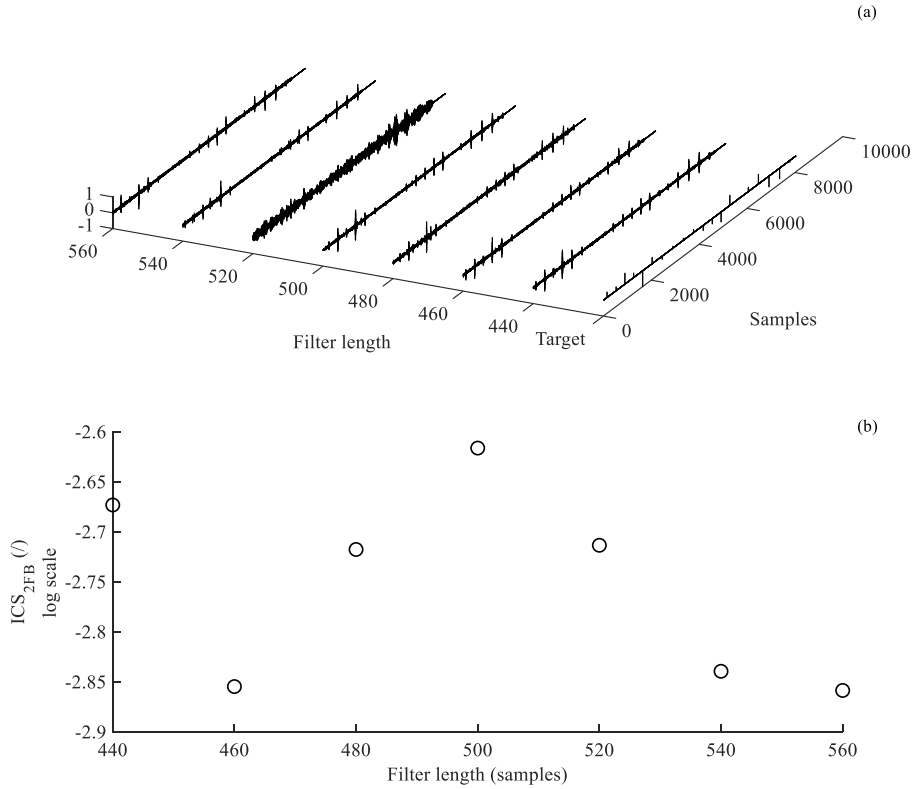

Figure 16: Effect of filter length N on the AngleFBBD analysis: (a) estimated sources, (b) ICS_{2FB} values

analysis of $\mathbf{x}_{angle}$, since the maximum period is 440 samples, the optimal filter length for AngleFBBD has been retrieved into a range from $N_{min} = 440$ to $N_{max} = 560$ samples. Fig.16 describes the ICS_{2FB} values and the respective estimated sources as function of the filter length N. In the stationary case, an higher influence on the criterion rather than on the estimated pattern has been demonstrated. For the non-stationary case, also the resulting waveform is strongly affected by the considered filter length. However, comparing Fig.16(a) and Fig.16(b) it is clear how the N that maximizes the ICS_{2FB} leads to the estimation of the source that better reproduces the target signature. According to aforementioned considerations, the procedure described in Eq.22 retains its validity for the angular domain analysis.

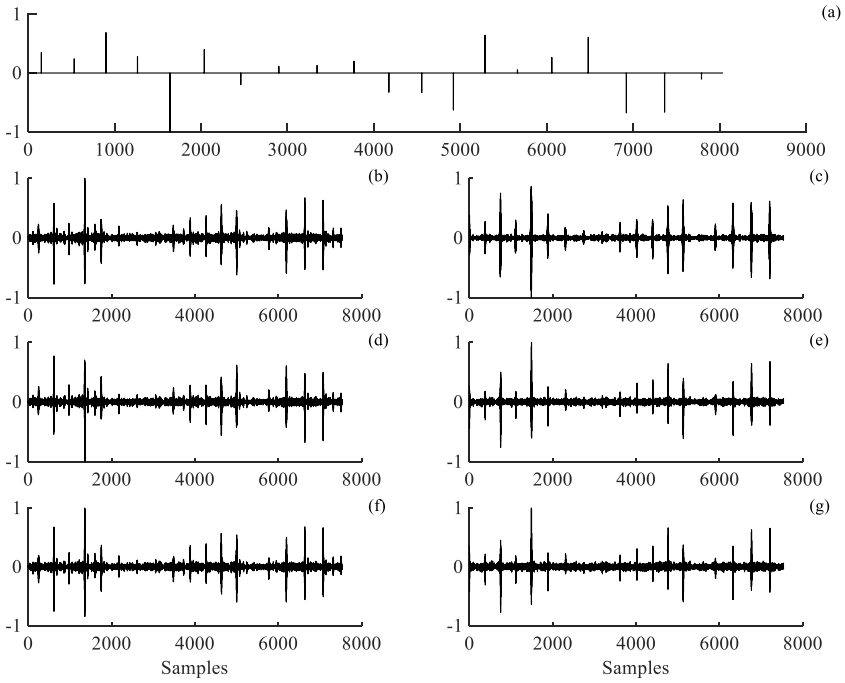

Figure 17: Estimated sources for the analysis of $\mathbf{x}_{angle}$ for increasing number of considered cyclic harmonics (from top to bottom): (a) target impulsive pattern, (b-d-f) results for FBBD method, (c-e-g) results for CYCBD method

The target waveform and the resulting patterns deconvolved through both FBBD (left side) and CYCBD (right side) taking into account of an increasing number of considered cyclic harmonics, i.e. 10, 30, 50 from top to bottom, are compared in Fig.17. It can be noticed that FBBD enables the estimation of the non-stationary excitation source through the analysis in the angular domain, releasing it from the period fluctuations. Furthermore, the main FBBD's cons highlighted in the previous section, i.e. the inability on recovering the real peak sign, is less significant in the angular domain. In fact, Fig.17(c-e-g) shows the inability of CYCBD to recover the correct peak sign regardless of the number of considered harmonic orders, despite the estimation of a non symmetrical waveform. The comparison between Fig.17(b-d-f) and Fig.17(c-e-g) confirms the physical implications of the different mathematical natures of FBSE and FSE. The FBBD reaches the maximum reconstruction quality already considering 10 harmonics and a further increment does not improve the results. On the

other hand, for the same purposes the CYCBD requires at least 30 harmonics confirming the considerations described in the stationary case. These results prove how the ICS_{2FB} allows the extraction of hidden cyclostationary sources with shorter cyclic frequency sets also in the angular domain. The quantitative

Table 4: Computational times for $\mathbf{x}_{angle}$

Harmonics	10 (s)	30 (s)	50 (s)
CYCBD	3.06	4.39	5.94
FBBD	3.51	4.83	6.17

consequence of this aspect is clearly explained by the computational times summarized in Tab.4. These values confirm the consideration of the stationary case: although the higher computational time required by the FBBD being equal the number of harmonics, the lower number of series terms needed for the pulse description, allows the reduction of the computational cost required for reaching the maximum reconstruction quality.

This preliminary benchmark takes into account ad hoc synthesized signals reproducing typical cyclostationary patterns of rotating machines in order to demonstrate the strictly relations between the mathematical nature of the series expansion basis functions and the number of cyclic harmonics required by the algorithm. Nevertheless, this analysis does not prove the effectiveness of the proposed criterion in the discrimination of the health conditions. Thus, in the next section a further validation is carried out on real vibration signals.

3.6 APPLICATION TO REAL SIGNALS

This section provides the application of the proposed BD method on real signals under both stationary and non stationary operating conditions in order to demonstrate the ability of the FBBD on the fault detection, since from the early damaging stages. The validation takes into account both datasets from academic laboratories and real industrial case studies.

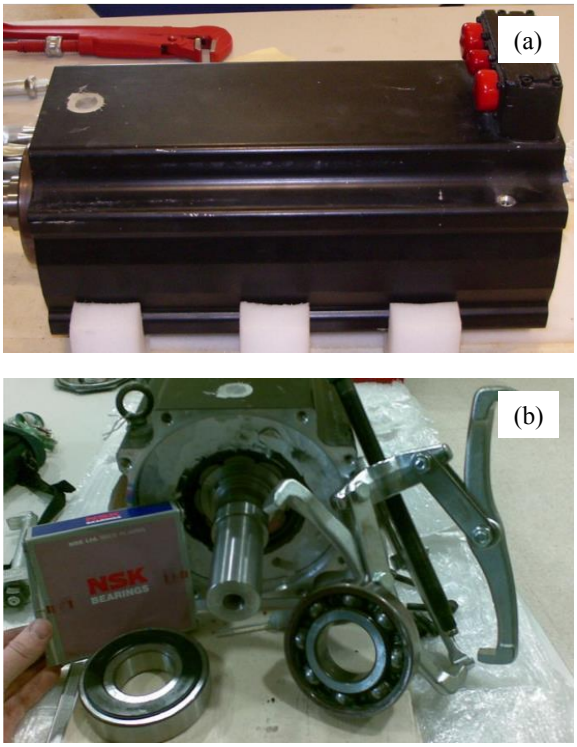

Figure 18: Rockwell Automation MPL-B680B AC brushless motor and NSK 6309 bearing. (a) lateral view of the motor, (b) Frontal view with bearing

3.6.1 *Detection of bearing faults under non stationary operating regime*

The first experimental validation of the proposed algorithm regards a real case of industrial diagnostics. The device under test is the front ball bearing of a servomotor used in a packaging machine. The servomotor is an MPL-B680B AC brushless motor by Rockwell Automation equipped with a SICK Hiperface ® encoder. The latter is a sin/cos encoder with 1024 periods per revolution and a resolution of 32768 steps per revolution. The velocity signal of the motor and the torque signal are retrieved using the analog outputs available in the motor drive, a Kinetix 6000 series BM-01 by Rockwell Automation. The signals are connected to a National Instruments acquisition board, made by a CDAQ-9172 backplane upon which a NI-9234 module collected an industrial accelerometer output (Entek 81001 - monoaxial) and a NI-9215 module is

Table 5: Bearing type NSK 6309 geometry parameters and characteristic frequencies

	NSK 6309 (/)	@Max rpm (Hz)	@Min rpm (Hz)	Order (/)
Outer diameter D (mm)	100	/	/	/
Inner diameter d (mm)	45	/	/	/
Width B (mm)	25	/	/	/
Number of spheres N	8	/	/	/
BPFI	/	41.6	10.8	5
BPFO	/	25.3	6.7	3
Cage	/	3.4	0.9	0.4

connected to the Kinetix analog output. This servomotor actuation provides rapid variations of speed and inversions in the rotation sense of the shaft. An external load of 3000 N is applied to the motor shaft in radial direction. The tested device is a NSK 6309 deep groove single-raw ball bearing, whose geometrical characteristics are reported in Tab.5. Fig.18 shows the brushless motor and the frontal bearing under test. The test motion profile is cyclic, following a polynomial profile, and it consists of both a clockwise and counter clockwise rotation of the shaft. Fig.19(d) shows the speed profile as returned by the drive controller of the motor, while the units have been normalized to the maximum value for confidentiality agreements with the industrial partner. The cyclic periodicity is equal to 0.972 seconds (1.029 Hz). Fig.19(e) shows the angular profile of the shaft and it is computed by integration of the speed profile. The sampling frequency chosen for all the experiments has been set 10 kHz; both the acquisition board and the accelerometer would have allowed an higher bandwidth, but the analogue output of the Kinetix, being generated by a DAC that converts digital information processed internally by the drive, has a bandwidth limited to 7.2 kHz. Thirteen bearings in different health conditions have been tested and the data has been collected in a database. In this thesis, three bearings have been chosen from the database: two bearings artificially damaged by using an electric drill (namely Bearing 4 and 5) and one healthy bearing (namely Bearing 8). A dent was made on the outer race surface of Bearing 4, and on the inner race surface of Bearing 5. Bearing 8 was in mint

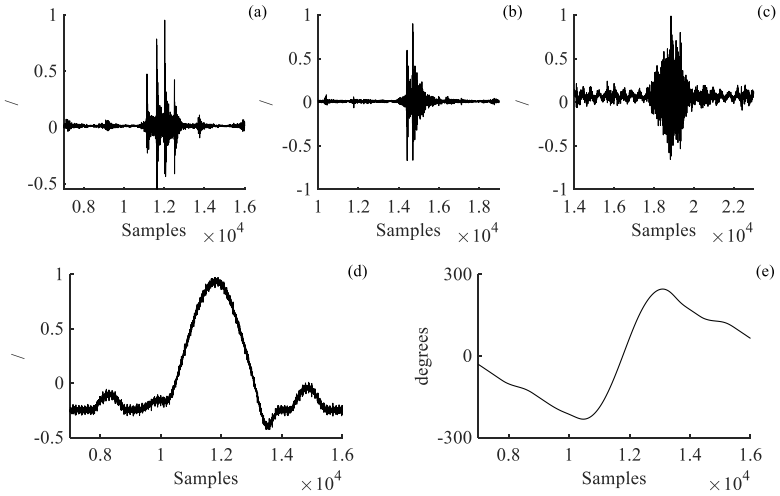

Figure 19: Observation showing part of the normalized test signals: (a) Bearing 4, (b) Bearing 5, (c) Bearing 8, (d) normalized speed profile, (e) instantaneous angular position

condition. Fig.19(a-c) shows the vibration signals corresponding to a machine cycle for the tested bearings. For confidentiality reasons, the amplitude values have been normalized with respect to the maximum value among the three signals to preserve the different excursion ranges. The Ball Pass Frequency Inner race (BPFI), the Ball Pass Frequency Outer race (BPFO) and the cage frequency - summarized in Tab.5 - correspond to normalized periods of 925, 1490 and 11110 samples, respectively, taking into account the lower rotation frequency, i.e. the higher fault period values. The proposed indicator has been calculated using the protocol proposed in the previous section, taking into account 10 order harmonics. The optimal filter length of each aforementioned fault period has been retrieved according to Eq.22 into a range from 900 to 1100 samples (BPFI), from 1500 to 1700 samples (BPFO) and from 11150 to 11350 samples (cage).

The ICS_{2FB} values for each characteristic frequency of the analysed bearings are displayed in Fig.20. The results confirm what have been detected on the physical systems:

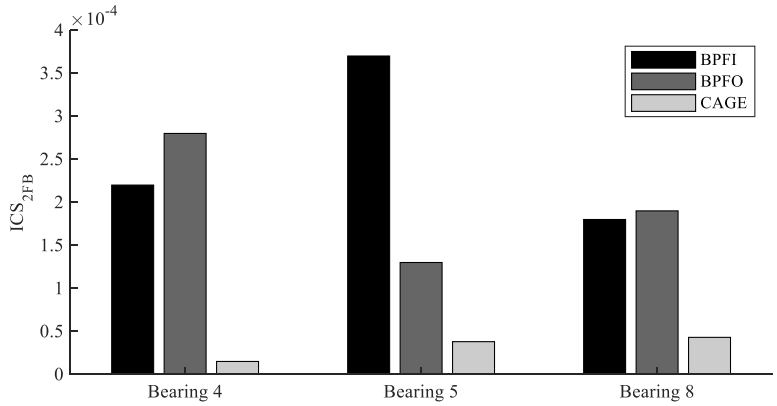

Figure 20: ICS_{2FB} values for Bearing 4, Bearing 5 and Bearing 8 taking into account BPFI, BPFO and cage frequency

BEARING 4: The highest ICS_{2FB} value (2.8e-4) is related to the BPFO. However, the difference between the fault related value and the highest value related to a non-faulty characteristic frequency (about 25 %) for this bearing is sensibly lower with respect to Bearing 5. This particular aspect can be explained through the analysis of the bearing geometry: remembering the fault orders given in Tab.5, between the 10 harmonics of the BPFI considered for the analysis it is possible to find some orders, i.e. 15, 30 and 45, that at the same time are harmonics of the BPFO, i.e. represent the least common multiple between BPFI and BPFO and its first integer multiples. Consequently, these orders contribute to slightly increase the ICS_{2FB} value also for the BPFI analysis, although no fault is related to this order. This interesting industrial case study demonstrates the robustness of the proposed indicator to possible unfavourable bearing geometries that may lead to mistaken fault detection.

BEARING 5: It is possible to note that the BPFI related ICS_{2FB} clearly prevails on the other values which can be considered as negligible. The issue previously explained for Bearing 4 not seems to affect the different frequency related values in this case. This can be understood by considering that for this bearing the non-faulty frequency, i.e. the BPFO, is the lower one and consequently in the 10 harmonics range a lower number of integer

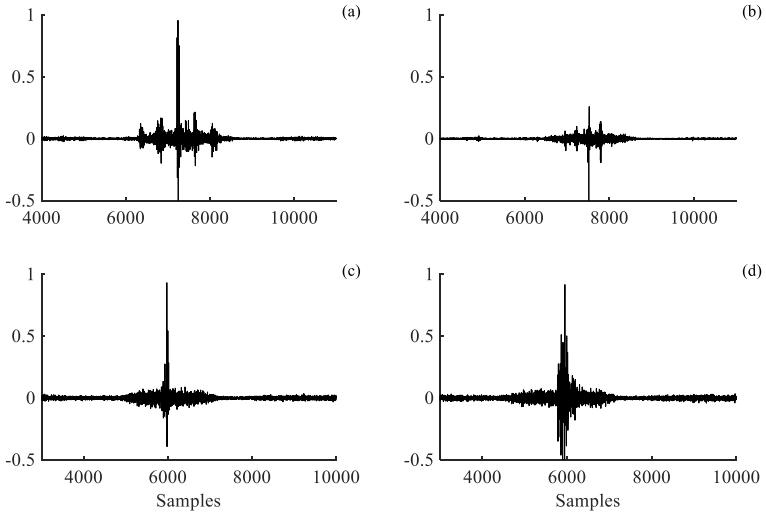

Figure 21: Estimated sources: (a) BPFO of Bearing 4, (b) BPFI of Bearing 5, (c) BPFI of Bearing 8, (d) BPFO of Bearing 8

multiples of the least common multiple are considered and contributes to increase the ICS_{2FB} value for the non-faulty characteristic frequency.

BEARING 8: The comparability of ICS_{2FB} values, combined with their lower values with respect to the fault related cases, well describes the bearing healthy conditions.

This quantitative analysis demonstrates the sensitivity of the proposed indicator on the emergence of cyclostationarity hidden into the vibration signal, proving its effectiveness on the discrimination between healthy and faulty bearing conditions, also under strongly non-stationary working conditions.

Fig.21 illustrates the estimated patterns related to the maximum value of ICS_{2FB}, i.e. the faulthy characteristic frequency, for Bearing 4 and Bearing 5 and the BPFI and BPFO related sources for Bearing 8. Considering both Bearing 4 and Bearing 5 (see Fig.21(a-b)), it can be clearly seen that the impulsive source given by the local fault is detected by the proposed method. Although for the BPFO the defect is visible already from the raw time signal (Fig.19(a)), the most interesting result is given by the BPFI fault detection. Due to the measurement sensor position closer to the outer ring, the direct detection of

the fault vibration signature from the acquired signal represents a challenging task. In fact, the impulsive excitation given by the impact between defect and rolling elements is damped by the system transfer function between inner and outer race and this filtered signal is directly measured by the accelerometer. Nevertheless, in the source estimated with the AngleFBBD method, the impulse train is clearly highlighted with respect to the residual background noise.

Another particular aspect of FBBD and, more in general, of BD theory that has been neglected in the synthesized signals analysis, may be underlined comparing the raw signal and the deconvolved sources from Bearing 8 in Fig.19(c) and Fig.21(c-d). The time signal presents a clearly visible periodic component, contemplated in the general scheme of BD given in Eq.1, despite the healthy bearing conditions. The estimated waveforms show how the proposed method allows the elimination of periodic patterns, e.g. related to electric interference, in addition to the background noise. This result proves the robustness of the proposed method also with respect to possible interference on the tested system or on the measuring chain.

3.6.2 *Identification of different levels of bearing damage under stationary working conditions*

The second experimental validation of the proposed BD algorithm regards the analysis of the Dataset 2 (previously described in Sec.2.3) acquired from the bearing test bench at the University of Ferrara. The goal of this validation is the demonstration of the sensitivity of the ICS_{2FB} to the damage severity through the analysis of three different artificial defects with increasing size on bearings operating under stationary conditions.

