

POLITECNICO DI TORINO  
INGEGNERIA INDUSTRIALE  
SSD ING/IND13 MECCANICA APPLICATA ALLE MACCHINE

SCUOLA DI DOTTORATO  
PhD Course in Mechanical Engineering – XXV cycle

PhD Thesis

**Advanced techniques for aircraft  
bearing diagnostics**

**Miriam Pirra**

**Tutor**  
Prof. L. Garibaldi

February 2013

# Contents

<b>Introduction</b>	<b>1</b>
<b>1 Machine diagnostics</b>	<b>4</b>
1.1 Bearings . . . . .	4
1.1.1 Bearing damages . . . . .	8
1.2 Bearing diagnostics steps . . . . .	10
1.3 Data acquisition . . . . .	13
1.4 Signal processing . . . . .	16
1.4.1 Time domain techniques . . . . .	17
1.4.2 Frequency domain techniques . . . . .	19
1.4.3 Time-frequency domain techniques . . . . .	24
1.5 Features selection . . . . .	26
<b>2 Bearing diagnostics techniques</b>	<b>31</b>
2.1 State of the art . . . . .	32
2.2 Wavelet transform . . . . .	39
2.3 Empirical Mode Decomposition . . . . .	45
2.4 Spectral Kurtosis . . . . .	53
2.5 Speed and load variations . . . . .	60
<b>3 Anomaly detection</b>	<b>65</b>
3.1 Anomaly detection . . . . .	65
3.2 Classification based techniques . . . . .	67
3.2.1 Neural Network based . . . . .	70
3.2.2 Support Vector Machines based . . . . .	72
3.3 Nearest Neighbour based techniques . . . . .	76
3.4 Clustering based techniques . . . . .	78
3.5 Examples in machine diagnostics . . . . .	81

<b>4</b>	<b>Experimental tests</b>	<b>83</b>
4.1	Bearing test rig . . . . .	84
4.2	Sensors and acquisition system . . . . .	86
4.3	Bearings . . . . .	89
4.4	Test setup . . . . .	90
4.4.1	Setup I . . . . .	91
4.4.1.1	Different type of damage . . . . .	92
4.4.1.2	Endurance . . . . .	92
4.4.2	Setup II . . . . .	93
4.4.2.1	Temperature variations . . . . .	94
4.4.2.2	Speed and load small variations . . . . .	95
<b>5</b>	<b>Test results</b>	<b>97</b>
5.1	Damage detection through Spectral Kurtosis . . . . .	97
5.1.1	The model . . . . .	98
5.1.2	Fast Kurtogram . . . . .	100
5.1.3	Application to bearings data . . . . .	102
5.1.4	Methodology . . . . .	103
5.1.5	Damage identification . . . . .	106
5.1.6	Various disassemblies . . . . .	109
5.1.7	Damage evolution . . . . .	115
5.2	External condition removal in bearing diagnostics through EMD and One- Class SVM . . . . .	117
5.2.1	One-Class Support Vector Machine . . . . .	117
5.2.2	Empirical Mode Decomposition application . . . . .	119
5.2.3	Methodology I . . . . .	122
5.2.4	Application to bearings data . . . . .	123
5.2.5	Methodology II . . . . .	129
5.2.6	Application to bearing data . . . . .	130
5.2.7	Comparison with Wavelet Decomposition . . . . .	134
	<b>Conclusion</b>	<b>137</b>
	<b>A Bearing failures</b>	<b>139</b>
	<b>B Market analysis</b>	<b>148</b>
	<b>Bibliography</b>	<b>157</b>

# Introduction

Since rotating machinery is widely used in various industrial, military, and commercial processes, fault diagnosis and failure prognosis have gained significant attention in the last years. At the same time, rolling bearings are certainly among the most widely used components in machines. Thus, their condition monitoring and fault diagnosis are very important aspects to treat in order to prevent the occurrence of breakdowns.

It is well known, in fact, that bearings are the best location for measuring machinery vibrations since this is where the basic dynamic loads and forces of machines are applied and they are a critical component of machinery. This means that condition monitoring and fault diagnosis of bearings can represent properly the condition of machine and their failures often result in a critical damage, downtime, and costly repair.

A proper fault diagnosis and failure prognosis providing a condition based maintenance strategy to either machinery or components is, therefore, very important to the safety of the system and results in substantial economic benefits. A fundamental aspect related to this topic consists, then, in correctly identifying a degraded behaviour of a bearing with a reasonable level of confidence. This is one of the main requirements a Health and Usage Monitoring System (HUMS) should have.

The topics treated in this thesis are related to these concepts and their application in a real situation. In fact, our research group Dynamics & Identification Research Group (DIRG) has collaborated with Avio S.p.A. industry within a Piemonte region project named “Piattaforme per l’Aerospazio – Progetto GREAT 2020”. The main topic of the project is “Progetto McLab: Unità prognostica per trasmissioni meccaniche di potenza (Gearbox)”, whose goal is the definition of some algorithms able in detecting the damage on a gearbox with a certain confidence level.

During the analysis on real industrial gearboxes, it has been noticed that the damage detection methods could be influenced, in their diagnostic ability, by some external factors. For example, a variation in the temperature of the oil circulating or in the system configuration may furnish misleading indications on the healthy of the structure. So, it is important to develop a method that gives proper results, reducing the number of False Alarms, in such cases. The same thoughts regards the speed and load variations that sometimes characterise the system. It is clear that a reliable diagnostic technique should be able to produce the right

---

fault warning independently by these fluctuations.

To analyse in depth these aspects, a test rig has been designed and set up in the laboratory of the Department of Mechanical and Aerospace Engineering (DIMEAS) of Politecnico di Torino. A wide campaign of data acquisitions has started whose goal has been the creation of a consistent data base taking into account many varying conditions. The laboratory condition helps in assessing and verifying the algorithms to work in certain different situations thanks to the fact that here we can work in controlled conditions.

Two main methods are developed in the thesis. The first one is based on the Spectral Kurtosis, a statistical tool enabling the analysis of non-stationary signals, as those originated by bearings. Its ability in bearing damage detection is demonstrated through the application of the method to the data acquired on the laboratory test rig. The behaviour of damage indicators is analysed for an endurance test made on a roller element faulty bearing too. The other method exploited another technique which was shown to be highly efficient in non-stationary data analysis. It is called Empirical Mode Decomposition and it is self-adaptive because it enables the decomposition of any complicated signal into a collection of Intrinsic Mode Functions. A methodology allowing the identification of damages removing the influence of small variations in speed and load is developed. The feature vector characterising each acquisition is computed applying the EMD to the signals acquired on the test rig. The results of these works have been published in various conference and journal papers, such as [FMP<sup>+</sup>10], [PGT<sup>+</sup>11], [PFGM12] and [PMFG12].

## **Outline of the thesis**

This thesis aims to develop and present some methods in the field of bearing damage diagnostics. As already said, during the analysis of data acquired on industrial gearboxes, it has been noticed as conditions such as temperature and assemblies could influence the diagnostic results. In fact, a variation in these factors may be erroneously interpreted as a fault, while a simple changing has been done in the system.

Since the main topic of this thesis are bearings, the fundamental concept related to this important machine components are outlined in Chapter 1. Their main features are presented, together with the most common damages that characterise them. The steps involved in a diagnostic method are developed too, starting from the data acquisition. Then, since the signals obtained are not so easy to analyse as they are, a signal processing technique must be applied. The most widespread in the domain of time, frequency and time-frequency are showed. The final stage, is, then, the selection of proper features that own the most informative content about the health of the system are analysed.

---

Chapter 2 enhances the aspects related to the diagnostic techniques that allows the recognition of fault in rotating machines and bearings. Since a lot of research works have been published in the last 40 years on the detection and diagnosis, a description of the evolution in these methods is presented. A particular aspect that is treated is the rather recent challenging topic concerning the analysis of vibration in case of varying load and speed. In many machines, in fact, these two conditions could vary in uncontrolled way, so it is important to develop techniques enabling the detection of the damages also in these variable states.

When a fault originates in a machine, a variation in the normal behaviour of the system arises. In a pattern recognition context, this is known as anomaly detection, that refers to the identification of patterns in a given data set that does not conform to an established condition. Approaches adopted with this target are presented in Chapter 3. Some applications in machine diagnostics are showed too.

Chapter 4 deals with the procedure lied to the whole process of data acquisition on the bearing test rig assembled by our group, Dynamics & Identification Research Group (DIRG). It is possible the study the vibrations due to the bearings in a controlled situation, with the reduction of spurious signals coming from other rolling and vibrating elements. The main task is the acquisition of signals produced by various damages, referring to different load applied on the system, that could rotate at many speeds. Moreover, interesting conditions, such as oil and bearing temperatures, could be registered and monitored.