Fig.22 shows the raw time signals contained into Dataset 2. The actual response of the system to the excitation given by the localized artificial fault on the outer race is represented by a train of impulses and this aspect can be considered as a validation of the simulated signal exploited in Sec.3.5.2. Being the artificial fault overstated, it has to be underlined that the fault related signatures are visible already from the raw signals. Nevertheless, the target of this experimental validation is the discrimination between different damaging levels instead of the extraction of the fault related excitation from a noisy

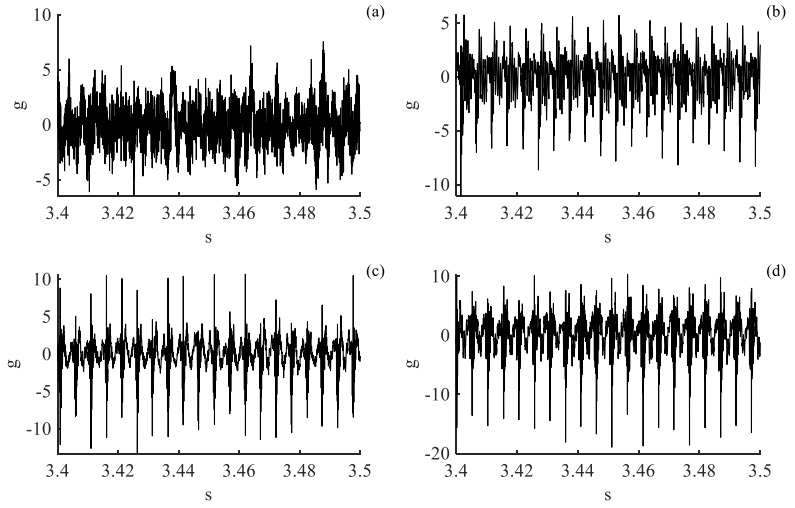

Figure 22: Observation showing part of the test signals of Dataset 2: (a) healthy bearing, (b) 0.8mm fault, (c) 1.6mm fault, (d) 2.4mm fault

observation as demonstrated in Sec.3.6.1. In this direction, the comparison between Fig. 22(b-c-d) clearly explains how an increment in the fault severity does not directly reflect on the pulse amplitude, i.e. the relation between damaging level and signal amplitude is not represented by a linear law. This behaviour is related to the non-linear physical filter between the excitation, i.e. the impacts between defect and rolling elements, and the response measured by the sensor. It has to be kept in mind that one of the target of BD is the reduction of the effect of the structure dependent filter in order to highlights the fault related signature and consequently enable the discrimination of different fault severities.

Being the main attention pointed out on the indicator value instead of the estimated source, the analysis has been performed considering five cyclic harmonics in order to reduce the computational time required by the algorithm. The characteristic frequencies (BPFO, BPFI and Ball Spin Frequency (BSF)) at the rotational speed of 2400rpm correspond to normalized fault periods of 262 samples, 180 samples and 497 samples, respectively, being the sample frequency set at 51.2kHz. Consequently the FBBD has been performed by retrieving the optimal filter length into a range from $N_{min} = 265$ to $N_{max} = 320$

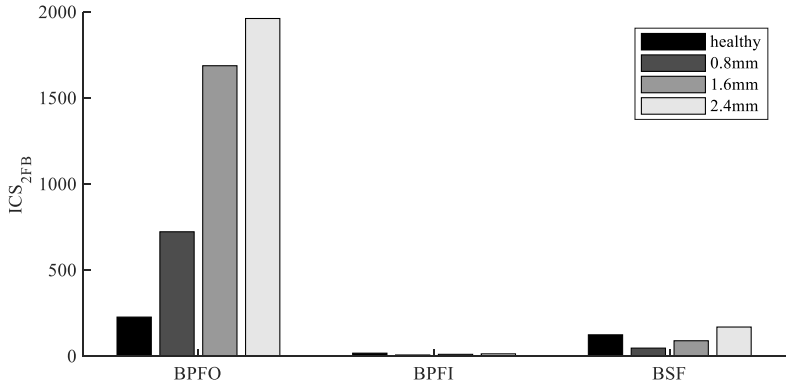

Figure 23: ICS_{2FB} values considering BPFO, BPFI and BSF for the four tested bearings of Dataset 2

for BPFO, from $N_{min} = 180$ to $N_{max} = 250$ for BPFI and from $N_{min} = 500$ to $N_{max} = 560$ for BSF. The ICS_{2FB} values for the BPFO, BPFI and BSF for each tested bearing of Dataset 2 are summarized in Fig.23. Despite the significant difference between the BPFO related values and the others, the main attention for the purpose of this analysis has to be pointed out on the values inside the BPFO family. In presence of a real damage the criterion value tend to significantly increase for increasing defect size. This increment reveals a quite linear trend, not visible in the raw signals due to the filtered excitation measured by the sensor. For the other characteristic frequencies, besides values clearly lower with respect to the fault related one, no relation between the different bearings can be identified. This aspect underlines the capability of FBBD of discriminates different fault severities, pivotal aspect in the real time condition monitoring.

However, this validation has been performed on a discrete set of damaging levels. The nowadays industrial scenario requires the continuous monitoring of the degradation trend for the real time assessment of the system conditions and the application of prognostic models for predictive maintenance purposes. From this standpoint, another experimental validation on a run to failure test needs to be performed in order to prove the effectiveness of the FBBD for the real time damaging level assessment and propose the ICS_{2FB} as suitable indicator for the development of prognostic models.

Table 6: Bearing type Rexnord ZA-2115 characteristic frequencies, normalized periods and filter length ranges

	α @2000 rpm (Hz)	N (Samples)	N_{min} (Samples)	N_{max} (Samples)
BPFI	297	68	70	130
BPFO	236	85	85	150
BSF	278	73	75	140

3.6.3 *Real time assessment of the damaging level: run to failure bearing test*

The last experimental validation deals with the analysis of two run to failure tests from the IMS dataset describe in Sec.2.1 in order to discuss the capability of the FBBD on the fault development monitoring. In particular, in this section two diffent datasets are considered:

- Campaign 2: Channel 1, related to a 7 days length endurance test leading to the appearance of an outer race fault on Bearing 1.
- Campaign 3: Channel 3, related to a 30 days length endurance test leading to the appearance of an outer race fault on Bearing 3.

The characteristic frequencies, the related normalized periods and the ranges of investigations for the optimal filter length identification are reported in Tab.6.

The evolution of the ICS_{2FB} related to all the characteristic frequencies for both tests are displayed in Fig.24 and Fig.25 together with the respective statistical thresholds. The thresholds can be designed in order to detect the fault appearance and at the same time identify the fault position. For diagnostic purposes, a suitable choice is represented by thresholds based on possible outliers, i.e. observations distant from the distributions of the value trend. In fact, usually the presence of outliers is related to fault appearances or changes on the working conditions. For this reason, for the fault detection through thresholds the necessary condition is the stationarity of the working parameters in order to associate all the changes inside the vibration signature

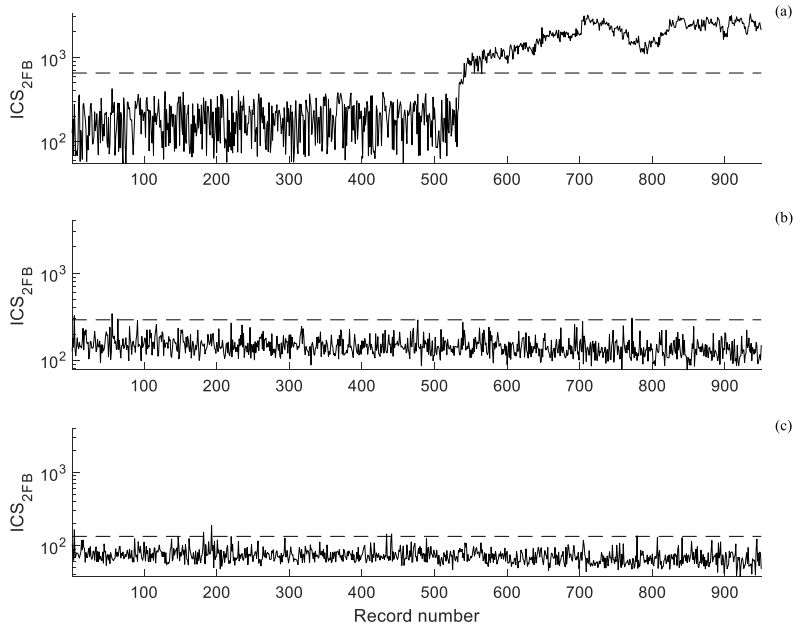

Figure 24: Campaign 2 - ICS_{2FB} values together with the statistical thresholds considering: (a) BPFO, (b) BPFI, (c) BSF

to the damaging process rather than to regime modifications. Hence, the bases idea is to define a threshold starting from observations under the hypothesis of healthy conditions (e.g. the first values at the beginning of the test) and check if the following samples are coherent with the initial distribution. In this contest, a suitable tools for the definition of outlier based thresholds is represented by the Tukey's method [67]. This method has been already successfully exploited for the fault detection through vibration based scalar indicator [68, 69] due to its general nature and its independence from the a priori knowledge of the data distribution. The only limitation is that the Tukey's method may no longer be effective if the data distribution is non-symmetric. The method is based on the interquartile range (IQR), i.e. the distance between the first and the third quartile, and defines two families of outliers, named mild outliers and extreme outliers. The first class regards the data not so far from the data distribution and is defined as 1.5 times the IQR range. The latter refers to values significantly distant from the reference distribution and is defined as 3 time the IQR distance. For this experimental validation only the extreme

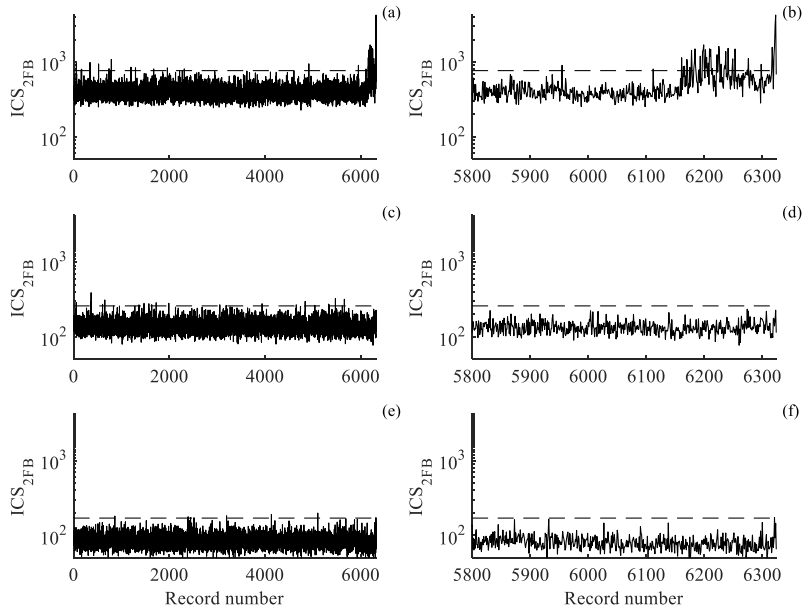

Figure 25: Campaign 3 - ICS_{2FB} values together with the statistical thresholds considering, from top to bottom, BPFO, BPFI, BSF: (a-c-e) entire test, (b-d-f) focus on the last stage

outliers threshold has been considered since the proposed indicator starts to rise only if the fault occurs. The threshold has been calculated starting from the ICS_{2FB} values from the first day of test under the hypothesis of healthy conditions for all the bearings in this time span.

The trends of the proposed agree with the physical observation of the tested bearings degradation:

CAMPAIGN 2: Fig.24(b-c) clearly show how the ICS_{2FB} values related to BPFI and BSF stand below the respective thresholds until the end except for some sparse points that cross the thresholds although no fault has occurred. This aspect is related to the data dispersion given by unpredictable phenomena related to the mechanical system or to the measuring chain and thus unrelated to the bearing fault. This issue can be overcome through a data smoothing, e.g. by means of a moving average [70] convolving the time series with a fixed rectangular window. On the contrary,

the BPFO related ICS_{2FB} values (Fig.24(a)) remain under the threshold until sample 530 (corresponding to 3.7 days). After this acquisition the trend follows the typical propagation phenomena of the bearing fault, i.e. consecutive propagation and smoothing after the defect appearance [71].

CAMPAIGN 3: this test represents a more challenging benchmark for the proposed criterion. In fact, despite the Campaign 2 where the fault related signature is clear visible in the vibration signals acquired after the fault appearance, for Campaign 3 the excitation pattern is strongly masked and barely identifiable in the last signal acquired. As well as in Campaign 2, the BPFI and BSF related ICS_{2FB} values (see Fig.25(c-d-e-f)) remain under the respective threshold for the entire test length. Moving to the BPFO (Fig.25(a-b)) the criterion values overcome the thresholds only in the last stage of the test (sample 6150) according to the issues already underlined for this dataset. Although the indicator evolution in this case is not clear as for the Campaign 2, the focus on the last stage (Fig.25(b)) shows the degradation propagation previously described, highlighting the robustness of the ICS_{2FB} also under strong interferences on the measured signals.

This experimental validation demonstrates that the bearing damaging process reflects on an transition from stationarity to cyclostationarity, confirming the need of exploiting this theory in order to improve the BD effectiveness in this field. Moreover, the above shown results prove the sensitivity of ICS_{2FB} to the fault severity, making the FBBD method a suitable tool for the assessment of the degradation level for monitoring purposes, even under strong interferences inside the vibration signal. This aspect is pivotal on the real time condition monitoring perspective but at the same time allows the idea of future exploitations of the proposed indicator as the observations for prognostic purposes, e.g. real time RUL estimation through prognostic models.

3.7 SUMMARIZING REMARKS

In this chapter, a novel BD criterion has been defined starting from the redefinition of the ICS_2 through the FBSE in order to improve the effectiveness of the cyclostationarity bases BD. The resulting method, called FBBD, has

been firstly introduced for the analysis of stationary working conditions, e.g. constant speed, then it has been extended to the angular domain (in this case named AngleFBBD) enabling the fault detection under non-stationary operating conditions. An exhaustive experimental validation has been carried out on both simulated signals and real case studies. This activity led to the following results:

- A new BD indicator (called ICS_{2FB}) has been proposed re-writing the ICS_2 through the FBSE in order to better fit the fault related waveform. This criterion has been written in form of generalized Rayleigh quotient and its maximization is the base of the proposed BD algorithm.
- The improvement given by the FBBD to the existing cyclostationarity based BD method (known as CYCBD) has been demonstrated through the analysis of simulated signals reproducing ciclostationary patterns under both stationary and non-stationary conditions. This analysis highlighted how the different mathematical natures of FBSE and FSE reflect on the BD effectiveness. The FBBD requires a lower number of terms for the fault related pattern reconstruction and consequently the computational times is reduced with respect to the CYCBD. This analysis also underlined the robustness of the proposed method with respect to possible masking interferences comprised into the measured signals.
- The analysis of real vibration signals from both academic and industrial environment highlights the capability of the proposed indicator to detect incipient faults, even in the early stages of the damaging process, under different operating conditions. Therefore, the sensitivity of the ICS_{2FB} to the damaging level has been demonstrated through the analysis of artificially damaged bearings with different fault sizes.
- The study of a run to failure test demonstrated the effectiveness of the proposed method for the real time condition monitoring on real mechanical systems, enabling its possible exploitation as the observation for prognostic purposes.

The main aspect of originality in this chapter is represented by the combination of cyclostationarity based BD and the FBSE. The modulated nature of the latter allows the faster reconstruction of the excitation related pattern

without losing the diagnostic capability of the cyclostationary BD method. All the theoretical consideration proposed in this chapter have been supported by an extended experimental validation on the analysis of several bearing faults.

4

PROGNOSTICS OF ROTATING MACHINES THROUGH GENERALIZED GAUSSIAN HIDDEN MARKOV MODELS

4.1 INTRODUCTION

In the last decades, reliability has been playing an even more fundamental role in the industrial economy, conditioning all the aspects of the productive line from the concept design to the final quality control. In particular, the need for time (and consequently costs) reduction drives the research of even more effective maintenance strategies for the failure prediction. In this contest, the Condition Based Maintenance (CBM) can be seen as a suitable maintenance strategy being it based on the real time health condition monitoring of a mechanical system. This methodology enables the optimal maintenance decision based on the actual system conditions allowing the reduction of unnecessary maintenance operations (and consequently of the machine downtime) and improving the reliability of the system. For this reason its exploitation in the industrial environment has been exponentially increasing in the last years.

One of the main tasks in CBM is represented by the health prognostics which aims to estimate the RUL of the system starting from the historical and on going degradation trends observed through some diagnostic indicator [72]. The machinery health prognostics is a wide field exhaustively studied starting from the 1960s but all the existing methods can be divided into three main families: physics model based approaches, Artificial Intelligence (AI) approaches and statistical model based approaches. The physics model based approaches describe the degradation process through mathematical models that reproduce the failure mechanism exploiting parameters related

to the material properties and the stress levels. These parameters are usually identified through experimental campaigns or finite elements analysis. The most widely exploited physics model is the Paris-Erdogan model [73]. This model has been proposed in 1963 in order to describe the crack growth. Starting form the original model, many modified versions have been proposed in order to better fit the machinery prognostics purposes. Between them, a particular mention has to be given to the work of Li at al. [74] that applied the Paris law for describing the rolling element bearing damaging process. Another interesting application has been proposed by Sun et al. [75] that enhanced the Paris-Erdogan model transforming it into a state space model. An exhaustive overview about physics model based prognostic approaches applied on rotating machines is provided in Ref. [76]. This family of methods guarantees an accurate description of the damaging process inside simple mechanical systems. However, the difficult representation of the degradation evolutions in complex machinery restricts the field of application of these approaches.

In this contest, the AI approaches better deal with complex systems being them based on the description of the degradation process through the intelligent analysis of the available observations directly from the physical system, without building mathematical models. The most famous AI prognostic approach is represented by the Artificial Neural Network (ANN) based on several nodes linked in a complex structure that aims to reproduce the working process of the human brain. In the last years, several applications of ANN to the machine prognostics have been presented in particular for the bearing RUL estimation. In this contest it is worth mentioning the work of Gebraeel et al. [77], the work of Pan et al. [78] and the work of Xiao et al. [79]. Inside the AI methodologies, interesting results have been reached through the Support Vector Machine (SVM) based approaches. These methods are based on the statistical learning theory proposed by Vapnik [80] and their effectiveness for the prognostics of rotating machinery has been widely demonstrated in the last decades. In this contest, the common application of the SVM for prognostic purposes is represented by the support vector regression [81, 82, 83]. Despite the suitability for complex system, the main weak point of the AI based prognostic approaches is represented by the need of a high amount of high-quality training data, not even possible to obtain in real case studies.

Moving to the statistical model based prognostic approaches, also known as empirical model based approaches, they try to estimate the RUL through statistical models based on the empirical knowledge of the system. Usually, the resulting RUL is given in form of conditional Probability Density Function (PDF) depending on the observations, i.e. diagnostic criteria that describe the damaging trend. The relation between observations and health condition is given in a probabilistic way taking into account random variances that allows the description of the uncertainly related to the degradation phenomena. For this reason in the last years the statistical approaches have become the most popular in the rotating machinery prognostic field. One of the most common statistical method is based on the AR models and considers the future state of the degradation trend as a linear function of the previous observations (except for some random errors). In this contest, Barraza-Barraza et al. [84] proposed an improvement of the AR model based RUL estimation through the study of a crack growth. Speaking about rolling element bearing prognostics, Qian et al. [85] improved the effectiveness of the AR based prognostics through the combination with the particle filter theory. Despite the simplicity of calculus, the performance of AR model based approaches strongly depends on the historical trends of observations leading to possible inaccurate estimations, in particular in the last stages of the working life.

In order to overcome this drawback at the expense of the computational time, Cox proposed a new statistical method, named Proportional Hazards Model (PHM) [86], that divides the hazard rate, i.e. the rate of fault for a system at a given operating age, into a base hazard function, i.e. the deterministic component, and a covariate function, i.e. the stochastic component. A particular mention has to be given to the work of Makis et al. [87, 88] that exploits the PHM in order to describe the failure rate of a system taking also into account the maintenance cost during the long run. From a more applicative standpoint, Jardine et al. [89] developed a PHM based condition monitoring software that has further been applied for the analysis of different types of machines [90]. An exhaustive review about other PHM based approaches and, more in general, other stochastic models can be found in Ref. [91]. The PHM based prognostic models enable a more accurate prediction of the damaging trend but the covariate function needs to be described through other stochastic processes, e.g. the Markov models, and consequently requires a computational time higher with respect to the single model based approaches.

The compromise between the effectiveness in failure prediction and low computational cost explains the huge exploitation of another statistical method, the Hidden Markov Model (HMM), for prognostic purposes. The basis idea is that the degradation process of a system can be described through a discrete number of finite states and the transition between two states follows the principle of the Markov chain [92]. The term hidden refers to the impossibility to directly observe the health state and consequently the relation between the observations and the damage level is given in a probabilistic way [93]. The basic theory of HMMs has been published in the late 1960s by Baum et al. [94] and has been firstly applied in the speech recognition field by several authors [95, 96]. In the last decades, the Markov chain has been exploited for describing the degradation mechanism of mechanical systems and the HMMs have been applied for machine prognostics and RUL estimation [97, 98]. The conditional distribution that relates the observations to each discrete state of the HMM is the single component of a mixture distribution and consequently the probability distributions of each state belong to the same distribution family, e.g. Gaussian or Bernoulli [99].

Unfortunately, in particular in the last stages of the working life, the relation between observations and health state is different with respect to the one at the first stages, i.e. where the system should be in healthy conditions. Therefore, the consideration of a generalized distribution as the component of the mixture distribution may allow to take into account these unavoidable modifications and better fit the real connections between the model and the observations. This research activity try to fill this gap through the realization of an HMM based on a mixture of Generalized Gaussian Distributions (GGDs) that enables the changes between the distributions of each state through different distribution parameters. In this direction, this chapter proposes an iterative algorithm for the estimation of the model parameters starting from the likelihood function and the observations measured on the physical system.

Firstly, a brief overview about the theory of HMM and the parameter estimation in case of GGD mixture is given. Then, the new parameter estimation algorithm is given for the case of univariate mixture distribution, i.e. considering a single series of observations. The algorithm is then extended for the more general case of multivariate mixture distribution in order to improve the results for complex system through the exploitation of several parameters as

model observations. The proposed model, named Generalized Gaussian Hidden Markov Model (GGHMM), is then validated taking into account several datasets regarding both academic test benches and real industrial applications. Dataset 1.2 of the University of Ferrara is exploited for demonstrating the better results reachable through the proposed model with respect to the classic Gaussian HMM. The suitability of the ICS_{2FB} described in the previous chapter for bearing prognostics through HMMs is proved through the application on the IMS dataset. The extension on more complex systems by means of a multivariate HMM is illustrated through the analysis of a run to failure test on a planetary gearbox. Finally, some final considerations are given.

4.2 GENERALITIES ABOUT GAUSSIAN HIDDEN MARKOV MODELS

Nowadays, HMMs are widely used for the treatment of sequential data in several fields. Consider a time series data, also known as model observations, i.e. a diagnostic indicator describing the degradation trend, $Y = \{Y_1, Y_2, \dots, Y_T\}$ that follows a continuous or discrete distribution. For the sake of clarity, hereafter capital letters refer to vectors and bold capital letters refer to matrices. A discrete HMM consists of a state sequence $S = \{S_1, S_2, \dots, S_T\}$ where the probabilistic dependence between two consecutive states follows a first order discrete Markov process, i.e. the current state only depends on the previous one, viz:

$$P(S_t|S_1, \dots, S_{t-1}) = P(S_t|S_{t-1}) \tag{23}$$

The limitation at the first order is a simplifying assumption but the Markov chain can be easily extended to higher orders [100]. If N is the number of model states, the state variables S_t are taken from a finite set, known as state-space of the HMM, $S = \{1, \dots, N\}$ such that $S_t = i, i \in S$, i.e. the HMM is discrete. The probabilities defined in Eq.23 drive the transitions between consecutive states, thus are called transition probabilities and comprised into a transition matrix $\mathbf{A}(t)$ with entries:

$$a_{ij}(t) = P[S_{t+1} = j|S_t = i], \qquad i, j = 1, \dots, N \tag{24}$$

The term hidden refers to the fact that each element of the state sequence can not be directly observable, thus the observations Y_t are related to the states

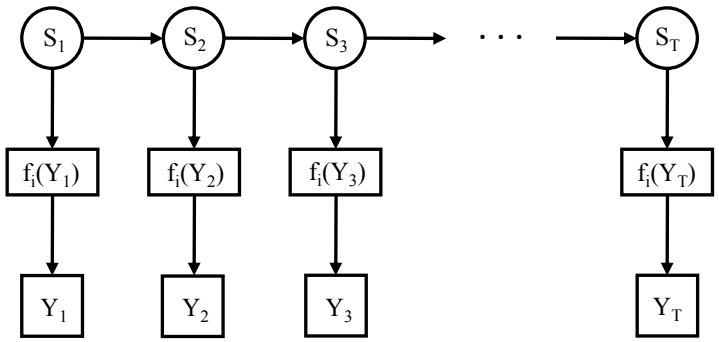

Figure 26: Dependence graph for observations inside an HMM

S_t (see Fig.26) through distribution functions, i.e. PDFs that map the S_t into Y_t making the states observable, defined as:

$$f_i(Y_t) = f(Y_t|S_t = i), \qquad i = 1, \ldots, N \tag{25}$$

Being the set of S finite the marginal distribution of the data is a mixture of N contribution [101], viz:

$$f(Y_t) = \sum_{i=1}^{N} p_i f_i(Y_t) \tag{26}$$

where p_i are the component proportions. From Eq.26 it is clear that the number of model states should be chosen as the optimal number of mixture components that better fits the observation distribution. This aspect can be better explained through a simple example. Fig.27 shows the RMS value trend and its PDF for the Dataset 1.2 related to the bearing test bench at the University of Ferrara. It can be immediately noticed that the RMS values deviate considerably from a Gaussian distribution (or in general from other uni-modal distributions) and there are several modes in these data, i.e. the marginal distribution is a mixture distribution. As previously stated, the number of mixture components corresponds to the number of model states, i.e. the number of

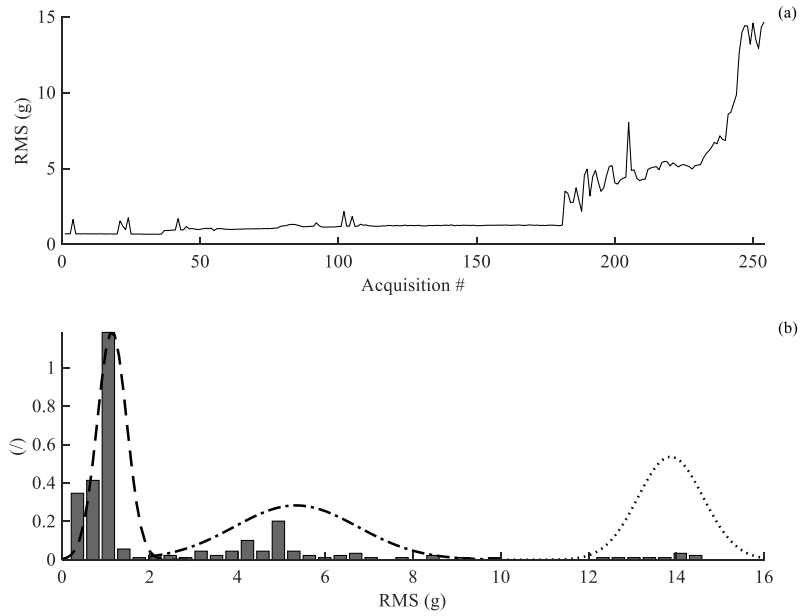

Figure 27: Results from Dataset 1.2 of the test bench at the University of Ferrara: (a) time RMS values from the raw signals, (b) PDF of the RMS

damaging stages for the working life of the mechanical system under test, and consequently has to be selected in order to describe the data distribution as accurately as possible. A suitable way to define the optimal state number is given by the Bayesian Information Criterion (BIC). This indicator has been proposed by Schwarz [102] and compares the maximized likelihood function with the actual distribution of data. The BIC function is increasing with respect to the number of model parameters, i.e. the number of variables required for defining the data distribution, and with respect to the variance error between the maximized likelihood function and the real distribution. Consequently, a lower BIC value indicates a model that better fits the considered data. The BIC should be applied for the comparison between models with the same number of parameters, in order to associate any variations only to different variance errors [103]. Tab.7 reports the BIC values computed for the data distribution given in Fig.27(b) taking into account the case of uni-modal Gaussian distribution and the Gaussian mixture with two and three components. It is clear how the higher improvement (one magnitude order) in the fitting quality is

Table 7: Selection of optimal state number: BIC values and distribution intervals for the RMS values on Dataset 1.2

Number of states	Distribution intervals	BIC
1	0-16	1285
2	0-3 , 3-16	131
3	0-3 , 3-10 , 10-16	39
4	0-1, 1-3 , 3-10 , 10-16	45

given by the consideration of a mixture distribution. The three components model represents the best approximation of the actual data distribution and consequently the optimal state number for the description of the working life is three, i.e. health stage, early damaging stage and high damaging stage.

Moving back to the element of an HMM, for completely defining a model it is necessary to define another element, named initial state (or prior) probability, that describes the probability of the system to be in a given state at the first time span:

$$\pi_i = P[S_1 = i], \qquad i = 1, \dots, N \tag{27}$$

Starting from all the aforementioned model parameters, i.e. the transition probabilities matrix $\mathbf{A}$, the initial state probability π and the distribution parameters vector B, the optimal state sequence, i.e. the new sequence state derived from the model, can be obtained through the Viterbi algorithm [104]:

$$S_T = \underset{i}{\operatorname{argmax}} \delta_T(i) \tag{28a}$$

$$S_t = \psi_{t+1}(S_{t+1}), \qquad t = T-1, \dots, 1 \tag{28b}$$

where

$$\delta_1(i) = \pi_i f_i(Y_1), \qquad i = 1, \dots, N \tag{29a}$$

$$\psi_1(i) = 0, \qquad i = 1, \dots, N \tag{29b}$$

$$\delta_t(j) = \max_{i \in [1,N]} \delta_{t-1}(i) a_{ij} f_i(Y_t), \qquad i = 1, \dots, N, t = 2, \dots, T \tag{29c}$$

$$\psi_t(j) = \underset{i \in [1,N]}{\operatorname{argmax}} \delta_{t-1}(i) a_{ij}, \qquad i = 1, \dots, N, t = 2, \dots, T \tag{29d}$$

The estimation of all the model parameters depends on the distribution family of the PDFs $f_i(Y_t)$. Usually, the components of the mixture are considered as Gaussian and thus, a brief description on the estimation algorithm for this distribution is mandatory.

4.2.1 *Parameters estimation for Gaussian Hidden Markov Models*

In general, the estimation of the model parameters means the computation of the maximized likelihood function of the HMM, i.e. the likelihood of the observations given a certain state sequence [105]. First of all, it is necessary to define two different variables, named forward variables and backward variables, respectively. According to Rabiner [99], the forward variables can be defined as the combination of the probabilities of different state sequences that lead to a certain state S_t, viz:

$$\alpha_1(i) = \pi_i f_i(Y_1), \qquad i = 1, \ldots, N \tag{30a}$$

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} f_j(Y_t), \qquad t = 2, \ldots, T, j = 1, \ldots, N \tag{30b}$$

Analogously, the backward variables are defined as the probabilities of the observations from $t+1$ to T given the state S_t such as:

$$\beta_T(i) = 1, \qquad i = 1, \ldots, N \tag{31a}$$

$$\beta_t(i) = \sum_{j=1}^{N} \beta_{t+1}(j) a_{ij} f_j(Y_{t+1}), \qquad t = T-1, \ldots, 1, i = 1, \ldots, N \tag{31b}$$

In order to prevent underflow and other computational artefacts, it is necessary to scale both forward and backward variables according to a given factor [93].

Starting from the forward and backward variables it is possible to define the probability of being in state i at the time t as:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \tag{32}$$

and, analogously, the probability of making a transition from state i to state j at time t as:

$$\epsilon_t(i,j) = \frac{\alpha_t(i)a_{ij}f_j(Y_{t+1})\beta_{t+1}(j)}{\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_t(i)a_{ij}f_j(Y_{t+1})\beta_{t+1}(j)} \tag{33}$$

It should be noticed that γ_t and ϵ_t are directly related, viz:

$$\gamma_t(i) = \sum_{j=1}^{N}\epsilon_t(i,j) \tag{34}$$

Finally, due to the ellipsoidal symmetry of the Gaussian distribution Liporace [105] demonstrated that the distribution parameters, i.e. means μ_i and standard deviations σ_i, that maximize the model likelihood function are given by:

$$\mu_i = \frac{\sum_{t=1}^{T}\gamma_t(i)Y_t}{\sum_{t=1}^{T}\gamma_t(i)}, \qquad i = 1,\ldots,N \tag{35a}$$

$$\sigma_i = \frac{\sum_{t=1}^{T}\gamma_t(i)(Y_t-\mu_i)^2}{\sum_{t=1}^{T}\gamma_t(i)}, \qquad i = 1,\ldots,N \tag{35b}$$

Starting from γ_t and ϵ_t it is possible to calculate the transition probabilities and the initial state probabilities, such as:

$$a_{ij} = \frac{\sum_{t=1}^{T-1}\epsilon_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)}, \qquad i,j = 1,\ldots,N \tag{36a}$$

$$\pi_i = \gamma_1(i), \qquad i = 1,\ldots,N \tag{36b}$$

The parameters estimation procedure is better known as Expectation-Maximization (EM) algorithm [106] and can be summarized as follows:

STEP 1: Choose initial values for the model parameters;

STEP 2: Compute the expected state sequence through the Viterbi algorithm;

STEP 3: Re-estimate the model parameters conditioned on the new hidden state sequence;

STEP 4: Repeat Step 2 and Step 3 until convergence.

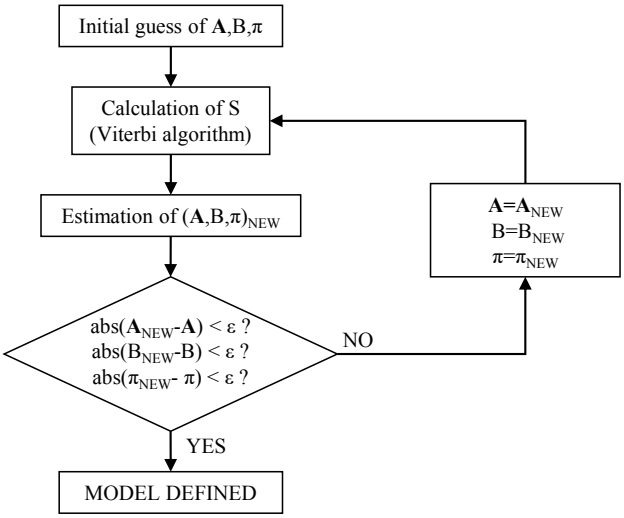

Figure 28: Flow chart of the EM algorithm for the estimation of HMM parameters

The EM algorithm is detailed described with the flow chart in Fig.28.

Fig.27(b) clearly shows the main drawback related to hypothesis of Gaussian PDFs for all the model states. It is possible to note that in the first states, i.e. system in healthy conditions and early fault appearance, the Gaussian PDFs fit the actual observation distributions. However, the heavy damage state presents an observation distribution that moves away from the Gaussian to the uniform distribution, thus the estimated model PDF is no longer able to describe the real data distribution. For this reason, the basis idea of the research activity proposed in this chapter regards the consideration a PDF family that allows the modifications within different states in order to improve the model capability of assessing the actual damaging level of the system.

4.3 GENERALIZED GAUSSIAN BASED HIDDEN MARKOV MODELS

A suitable way to overcome the aforementioned limitation is to consider a generalized Gaussian mixture as the model PDF. This generalized distribution allows the modifications within the model states through the values of few distribution parameters. In this section a novel algorithm for the estimation

of the GGD parameters based on the actual observations is proposed. The algorithm is firstly defined in case of mono-variate distributions, i.e. the related HMM considers only a single set of observations, and then is extended to the case of multi-variate PDFs, i.e. the model takes into account several different sets of diagnostic parameters as observations.

4.3.1 *Parameters estimation for mono-variate generalized Gaussian Hidden Markov Models*

First of all, it might be useful to remind the definition of the generalized Gaussian PDF:

$$f(x) = \frac{p}{2\Sigma^{\frac{1}{2}}\Gamma\left(\frac{1}{p}\right)} e^{\left[-\left(\frac{|x-\mu|}{\Sigma^{\frac{1}{2}}}\right)^p\right]} \tag{37}$$

where μ is the mean value, p is the shape factor, Σ is the scaling factor and Γ represents the Gamma function. This family of distributions includes the Gaussian distribution ($p = 2$ and $\sigma^2 = \Sigma^2/2$), the Laplace distribution ($p = 1$) and converges pointwise to the uniform distribution for $p \to \infty$. In general the shape factor p describes the exponential rate of decay of the distribution, i.e. the higher p the flatter is the PDF.

The methodology proposed by Liporace [105] for the model parameters estimation based on the observations can be generalized for all the ellipsoidal symmetrical PDFs that can be expressed in the following form:

$$|\Sigma_i|^{-\frac{1}{2}} f_i(q_i(x)), \qquad i = 1, \ldots, N \tag{38}$$

where $q_i(x)$ is a quadratic form, viz:

$$q_i(x) = \frac{(x-\mu_i)^2}{\Sigma_i}, \qquad i = 1, \ldots, N \tag{39}$$

Under this assumption, being $Y = \{Y_1, Y_2, \dots, Y_T\}$ the observation dataset, the mean value and the scale factor that maximizes the likelihood function for each state are given by the following:

$$\mu_i = \frac{\sum_{t=1}^{T} \rho_t(i)\beta_t(i)Y_t}{\sum_{t=1}^{T} \rho_t(i)\beta_t(i)}, \qquad i = 1, \dots, N \tag{40a}$$

$$\Sigma_i = \frac{\sum_{t=1}^{T} \rho_t(i)\beta_t(i)(Y_t - \mu_i)^2}{\sum_{t=1}^{T} \alpha_t(i)\beta_t(i)}, \qquad i = 1, \dots, N \tag{40b}$$

where α_t and β_t are the previous defined forward and backward variables and ρ_t is defined as:

$$\rho_t(i) = \sum_{j=1}^{N} \alpha_{t-1}(j)a_{ji}\left[-2\frac{\partial f_i(x)}{\partial q_i(x)}\bigg|_{x=Y_t}\right], \qquad i = 1, \dots, N \tag{41}$$

It should be noticed that for the Gaussian case, the sobstitution of Eq.41 inside Eq.40a and Eq.40b leads to the expression given in Eq.35a and Eq.35b.

In order to apply this method, Eq.37 has to be rewritten in a form compatible with Eq.38,viz:

$$f(x) = \frac{p}{2\Sigma^{\frac{1}{2}}\Gamma\left(\frac{1}{p}\right)} e^{\left\{-\left[\left(\frac{(x-\mu)^2}{\Sigma}\right)^{\frac{p}{2}}\right]\right.} \tag{42}$$

Starting from Eq.42, the derivative inside the square brackets in Eq.41 is given by:

$$-2\frac{\partial f_i(x)}{\partial q_i(x)}\bigg|_{x=Y_t} = f_i(x)p_i q_i(Y_t)^{\frac{p_i}{2}-1}, \qquad i = 1, \dots, N \tag{43}$$

Substituting Eq.43 into Eq.41 and remembering the definition of forward variables in Eq.30b, after a simple manipulation ρ_t is given by:

$$\rho_t(i) = \alpha_t(i)p_i q_i(Y_t)^{\frac{p_i}{2}-1}, \qquad i = 1, \dots, N \tag{44}$$

Finally, substituting Eq.44 into Eq.40a and Eq.40b the expression of mean value and scale factor can be written (taking also into account Eq.32) as:

$$\mu_i = \frac{\sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1} Y_t}{\sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1}}, \qquad i = 1, \ldots, N \tag{45a}$$

$$\Sigma_i = \frac{p_i \sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1} (Y_t - \mu_i)^2}{\sum_{t=1}^{T} \gamma_t(i)}, \qquad i = 1, \ldots, N \tag{45b}$$

It should be noticed that both the mean values and the scale factors depend on the shape factor and consequently a third equation is mandatory for the resolution of the problem.