Some diagnostic methods are, then, applied on the data acquired on the test rig and Chapter 5 presents the results obtained. In the first part, Spectral Kurtosis is the base concept to develop a technique whose target is the proper identification of various kind of damages on bearings. It is verified how our procedure is able to find the damage, also when it is not so wide. Some considerations on the damage evolution and on the behaviour of this technique when the system is taken apart are presented too.

The second part of Chapter 5 concentrates on the removal of effects produced in vibrations by the variation of speed and load. Empirical Mode Decomposition is applied to the data to extract the features contained in the signal. Then, the One Class SVM classifier is trained to check the ability of the features in effectively identify the damage independently from the operating conditions. The results show how these features allow a proper damage recognition avoiding the errors that could be caused by the speed or load variations.

# Chapter 1

## Machine diagnostics

Rolling element bearings are certainly among the most widely mechanical elements used in machines and their failure is one of the most frequent reasons for machine breakdown. In this chapter the fundamental concepts related to this component are presented in Section 1.1 with the description of their typical damages too. In Section 1.2 the main steps composing the diagnostic process are explained and then developed in the following Section 1.3, 1.4 and 1.5. In details, various kind of data acquisition methods are presented in Section 1.3, while the three biggest domains of signal processing techniques form Section 1.4. Finally, some concepts related to feature extraction are showed (Section 1.5). The main aspects presented in this Chapter are part of [PG13].

### 1.1 Bearings

The term “bearing” derives originally from the verb “to bear” since it is a machine components whose goal is to bear some other components. In fact, the bearing is seen, mechanically, as a machine element that constrains relative motion between moving parts to a certain desired motion. On the whole, the most widespread types of bearings could be grouped in six great categories:

- I. plain bearings
- II. rolling element bearings
- III. jewel bearings
- IV. fluid bearings
- V. magnetic bearings
- VI. flexure bearings

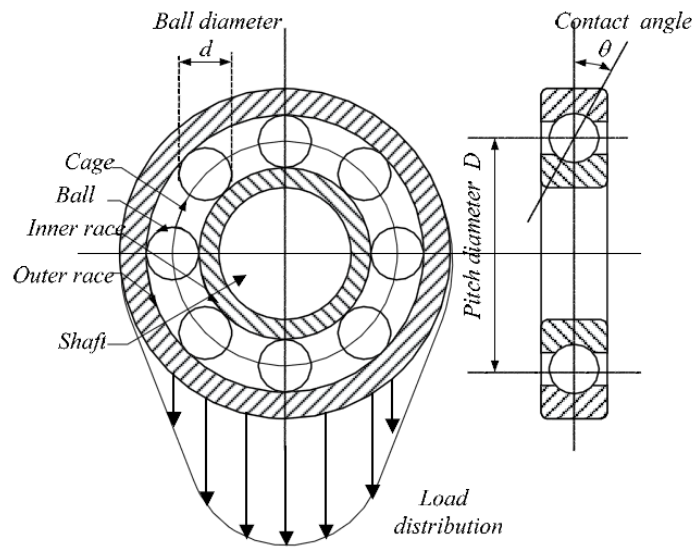


Figure 1.1. Rolling element bearing components and the load distribution (Figure taken from[LCD<sup>+</sup>12]).

The simplest type of bearing is certainly the *plain* one (Figure 1.2(a)) which is merely composed by a bearing surface without the presence of any rolling element. In this way, the journal could slide over the bearing surface, usually with the aid of a proper lubricant. They are greatly common in electric motors, in automotive applications, for home appliances and in industrial equipment. A possible drawback of this kind of bearings is the static and friction effect that originates in some applications. Moreover, their lifetime depends upon the application and can be higher or lower than that of rolling element bearings.

The main characteristic of *rolling element bearings* stands in the ability of transferring the main load through the elements that are in rolling contact rather than in sliding contact. Thus, the relative motion of pieces causes the round elements to roll with very little rolling resistance and with little sliding in order to prevent or minimise rubbing. An example of this type of bearing is Figure 1.2(b), while the main fundamental components are shown in Figure 1.1. Here it is possible to observe the outer and the inner race, the cage, the rolling elements, along with the load zone associated with the unidirectional vertical load. Some important parameters in rolling element bearings are the pitch diameter, the ball diameter and the bearing contact angle ( $D$ ,  $d$  and  $\theta$  respectively in Figure 1.1). They are usually used for higher moment loads and have the advantage of a good tradeoff between cost, size, weight, carrying capacity, durability, accuracy and friction. All this elements contributes to its diffusion which is almost as wide as those of plain bearings.

The rolling elements could be classified into five types: balls, cylindrical rollers, tapered rollers, spherical rollers, and needles (Figure 1.2). As said by the name itself, ball bearings





Figure 1.2. Various type of bearing: (a) plain, (b) rolling element (ball), (c) jewel, (d) fluid, (e) magnetic, (f) flexure.

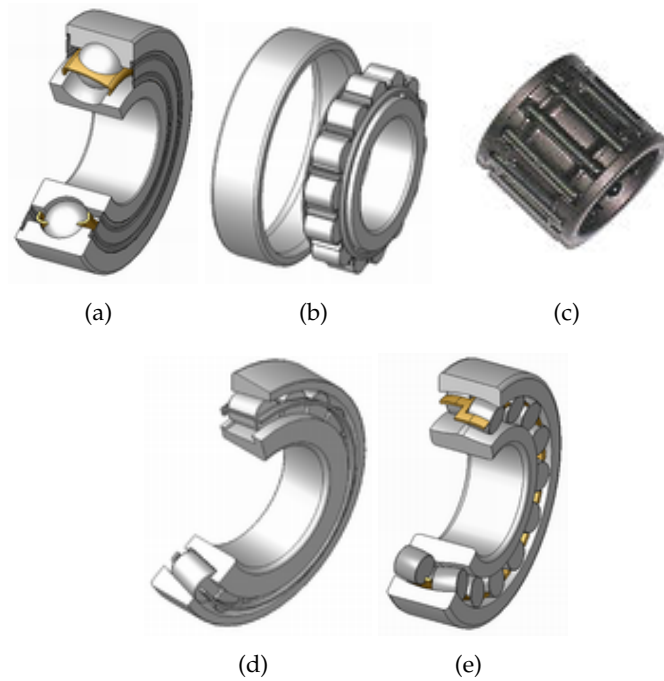


Figure 1.3. Various type of rolling elements in rolling elements bearing: (a) ball, (b) cylindrical, (c) needle, (d) tapered, (e) spherical.

(Figure 1.3(a)) use balls to carry the applied load and which could support both radial (perpendicular to the shaft) and axial loads. The roller bearings are those shown in Figure 1.3(b), whose components are usually cylinders with length slightly greater than diameter. The main difference with ball bearings is that roller bearings have typically an higher load capacity, but they are affected by a lower capacity and higher friction under loads perpendicular to the primary supported direction.

Figure 1.3(c) presents the needle roller bearings, that are a kind of rolling elements characterised by very long and thin cylinders. In spite of their low cross section, they have a high load carrying capacity and thus they are extremely suitable when radial space is limited. They are able to feature smaller cross-section, higher load-carrying capacity and are characterised by greater rigidity and lower inertia forces that facilitate size and weight reductions in machinery. Moreover, they could withstand oscillation and perform under severe conditions. On the whole, they could interchange with sliding bearings.

When conical rollers running on conical races are used we talk about the tapered roller that are shown in Figure 1.3(d). Their particularity stands in their ability of supporting both radial and axial load, while most of the other roller bearings could do it for only one load direction. Moreover, they are able to carry higher loads than ball bearings thanks to the greater contact area. These facts allows them to be used, for example, as the wheel bearings of most wheeled land vehicles. However, some drawbacks concerns their manufacturing complexities and their cost and some tendency to eject the roller in presence of heavy loads.

The last type of rolling element bearing is the spherical roller one showed in Figure 1.3(e). Their main characteristics are the outer ring provided with an internal spherical shape and the particular rollers shape. In fact, they are thicker in the middle and thinner at the ends and this fact helps the support of both static and dynamic misalignment. However, they are difficult to produce and, thus, they are quite expensive. Moreover, these bearings have higher friction than an ideal cylindrical or tapered roller bearing since there would be a certain amount of sliding between rolling elements and rings.

A particular type of plain bearing is the *jewel* one. Its peculiarity stands in the metal spindle that turns in a jewel-lined pivot hole. An example of ruby jewel bearings is presented in Figure 1.2(c), where the one used for a balance wheel in a mechanical watch movement is shown. In fact, this kind of bearing has been invented for watches, and this is their largest used joined with that in sensitive measuring instruments. In fact, they are known for their high accuracy, very small size and weight, low and predictable friction, good temperature stability, and the ability to operate without lubrication and in corrosive environments.