According to Varanasi and Aazhang [107], the relation between scale factor and variance for a generalized distribution is given as follows:

$$|\Sigma|^{\frac{1}{2}} = \left(\frac{p}{T} \sum_{t=1}^{T} |Y_t - \mu|^p \right)^{\frac{1}{p}} \tag{46}$$

The same relation can be expressed depending on the Gamma function [108]:

$$\Sigma^{\frac{1}{2}} = \left[\sigma^2 \frac{\Gamma\left(\frac{1}{p}\right)}{\Gamma\left(\frac{3}{p}\right)} \right]^{\frac{1}{2}} \tag{47}$$

A title of example, for the Gaussian case ($p = 2$) both Eq.46 and Eq.47 describe the well established relation $\sigma^2 = \Sigma^2/2$. Moving back to the parameters estimation through the system observations, the left side of Eq.46 can be rewritten through Eq.45b. Eq.47 clearly explains how variance and scale factor are proportional given a value of the shape factor. Consequently, the right side of Eq.46 must be weighted in the same way of the left side in order to respect the proportionality between the two parameters. This consideration

leads to a reinterpretation of Eq.46 that can be expressed, after some simple manipulation, in the following form:

$$\left[\frac{p_i \sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1}(Y_t-\mu_i)^2}{\sum_{t=1}^{T}\gamma_t(i)}\right]^{\frac{1}{2}} = \left[\frac{p_i^2 \sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1}|Y_t-\mu_i|^{p_i}}{\sum_{t=1}^{T}\gamma_t(i)}\right]^{\frac{1}{p_i}}, \quad i=1,\ldots,N \tag{48}$$

Finally, the value of the shape factor can be found as the zero of the following function:

$$\left[\frac{p_i \sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1}(Y_t-\mu_i)^2}{\sum_{t=1}^{T}\gamma_t(i)}\right]^{\frac{1}{2}} - \left[\frac{p_i^2 \sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1}|Y_t-\mu_i|^{p_i}}{\sum_{t=1}^{T}\gamma_t(i)}\right]^{\frac{1}{p_i}} = 0, \quad i=1,\ldots,N \tag{49}$$

Eq.49 depends on the scale factor and the mean value, thus, the model parameters have to be estimated through an iterative algorithm that can be summarized as follows:

STEP 1: Assume an initial guess for μ_i and σ_i;

STEP 2: Calculate the shape factor p_i through Eq.49 by means of a zero finding algorithm;

STEP 3: Re-estimate μ_i and σ_i through Eq.40a and Eq.40b, respectively;

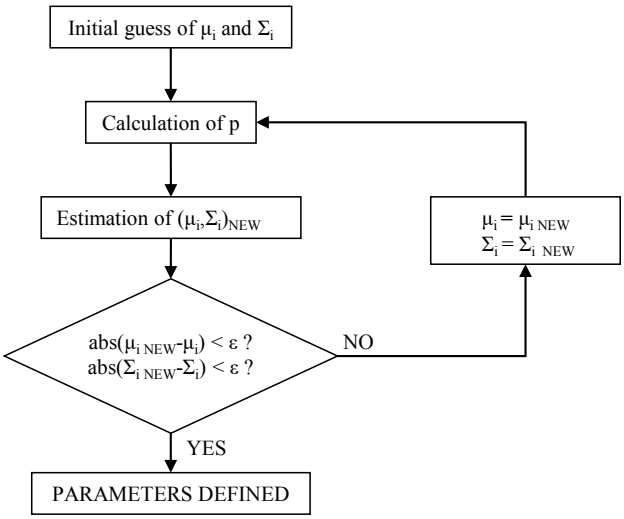

Figure 29: Flow chart of the algorithm for the estimation of the generalized Gaussian parameters

STEP 4: Repeat Step 2 and Step 3 until convergence.

This iterative parameter estimation algorithm (described with the flow chart in Fig.29) goes inside Step 3 of the main model iterative algorithm (Fig.28) and, starting from the computed parameters it is possible to extract the optimal state sequence for the operating life of the considered system.

4.3.2 *Extension to multi-variate generalized Gaussian Hidden Markov Models*

The effectiveness of the HMMs for prognostic purposes can be highly improved through the exploitation of several different observation sets, i.e. considering more than a single diagnostic parameter for describing the damaging process, moving from an observation vector to an observation matrix as:

$$\mathbf{Y} = \begin{bmatrix} Y_{11} & \dots & Y_{1K} \\ \vdots & \ddots & \vdots \\ Y_{T1} & \dots & Y_{TK} \end{bmatrix} \tag{50}$$

where T is the dataset length and K is the number of components. The analysis of different parameters for the same HMMs allows the extension of the model

to more complex systems where the interaction between many components has to be taken into account. For this general case, each component of the mixture distribution described in Eq. 26 is represented by a multivariate PDF which dimension is defined by the number of considered observation vector.

For a multivariate GGD which components are s-independent, i.e. describing different characteristics of the system, the PDF is defined as:

$$f(X) = \frac{\prod_{k=1}^{K} p_k}{2^K |\boldsymbol{\Sigma}|^{\frac{1}{2}} \prod_{k=1}^{K} \Gamma\left(\frac{1}{p_k}\right)} e^{\left[-\sum_{k=1}^{K} \left(\frac{|X_k - \mu_k|}{\frac{1}{\Sigma_k^2}} \right)^{p_k} \right]} \tag{51}$$

where the shape factor vector, the covariance matrix, i.e. the scale factor matrix, and the mean vector are defined as:

$$P = [p_1 \dots p_k \dots p_K] \tag{52a}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \ddots & & 0 \\ & \Sigma_k & \\ 0 & & \ddots \end{bmatrix} \tag{52b}$$

$$\mu = [\mu_1 \dots \mu_k \dots \mu_K] \tag{52c}$$

respectively. According to the procedure described for the monovariate distribution case, after expressing the PDF through the quadratic form, it is possible to calculate the distribution's parameters through the following equations:

$$\mu_k^i = \frac{\sum_{t=1}^{T} \gamma_t(i) \left(\sum_{k=1}^{K} p_k^i q_k^i (Y_{tk})^{\frac{p_k^i}{2}-1} \right) Y_{tk}}{\sum_{t=1}^{T} \gamma_t(i) \sum_{k=1}^{K} p_k^i q_k^i (Y_{tk})^{\frac{p_k^i}{2}-1}}, \qquad i = 1, \ldots, N \tag{53a}$$

$$\Sigma_k^i = \frac{\sum_{t=1}^{T} \gamma_t(i) \left(\sum_{k=1}^{K} p_k^i q_k^i (Y_{tk})^{\frac{p_k^i}{2}-1} \right) (Y_{tk} - \mu_k^i)^2}{\sum_{t=1}^{T} \gamma_t(i)}, \qquad i = 1, \ldots, N \tag{53b}$$

$$\begin{aligned} &\left[\frac{\sum_{t=1}^{T} \gamma_t(i) \left(\sum_{k=1}^{K} p_k^i q_k^i (Y_{tk})^{\frac{p_k^i}{2}-1} \right) (Y_{tk} - \mu_k^i)^2}{\sum_{t=1}^{T} \gamma_t(i)} \right]^{\frac{1}{2}} = \\ &\left[p_k^i \frac{\sum_{t=1}^{T} \gamma_t(i) \left(\sum_{k=1}^{K} p_k^i q_k^i (Y_{tk})^{\frac{p_k^i}{2}-1} \right) |Y_{tk} - \mu_k^i|^{p_k^i}}{\sum_{t=1}^{T} \gamma_t(i)} \right]^{\frac{1}{p_k^i}}, \quad i = 1, \ldots, N \end{aligned} \tag{53c}$$

where the notation μ_k^i refers to the parameter of the k-th components of the i-th state related multivariate distribution. The complete proof for the case of multivariate distributions is detailed described in Appendix A.2.

4.4 EXPERIMENTAL VALIDATION ON REAL VIBRATION SIGNALS

This section provides the validation of the proposed HMM through the analysis of several run to failure tests taking into account both academic test benches and real industrial machines. The validation aims to demonstrate several aspects:

- The improvement given by the proposed approach based on the GGD with respect to the classic methods based on the GD in terms of fitting quality.

- The enhancement reachable through the exploitation of a cyclostationary index as model observation.
- The effectiveness of the proposed model also for the prognostics of complex systems through the consideration of the multivariate distributions

4.4.1 *Comparison between Gaussian and generalized Gaussian based HMMs*

The first experimental validation deals the analysis of the Dataset 1.2 related to the run to failure test performed on the bearing test bench at the University of Ferrara. For the sake of clarity, it has to be remembered that this test has been stopped after 13 days due to a rolling element defect as detailed described in Sec.2.3. The time RMS (which trend during the entire test is shown in Fig.30(a)) has been selected as the observation vector for the comparison between the results obtainable through both Gaussian based and generalized Gaussian based HMM.

Due to the availability of only one test, the same dataset has been exploited for both training phase and validation process. For this purpose the observation vector has been divided into two separated datasets with the same length and composed by samples picked alternatively from the main vector, as shown in Fig.30. For the Gaussian case, the training process has been performed through the algorithm described in Fig.28 and the parameters of the conditional PDFs have been estimated according to Eq.35a and Eq.35b. For the generalized Gaussian case, the model has been trained through to the iterative algorithm summarized in Fig.29 and the model parameters have been calculated according to Eq.45a, Eq.45b and Eq.49.

Tab.8 reports the estimated values of the transition probabilities matrix $\mathbf{A}$ and the prior probabilities vector π. It can be immediately note that the values obtained by applying the two bases distributions do not differ significantly in the two cases. The form of $\mathbf{A}$ depicts the main assumption made for the considered models: the system follows a first order left-right model (better known as Bakis model [109]), i.e. the degradation process is irreversible and the transition regards only successive states. The parameters of the conditional PDFs between observations and actual state are reported in Tab.9. These values,

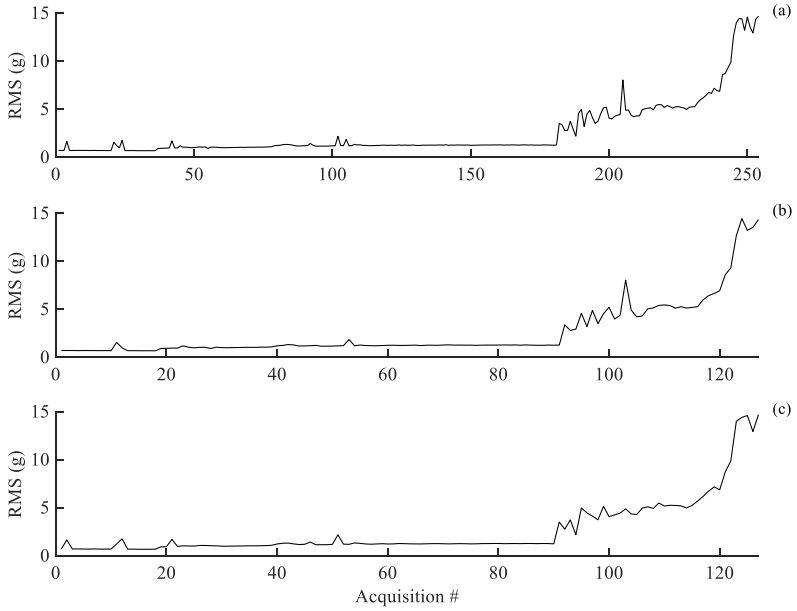

Figure 30: RMS values from Dataset 1.2 of the University of Ferrara test bench: (a) overal RMS trend, (b) training dataset, (c) validation dataset

Table 8: Model parameters: initial state and transition probabilities

	Gaussian			Generalized Gaussian		
	0.98	0.02	0	0.99	0.01	0
A	0	0.93	0.07	0	0.97	0.03
	0	0	1	0	0	1
π	1	0	0	1	0	0

combined with the comparison between estimated PDFs and observation distribution illustrated in Fig.31, explain the problem statement at the base of the proposed HMM. In the first state, i.e. system under healthy conditions, the data distribution is clearly Gaussian and consequently the results for the two analysis are strictly comparable. Moving to the second state, i.e. early damaging stage, the data distribution seems to still be Gaussian but it starts to

Table 9: Estimated parameters of the conditional PDFs

State	Gaussian		Generalized Gaussian		
#	μ	σ	μ	p	Σ
1	1.09	0.23	1.12	2.01	0.35
2	4.78	1.31	5.06	1.93	2.14
3	12.1	2.45	13.8	2.61	1.25

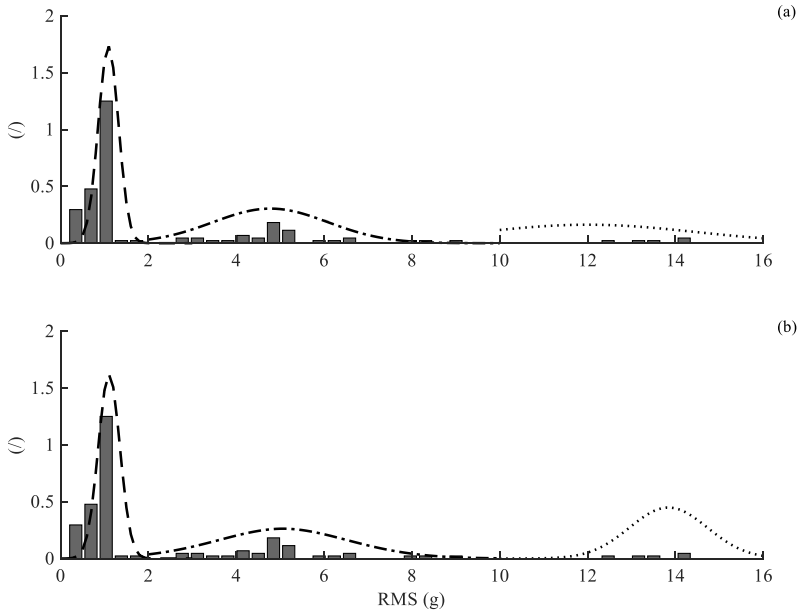

Figure 31: Results of the training step in term of estimated PDFs for: (a) Gaussian distribution, (b) generalized Gaussian distribution

move away from the ideal condition and consequently the two distributions presents a slight difference (the generalized Gaussian is flatter and with a slight higher mean value). The main difference regards the last state, i.e. the damaging process is in an advanced stage, where the generalized distribution presents a sensibly higher mean value and the distribution forms are strongly different. For this comparison, the fitting quality is no longer evaluable through the BIC value due to the different number of model parameters between

Table 10: Maximum log-likelihood and dimensionality for the compared HMMs

Model	Max Log-likelihood	Dimensionality
Gaussian	-26.78	6
Generalized Gaussian	-22.59	9

the distributions. For the same purpose, a suitable tool is represented by the Likelihood Ratio (LR) test [110]. This method compares two different statistical models with known parameters on the base of the ratio between their maximized likelihood functions over the entire space parameters. The main constrain of the method is given by the need of a relation between the families of the two models, i.e. one of the competing models must be obtained from the other one through the application of some constrains. Tab.10 reports the maximized log-likelihood and the dimensionality, i.e. the number of free parameters, of both models. The LR test statistic, defined as twice the difference between the maximized log-likelihoods, for the comparison of the Gaussian model with the generalized Gaussian one is 8.38. It has been demonstrated that the LR, for large number of samples, follows a χ^2 distribution with degrees of freedom equal to the difference among the number of parameters of the two models. The 95% quartile of the $\chi^2(3)$ distribution is 7.81 and consequently the GGD should be chosen as the one that better fits the observation data.

The models described by the parameters summarized in Tab.8 and Tab.9 have been validated through the part of observations illustrated in Fig.30c. Fig.32 describes the state sequences estimated on the validation dataset with both Gaussian and generalized Gaussian models. Both state sequences reproduce correctly the RMS trend depicted in Fig.30(c) also according to the data distribution shown in Fig.31. However, the comparison between the two sequences allows some important considerations. The main one regards the sparse points present in the first state (see Fig.32(a)). These points correspond to isolated peaks in the RMS trend (see samples 12 and 50 of Fig.30(c)) and the Gaussian based method understand these peaks as a state transition although the system remains in healthy conditions, i.e. the model should remain in state 1. In this context, the generalized Gaussian mixture allows the overcoming of this issue and these points are considered in state 1 although the slight RMS

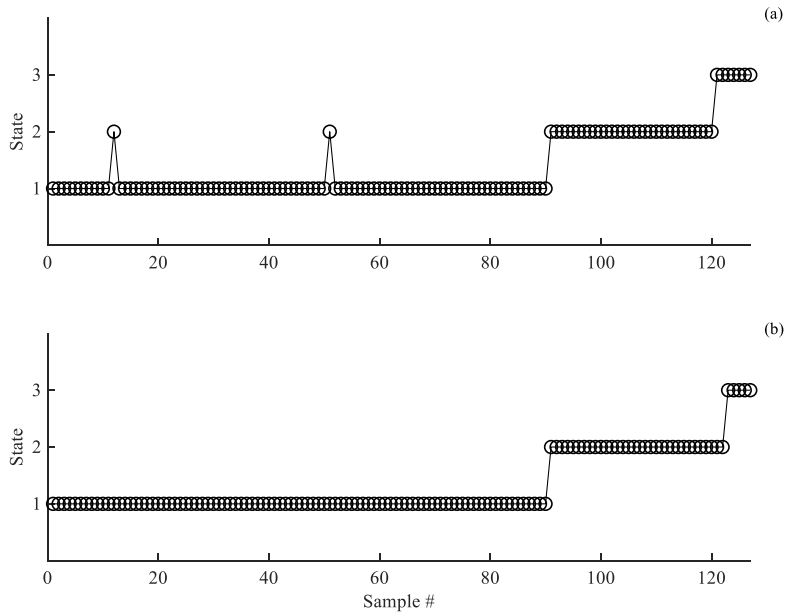

Figure 32: Estimated state sequences from: (a) Gaussian model, (b) generalized Gaussian model

increment avoiding an unnecessary alarm regarding the health state of the system. This aspect clearly explains the higher robustness of the proposed model with respect to outlier comprised into the observation vector related to uncertainties given by the physical system, the measuring chain and the signal processing techniques exploited for the calculation of the diagnostic indicators.

Another interesting aspect regards the state transitions. Taking into account the first transition, i.e. from healthy condition to incipient fault appearance, both methods move from the first state to the second in the same time span, i.e. with the same sample of the observation trend. This behaviour is easily understandable being the detection of the fault appearance strongly driven by the diagnostic capability of the indicator considered for the HMM. Consequently, the transition to the first damaging stage does not depend on the considered prognostic model and different distributions should not affect the results. On the other hand, the transition from early damage to heavy damage stage, i.e. from state 2 to state 3, highlights the departure of the data distribution

from the ideal Gaussian conditions in the last part of the working life. The Gaussian case estimates a distribution with a lower mean value (see Fig.31(a)) that does not reproduce the actual data distribution in the last stage. For this reason, the transition from state 2 to state 3 occurs earlier in the Gaussian case with respect to the generalized Gaussian case where the state sequence better reproduces the effective damage development. The transition to the heavy damage stage is pivotal in order to program maintenance operations, thus the correct transition estimation enables a more effective maintenance program, avoiding unnecessary machine downtimes related to an ahead of schedule system fixing.

From the same point of view of predictive maintenance, the main target of each prognostic models is the estimation of the RUL. In general, a system should be considered as in failure if some diagnostic indicator crosses a critical threshold or the system reaches the last state of the prognostic model. Over the years, several RUL prediction methods have been proposed. They can be mainly divided into two main families [111]: observation based method, i.e. based on the trend of the diagnostic indicator considered for developing the model, and state-based method, i.e. based on the state sequence estimated through the Viterbi algorithm. In order to highlight the different behaviour of the two base distributions (described by the estimated state sequences in Fig.32) it should be suitable to estimate the RUL on the validation dataset through a state-based method. In this direction, Medjaher et al. [112] proposed an estimation algorithm based on the stay time of the system into each state. According to this method, the RUL is defined as:

$$RUL_t = \sum_{i=currentstate}^{N} \upsilon_i \mu_d(S_i) - t_{ac}(t) \tag{54}$$

where υ_i is the number of possible visits to the state S_i, i.e. the number of blocks of consecutive samples in the same state, $\mu_d(S_i)$ is the mean stay time

Table 11: Mean time and number of visits in each state for the compared HMMs

	S_1		S_2		S_3	
	$\mu_d(h)$	Z	$\mu_d(h)$	Z	$\mu_d(h)$	Z
Gaussian	58.7	3	21.3	3	14	1
Generalized Gaussian	90	1	32	1	10	1

in each state and $t_{ac}(t)$ represents the time spent in the active state at the time span t. $\mu_d(S_i)$ and $t_{ac}(t)$ are defined as:

$$\mu_d(S_i) = \frac{1}{Z_i} \sum_{z=1}^{Z} \Delta(S_{iz}) \tag{55a}$$

$$t_{ac}(t) = \begin{cases} 0 & S_t \neq S_{t-1} \\ t_{ac}(t-1) + \Delta t & S_t = S_{t-1} \end{cases} \tag{55b}$$

where Z_i is the number of visits at state S_i, $\Delta(*)$ stands for the visit duration and Δt is the time span between two consecutive samples in the observation vector. Tab.11 summarizes the mean duration and the number of visits for each states starting from the state sequences estimated for the validation set through both Gaussian and generalized Gaussian HMM. The RUL definition given in Eq.54 leads to a result that underestimates the RUL in the first state, i.e. healthy conditions, but enhances the estimation quality for the last model states, i.e. the prediction error decreases with the damage evolution. This particular aspect is pivotal for industrial application where the critical stages are related to the appearance and the following evolution of the damage and consequently the effectiveness of the prediction has to be higher in the last part of the working life. The RUL estimated according Eq.54 taking into account both Gaussian and generalized Gaussian HMM are compared to the actual RUL in Fig.33. The previous depicted consideration about the main characteristics of the estimation method is clear visible. The difference between the predicted patterns (dotted and dash-dotted line) confirms the behaviour of the estimation method with a decreasing error for an increasing test progress. In particular its is interesting to note that the proposed HMM leads to an estimated state sequence that allows the estimation of the correct RUL in the

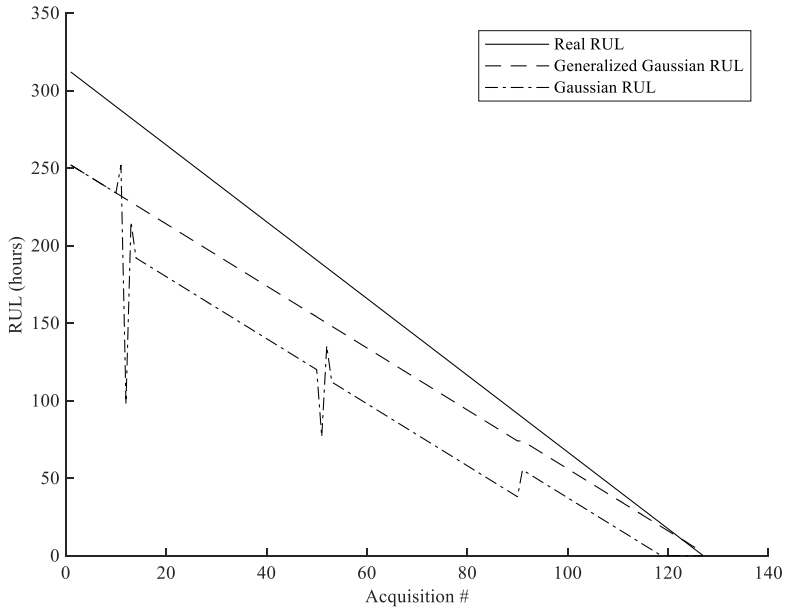

Figure 33: Comparison between estimated RUL and real RUL for the validation dataset

last stage of the damaging process, i.e. after sample 120, despite an initial prediction error not negligible (around 15%). The effect of the different state sequences estimated with the Gaussian HMM and the generalized Gaussian HMM can be immediately understood by comparing the respective predicted RUL. The previous discussion highlights two main points of improvement: the robustness with respect to possible sparse points inside the observation trend and the better fit of the data distribution in the last state. Between them, the first one is the aspect with more specific weight on the RUL prediction. In the first part (until sample 12) the two base distributions lead to the same result. From this point, the presence of a sparse value that is detected in the second state (despite the healthy conditions of the system) reflects on a peak in the RUL trend. Starting from this discontinuity the trend is parallel to the one related to the generalized case. Consequently, the overall trend from sample 12 to the end is shifted down and presents another discontinuity at sample 51, i.e. the second state error in the estimated sequence for the Gaussian HMM. Moving to the last part of the RUL (from sample 118 of Fig.33 corresponding to the 0 of the Gaussian RUL), the underestimation obtained by the Gaussian

Table 12: Metrics values for evaluating the prediction performance

	RMSE (/)	Precision (/)	MAPER (/)
Gaussian	62.4	26.1	45.1
Generalized Gaussian	34.5	18.2	17.8

analysis, combined with the earlier transition to the last state, leads to an earlier prediction of the end life, confirming how previous stated analysing Fig.32.

The prediction quality can be quantitatively measured through several indicators [113]. For this experimental validation three different criteria have been considered: the Root Mean Square Error (RMSE), the Precision and the Mean Absolute Percentage Error (MAPER), defined as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{T} (E_t)}{T}} \tag{56a}$$

$$\mathrm{Precision} = \sqrt{\frac{\sum_{t=1}^{T} (E_t - \bar{E})^2}{T}} \tag{56b}$$

$$\mathrm{MAPER} = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{100 * E_t}{RUL_t} \right| \tag{56c}$$

where RUL_t is the real RUL at time t, $\hat{RUL}_t$ is the estimated RUL, $E_t = RUL_t - \hat{RUL}_t$ is the error and $\bar{E}$ is the mean error defined as:

$$\bar{E} = \frac{1}{T} \sum_{t=1}^{T} E_t \tag{57}$$

Tab.12 reports the values computed for all the metrics in both cases. All the indicators confirm the qualitative consideration described starting from Fig.33 highlighting the better performance of the proposed HMM. In particular the more considerable differences are related to the RMSE and the MAPER being

them directly related to the error function without the introduction of the mean error considered for the calculation of the Precision. In fact, the higher mean error for the Gaussian case compensates for the higher instantaneous error, leading to a lower difference between the cases for this particular criterion.

The experimental validation described in this section demonstrates the improvement given by the exploitation of the generalized Gaussian mixture distribution as the base PDF for the construction of the HMM. The effectiveness enhancement regards in particular the better fitting quality in the last states, confirming the departure of the observation distribution from the ideal Gaussian case in these stages. The proposed method also proves to avoid possible estimation errors due to state outliers in the observations vector, simplifying the RUL estimation through state-based prediction algorithms. Being the comparison between different base PDFs the target of this validation, the RMS trend has been considered as the observations of the model. However, the exploitation of criteria particularly designed for the fault detection may improve the prognostic capability of the proposed HMM through a better description of the degradation process.

4.4.2 *Exploitation of cyclostationarity for bearing prognostics through generalized Gaussian hidden Markov Models*

In general, the prognostic effectiveness of the HMM is directly related to the capability of the diagnostic indicator, i.e. the observation vector, to describe the appearance and the evolution of a possible defect. For this reason, in this experimental validation the ICS_{2FB} described in Chapter3 is considered as the physical observation for training and validating a GGHMM for the prognostics of a rolling element bearing. The analysis presented in this section has been performed on the Campaign 2 provided by the IMS. The BPFO related ICS_{2FB} vector (already shown in Fig.24) has been normalized with respect to its maximum value in order to speed up the convergence of the algorithm. The target of this experimental analysis is the demonstration of the ability of the proposed method to describe the degradation process even for observation trends not monotonically increasing and consequently the main attention has to be pointed out on the estimated state sequence justifying the decision of normalizing the data. The data has been divided into a training dataset and

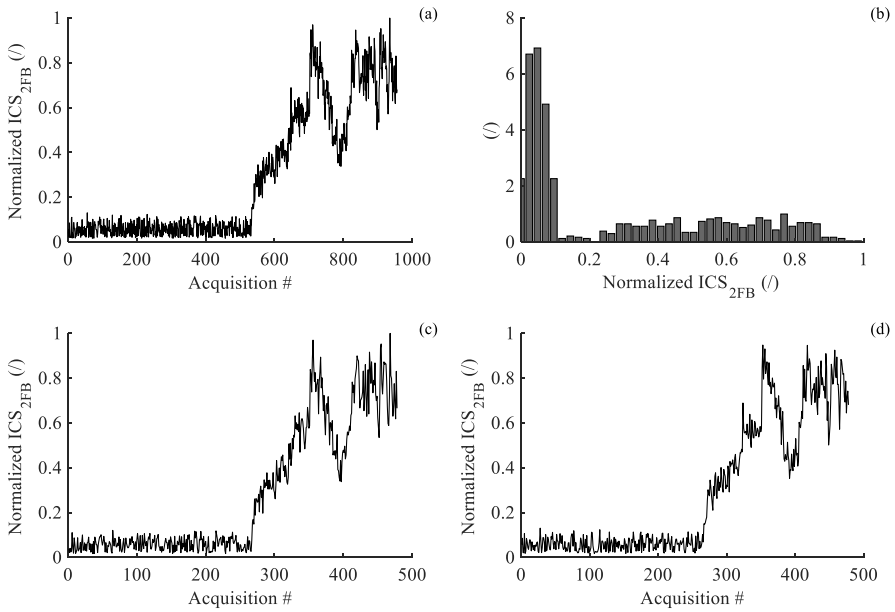

Figure 34: BPFO related ICS_{2FB} values for Campaign 2 of IMS dataset: (a) Overall dataset, (b) data distribution, (c) training dataset, (d) validation dataset

a validation dataset with the same length with the protocol described in the previous section. The overall dataset, the training set and the validation set are displayed in Fig.34 together with the data distribution. The trend described in Fig.34(a-c-d) describes the idea at the base of this analysis. The degradation process reaches a first peak due to development of pitting on the outer race. Then, the damaging level decreases due to the smooth of the defect area before a final stage of increment that leads to the final failure. According to the data distribution shown in Fig.34(b) the model has been built taking into account three finite states, i.e. healthy, early damage and heavy fault.

Tab.13 summarizes all the model parameters estimated through the iterative algorithm described in Fig.29. The form of the transition probabilities matrix **A** is directly related to the degradation trend on the outer ring. The decrease of the ICS_{2FB} values around sample 400 of the training dataset disables the exploitation of a pure left-to-right HMM due to the possibility of transitions from the last state to the second one before the final failure. Consequently, the

Table 13: Model parameters for the analysis of the IMS dataset: initial state and transition probabilities

	Model parameters		
A	0.99	0.01	0
	0	0.90	0.10
	0	0.08	0.92
π	1	0	0

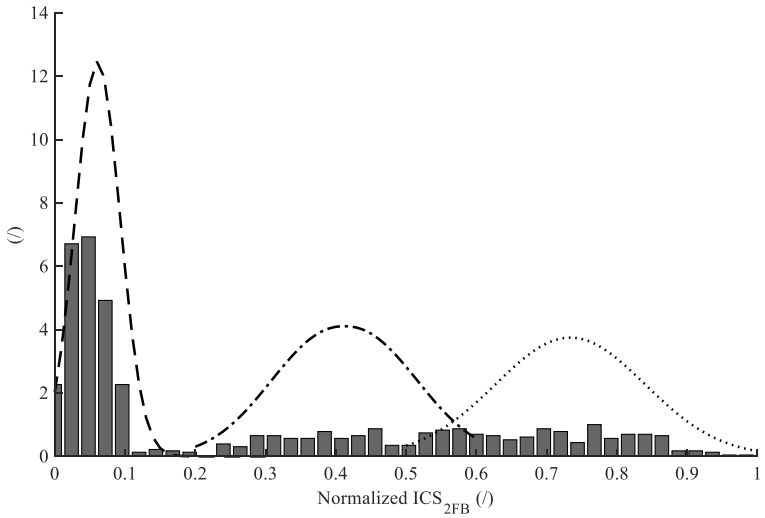

Figure 35: Comparison between original observation distribution and estimated PDFs for the analysis of the IMS dataset

element a_{32} can not be set to zero but has to take into account the possible reversible process. The same back transition can not be possible between the first two stages, i.e. once the fault appears it is impossible to come back to the healthy state, and the element a_{21} can be set to 0. The estimated PDF parameters are summarized in Tab.14 and the resulting functions are compared to the real observation distribution in Fig.35. It is interesting to underline how once again the values of the shape factor p describe the basis idea behind the exploitation of the GGD. In the healthy state the distribution remains Gaussian,

Table 14: Estimated parameters of the conditional PDFs for the analysis of the IMS dataset

State#	μ	p	Σ
1	0.06	2	0.04
2	0.41	2.2	0.14
3	0.73	2.2	0.15

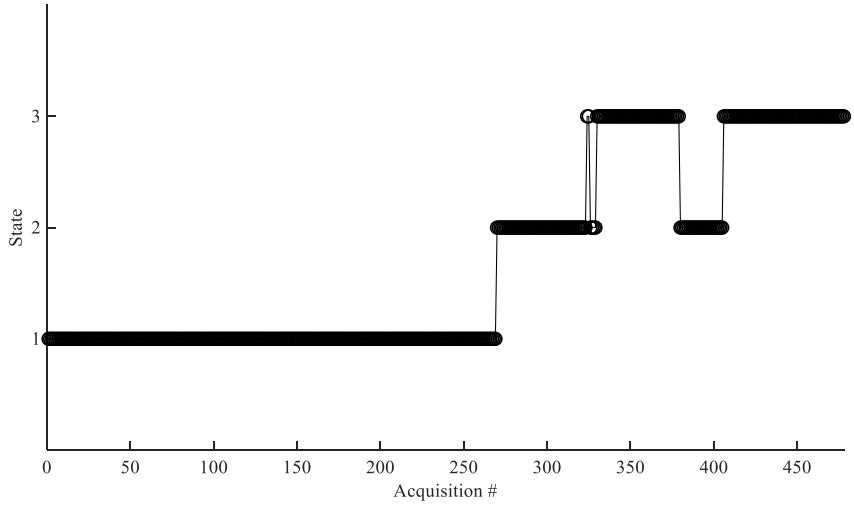

Figure 36: Estimated state sequence for the analysis of IMS validation dataset

i.e. $p = 2$, and the model seems to guarantee a good fitting quality both in term of mean value and data dispersion. Looking to the faulty stages, i.e. the second state and the third one, the data distribution moves away from the Gaussian case going towards a uniform distribution. This behaviour is confirmed by the increment of the shape factor and the scale factor for both distributions leading to flatter PDFs. It should also be noticed that the data distributions of the last states are similar both in form and amplitude and this aspect is caught by the estimated density functions that differ only in the mean value.

Fig.36 depicts the estimated state sequence for the validation dataset shown in Fig.34(d). The attention has to be focused on the last part of the state

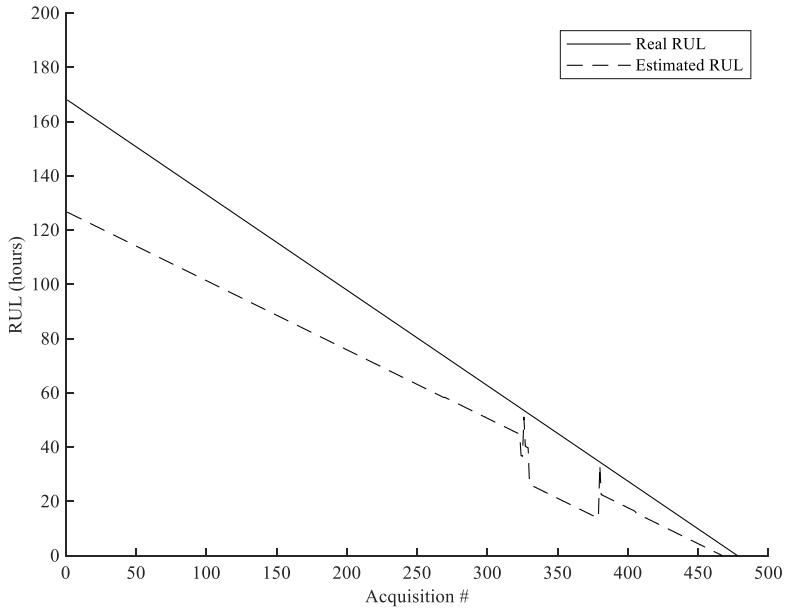

Figure 37: Estimated RUL for the analysis of IMS validation dataset

trend. The exploitation of a cyclostationary indicator as model observation enables the correct identification of the instantaneous health condition of the system. In fact, the slackness of the left-to-right model constrain allows the correct reproduction of the actual damaging process of the bearing. After reaching the state 3, i.e. the first heavy pitting on the outer race, at sample 330, the system return in state 2 due to the smoothing of the damaged area. Finally, after sample 410 the model moves for the last time to state 3 reaching the damaging level that leads to the bearing failure. The uncertainty on the identification of the transition between the real data distribution of state 2 and state 3 (around normalized $ICS_{2FB} = 0.55$ in Fig.35) reflects on a not clear transition from state 2 to state 3 at sample 330. However, the proposed model limits this critical aspect as confirmed by the low number of outlier samples, i.e. sparse points in a state that differs from the one related to the neighbouring samples. This possible issue regards only the first transition due to the steeper increasing trend related to the last transition that leads to the final failure, as demonstrated in Fig.36.

The estimated RUL is compared with the real RUL in Fig.37. The RUL has been estimated through Eq.54 starting from the state sequence illustrated in Fig.36. The result confirms the behaviour of the considered RUL prediction method characterized by the underestimation in the healthy stage and a increasing prediction quality for increasing damaging level. One of the critical points highlighted in the previous experimental validation is represented by the spikes in the RUL due to the sparse samples in the state sequence. Taking into account the uncertainty in the detection of the transition between state 2 and state 3 (sample 330), it is possible to note that the discontinuity in the RUL is reduced with respect to the previous analysed case. This behaviour can be understand through the different positions of the sparse points in the two analysis. For the current analysis, the sparse points are strictly connected to the transition phase and consequently are not related to anomalies in the observation vector. The previous analysis presented outliers in the first state far from the transition. As a result, the two types of outlier reflect on a different specific weight in the predicted RUL with a slighter effect if the outliers are the consequence of a physical uncertainty instead of a measuring issue.

A further validation of the cyclostationary application of the proposed HMM can be obtained through the analysis of the other bearing components, i.e. the BPFI and BSF, that remain under healthy conditions during the entire run to failure test being the outer race fault the only one detected on the physical system. The ICS_{2FB} trend and distribution taking into account BPFI and BSF are displayed in Fig.38. The cyclostationary indicator has been normalized through the maximum value of the BPFO related trend in order to enable the application of the model parameters previously described in Tab.13 and Tab.14. It is clear that both trends remain constant during the overall test (Fig.38(a-b)) and the data distribution is comprised in the state 1 related mixture component, in accordance with the effective health condition of these bearing components. The resulting state sequences are displayed in Fig.39. The results confirm how both components remain in health state, i.e. state 1, for the entire test according to the data distribution shown in Fig.38(c-d). Although the data dispersion for the BSF can be considered as irrelevant (Fig.38(b)), it is possible to note for the BPFI (Fig.38(a)) several possible outliers, i.e. the distribution is unbalanced in direction of the right tail. Despite this behaviour, the resulting state sequence does not present sparse outliers, confirming the robustness of the proposed method with respect to interferences in the measurement chain.