Another kind of bearing is the *fluid* type, characterised by the support of the bearing loads solely on a thin layer of liquid or gas (Figure 1.2(d)). Based on their different structure they could be broadly grouped into fluid dynamic and hydrostatic bearings. The first type rely on the high speed of the journal to pressurise the fluid in a wedge between the faces, while the latter are characterised by an external fluid (oil, water or air) pressurisation usually

executed through a pump. This kind of bearing is frequently used in high load, high speed or high precision applications where ordinary ball bearings would have short life or cause high noise and vibration also in order to reduce cost.

A *magnetic bearing* is, as the name itself suggests, a bearing which supports a load through magnetic levitation (Figure 1.2(e)). In fact, it could bear moving machinery without any bodily contact, it could levitate a rotating shaft and allow relative motion with really low friction and no mechanical wear. That means that they are characterised by zero friction at zero speed, but constant power for levitation. They are usually used in such industrial applications as electric power generation, petroleum refining, machine tool operation, and natural gas pipelines. Moreover, they could be seen in turbomolecular pumps, where oil-lubricated bearings would be a source of contamination. Another particularity is that they support the highest speeds of any kind of bearing and so no maximum relative speed is known.

The last type of bearing considered is the *flexure* one, which allows motion by bending a load element (Figure 1.2(f)). It relies on an element being made of a material which can be repeatedly flexed without disintegrating and this is the biggest advantage, combined to simplicity and low cost. They are also known for their compactness, their light weight and very low friction and their handiness in reparation. On the contrary, the disadvantage is a limited range of motion, especially for bearings that support high loads.

### 1.1.1 Bearing damages

Usually, the normal service life of rolling element bearing rotating under load is determined by material fatigue and wear at the running surfaces [How94]. However, premature failures could generate due to various factors, which could act autonomously, overlap or start to develop in one mode that then leads on to other failure modes.

The main factor causing damages is *fatigue*. In fact, after a certain running time, a bearing is subjected to material fatigue also if it is normally loaded, properly lubricated and correctly assembled. This fact is generally known as fatigue life of a bearing and is evaluated through the following formula [LP47]:

$$L_{10} = \left( \frac{C_D}{P_{eq}} \right)^p,$$

where  $C_D$ , named basic dynamic load capacity, is the load the bearing can carry for one million inner-race revolution with a 90% probability of survival.  $P_{eq}$ , instead, is the dynamic equivalent bearing load which depends on bearing geometry and on the radial and thrust load components. Finally,  $p$  represents the load-life exponent whose values are, according to International Organisation for Standardisation (ISO),  $p = 3$  for the ball bearing and  $p = 10/3$  for the roller one.

The evolution of the fatigue damage starts with the creation of minute cracks below the bearing surface until, because of the continuous loading, they reach it definitively. The damage takes the form of the so called pitting, spalling or flaking of the bearing races or

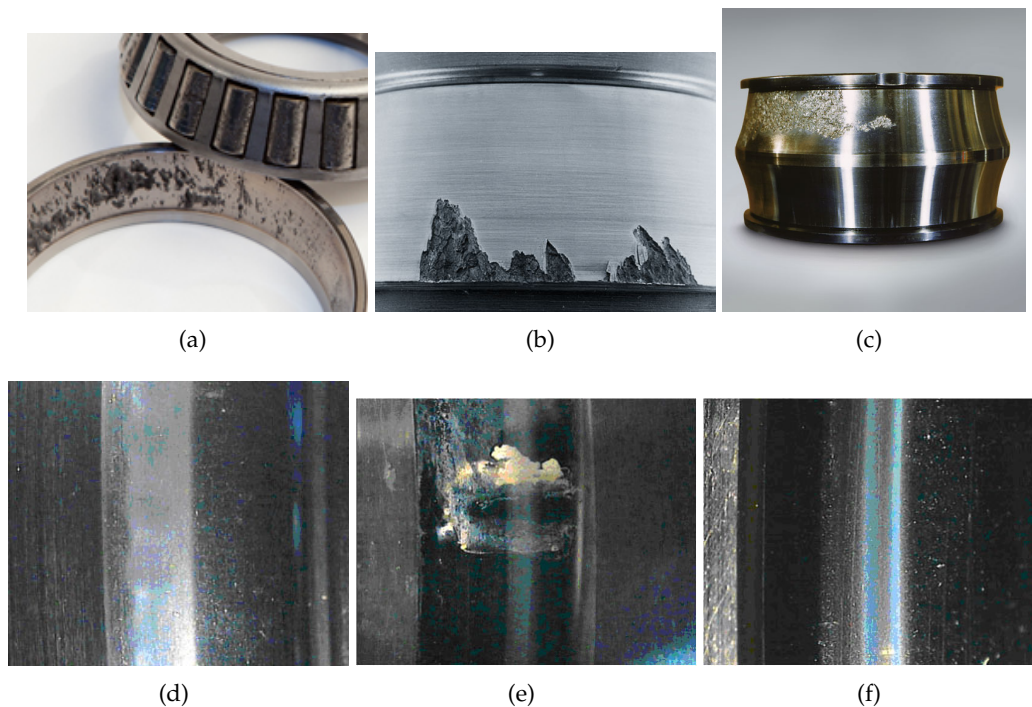


Figure 1.4. Bearing and races damages: (a) bearing and outer race pitting, (b) race spalling, (c) inner race flaking; damage produced by (d) brinelling, (e) corrosion and (f) misalignment on the race.

rolling elements (Figure 1.4). More the system continues working, more the failure increases disturbing the rolling elements motion which produces short time impacts repeated at certain defect frequency. However, this periodicity vanishes when the fault increases, due to the irregular motion of rolling elements and to the spread of the damage to the near components of the bearing.

Another common cause of failure is *wear*, usually due to dirt and foreign particles which enter into contact with bearing elements through contaminated lubricant and inadequate sealing. In this way, the contacting surfaces roughened generate changes in the raceway and rolling element profile and diameter, producing bigger bearing clearance. Also the rolling friction increases causing high level of slip and skidding until a complete breakdown. The action of excessive loading produces plastic deformation of bearing contacting surfaces and indentation of the raceway too. In this way, the deformed bearing rotates unevenly creating excessive vibration and poor service performs. If water, acids or other contaminants enter in contact with the bearing, corrosion or rust on the running surfaces generates. This rust particles have an abrasive effect, generate wear itself and could form the initiation sites for subsequent flaking and spalling (Figure 1.4).

Another possible fault is *brinelling*, which is characterised by regularly spaced indentations on the entire raceway circumference (Figure 1.4). It generates when static overloading

leads to plastic deformation of the raceways, when a stationary rolling bearing is subject to vibration and shock loads or when a bearing forms the loop for the passage of electric current. All this factors could produce repetitive indentations in the raceways, with noisy and uneven operations that could result in spalling and increasing distributed damage. Moreover, if the lubrication is inadequate, skidding, slip, increased friction, heat generation and sticking could generate. In particular, the contacting surfaces could weld together and be torn apart as the rolling element moves on. In addition, a lubricant starvation or an improper oil use could cause a temperature increase with hardness and fatigue life reduction. This factors cause excessive wear of the bearing elements and a final catastrophic failure.

Another important aspect is related to the bearing *components installation*. In fact, it has to be carefully designed in order to avoid effects as excessive preloading in either radial or axial directions, misalignment, loose fits or damage due to excessive force during the assembly (Figure 1.4). For example, the presence of radial preloading originates a noisy running together with temperature variations between the inner and the outer races. In this way, the undesirable preloads and contact pressures are going to increase. The direct consequence is the creation of premature fatigue, heavy rolling element wear and overheating. Similar effects and problems are reached when the excessive preload develops in axial direction. Correlated to the bearing components installation is also the choice of proper bearing type or size for the required operation.

As clearly explained in [TC99], a general distinction of bearing defects may be done between “distributed” and “local”. The first kind includes surface roughness, waviness, misaligned races and off-size rolling elements. They are usually due to manufacturing error, improper installation or abrasive wear. When distributed defects produces a variation in contact force between the rolling elements and raceways, an increased vibration level is produced. Thus, vibration response when this kind of damage is present is important for quality inspection as well as condition monitoring. Localised defects include, instead, cracks, pits and spalls on the rolling surfaces. As stated by Tandon and Choudhury, spalling of the races or of the rolling elements is the most common damage causing the great number of failure of rolling element bearings [TC99]. This fault arises when a fatigue crack begins below the surface of the metal and propagates towards the surface until a piece of metal breaks away to leave a small pit or spall. Moreover, electric pitting or cracks caused by excessive shock loading are also among the different types of bearing damage.

An exhaustive description of all possible bearing failures, with respective causes and countermeasures, is presented in Appendix A.