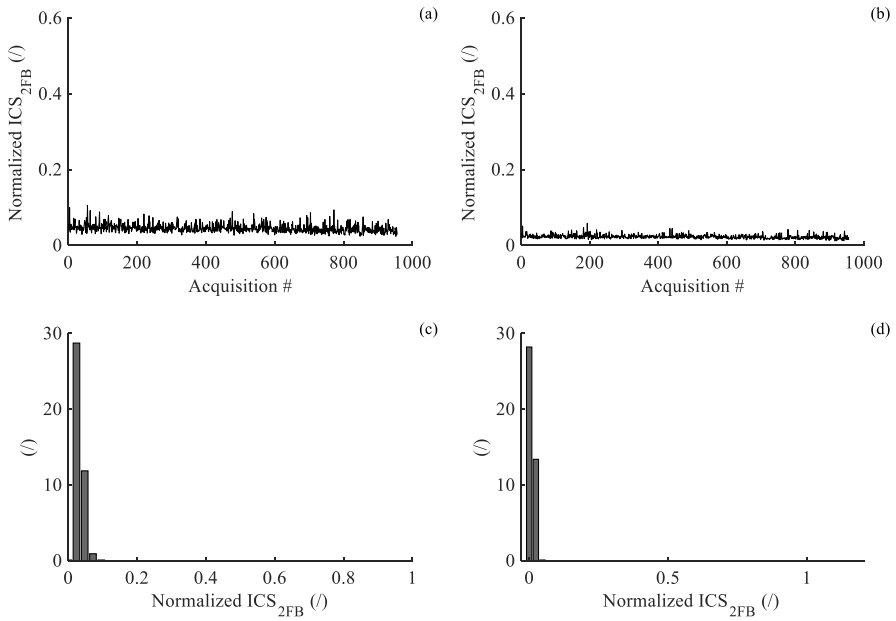

Figure 38: ICS_{2FB} trend and distributions for BPFI and BSF: (a) BPFI related normalized ICS_{2FB} trend, (b) BSF related normalizedICS_{2FB} trend, (c) BPFI related normalized ICS_{2FB} distribution, (d) BSF related normalized ICS_{2FB} distribution

The above described experimental validation illustrates how the effectiveness of the proposed HMM for the bearing prognostics can be improved considering a cyclostationary criterion as bases observation for the training of the model. The exploitation of this kind of indicator, combined with the removal of the constrain of pure left-to-right model, allows a better reproduction of the fault evolution improving the capability of the model on the RUL prediction. The analysis of the healthy parts proves how the GGHMM guarantees the robustness with respect to possible measuring interferences avoiding state transitions although no damage evolution occurred. Both validations already analysed regard a single component, e.g. a rolling element bearing, where a monovariate model, i.e. a single observation set, enables the correct damaging process description and proves its prognostic effectiveness. However, the industrial word requires the analysis of complex systems, where several components may reach failure. In this context, there is the need of exploiting more indicators

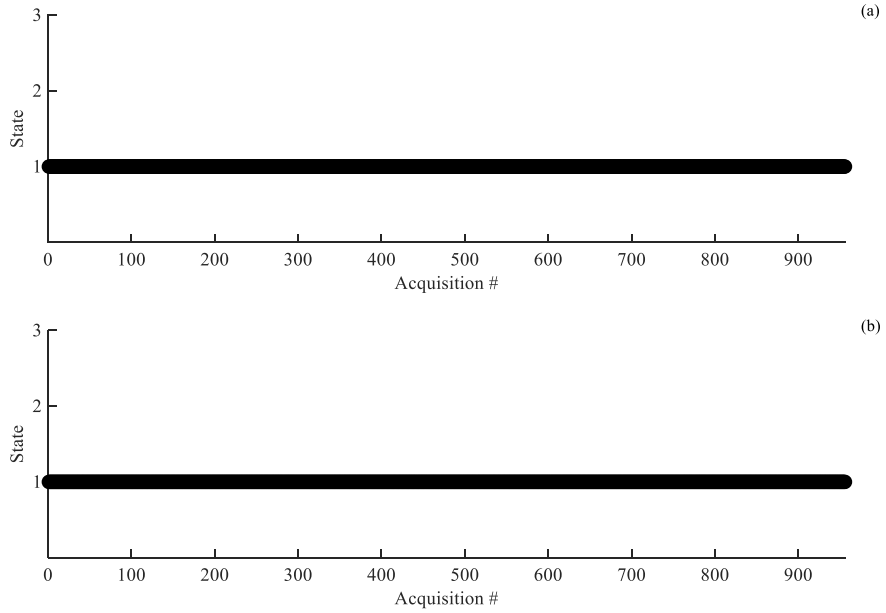

Figure 39: Estimated state sequences taking into account other bearing components: (a) BPFI, (b) BSF

for describing all the possible damaging sources, i.e. a multivariate model is required.

4.4.3 *Multivariate HMM for the prognostics of complex mechanical systems*

The last experimental validation deals with a real industrial case study regarding a planetary gearbox. The target of the analysis of a complex system is the demonstration of the possibility to exploit the GGHMM for the detection and prediction of several kind of faults through the training of the model with more than a single observation vector, i.e. with a multivariate generalized Gaussian mixture distribution. The experimental setup is described in Fig.40. The three stages planetary gearbox under test (hereafter called reducer) is driven by an electric motor and coupled to a second identical gearbox (from now named multiplier) driven by an electric motor working as brake for increasing the contact forces between the teeth. The test has been performed at constant

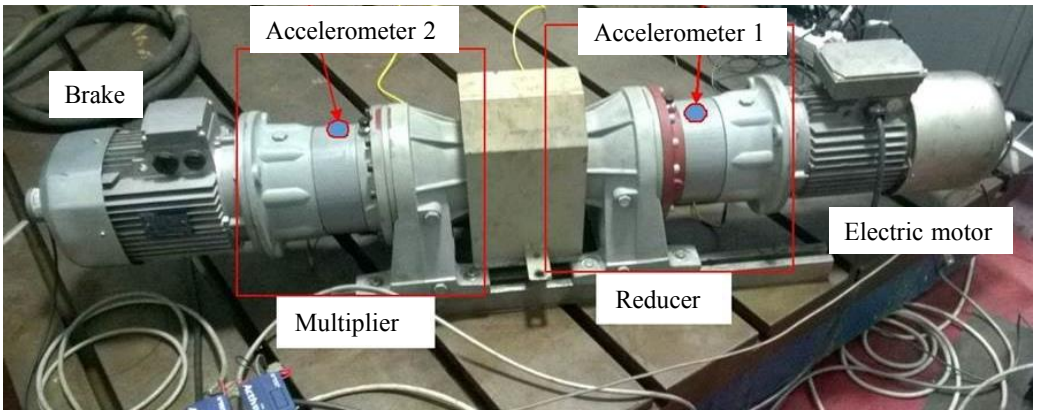

Figure 40: Experimental setup for the analysis of the gearbox run to failure test

speed with two different values of torque applied (for confidentiality reasons called low and high). The presence of different test conditions may bring to light another interesting aspect: each model state does not only represent a damaging level but may also discriminate several working conditions [97]. The vibration signal on the reducer has been measured with an industrial monoaxial accelerometer and the tachometer signal on the input shaft has been acquired through an encoder mounted on the electric motor shaft. The test has been stopped after 5 days and a deep pitting on the elements, i.e. sun gear, planets and anulus, of the second stage has been found after the inspection on the physical system.

The diagnostic indicators to be considered as physical observations for the training phase of the model should be able to detect the presence of different types of fault and thus a suitable solution is the selection of criteria that analyse the vibration signal in a general way, without focusing on a particular components (main characteristics of cyclostationary indicators or the TSA for example). For this purpose, for the HMM presented in this section, two observation dataset have been considered: the RMS and the Kullback-Leibler Divergence (KLD) calculated on the raw vibration signal. The KLD, also known as relative entropy, is a measure of distance between two distribution [114] where one of them is taken as reference. For this case study the KLD has been calculated for each acquisition considering the first sample as reference, under the hypothesis that the system is healthy in this time span. The system degradation should reflect on modifications of the signal distribution from

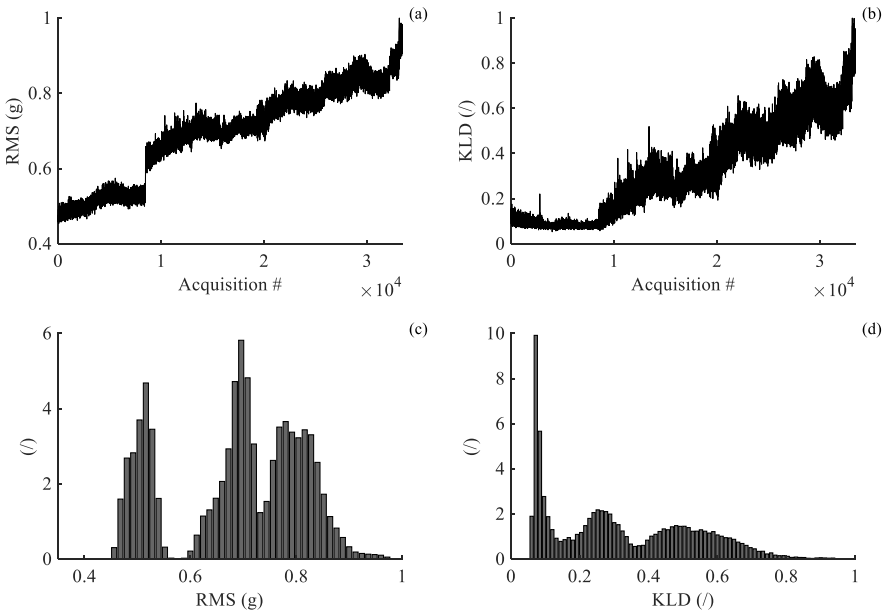

Figure 41: Experimental observation from Accelerometer 1: (a) RMS values, (b) KLD values, (c) RMS distribution, (d) KLD distribution

the ideal Gaussian condition and consequently the KLD should increase with the damage evolution. The evolution of the RMS and the KLD during the entire run to failure test is displayed in Fig.41(a-b). The RMS value is strongly dependent on the applied torque due to its energetic nature and this behaviour can be seen in the discontinuity around sample 9500. On the other hand, the load effect on the KLD value has a lower specific weight being this indicator based on the data distribution. These difference behaviours are clearly explainable also through the observation distribution shown in Fig.41(c-d). Both distributions are represented by a mixture of three components. The first one is related to the data acquired under the lower torque and the others refers to the test with the higher load applied. The RMS distribution (Fig.41(c)) is characterized by a clear division between the first state and the others related to the higher torque. On the contrary, in the KLD distribution the three components are connected each other, confirming the lower influence of the load on a generic distribution based diagnostic indicator. The exploitation of two observation datasets entail a bivariate basis PDF for the construction of

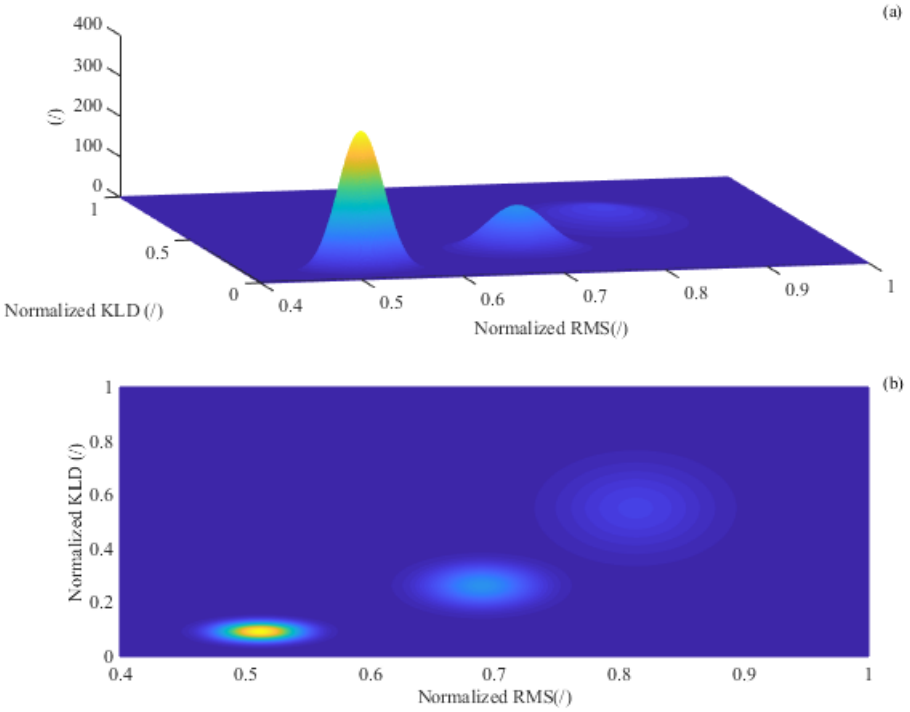

Figure 42: Bivariate PDFs from RMS and KLD distributions: (a) 3D view, (b) contour plot

the HMM. The different natures of the chosen criteria allows the assumption that these components can be considered as s-independent, thus the bivariate PDF can be expressed according to Eq.51. Fig.42 shows the bivariate mixture PDF obtained from the independent monovariate components described in 41(c-d). From the contour plot (Fig.42(b)) the three components of the mixture density are clearly identifiable and it is possible to split the graph into the low torque area (down-left) and high load area. In the latter, two states can be identified, each one related to a different damaging level (being the test conditions constant in this area). As already done in the previous chapter, the datasets have been separated in two different parts with the same number of samples, i.e. the training dataset and the validation dataset.

Table 15: Model parameters for the analysis of the planetary gearbox dataset: initial state and transition probabilities

	Model parameters		
	0.986	0.014	0
A	0	0.884	0.116
	0	0	0.1
π	1	0	0

Table 16: Estimated parameters of the conditional PDFs for the analysis of the planetary gearbox dataset

State	RMS			KLD		
#	μ	p	Σ	μ	p	Σ
1	0.52	2	0.05	0.088	2	0.04
2	0.68	2.3	0.12	0.26	2.4	0.14
3	0.82	2.4	0.15	0.53	2.6	0.16

The estimated model parameters are reported in Tab.15 and Tab.16. It should be noticed that the probability matrix **A** has been defined in the same form of a pure left-to-right model. This can be done because the two applied torques have been test in a sequential way without coming back to the lower value. If the two conditions had been tested alternatively, the element a_{21} could not be set to 0. Analysing the parameters of the bivariate PDFs (Tab.16), the main attention should be pointed out on the shape factor p. In the first state, i.e. the condition of lower load, the value is comparable to the Gaussian reference for both distributions confirming the PDF shape in Fig.41(c-d). Moving to the second state, i.e. the first state with higher torque, both distributions of RMS and KLD seem to depart from the ideal Gaussian shape, as confirmed by the higher shape factor. This behaviour is interesting because brings to light the possible appearance of an early damaging stage. In fact, as already discussed in the previous experimental validation, the observation distribution starts to move away from the Gaussian when the damaging process is ongoing. The

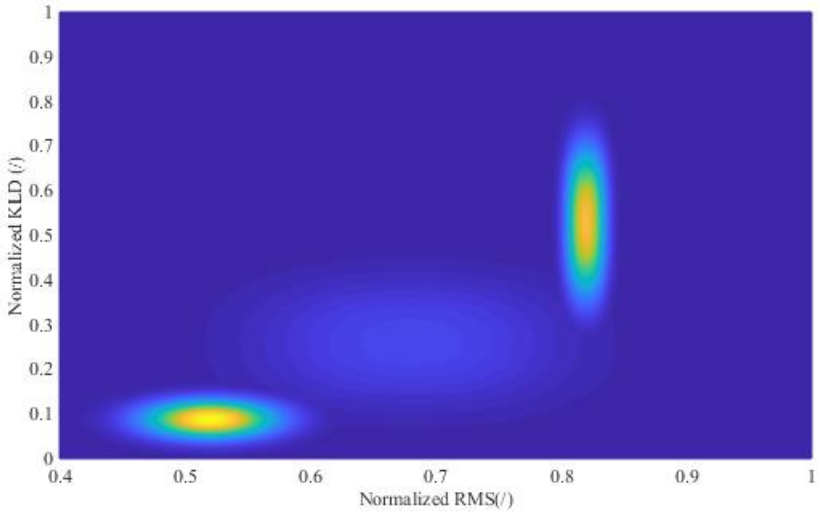

Figure 43: Estimated mixture of bivariate PDFs for the training dataset

application of an higher load should not changing the observation distribution and consequently the increasing of the shape factor should detect a possible fault appearance. Finally, in the last stage where the fault severity is relevant both distributions (in particular for the KLD) are flatter and consequently far from the Gaussian condition. The mixture of bivariate GGDs obtained from the values in Tab.16 is illustrated in the contour plot in Fig.43.

The above described model has been validated through the RMS and KLD datasets shown in Fig.44(a-b). The resulting state sequence is depicted in Fig.44(c). The sensitivity of the RMS with respect to the different loads applied enables the correct estimation of the transition between state 1 and state 2. On the other hand, it is possible to note how both diagnostic indicators do not present a sudden increment on the trend related to the fault appearance but the damage developments seems to be gradual. This system behaviour reflects on the presence of an uncertainty phase during the transition from state 2 to state 3. This unfavourable conditions is increased by the physiological fluctuation of the KLD values being it based on the comparison between two data distributions instead of a direct calculation on the vibration signal. However, despite the increment of the fluctuation amplitude in the last stages of the degradation process, the model demonstrate its robustness with respect

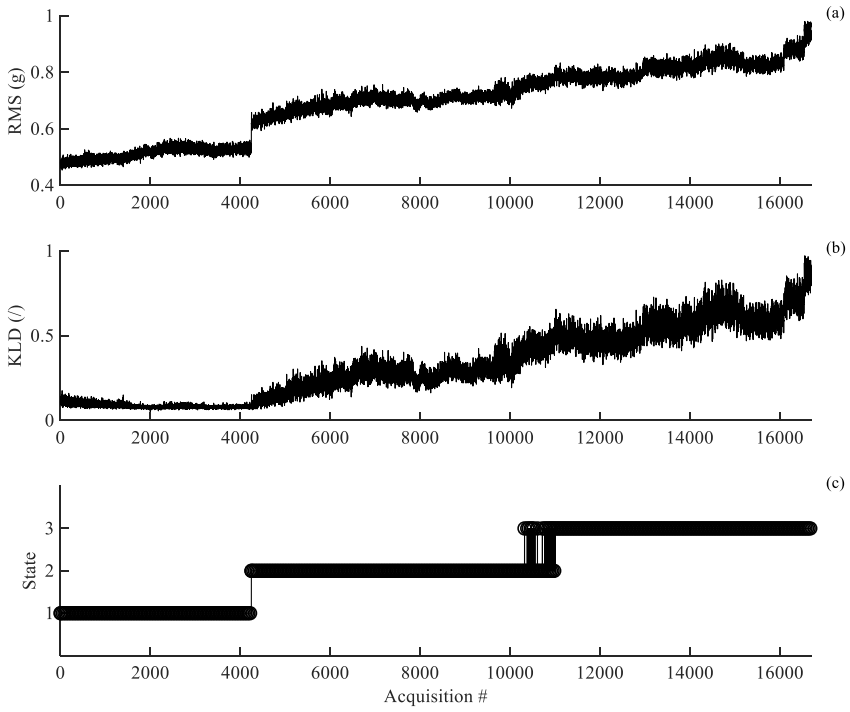

Figure 44: Estimated state sequence for the validation dataset: (a) RMS values, (b) KLD values, (c) estimated state sequence

to possible outliers in the last part of the test that may lead to an overestimation of the RUL.

This experimental validation demonstrated how the application of a multivariate GGHMM enables the prognostics of complex system characterized by several possible fault sources. The exploitation of observation vectors represented by diagnostic criteria that describe the overall damaging level of the system increments the model effectiveness embracing all the possible failure modes. In addition, the analysis of a run to failure test performed with different test specifications proved the ability of the proposed HMM to discriminate the different working condition even if applied for the system under the same health state.

4.5 SUMMARIZING REMARKS

In this chapter, a novel approach for the development of HMMs for prognostic purposes has been proposed. The model is based on the GGD as the mixture components for describing the fault appearance and evolution. The exploitation of a generalized PDF enables a better fitting quality for the distribution of the physical observation, i.e. the trend of some diagnostic indicators, specially in the last stages of the damaging process where these PDFs depart from the ideal Gaussian shape. An iterative algorithm for the estimation of the model parameters has been defined and the resulting model has been validated on real systems from both academic and industrial word. The aforementioned activity led to the following interesting results:

- A new HMM has been proposed exploiting a generalized basis distribution in order to take into account possible modification during the working life due to the evolution of the damaging process. An iterative algorithm for the estimation of the model parameters has been proposed both for the case of monovariate distribution, i.e. a single observation vector, and multivariate PDFs, i.e. several observation vectors.

- The practical implication of the different basis distribution has been demonstrated through the analysis of a run to failure test performed on the bearing test bench of the University of Ferrara. The comparison has been carried out in order to prove how the proposed model allows a better description of the damaging evolution, in particular in the last stages near the final failure. This aspect has been quantitatively demonstrated through a statistical test. Moreover, this experimental validation demonstrated how the robustness of the GGD avoids possible errors in the state estimation due to measuring problems, leading to a better estimation of the RUL.

- The analysis of the IMS run to failure test highlights how the exploitation of indicators fitted for the damaging description, e.g. cyclostationary indicators, as physical observations improves the model effectiveness, in particular for the analysis of not monotonically increasing damaging processes. This analysis underlines the close link between diagnostics and prognostics demonstrating the importance on selecting the correct

observation set for training the model in order to increase the model capability on following the fault evolution.

- The study of a real industrial case on a planetary gearbox has been taken into account for demonstrating how the exploitation of a multivariate model enables the prognostics of complex system characterized by several possible damaging sources. This validation demonstrated also how the model states may not only be related to different damaging level but also to different working conditions.

The main aspect of originality proposed in this chapter regards the definition of an iterative algorithm for the estimation of the model parameters for the GGD as basis function for the construction of an HMM. All the theoretical consideration depicted in this chapter have been supported by several experimental validation on run to failure test performed on single mechanical components or more complex system from both the academic and industrial word.

5

FINAL REMARKS

This thesis has been focused on the diagnostics and prognostics of rotating machines through cyclostationary techniques, highlighting the strict connection between the two fields.

After a brief introduction regarding the state of art of diagnostic signal processing techniques in Chapter 1, Chapter 2 described an overview about the experimental datasets used for validating the algorithms proposed in the thesis. These datasets are composed by vibration signals acquired on the bearing test bench at the University of Ferrara during several typologies of test, e.g. run to failure or stationary tests.

Chapter 3 has been focused on the development of a new BD method based on the combination between the cyclostationary theory and the FBSE. The FBSE has been taken into account for improving the diagnostic effectiveness of the cyclostationary based BD through a series expansion that better describes the mathematical nature of the fault related impulsive signal. The proposed cyclostationary criterion, named ICS_{2FB}, has been defined and the resulting method, named FBBD, has been compared with the other cyclostationary based BD method considering both simulated and real vibration signals.

Taking into account simulated signals reproducing cyclostationary patterns under both stationary and non-stationary working conditions, the effectiveness of the FBBD has been investigated in comparison to the other cyclostationary BD method both in qualitative and quantitative terms. This analysis demonstrated how the exploitation of the FBSE enables the reduction of the number of harmonics of the cyclic frequency required for the description of the impulsive

excitation, consequently reducing the computational costs of the algorithm. In this contest, the proposed method also proved to guarantee a lower residual background noise in the estimated source, enabling the detection of the weakest peaks. The analysis of simulated signals proved also the robustness of the proposed method to the presence of possible interferences that may mask the target fault related pattern. This experimental validation also demonstrated the robustness of the method with respect to the free parameters of the algorithm such as the FIR filter length and the cyclic period to be investigated.

Moving to the real signals, the diagnostic capability of the proposed BD indicator has been investigated through the analysis of several applications on faulty bearings both from the academic and industrial world. The analysis of several bearings with different fault sizes under both constant and variable speed conditions has been taken into account for demonstrating the sensitivity of the novel criterion to the damage severity. The analysis of a bearing run to failure test demonstrated how the ICS_{2FB} allows the real time assessment of the damaging level during the entire operating life, describing the real degradation evolution in all its stages. This validation has been also exploited for describing the availability of the novel BD indicator as physical observation for building a prognostic model, demonstrating the link between the two fields.

The main aspect of novelty of this research activity is represented by the rewriting of the existing ICS_2 through the FBSE in order to define a novel cyclostationary indicator that better fits the modulated nature of the classic impulsive pattern related to a faulty rotating machine. The direct exploitation of cyclostationarity in BD has been recently proposed and the only existing method, named CYCBD, is based on the classic FSE. Consequently, the proposed criterion improves the diagnostic capability allowing a better fitting quality of the target sources. A complete study is provided, from the analytical formulation of the proposed method to the numerical and experimental validation. Moreover, the mathematical definition has been extended to the angular domain enabling the fault detection under variable regimes in order to avoid possible issues related to the speed fluctuations, pivotal aspect in the real industrial scenario.

Chapter 4 has been devoted to the development and validation of a novel prognostic model based on the HMM. The proposed model considers the

relation between physical observations and damaging state as driven by a generalized Gaussian mixture distribution in order to allows distribution modifications from the first states (healthy conditions) and the last states (damaged system) based on the values of some distribution parameters. In Chapter 4 a novel iterative algorithm for the estimation of the model parameters starting from some diagnostic indicator as observation has been defined. The proposed HMM has been validated through the analysis of several run to failure tests regarding both single components, e.g. bearings, or more complex systems, e.g. planetary gearboxes.

The analysis of single components, i.e. the validation of the monovariate GGHMM, demonstrated the improvement given by the proposed model with respect to the classic Gaussian HMM in terms of fitting quality of the real observation distribution. This validation proved how the observation distribution in the last states moves away from the ideal Gaussian form. Consequently a density function that enables the modification of the function form among all states allows a better assessment of the actual damaging level leading to a more accurate estimation of the RUL. The comparison highlighted the robustness of the GGHMM with respect to possible outliers on the observation vector that may reflect on issues in the state estimation and consequently in the RUL prediction.

The exploitation of the ICS_{2FB} as model observation demonstrated the effectiveness of the proposed method in the assessment of the damaging evolution in case of not monotonically increasing fault severity (like the classical pitting evolution on rolling element bearings). This particular experimental validation explained how the selection of a diagnostic indicator particularly designed for the application object of study strongly improves the effectiveness of the prognostic model justifying the aforementioned relations between diagnostics and prognostics.

The analysis of a planetary gearbox demonstrated the effectiveness of multivariate GGHMMs for the prognostics of complex systems where several different damaging sources may lead to the final failure. In this contest, the selection of diagnostic indicators sensitive to the overall damaging level as model observations proved to be effective for the application of the proposed HMM. In this contest the GGHMM demonstrated its ability also in discriminating between different test conditions, e.g. different loads applied on the

system, removing the requirement of stationary conditions for the application of the prognostic model.

For this second research activity, the originality is represented by the estimation algorithm proposed for the estimation of the model parameters in case of generalized Gaussian density function. As the author is aware, the existing HMMs consider the same distribution for all the model states and consequently do not take into account the changes of distribution related to the appearance and evolution of the fault. The proposed model overcomes this limitation through a generalized distribution that may assume several different forms according to the values of the model parameters. The mathematical formulation of the novel iterative algorithm has been also defined for the case of multivariate distribution in order to extend the model applicability for complex systems where a single observation set may no longer be able to describe correctly the damaging process.

A

APPENDIX: PROOF OF EQUATIONS

A.1 BD INDICATOR IN THE ANGULAR DOMAIN

The Fourier-Bessel coefficients for a generic analog signal $x(t)$ are defined as [58]:

$$c_F^k = \frac{2}{T^2 J_1^2(\beta_k)} \int_0^T t x(t) J_0\Big(\frac{\beta_k t}{T}\Big) dt \tag{58}$$

where T is the signal time period and J_0 and J_1 the zero order and first order Bessel functions, respectively. Under the assumption of time/angle dependent signal, i.e. $x = x(\theta(t))$, it is possible to apply a variable change:

$$c_F^k = \frac{2}{\Theta^2 J_1^2(\beta_k)} \int_0^\Theta \theta(t) x(\theta(t)) J_0\Big(\frac{\beta_k \theta(t)}{\Theta}\Big) \frac{d\theta}{dt} dt \tag{59}$$

where Θ is the angular position related to the time period T. This expression can be simplified as follows:

$$c_F^k = \frac{2}{\Theta^2 J_1^2(\beta_k)} \int_0^\Theta \theta(t) x(\theta(t)) J_0\Big(\frac{\beta_k \theta(t)}{\Theta}\Big) \dot{\theta}(t) dt \tag{60}$$

where $\dot{\theta}(t)$ is the angular velocity, e.g. measured by a tachometer signal. Eq.60 can be rewritten for a generic digital signal $x(n)$, viz:

$$c_F^k = \frac{2\Delta t}{\Theta^2 J_1^2(\beta_k)} \sum_{n=0}^{N-1} \theta(n) x(n) J_0\Big(\frac{\beta_k \theta(n)}{\Theta}\Big) \dot{\theta}(n) \tag{61}$$

where N is the number of samples of $x(n)$ and Δt is the inverse of the sampling frequency. After some simple manipulation, the final form is given by the following:

$$c_F^k = \frac{2}{\Theta\dot{\Theta}J_1^2(\beta_k)} \sum_{n=0}^{N-1} \theta(n)x(n)J_0\Big(\frac{\beta_k\theta(n)}{\Theta}\Big)\dot{\theta}(n) \tag{62a}$$

$$\Theta = \sum_{n=0}^{N-1} \Delta\theta_n \tag{62b}$$

$$\dot{\Theta} = \sum_{n=0}^{N-1} \dot{\theta}(n) = \frac{\Theta}{\Delta t} \tag{62c}$$

A.2 HMM PARAMETER ESTIMATION FOR MULTIVARIATE GENERALIZED GAUSSIAN DISTRIBUTIONS

Under the hypothesis of s-independent components, the multivariate generalized Gaussian PDF for a random multivariate variable $X = [x_1, \ldots, x_K]$ is defined as:

$$f(X) = \frac{\prod_{k=1}^{K} p_k}{2^K|\boldsymbol{\Sigma}|^{\frac{1}{2}} \prod_{k=1}^{K} \Gamma\Big(\frac{1}{p_k}\Big)} e^{\left[-\sum_{k=1}^{K}\left(\frac{|x_k-\mu_k|}{\sigma_k^2}\right)^{p_k}\right]} \tag{63}$$

Eq.64 can be rewritten in order to express the PDF as function of a quadratic form, viz:

$$f(X) = \frac{\prod_{k=1}^{K} p_k}{2^K|\boldsymbol{\Sigma}|^{\frac{1}{2}} \prod_{k=1}^{K} \Gamma\Big(\frac{1}{p_k}\Big)} e^{\left\{-\sum_{k=1}^{K}\left[\left(\frac{|x_k-\mu_k|}{\sigma_k^2}\right)^2\right]^{\frac{p_k}{2}}\right\}} \tag{64}$$

Remembering Eq.40a, Eq.40b and Eq.41, the derivative inside the square bracket of Eq.41 is given by:

$$-2\frac{\partial f_i(X)}{\partial q_k^i(x)}\bigg|_{x=Y_{tk}} = f_i(X)\sum_{k=1}^{K} p_k^i q_k^i (Y_{tk})^{\frac{p_k^i}{2}-1}, \qquad i = 1,\dots,N \tag{65}$$

Substituting Eq.65 into Eq.41 and taking into account the definition of forward variable given in Eq.30b, the term ρ_t is:

$$\rho_t(i) = \alpha_t(i)\sum_{k=1}^{K} p_k^i q_k^i (Y_{tk})^{\frac{p_k^i}{2}-1}, \qquad i = 1,\dots,N \tag{66}$$

Substituting Eq.66 into Eq.40a and Eq.40b the final form of mean values and scale factors are defined as:

$$\mu_k^i = \frac{\sum_{t=1}^{T}\gamma_t(i)\left(\sum_{k=1}^{K} p_k^i q_k^i (Y_{tk})^{\frac{p_k^i}{2}-1}\right) Y_{tk}}{\sum_{t=1}^{T}\gamma_t(i)\sum_{k=1}^{K} p_k^i q_k^i (Y_{tk})^{\frac{p_k^i}{2}-1}}, \qquad i = 1,\dots,N \tag{67a}$$

$$\Sigma_k^i = \frac{\sum_{t=1}^{T}\gamma_t(i)\left(\sum_{k=1}^{K} p_k^i q_k^i (Y_{tk})^{\frac{p_k^i}{2}-1}\right)(Y_{tk}-\mu_k^i)^2}{\sum_{t=1}^{T}\gamma_t(i)}, \qquad i = 1,\dots,N \tag{67b}$$

Finally, according to the consideration depicted in Sec.4 the relation between scale factor and variance can be written as function of the observations as:

$$\left[\frac{\sum_{t=1}^{T}\gamma_t(i)\left(\sum_{k=1}^{K} p_k^i q_k^i (Y_{tk})^{\frac{p_k^i}{2}-1}\right)(Y_{tk}-\mu_k^i)^2}{\sum_{t=1}^{T}\gamma_t(i)}\right]^{\frac{1}{2}} =$$

$$\left[p_k^i\frac{\sum_{t=1}^{T}\gamma_t(i)\left(\sum_{k=1}^{K} p_k^i q_k^i (Y_{tk})^{\frac{p_k^i}{2}-1}\right)|Y_{tk}-\mu_k^i|^{p_k^i}}{\sum_{t=1}^{T}\gamma_t(i)}\right]^{\frac{1}{p_k^i}}, \quad i = 1,\dots,N \tag{68}$$

For the sake of clarity it has to be remembered that the notation μ_k^i refers to the parameter of the k-th components of the i-th state related multivariate distribution.

BIBLIOGRAPHY

[1] M. Buzzoni, E. Soave, G. D'Elia, E. Mucchi, and G. Dalpiaz. "A diagnostic protocol for the monitoring of bearing fault evolution based on blind deconvolution algorithms." In: *ISMA 2018 and USD 2018*. Leuven, Belgium, 2018, pp. 809–821.

[2] M. Buzzoni, E. Soave, G. D'Elia, E. Mucchi, and G. Dalpiaz. "Development of an indicator for the assessment of damage level in rolling element bearings based on blind deconvolution methods." In: *Shock and Vibration* 2018 (2018). DOI: 10.1155/2018/5384358.

[3] E. Soave, G D'Elia, and G. Dalpiaz. "Fourier-Bessel series expansion based blind deconvolution method for bearing fault detection." In: *Sourveillance 2019*. Lyon, France, 2019.

[4] E. Soave, G. D'Elia, M. Cocconcelli, and M. Battarra. "Blind deconvolution criterion based on Fourier-Bessel series expansion for rolling element bearing diagnostics." In: *Mechanical System and Signal Processing* 169 (2022), pp. 83–101. DOI: 10.1016/j.ymssp.2021.108588.

[5] E. Soave, G. D'Elia, and E. Mucchi. "A laser triangulation sensor for vibrational structural analysis and diagnostics." In: *Measurement and Control* 53 (2020), pp. 73–82. DOI: 10.1177/0020294019877484.

[6] A. Gabrielli, F. Pizzolante, E. Soave, M. Battarra, C. Mazzeo, M. Tarabra, E. Fava, and E. Mucchi. "A numerical model for NVH analysis of gearboxes employed on agricultural equipment." In: *ISMA 2020 and USD 2020*. Leuven, Belgium, 2020, pp. 3191–3203.

[7] R. B. Randall. *Vibration-based Condition Monitoring*. Ed. by John Wiley and Sons Ltd. 2011.

[8] R. B. Randall. "A new method of modeling gear faults." In: *Trans. Asme J. Mech. Des.* 104.2 (1982), pp. 259–267. DOI: 10.1115/1.3256334.

[9] J. Antoni and R. B. Randall. "Differential diagnosis of gear and bearing faults." In: *Journal of Vibration and Acoustic* 124.2 (2002), p. 165. DOI: 10.1115/1.1456906.

[10] W. D. Mark. "Analysis of the vibratory excitation of gear systems: Basic theory." In: *The Journal of the Acoustical Society of America* 63.5 (1978), p. 1409. DOI: 10.1121/1.381876.

[11] W. D. Mark, H. Lee, R. Patrick, and J. D. Coker. "A simple frequency-domain algorithm for early detection of damaged gear teeth." In: *Mechanical Systems and Signal Processing* 24.8 (2010), pp. 2807–2823. DOI: 10.1016/j.ymssp.2010.04.004.

[12] Z. Man, W. Wang, S. Khoo, and J. Yin. "Optimal sinusoidal modelling of gear mesh vibration signals for gear diagnosis and prognosis." In: *Mechanical Systems and Signal Processing* 33 (2012), pp. 256–274. DOI: 10.1016/j.ymssp.2012.07.004.

[13] W. Wang. "Early detection of gear tooth cracking using the resonance demodulation technique." In: *Mechanical Systems and Signal Processing* 15.5 (2011), pp. 887–903. DOI: 10.1006/mssp.2001.1416.

[14] P. McFadden and J. smith. ""Model for the vibration produced by a single point defect in a rolling element bearing." In: *Journal of Sound and Vibration* 96.1 (1984), pp. 69–82. DOI: 10.1016/0022-460X(84)90595-9.

[15] R. B. Randall and J. Antoni. "Rolling element bearing diagnostics-A tutorial." In: *Mechanical Systems and Signal Processing* 25.2 (2011), pp. 485–520. DOI: 10.1016/j.ymssp.2010.07.017.

[16] J. Antoni and R. B. Randall. "A stochastic model for simulation and diagnostics of rolling element bearings with localized faults." In: *Journal of Vibration and Acoustic* 125.3 (2003), p. 282. DOI: 10.1115/1.1569940.

[17] J. Antoni, F. Bonnardot, A. Raad, and M. El Badaoui. "Cyclostationary modelling of rotating machine vibration signals." In: *Mechanical Systems and Signal Processing* 18.6 (2004), pp. 1285–1314. DOI: 10.1016/S0888-3270(03)00088-8.

[18] C. Capdessus, M. Sidhamed, and J. Lacoume. "Cyclostationary processes: application in gear faults early diagnosis." In: *Mechanical Systems and Signal Processing* 14.3 (2000), pp. 371–385. DOI: 10.1006/mssp.1999.1260.

[19] V. Sharma and A. Parey. "Gear crack detection using modified TSA and proposed fault indicators for fluctuating speed conditions." In: *Measurement* 90 (2016), pp. 560–575. DOI: 10.1016/j.measurement.2016.04.076.

[20] S. Braun. "The synchronous (time domain) average revisited." In: *Mechanical Systems and Signal Processing* 25.4 (2011), pp. 1087–1102. DOI: 10.1016/j.ymssp.2010.07.016.

[21] P.D. McFadden. "Examination of a technique for the early detection of failure in gears by signal processing of the time domain average of the meshing vibration." In: *Mechanical Systems and Signal Processing* 1.2 (1987), pp. 173–183. DOI: 10.1016/0888-3270(87)90069-0.

[22] P.D. McFadden. "Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration." In: *Journal of vibration, acoustics, stress, and reliability in design* 108.2 (1986), pp. 165–170. DOI: 10.1115/1.3269317.

[23] G. Dalpiaz, A. Rivola, and R. Rubini. "Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears." In: *Mechanical Systems and Signal Processing* 14.3 (2000), pp. 387–412. DOI: 10.1006/mssp.1999.1294.

[24] R. B. Randall. "A history of cepstrum analysis and its application to mechanical problems." In: *Mechanical Systems and Signal Processing* 97 (2017), pp. 3–19. DOI: 10.1016/j.ymssp.2016.12.026.

[25] H. Endo and R. B. Randall. "Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter." In: *Mechanical Systems and Signal Processing* 21.2 (2007), pp. 906–919. DOI: 10.1016/j.ymssp.2006.02.005.

[26] W. Wang and A.K. Wong. "Autoregressive model-based gear fault diagnosis." In: *Journal of Vibration and Acoustic* 124.2 (2002), p. 172.

[27] P.D. McFadden and J. Smith. "Vibration monitoring of rolling element bearings by the high-frequency resonance technique — a review." In: *Tribology International* 17.1 (1984), pp. 3–10. DOI: 10.1016/0301-679X(84)90076-8.

[28] J. Antoni. "Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions." In: *Journal of Sound and Vibration* 304.3-5 (2007), pp. 497–529. DOI: 10.1016/j.jsv.2007.02.029.

[29] J. Antoni. "Cyclostationarity by examples." In: *Mechanical Systems and Signal Processing* 23.4 (2008), pp. 987–1036. DOI: 10.1016/j.ymssp.2008.10.010.

[30] J. Antoni, G. Xin, and N. Hamzaoui. "Fast computation of the spectral correlation." In: *Mechanical Systems and Signal Processing* 92 (2017), pp. 248–277. DOI: 10.1016/j.ymssp.2017.01.011.