## 1.2 Bearing diagnostics steps

In the last forty years there have been many developments in the use of vibration measurement and analysis for monitoring the condition of rotating machinery. In particular, there is

diagnostics {	anomaly detection fault identification fault monitoring RUL estimation	} prognostics
---------------	---	---------------

Table 1.1. Diagnostics and prognostics tasks.

a strong need for accurate diagnostics and prognostics in order to have efficient, safe, reliable and profitable operation of machines. Since machines operate in complicated, uncertain and varying environments, a good diagnostic methodology is expected to work accurately also in these tough conditions. It is worth of interest, thus, an intelligent use of techniques coming from various domains such as machinery dynamics, signal processing, statistical analysis and machine intelligence.

Table 1.1 shows the general tasks required to a diagnostic and prognostic methodology, which could be also listed as [NK11]:

- detect any change in performance;
- determine if the change is a degradation in the system;
- determine the cause of degradation;
- identify the amount of degradation;
- estimate the remaining useful life.

As could be imagined, these tasks need to be performed in sequence. In fact, at the beginning, there is no reason to expect failure as so it is sufficient to determine if the performance is nominal. Usually, this involves the use of anomaly detection algorithms and the deviation from a nominal performance is measured. Some variations from nominal system behaviour could be due to temporary changes in the operating conditions or due to wear, which is expected over time but not regarded as degradation. Thus, the anomaly detection algorithm needs to take into consideration these aspects before classifying the anomalies as faults. When, afterwards, the anomaly is assessed as a real damage, it is necessary to operate with the fault identification algorithm, whose goal is the correct identification of the faulty components and the failure modes. The prognostics starts, instead, when a sufficient confidence of knowledge related to a certain type of failure is reached. This domain of analysis is based on two main steps: the first relies on the identification and the tracking of the amount of fault, while the second has to understand the damage progression and to make an estimate of the Remaining Useful Life (RUL) of the machine according to its future usage. As could be imagined, the main difference between diagnostics and prognostics stands in the schedule, because the former is posterior event analysis while latter is prior one. Moreover, major economic benefits come from being able to predict with reasonable certainty how much

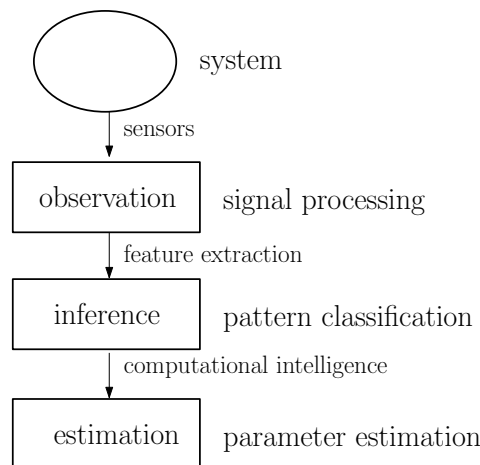


Figure 1.5. Diagnostics methodology.

longer a machine can safely operate.

The basic steps characterising a diagnostic method could be seen in the diagram in Figure 1.5. The first task is the recording of the signals produced by the machine through proper sensors. Some kind of sensors that could be used during the acquisition of signals is presented in Section 1.3. However, the signals can not be directly used as they are acquired due to their very large size and because of the noise that affects them. So, signal processing techniques are necessary to obtain useful and compact information about the system from the measurements (Section 1.4). In particular, these “information packets” are usually known as features and could be used to make inferences about the system state. More details about this process, called feature extraction, are given in Section 1.5. As it could be imagined, machines affected by damages behave in a different way compared to those without defects and so features obtained in the first case may follow another pattern compared to the others. In order to understand this pattern and to infer the state of the system, computational intelligence algorithms could be used, as presented in Chapter 3. The final step in the diagnostic analysis concerns the analysis of the variations in the patterns and features, that allows the estimation of the future state of the machine through the so called parameter estimation algorithms.

A further important and well developed topic in machinery dynamics is the modelling of rotating machines. These systems are characterised by many sources of excitations and nonlinearity and by a residual unbalance which is an external time dependent force on the system. One of the primary sources of nonlinearity in rotating systems is the restoring force from the bearing. In particular, rolling element bearings are highly nonlinear and have three important causes of nonlinearity: the nonlinear restoring forces between various curved surfaces in contact, the radial clearance between the races and the defects themselves. The first is usually modelled considering the Hertzian contact theory [NK11]. Clearance, instead, is usually modelled as a dead band that is subtracted from the effective displacement

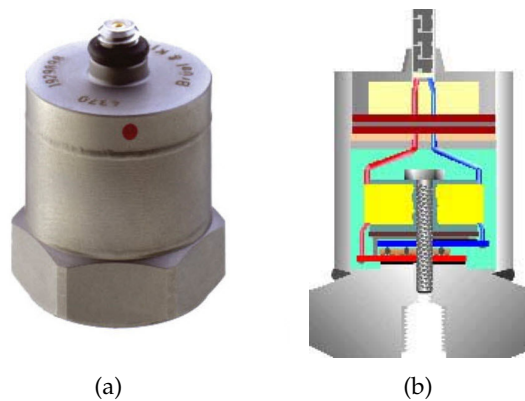


Figure 1.6. A piezoelectric accelerometer (a) and its internal components (b).

of the shaft. An example that starts from these basis is presented in [SR08], in order to analyse rotating shafts on rolling element bearings and reported interesting behaviour. Many approaches could be assumed to study nonlinear phenomena due to clearance, such as bifurcation and chaos, sub-harmonics, intermittency and quasi-periodic routes to chaos, numerical techniques and nonlinear phenomena like quasiperiodicity [NK11]. Referring, instead, to the localisation of defects in rolling element bearings, they are usually seen and included as pits in the race or as impulse forces. Some examples are presented in [NP05], where the effect of defects on the rotor bearing system is modelled as a change in the deformation of the rolling element bearing when it enters a pit. In [SR08], instead, a two degree of freedom model of shaft is considered and the defects on a rolling element bearing are modelled as smooth pits whose shape is an assumed curve.

### 1.3 Data acquisition

A very important task in the health monitoring of a system is surely the data acquisition, also known as information collecting. On the whole, there are many techniques related to machine monitoring, such as an usually difficult visual check, temperature measure, ferromagnetic elements in the lubricant inspection and vibrations measure. For example, a simple infrared thermography could identify a damage since it is usually correlated to an unexpected variation in the temperature. Unfortunately, this changing is only reliable when the damage is already in advanced state. Usually, the analysis of materials circulating in the lubricant, and specifically their quality and dimensions, allows the classification and the localisation of the damages. However, one of the most reliable monitoring techniques is the vibration measure, thanks to its sensibility to damage and the bypassing of machinery stop for inspection. In fact, the vibrations measured on operating machines could produce information about system condition since the normal condition characteristic “vibration signature” changes in a well-defined way. Thus, vibration analysis is a powerful and reliable



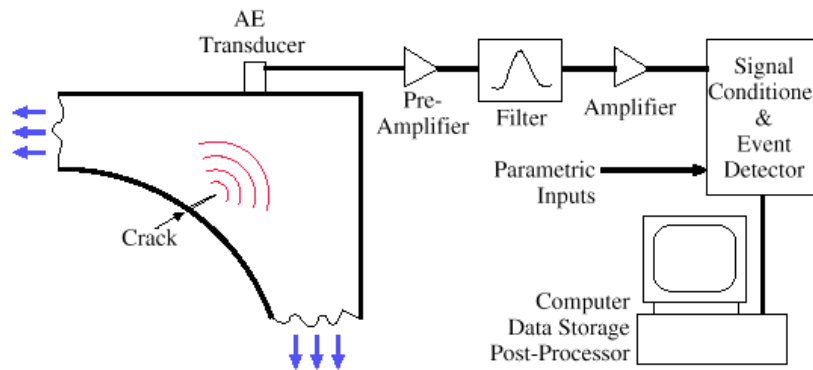


Figure 1.7. A typical AE system setup. [HJL<sup>+</sup>98]).

way of getting information from the inside of operating machines without shutting them down.

Since all vibrations represent an alternation between potential energy (in the form of strain energy) and kinetic energy, their velocities could be seen as the parameters most closely related to stress and, thus, could be used to evaluate severity in most vibration criteria [Ran04a]. For this reason, the velocity spectrum is usually “flattest” over a wide frequency range and it requires the minimum dynamic range to represent all important components. However, vibration displacement tends to overemphasise low frequencies contents while vibration acceleration tends to overemphasise the highest ones. So, this latter could be useful for faults, such as in rolling element bearings that usually originates at high frequencies, but may disguise changes at low frequencies.

Usually, the best and most common transducer used to measure absolute casing vibration is a piezoelectric accelerometer originating a signal proportional to accelerations (Figure 1.6). Inside this accelerometer the sensing element is a crystal that emits a charge when subjected to a compressive force. It is bonded to a mass such that, when the accelerometer is subjected to an acceleration, this mass compresses the crystal emitting a signal. The sensing element is housed in a suitable sensor body, usually made in stainless, to withstand the environmental conditions. The electrical connection that transmits the signal could be a sealed cable or a plug/socket arrangement. A key point is then the mechanical fixing of the sensor: it is, in fact, very important to assure true transfer of the vibration or acceleration. This target could be reached through various fixing methods such as beeswax, hard glues, threaded stud (male or female) and magnetic mounts. Since usually the dynamic range of the accelerometer is quite large (160 dB), it could be combined with electronic integration to give a velocity signal with more than 60 dB dynamic range over three decades in frequency. Unfortunately, typical velocity transducers are limited by an upper frequency limit of 1000-2000 Hz, so this results could not be reached. Thanks to these facts, the measurement of signals with a frequency range of more than three decades, such as 5-5000 Hz or 20-20000 Hz, with very good dynamic

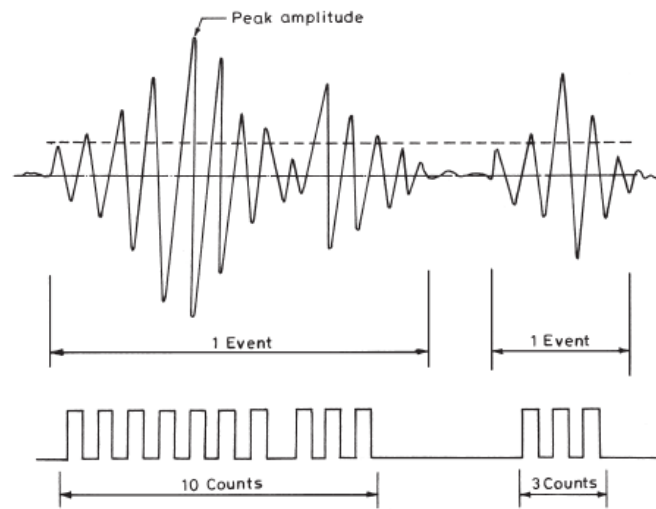


Figure 1.8. A typical acoustic emission burst signal. (Figure taken from [TC99]).

range could be obtained through accelerometers. Such a range would be necessary to detect the full range of possible faults.

Between the several techniques that have been used to measure the vibration and acoustic responses from defective bearings since now, an interesting and useful approach is those represented by Acoustic Emission (AE) [TC99]. This is based on the creation of transient elastic wave due to a rapid release of strain energy caused by a structural alteration in a solid material under mechanical or thermal stresses. This allows a high frequency analysis technique which was initially developed as a non-destructive testing (NDT) tool to detect crack growth in materials and structures. This main target has been developed since the 1980s, but in most recent years, this method has been increasingly used as condition monitoring tool of engineering assets such as structures and industrial machines. In addition, it is also exploited in machine tool monitoring, tribological and wear process monitoring, gear defects monitoring and bearing fault monitoring. Widodo *et al.* present an interesting review of AE applications to bearing diagnostics in [WKS<sup>+</sup>09], in particular related to the monitoring of various speed machines. The main advantage of AE is twofold: firstly, it is a dynamic method capable of continuously testing and *in situ* monitoring structures under supervision without the need of their stoppage or shutdown, usually through a few location monitoring. Moreover, it is a passive means without the need of injecting signals from outside to the object under investigation, thus causing little disturbance to the object. On the other hand, the background noise interferences, the difficulty in realising the quantitative evaluation of the defect because of the lack of understanding of the AE source mechanism and the distortion of sound wave during propagation are some of the main drawbacks. Usually, generation and propagation of cracks, growth of twins associated with plastic deformation are among the primary sources of AE.

As outlined previously, the AE technique is based on the detection and conversion of high frequency elastic waves emanating from the source to electrical signals. This is accomplished by directly coupling a transducer, mostly of the piezoelectric type, on the surface of the structure under test and loading the structure. These transducers are characterised by a very high natural frequency and have a resonant-type response. The output of these sensors (during stimulus) is amplified through a low-noise preamplifier, then filtered to remove any extraneous noise and further processed by suitable electronics. In this way, the bandwidth of the AE signal can be controlled by using a suitable filter in the preamplifier (Figure 1.7). Acoustic emissions could broadly be classified into two types: the continuous and the burst one. The waveform of continuous type AE signal is similar to Gaussian random noise, while the amplitude varies with acoustic emission activity. The burst type of emissions is formed by short duration pulses and is usually associated with discrete release of high amplitude strain energy. Some of the most commonly used AE parameters are AE burst, threshold, ring down count, cumulative counts, event duration, peak amplitude, rise time, energy and RMS voltage. For example, Tandon and Choudhury presents the ringdown counts, events and peak amplitude of the signal on a typical AE signal in Figure 1.8 [TC99]. In particular, the ringdown counts regards the counting of the number of times the amplitude exceeds a preset voltage level, chosen as threshold level, in a given time and gives a simple number characteristic of the signal. An event, instead, consists of a group of ringdown counts and signifies a transient wave. AE approach presents various advantages compared to vibration monitoring. Firstly, it is able to detect the growth of subsurface cracks, whereas the latter can detect defects only when they appear on the surface. Moreover, it is insensitive to structural resonances and unaffected by typical mechanical background noise while it is more sensitive to activity from faults. Furthermore, it is also interesting to note that the energy released by neighbouring components in the vibrational frequency range (up to 50000 Hz), which often masks the vibrational energy released from a defective rolling element bearing, do not affect the AE signal released in the very high frequency range [TC99].

## 1.4 Signal processing

In the previous sections some techniques and types of sensors allowing the signals acquisition from machines have been presented. However, these signals can not be directly used since they are usually very large in size and noisy. So, it is necessary to analyse them through some proper processing techniques in order to derive useful and compact information about the system under analysis. On the whole, these techniques could be grouped into three big domains: time, frequency and time-frequency.

### 1.4.1 Time domain techniques

The time domain techniques are among the first feature extraction techniques developed for bearing fault detection since they are directly based on the time waveform itself. The first considerations started with the observations that rolling element bearings with faults show higher peak to peak vibration compared to a healthy bearing ([Tan94]). The main advantages of time domain analysis is the simple calculations, straightforward signal pre-processing, and speed independence. However, it is instead usually insensitive to early stage faults and to deeply distributed defects.

Certainly, the simplest approach in the time domain is the measure of elementary parameter such as the overall Root Mean Square (RMS) level and the crest factor:

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{k=1}^N x_k^2} \quad \text{crest factor} = \frac{\max(\text{abs}(x))}{\text{RMS}}, \quad (1.1)$$

where  $x$  is the signal acquired composed by  $N$  elements.

RMS is a measure of the overall vibration level and, unfortunately, often increases only at the final stages of failure and it is not so sensitive to small or early-stage defects. A further parameter related to RMS is the so called Delta RMS, which represents the difference between a current RMS value and the previous. The crest factor could give instead an earlier warning of bearing failure since the vibration pulses produced by the defect are taken into account. Usually, the level of the crest factor for a normal bearing is approximately 5 and so a different value could be a possible damage indicator. In general, in spite of RMS, this parameter could be a good indicator of small size defects but when localised damage propagates, it decreases significantly due to the increasing RMS.

A statistical parameter such as probability density has been proposed for bearing defect detection, but instead of studying the probability density curves, it is often more informative to examine the statistical moments of the data, defined as

$$M_x = \int_{-\infty}^{+\infty} x^n P(x) dx \quad n = 1, 2, \dots, m, \quad (1.2)$$

with  $P(x)$  is the probability density function with instantaneous amplitude  $x$ .

The first and second moments are well known as mean value and variance respectively, and not so much impressive. The third moment and the fourth moments normalised, also called skewness and kurtosis, are instead quite useful in bearing defect detection, especially if related to healthy values. These are their expression for discrete signals:

$$\text{skewness} = \frac{\frac{1}{N} \sum_{k=1}^N (x_k - \bar{x})^3}{\left( \frac{1}{N} \sum_{k=1}^N (x_k - \bar{x})^2 \right)^{3/2}} \quad \text{kurtosis} = \frac{\frac{1}{N} \sum_{k=1}^N (x_k - \bar{x})^4}{\left( \frac{1}{N} \sum_{k=1}^N (x_k - \bar{x})^2 \right)^2}, \quad (1.3)$$

parameter	healthy	damaged
RMS	35.75	65.84
crest factor	4.19	8.08
skewness	0.04	-0.15
kurtosis	3.04	7.60

Table 1.2. Parameters evaluation on two signals, where the damage is a 450  $\mu\text{m}$  indentation on a rolling element.

where  $\bar{x}$  is the mean value of the signal  $x$ .