[31] P. Borghesani and M. Shahriar. "Cyclostationary analysis with logarithmic variance stabilisation." In: *Mechanical Systems and Signal Processing* 70-71 (2016), pp. 51–72. DOI: 10.1016/j.ymssp.2015.08.014.

[32] P. Borghesani and J. Antoni. "CS2 analysis in presence of non-Gaussian background noise – Effect on traditional estimators and resilience of log-envelope indicators." In: *Mechanical Systems and Signal Processing* 90 (2017), pp. 378–398. DOI: 10.1016/j.ymssp.2016.12.033.

[33] A. Raad, J. Antoni, and M. Sidahmed. "Indicators of cyclostationarity: Theory and application to gear fault monitoring." In: *Mechanical Systems and Signal Processing* 22.3 (2008), pp. 574–587. DOI: 10.1016/j.ymssp.2007.09.011.

[34] R.B. Randall, J. Antoni, and Chobosaard. "The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals." In: *Mechanical Systems and Signal Processing* 15.5 (2001), pp. 945–962. DOI: 10.1006/mssp.2001.1415.

[35] P. Borghesani, P. Pennacchi, and S. Chatterton. "The relationship between kurtosisand envelope-based indexes for the diagnostic of rolling element bearings." In: *Mechanical Systems and Signal Processing* 43.1-2 (2014), pp. 25–43. DOI: 10.1016/j.ymssp.2013.10.007.

[36] J. Lee, H. Qju, G. Yu, and J. Lin. "Rexnord Technical Services, 'Bearing Data Set', IMS, University of Cincinnati, NASA Ames Prognostics Data Repository." In: (2007). URL: http://ti.arc.nasa.gov/project/prognostic-data-repository..

[37] W.A. Smith and R.B. Randall. "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study." In: *Mechanical Systems and Signal Processing* 64-65 (2015), pp. 100–131. DOI: 10.1016/j.ymssp.2013.10.007.

[38] D.. Neupane and J. Seok. "Bearing fault detection and diagnosis using Case Western Reserve University dataset with deep learning approaches: A Review." In: *IEEE Access* 8 (2020), pp. 93155–93178. DOI: 10.1109/ACCESS.2020.2990528.

[39] A.H. Zamanian and A. Ohadi. "Gear fault diagnosis based on gaussiancorrelation of vibrations signals and wavelet coefficients." In: *Applied Soft Computing* 11.8 (2011), pp. 4807–4819.

[40] C. Bunks, D. McCarthy, and T. Al-Ani. "Condition-based maintenance of machines using hidden markov models." In: *Mechanical Systems and Signal Processing* 14.4 (2000), pp. 597–612.

[41] S.R. Saufi, Z.A.B. Ahmad, M.S. Leong, and M.H. Lim. "Gearbox fault diagnosis using a deep learning model with limited data sample." In: *IEEE Transactions on Industrial Informatics* 1.1 (2020), p. 99.

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

[43] J. Ding, Y. Lin, and S. Yu. "Dynamic unbalance detection of Cardan shaft in high-speed train applying double decomposition and double reconstruction method." In: *Measurement* 73 (2015), pp. 111–120. DOI: 10.1016/j.measurement.2015.05.016.

[44] P. Pennacchi and A. Vania. "Diagnosis and model based identification of a coupling misalignment." In: *Shock and Vibration* 12 (2005), pp. 293–308. DOI: 10.1155/2005/607319.

[45] R.A. Wiggins. "Minimum entropy deconvolution." In: *Geoexploration* 16.1-2 (1976), pp. 21–35. DOI: 10.1016/0016-7142(78)90005-4.

[46] C.A. cabrelli. "Minimum entropy deconvolution and simplicity: A noniterative algorithm." In: *Geophysics* 50.3 (1985), pp. 394–413. DOI: 10.1190/1.1441919.

[47] H. Endo and C. Randall R.B.and Gosselin. "Differential diagnosis of spall vs. cracks in the gear tooth fillet region: Experimental validation." In: *Mechanical Systems and Signal Processing* 23.3 (2009), pp. 636–651. DOI: 10.1016/j.ymssp.2008.08.015.

[48] L. Zhang and N. Hu. "Fault diagnosis of sun gear based on continuous vibration separation and minimum entropy deconvolution." In: *Measurement* 141 (2019), pp. 332–344. DOI: 10.1016/J.MEASUREMENT.2019.04.049.

[49] N. Sawalhi and H. Randall R.B.and Endo. "The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis." In: *Mechanical Systems and Signal Processing* 21.6 (2007), pp. 2616–2633. DOI: 10.1016/j.ymssp.2006.12.002.

[50] D. He, X. Wang, S. Li, J. Lin, and M. Zhao. "Identification of multiple faults in rotating machinery based on minimum entropy deconvolution combined with spectral kurtosis." In: *Mechanical Systems and Signal Processing* 81 (2016), pp. 235–249. DOI: 10.1016/j.ymssp.2016.03.016.

[51] G.L. McDonald, Q. Zhao, and M.J. Zuo. "Maximum Correlated Kurtosis Deconvolution and application on gear tooth chip fault detection." In: *Mechanical Systems and Signal Processing* 33 (2012), pp. 237–255. DOI: 10.1016/j.ymssp.2012.06.010.

[52] G.L. McDonald and Q. Zhao. "Multipoint Optimal Minimum Entropy Deconvolution and convolution fix: Application to vibration fault detection." In: *Mechanical Systems and Signal Processing* 82 (2017), pp. 461–477. DOI: 10.1016/j.ymssp.2016.05.036.

[53] Y. Miao, M. Zhao, J. Lin, and Y. Lei. "Application of an improved Maximum Correlated Kurtosis Deconvolution method for fault diagnosis of rolling element bearings." In: *Mechanical Systems and Signal Processing* 92 (2017), pp. 173–195. DOI: 10.1016/j.ymssp.2017.01.033.

[54] J. Antoni and P. Borghesani. "A statistical methodology for the design of condition indicators." In: *Mechanical Systems and Signal Processing* 114 (2019), pp. 290–327. DOI: 10.1016/j.ymssp.2018.05.012.

[55] M. Buzzoni, J. Antoni, and G. D'Elia. "Blind deconvolution based on cyclostationarity maximization and its application to fault identification." In: *Journal of Sound and Vibration* 432 (2018), pp. 569–601. DOI: 10.1016/j.jsv.2018.06.055.

[56] G. D'Elia, S. Delvecchio, and G. Dalpiaz. "On the use of fourier-bessel series expansion for gear diagnostics." In: *Proc. of second CMMNO*. Hammamet, Tunisia, 2012, pp. 267–275.

[57] J. Schroeder. "Signal processing via Fourier-Bessel series expansion." In: *Digital Signal Processing* 3 (1993), pp. 112–124. DOI: 10.1006/DSPR.1993.1016.

[58] R.B. Pachori and P. Sircar. "A new technique to reduce cross terms in the Wigner distribution." In: *Digital Signal Processing* 17 (2007), pp. 466–474. DOI: 10.1016/j.dsp.2006.10.004.

[59] B. Kilundu, X. Chiementin, J. Duez, and D. Mba. "Cyclostationarity of Acoustic Emissions (AE) for monitoring bearing defects." In: *Mechanical Systems and Signal Processing* 25.6 (2011), pp. 2061–2072. DOI: 10.1016/j.ymssp.2011.01.020.

[60] B.N. Parlett. *The symmetric eigenvalue problem*. Ed. by Society for Industrial and Applied Mathematics. 1987. ISBN: 9780898714029.

[61] V.T. Tran, F. AlThobiani, A. Ball, and B.K. Choi. "An application to transient current signal based induction motor fault diagnosis of Fourier–Bessel expansion and simplified fuzzy ARTMAP." In: *Expert Systems with Applications* 40.13 (2013), pp. 5372–5384. DOI: 10.1016/j.eswa.2013.03.040.

[62] G. D'Elia, M. Cocconcelli, and E. Mucchi. "An algorithm for the simulation of faulted bearings in non-stationary conditions." In: *Meccanica* 53 (2018), pp. 1147–1166. DOI: 10.1007/s11012-017-0767-1.

[63] P. Borghesani, P. Pennacchi, S. Chatterton, and R. Ricci. "The velocity synchronous discrete Fourier transform for order tracking in the field of rotating machinery." In: *Mechanical Systems and Signal Processing* 44.1-2 (2014), pp. 118–133. DOI: 10.1016/j.ymssp.2013.03.026.

[64] J.P. Den Hartog. *Mechanical vibrations*. Ed. by Dover Pubns. 2003. ISBN: 978-0486647852.

[65] C.F.N. Cowan and P.M. Grant. *Adaptive filters*. Ed. by Prentice-Hall Inc. 1985. ISBN: 0-13-004031-1.

[66] C. Cassisi, P. Montalto, Aliotta M.A., A. Cannata, and A. Pulvirenti. "Similarity measures and dimensionality reduction techniques for time series data mining." In: *Advances in data mining knowledge discovery and applications*. Ed. by A. Karahoca. Intech, 2012. Chap. 3.

[67] J.W. Tukey. *Exploratory data analysis*. Ed. by Pearson. 1977. ISBN: 978-0201076165.

[68] S. Delvecchio, G. D'Elia, and G. Dalpiaz. "On the use of cyclostationary indicators in IC engine quality control by cold tests." In: *Mechanical Systems and Signal Processing* 60-61 (2015), pp. 208–228. DOI: 10.1016/j.ymssp.2014.09.015.

[69] M. Malagò, E. Mucchi, and G. Dalpiaz. "Fault detection in heavy duty wheels by advanced vibration processing techniques and lumped parameter modeling." In: *Mechanical Systems and Signal Processing* 70-71 (2016), pp. 141–160. DOI: 10.1016/j.ymssp.2015.09.043.

[70] F. Kozing. "Autoregressive moving average models of earthquake records." In: *Probabilistic Engineering Mechanics* 3.2 (1988), pp. 58–63. DOI: 10.1016/0266-8920(88)90016-1.

[71] T. Williams, X. Ribadeneira, S. Billinton, and T. Kurfess. "Rolling element bearing diagnostics in run-to-failure lifetime testing." In: *Mechanical Systems and Signal Processing* 15.5 (2001), pp. 979–993. DOI: 10.1006/mssp.2001.1418.

[72] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin. "Machinery health prognostics: A systematic review from data acquisition to RUL prediction." In: *Mechanical Systems and Signal Processing* 104 (2018), pp. 799–834. DOI: 10.1016/j.ymssp.2017.11.016.

[73] P.C. Paris and F. Erdogan. "A critical analysis of crack propagation laws." In: *Journal of Fluids Engineering* 85 (1963), pp. 528–533. DOI: 10.1115/1.3656900.

[74] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk, and S Liang. "Adaptive prognostics for rolling element bearing condition." In: *Mechanical Systems and Signal Processing* 13.1 (1999), pp. 103–113. DOI: 10.1006/mssp.1998.0183.

[75] J. Sun, H. Zuo, W. Wang, and M.G. Pecht. "Prognostics uncertainty reduction by fusing on-line monitoring data based on a state-space-based degradation model." In: *Mechanical Systems and Signal Processing* 45.2 (2014), pp. 396–407. DOI: 10.1016/j.ymssp.2013.08.022.

[76] A. Cubillo, S. Perinpanayagam, and M.E. Miguez. "A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery." In: *Advances in Mechanical Engineering* 8.8 (2016), pp. 1–21. DOI: 10.1177/1687814016664660.

[77] N.Z. Gebraeel and M.A. Lawley. "A neural network degradation model for computing and updating residual life distributions." In: *IEEE Transactions on Automation Science and Engineering* 5.1 (2008), pp. 154–163. DOI: 10.1109/TASE.2007.910302.

[78] Y. Pan, M.J. Er, X. Li, H. Yu, and R. Gouriveau. "Machine health condition prediction via online dynamic fuzzy neural networks." In: *Engineering Applications of Artificial Intelligence* 35 (2014), pp. 105–113. DOI: 10.1016/j.engappai.2014.05.015.

[79] L. Xiao, X. Chen, X. Zhang, and M. Liu. "A novel approach for bearing remaining useful life estimation under neither failure nor suspension histories condition." In: *Journal of Intelligent Manufacturing* 28.8 (2015), pp. 1–22. DOI: 10.1007/s10845-015-1077-x.

[80] V.N Vapnik. "An overview of statistical learning theory." In: *IEEE Transactions on Neural Networks* 10.5 (1999), pp. 988–999. DOI: 10.1109/72.788640.

[81] R. Khelif, B. Chebel-Morello, S. Malinowski, E. Laajili, F. Fnaiech, and N. Zerhouni. "Direct remaining useful life estimation based on support vector regression." In: *IEEE Transactions on Industrial Electronics* 64 (2017), pp. 2276–2285. DOI: 10.1109/TIE.2016.2623260.

[82] T.H. loutas, D. Roulias, and G. Georgoulas. "Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic E-support vectors regression." In: *IEEE Transactions on Reliability* 62.4 (2013), pp. 821–832. DOI: 10.1109/TR.2013.2285318.

[83] J. Liu and E. Zio. "An adaptive online learning approach for support vector regression: Online-SVR-FID." In: *Mechanical Systems and Signal Processing* 76-77 (2016), pp. 796–809. DOI: 10.1016/j.ymssp.2016.02.056.

[84] D. Barraza-Barraza, V.G. Tercero-Gomez, M.G. Beruvides, and J. Limon-Robles. "An adaptive ARX model to estimate the RUL of aluminum plates based on its crack growth." In: *Mechanical Systems and Signal Processing* 82 (2017), pp. 519–536. DOI: 10.1016/j.ymssp.2016.05.041.

[85] Y. Qian, R. Yan, and H. Shijie. "Bearing degradation evaluation using recurrence quantification analysis and Kalman filter." In: *IEEE Transactions on Instrumentation and Measurement* 63 (2014), pp. 2599–2610. DOI: 10.1109/TIM.2014.2313034.

[86] D. Cox. "Regression models and life-tables." In: *Journal of the Royal Statistical Society* 34.2 (1972), pp. 187–220.

[87] V. Makis and A.K.S. Jardine. "Optimal replacement in the proportional hazards model." In: *INFOR: Information Systems and Operational Research* 30.1 (1992), pp. 172–183. DOI: 10.1080/03155986.1992.11732183.

[88] V. Makis and X. Jiang. "Optimal replacement under partial observations." In: *Mathematics of Operations Research* 28.2 (2003), pp. 382–394.

[89] A.K.S. Jardine, D. Banjevic, and V. Makis. "Optimal replacement policy and the structure of software for condition-based maintenance." In: *Journal of Quality in Maintenance Engineering* 3.2 (1997), pp. 109–119. DOI: 10.1108/13552519710167728.

[90] P.J. Vlok, J.L. Coetzee, D. Banjevic, A.K.S. Jardine, and V. Makis. "Optimal component replacement decisions using vibration monitoring and the proportional-hazards model." In: *Journal of the Operational Research Society* 53.2 (2002), pp. 193–202. DOI: 10.1057/palgrave.jors.2601261.

[91] Z.S. Ye and M. Xie. "Stochastic modelling and analysis of degradation for highly reliable products." In: *Applied Stochastic Modelling in Business and Industry* 31.1 (2015), pp. 16–32. DOI: 10.1002/asmb.2063.

[92] J.P. Kharoufeh and S.M. Cox. "Stochastic models for degradation-based reliability." In: *IIE Transactions* 37.6 (2005), pp. 533–542. DOI: 10.1080/07408170590929009.

[93] I. Visser. "Seven things to remember about hidden Markov models: A tutorial on Markovian models for time series." In: *Journal of Mathematical Psychology* 55.6 (2011), pp. 403–415. DOI: 10.1016/j.jmp.2011.08.002.

[94] L.E. Baum and T. Petrie. "Statistical inference for probabilistic functions of finite state Markov chains." In: *The Annals of Mathematical Statistics* 37.6 (1966), pp. 1554–1563.

[95] J. Baker. "The Dragon system - An overview." In: *IEEE Transactions on Acoustics, Speech and Signal Processing* 23.1 (1975), pp. 24–29.

[96] F. Jelinek, L. Bahl, and R. Mercer. "Design of a linguistic statistical decoder for the recognition of continuous speech." In: *IEEE Transactions on Information Theory* 21.3 (1975), pp. 250–256. DOI: 10.1109/TIT.1975.1055384.

[97] C. Bunks, D. McCarthy, and T. Al-Ani. "Condition-based maintenance of machines using hidden Markov models." In: *Mechanical Systems and Signal Processing* 14.4 (2000), pp. 597–612. DOI: 10.1006/mssp.2000.1309.

[98] S.S.H. Zaidi, S. Aviyente, M. Salman, and K.K. Shin. "Prognosis of gear failures in DC starter motors using hidden Markov models." In: *IEEE Transactions on Industrial Electronics* 58.5 (2011), pp. 1695–1706. DOI: 10.1109/TIE.2010.2052540.

[99] L. Rabiner. "A tutorial on hidden Markov models and selected applications in speech recognition." In: *Proceedings of the IEEE* 77 (1989), pp. 257–286. DOI: 10.1109/5.18626.

[100] M. Karny. "Recursive estimation of high-order Markov chains: Approximation by finite mixtures." In: *Information Sciences* 326 (2015), pp. 188–201. DOI: 10.1016/j.ins.2015.07.038.

[101] C.M. Bishop. *Pattern recognition and machine learning*. Ed. by Springer. 2006. ISBN: 978-0387-31073-2.

[102] G. Schwarz. "Estimating the dimension of a model." In: *Annals of Statistics* 6.2 (1978), pp. 461–464. DOI: 10.1214/aos/1176344136.

[103] W. Zucchini. "An introduction to model selection." In: *Journal of Mathematical Psychology* 44.1 (2000), pp. 41–61. DOI: 10.1006/jmps.1999.1276.

[104] A. Viterbi. "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm." In: *IEEE Transactions on Information Theory* 13.2 (1967), pp. 260–269. DOI: 10.1109/TIT.1967.1054010.

[105] L.A. Liporace. "Maximum likelihood estimation for multivariate observations of Markov sources." In: *IEEE Transactions on Information Theory* 28.5 (1982), pp. 729–734. DOI: 10.1109/TIT.1982.1056544.

[106] A.P. Dempster, N.M. Laird, and D.B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." In: *Journal of the Royal Statistical Society* 39.1 (1977), pp. 1–38.

[107] M.K. Varanasi and B. Aazhang. "Parametric generalized Gaussian density estimation." In: *Journal of the Acoustical Society of America* 86.4 (1989), pp. 1404–1415. DOI: 10.1121/1.398700.

[108] R. Krupinski and J. Purczynski. "Approximated fast estimator for the shape parameter of generalized Gaussian distribution." In: *Signal Processing* 86.2 (2006), pp. 205–211. DOI: 10.1016/j.sigpro.2005.05.003.

[109] R. Bakis. "Continuous speech recognition via centisecond acoustic states." In: *Journal of the Acoustical Society of America* 59 (1976). DOI: 10.1121/1.2003011.

[110] P. Giudici, T. Ryden, and P. Vandekerkhove. "Likelihood-ratio tests for hidden Markov models." In: *Biometrics* 56.3 (2000), pp. 742–747. DOI: 10.1111/j.0006-341X.2000.00742.

[111] Z. Chen, Y. Li, T. Xia, and E. Pan. "Hidden Markov model with auto-correlated observations for remaining useful life prediction and optimal maintenance policy." In: *Reliability Engineering and System Safety* 184 (2019), pp. 123–136. DOI: 10.1016/j.ress.2017.09.002.

[112] K. Medjaher, D. Tobon-Mejia, and N. Zerhouni. "Remaining useful life estimation of critical components with application to bearings." In: *IEEE Transactions on Reliability* 61.2 (2012), pp. 292–302. DOI: 10.1109/TR.2012.2194175.

[113] A. Saxena, J. Celaya, B. Saha, and S. Saha. "Metrics for offline evaluation of prognostic performance." In: *International Journal of Prognostics and Health Management* 1.1 (2010), pp. 2153–2173. DOI: 10.36001/ijphm.2010.v1i1.1336.

[114] S. Kullback and R.A. Leibler. "On information and sufficiency." In: *Annals of Mathematical Statistics* 22.1 (1951), pp. 79–86. DOI: 10.1214/aoms/1177729694.