Related to kurtosis, its value for an undamaged bearing with Gaussian distribution is close to 3 and so a greater value could be considered as an indication of impending failure. Several studies have also shown its effectiveness in bearing defect detection but, unfortunately, in some cases this method could not detect the incipient damage effectively.

Besides the cited parameters, there are other characteristic features that could be calculated from time waveform signals as descriptive statistics such as mean, peak, peak-to-peak interval, standard deviation and they are usually known as time domain features. An example of the use of parameters described in Equation (1.1) and (1.3) could be seen in Table 1.2. The RMS value changes a lot between the two states and is bigger in the faulty case as could be expected. The crest factor and the kurtosis confirm what said previously: in fact the former is approximately equal to 5, while the latter is around 3. The skewness instead does not seem to be a relevant indicator since it assumes almost equal values in both cases.

A popular time-domain analysis approach, rather common for gears failure early detection, is Time Synchronous Average (TSA) [JLB06]. This TSA signal is obtained through the synchronisation of the sampling of the vibration signal with the rotation of a particular gear and the following evaluation of the ensemble average over many revolutions with the start of each frame at the same angular position [DRR00]. Another approach related to the time domain analysis is the application of time series models to waveform data [JLB06]. The main idea is to fit the waveform data to a parametric time series model and to extract features based on this parametric model. Two examples are the so called autoregressive (AR) model and the autoregressive moving average (ARMA) model. A general ARMA model of order  $p$  and  $q$  is given by

$$x_t = a_1x_{t-1} + \dots + a_px_{t-p} + \varepsilon_t - b_1\varepsilon_{t-1} - \dots - b_q\varepsilon_{t-q},$$

where  $x$  represents the waveform signal,  $a_i$  and  $b_i$  are the model coefficients and  $\varepsilon$  are normally distributed.

An AR model of order  $p$  is, instead, a special case of the ARMA model where  $q = 0$ . The key steps of this procedure are the estimation of the model order, that means the selection of the number of past inputs required to model the vibration signal accurately, the estimation of the model parameters and the model validation. This type of models could be applied

to vibration signals, for example, when the goal is the extraction of features from model coefficients. An application is presented in [BM96], where three distinct techniques of AR model are considered and compared experimentally with diagnostic intents. A number of induced faults in a rolling element bearing are, in fact, analysed and identified using the measured time series vibration signal. Another interesting application of the AR, this time operated through the application of an inverse filter, is presented in [GWMF09]. This method consists of two steps: firstly the inverse filter based on the signal of a healthy system is constructed and then the residual signal is analysed. The use of simulated and real data allows the conclusion that the AR inverse filter could enhance the feature of impulse signal.

The advantages of the AR model stands in its simplicity, its efficiency and the low prior requirements for its application. The system of equations for the parameter estimation is linear and the solution is straightforward. Moreover, it is possible to test the quality of the model and the parameter  $p$ , representing the model order and being the only parameter to adjust, could be optimised for a best fit too. Various criteria have been developed to select the optimum model order, such as the Akaike information criterion (AIC) [SRE07], the Marple method, the covariance method and the modified covariance methods [GWMF09]. Another approach, described by [SRE07], takes into account the fact that the main objective of using the AR technique for rolling element bearings resides in separating the impulses from the original signal. So, they base the order selection criterion on the maximisation of the kurtosis of the residual signal, which contains both impulses and stationary noise. According to this opinion, the optimum model order  $p$  should be less than the spacing between two consecutive impulses so that this guarantees the model not to adapt to the impacts as being a part of the deterministic signal and that they will be contained in the residual signal.

### 1.4.2 Frequency domain techniques

Frequency domain methods are certainly among the most used feature extraction techniques for bearing fault detection. This high diffusion is mainly due to its ability in identifying and isolating certain frequency components of interest, which is a great advantage over time domain analysis. In this way, these approaches are sensitive and robust in bearing defects detection and localisation. However, their accuracy highly depends on the bearing dimensions and rotational speed. In addition, these methods require an intelligent selection of the frequency band in order to be effective.

The main characteristic of the frequency domain is related to the possibility of estimating the typical defect frequencies of the bearing vibration. When a defect is present, in fact, its interaction in rolling element bearings produces pulses of very short duration whenever the fault strikes or is struck owing to the rotational motion of the system. These pulses excite the natural frequencies of bearing elements and housing structures and these facts produces an increase in the vibrational energy at the high frequencies. These particular

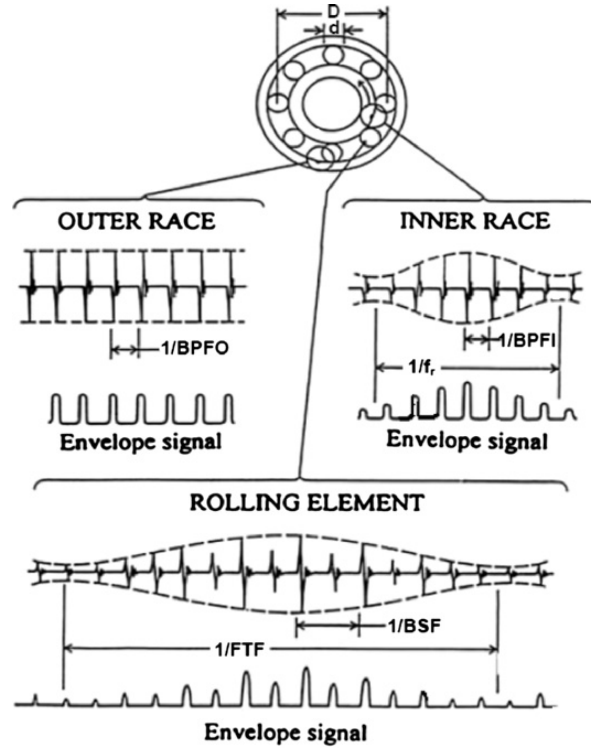


Figure 1.9. Typical signals and envelope signals of local faults in a rolling element bearing (Figure taken from [RA11]).

values characterising the defects could be calculated considering the geometry of the bearing and the rotating speed according to following equations [TC99]:

$$\text{BPFO} = \frac{nf_r}{2} \left( 1 - \frac{d}{D} \cos \theta \right), \quad (1.4)$$

$$\text{BPFI} = \frac{nf_r}{2} \left( 1 + \frac{d}{D} \cos \theta \right), \quad (1.5)$$

$$\text{FTF} = \frac{f_r}{2} \left( 1 - \frac{d}{D} \cos \theta \right), \quad (1.6)$$

$$\text{BSF(RSF)} = \frac{D}{2d} \left[ 1 - \left( \frac{d}{D} \cos \theta \right)^2 \right], \quad (1.7)$$

where  $n$  is the number of rolling elements,  $f_r$  the shaft speed,  $d$  the diameter of the rolling element,  $D$  the pitch diameter and  $\theta$  the contact angle (Figure 1.1).

Equation (1.4) refers to the ball pass frequency on the outer race (BPFO), Equation (1.5) to that on the inner race (BPFI), while Equation (1.6) is the fundamental train frequency related to the cage speed (FTF) and Equation (1.7) is the ball (or roller) spin frequency (BSF). These frequencies could be, theoretically, noticed observing the envelope of the vibration signal acquired for a damaged bearing. In addition to these evaluations, there are other frequency

components present in a typical bearing signal such as the 1X response, its harmonics and sub-harmonics which are indicators of nonlinear behaviour in general. An example is in Figure 1.9, where typical acceleration signals produced by localised faults in the various components of a rolling element bearing, and the corresponding envelope signals produced by amplitude demodulation, are presented. As explained in [RA11], these frequencies are determined assuming no slip. However, to approach real situations, some slip is always considered because the angle  $\phi$  varies with the position of each rolling element in the bearing, as the ratio of local radial to axial load changes. So, each rolling element tries to roll at a different speed while the cage limits its deviation from its mean position. In this way some random slip is produced and a variation of about 1-2% in the bearing frequencies is observed, due both to the deviation from the calculated value and also to the random variation around the mean frequency. Some particular attentions must be taken when the goal is the detection of the inner race defects [NK11]. In fact, in this case, the excitation signals need to travel through the both rolling element and the casing before the sensor detection. The problem then the presence of other excitation that could mask the signal and, so, the frequency. In order to overcome this, an envelope spectrum technique is often used.

One of the eldest, and probably the most common, method to extract the frequency components in a signal is Fast Fourier Transform (FFT). The spectrum obtained, in fact, usually contains a peak at the defect frequency also if it is not always clearly observable because of slip and masking by other stronger vibrations. FFT was firstly presented in 1965 in [CT65] and the main innovation stands in the lower computation time requested for the evaluation of a Fourier transform of size  $N$ . In fact, this new procedure decreases the computational order from  $N^2$  to  $N \log_2 N$  [Ran10]. As stated by Randall, FFT was firstly implemented on mini-computers, then as firmware in frequency analysers, and by the early 1980s has been available in portable data collectors, which could perform spectrum comparisons in the field.

The FFT is based, as said before, on an efficient method to evaluate the Discrete Fourier Transform (DFT) of signal  $x$ , which is a succession  $X$  of the form:

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-\frac{j2\pi kn}{N}}. \quad (1.8)$$

The previous equation represents the forward transform, while the inverse transform is:

$$x(n) = \sum_{k=0}^{N-1} X(k) e^{\frac{j2\pi kn}{N}}. \quad (1.9)$$

To evaluate the forward DFT transform in Equation (1.8) it could be introduced the term  $W = e^{-j2\pi/N}$ , so that it becomes:

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) W^{kn}.$$



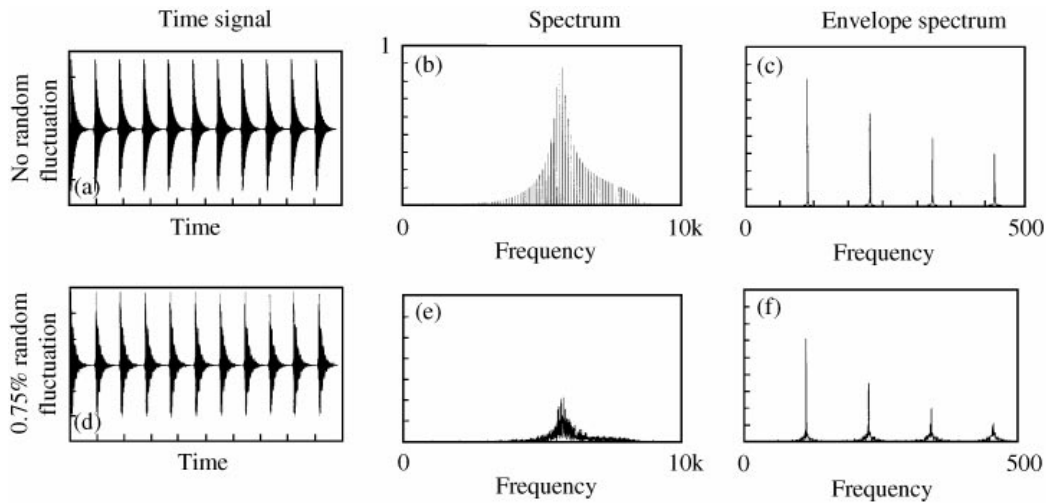


Figure 1.10. Bearing fault pulses with and without random fluctuations: (a) a series of bearing fault impulses at equal time spacing, (b) its spectrum and (c) the envelope spectrum of (b) demodulated around the resonance frequency; (d) series of bearing fault impulses with 0.75% random fluctuation, (e) its spectrum and (f) the envelope spectrum of (ef) demodulated around the resonance frequency (Figure taken from [HR00]).

Thus, this powerful algorithm FFT is based on the term  $W$  properties and on the binary decomposition of index  $k$  and  $n$ .

FFT it is usually used in association with some spectrum analysis tools and one the most commonly widespread is power spectrum [JLB06]. It is defined as the distribution of the energy of a waveform among its different frequency components:

$$E[X(f)X^*(f)],$$

with  $X(f)$  being the Fourier transform of the original signal  $x(t)$ ,  $E[\ ]$  the expected value and “\*” the complex conjugate.

Since results are averaged over the entire duration of the signal, the FFT is not able to provide any information about the time dependence of the spectrum of the signal analysed. This aspect is a consistent weak point in non-stationary signals analysis, since these are the type of signal produced by bearings. In such cases, it becomes necessary the acquisition of a correlation between the time and frequency contents of the signal. Moreover, the spectrum, which usually contains a peak at the defect frequency, is not always clearly observable because of slip and masking by other stronger vibrations. A way to overcome this problem is the application of an envelope spectrum technique.

The envelope of a signal  $x(t)$  is obtained through the computation of the complex signal amplitude, composed by the real part  $x(t)$  and its Hilbert transform as imaginary part. The Hilbert transform of a time domain signal can be obtained by shifting the phase of the original spectrum and then inverse Fourier transforming back to the time domain. Alternatively, it could be generated from the one-sided spectrum of a signal in order to avoid the manipulation

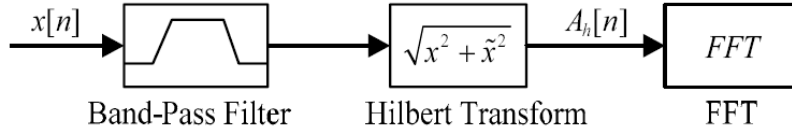


Figure 1.11. The process of High Frequency Resonance Technique (Figure taken from [CHY12]).

of the phase of the frequency components. The cited Hilbert transform of a time domain signal is expressed by

$$\text{Hilbert}[x(t)] = \tilde{x}(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau$$

and the envelope of the considered signal is

$$\text{Envelope}[x(t)] = \sqrt{x^2(t) + \tilde{x}^2(t)}.$$

A possible, and widely used, method to extract bearing defect frequency components from the signal is envelope analysis [HR00]. An example of this application is in Figure 1.10, where (a) is a simulated series of impulse responses without random fluctuation and (b) shows its spectrum. Here, the higher harmonics of the defect frequency are quite clear also if the low harmonics are very small. If it is compared with the spectrum for the series of impulse responses with slight random fluctuation in (e), it could be noticed that here the defect components are smeared into each other. The comparison between the envelope spectrum for the two cases represented in (c) and (f) shows, instead, how defect frequency components are not sensitive to random fluctuations and they are still clearly visible. As explained in [HR00], the simplest method to obtain envelope analysis is the removal of the low-frequency noise through the passage of the signal through an analogue high-pass filter. The next step is the rectification and the frequency analysis of the signal in order to determine the defect frequency components in the envelope spectrum. This analysis could be more efficient if the signal is digitalised and band-pass filtered around a frequency where the high signal-to-noise ratio is maximum, as, for example, around the resonance. Then, the Hilbert transform is applied to obtain the envelope spectrum with the advantages of a smaller size of the record that is being processed and a flexibility in cut-off frequencies for the band-pass filter choice. An appealing analysis is presented by Ho and Randall in [HR00], where they compare the squared envelope with the envelope. They found the former having the interesting advantage of not including the higher harmonics of the sum frequency components in the envelope spectrum.

Another way to call the previous procedure is High Frequency Resonance Technique (HFRT). As showed in Figure 1.11, this procedure involves three different steps. First, the measured signal is bandpass around a selected resonance frequency in order to separate this high frequency natural vibration from original signal. Then, the bandpass-filtered signal is

demodulated by an envelope demodulation detector, such as the already cited Hilbert transform, to eliminate its resonance frequency, so that the enveloped signal contains only defect information. At the end, a low pass filter is used to cancel high frequency components and retain the low frequency information associated with bearing defects. In this way, vibration signals generated by defect are demodulated and the characteristic defect frequencies are present in spectrum analysis of the enveloped signal.

Another useful method for signal processing is the cepstrum, which is able to detect harmonics and sideband patterns in power spectrum. It was first presented in publication in 1963 [BHT63] and it was defined as the power spectrum of the logarithmic power spectrum. Later, it has been redefined as the inverse Fourier transform of the logarithmic spectrum:

$$C(\tau) = \mathfrak{F}^{-1}[\log(X(f))],$$

where

$$X(f) = \mathfrak{F}[x(t)] = A(f) \exp(j\phi(f))$$

so that

$$\log(X(f)) = \ln(A(f)) + j\phi(f).$$

The abscissa  $\tau$  of the cepstrum has the dimensions of time but is named *quefrequency*. Between the several version of cepstrum such as complex cepstrum, real cepstrum, phase cepstrum and power cepstrum, this last one is the most commonly used. It is defined as the squared magnitude of the Fourier transform of the logarithm of the squared magnitude of the Fourier transform of a signal:

$$\text{power cepstrum}[x(t)] = |\mathfrak{F}[\log(|\mathfrak{F}[x(t)]|^2)]|^2.$$

### 1.4.3 Time-frequency domain techniques

Unfortunately, frequency domain analysis is not able to handle non-stationary waveform signals, which are very common when machinery faults occur. The solution to this problem is time–frequency analysis, because it could investigate waveform signals in both time and frequency domain. The basic problem that arises with this approach is the so called “uncertainty principle”, which states that the product  $\Delta t \Delta f$  is constant, where  $\Delta t$  is time resolution and  $\Delta f$  is frequency resolution. This means that a better resolution in one term is achieved at the expense of the other one.

The simplest, and also the first, approach in this domain is certainly the “Short Time Fourier Transform” (STFT), which has been applied by Gabor in 1946 to speech communication [Gab46]. It could be seen as a method that breaks down the non-stationary signal into many small segments, which can be assumed to be locally stationary, and applies the conventional FFT to each one of them. So, it could be said the it gives a measure of the

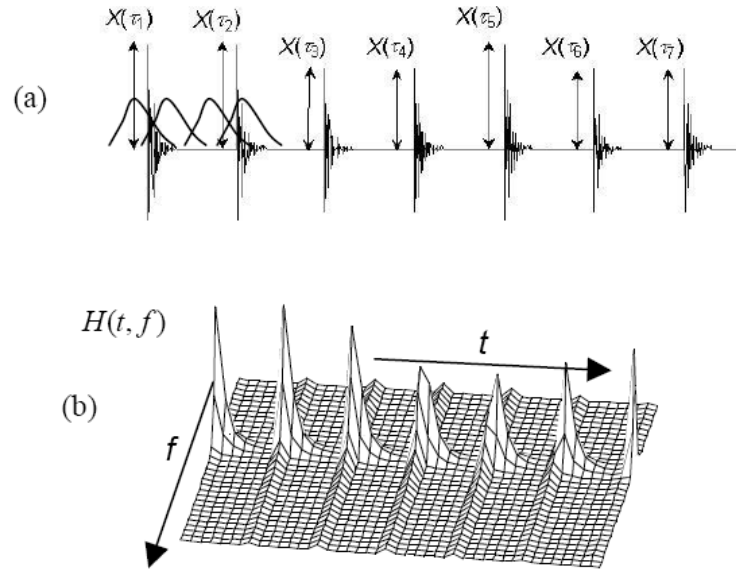


Figure 1.12. A time signal with (a) the moving time windows and (b) amplitude of STFT (Figure taken from [SR04]).

sinusoidal frequency and phase content of a signal as a function of time. Given a time signal  $x(t)$ , STFT is obtained by multiplying the signal itself by a window function  $\omega(t)$  centred in  $t$ :

$$\text{STFT}[x(t)] = X(\tau - \omega) = \int_{-\infty}^{+\infty} x(t)\omega(t - \tau)e^{-j\omega t} dt,$$

where  $\omega(t)$  is commonly a Hanning window.

A series of spectra could be produced by moving a window (such as Hanning) along a time record and drawing a waterfall plot of spectra corresponding to the centre positions of the window (Figure 1.12). So, the frequency resolution bandwidth is given by the reciprocal of the effective window length. The major disadvantage of the STFT is the resolution tradeoff between time and frequency. This resolution is determined by the width of window  $\omega(t)$  and, so, a large window width provides good resolution in the frequency domain, but poor resolution in the time one. On the contrary, if a small window width is considered, good resolution in the time domain and poor resolution in the frequency domain is achieved. This particular limitation is due to the use of a single window for all frequencies that cause the same resolution of analysis at all locations in the time-frequency plane.

A parameter describing how the energy of a signal or a time series is distributed with frequency is the energy density spectrum. At time  $t$ , it could be expressed as:

$$|\text{STFT}[x(t)]|^2 = \left| \int_{-\infty}^{+\infty} x(t)\omega(t - \tau)e^{-j\omega t} dt \right|^2.$$

For each time, different spectra are obtained and the ensemble of these spectra provides the time-frequency distribution which is called spectrogram. The spectrogram has some

limitations in time–frequency resolution due to signal segmentation since it can only be applied to non-stationary signals with slow change in their dynamics.

A better resolution is given by the Wigner-Ville distribution (WVD) which is able to overcome this spectrogram limitation since it is not based on signal segmentation. The Wigner distribution was first defined by Wigner in 1932 [Wig32], but it was greatly conditioned by interference components. In the following years, Ville recognised that at least the interference between positive and negative frequencies could be removed by analysing the analytic signal [Vil48]. More later versions have been devised with the task of reducing interference until Cohen has generalised them in the so-called “Cohen’s class” of time-frequency distributions [Coh95]. Finally, WVD could be defined as the Fourier transform with a time lag  $\tau$  of the non-averaged autocorrelation function  $R(t; \tau)$ :

$$\text{WVD}(t, f) = \mathfrak{F}[R(t; \tau)] = \int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau.$$

In particular,  $R(t; \tau)$  could be also defined as a weighted autocorrelation-like function if the goal is to obtain some smoothed versions:

$$R(t; \tau) = \int_{-\infty}^{+\infty} x\left(u + \frac{\tau}{2}\right) x^*\left(u - \frac{\tau}{2}\right) \phi((t - u), \tau) du,$$

where  $\phi(u, \tau)$  is a kernel function used to smooth the WVD.

While the previous distributions are linear, and so the similarity of the signal to a window function is measured using the correlation function, this distribution is a bilinear one which applies the Fourier transform to the instantaneous auto-correlation of the signal. In this way, its time-frequency representation is independent of the window function. On the whole, WVD reveals a better resolution of STFT but suffers from non-physical interference cross terms.

Certainly, the most widely used time-frequency technique, due to the flexibility and the efficient computational process, are the wavelet transforms. This method will be explained extensively in Section 2.2.

## 1.5 Features selection

A further key point related to machine diagnostics is the evaluation, and the selection, of proper features. In fact, the performance of a diagnostic methodology greatly depends on the quality of the features used. It is very important to take them in a proper combination that could contain as much pertinent information about the fault as possible. In particular, if the operating conditions are variable, the features should be able to hold all information about these variations, implicitly or explicitly. In this contest, feature extraction aims at transforming the existing features into a lower dimensional space which is useful for feature reduction to avoid the redundancy due to high-dimensional data. An example of a procedure

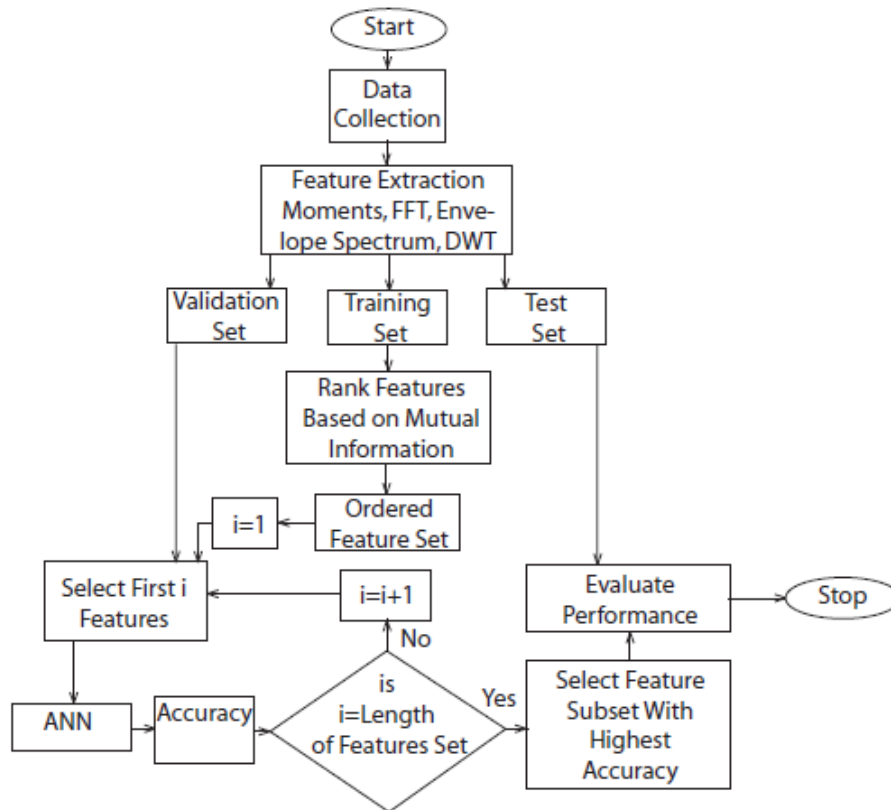


Figure 1.13. Algorithm for feature selection presented by Kappaganthu *et al.* (Figure taken from [KNS09]).

of feature extraction presented in [KNS09] could be observed in the flow chart in Figure 1.13.

The first step is the feature extraction through one among the various signal processing techniques, such as the time, the frequency and time-frequency domain presented so far. Each one of these techniques has its own sensitivity to various types of defects and operating conditions and this aspect must be taken into account. Moreover, the choice of the features depends on the system where the bearing is mounted and on the conditions in which they are expected to perform. Since often only one feature is not sufficient for a proper detection, multiple features need to be used for efficient classification and so the interaction between them is important for good diagnostic performance too. As could be imagined, the number itself of features is important, since it could degrade the classifier performance. Another key point is the content information, that has to be highest as possible in combination with a great accuracy.

Most of recent feature extraction techniques are based on linear technique such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) [WY07b]. The concept behind PCA is the creation of a set of basis functions that could optimally model the data in the sense of minimum error. PCA technique provides a linear transform of a certain

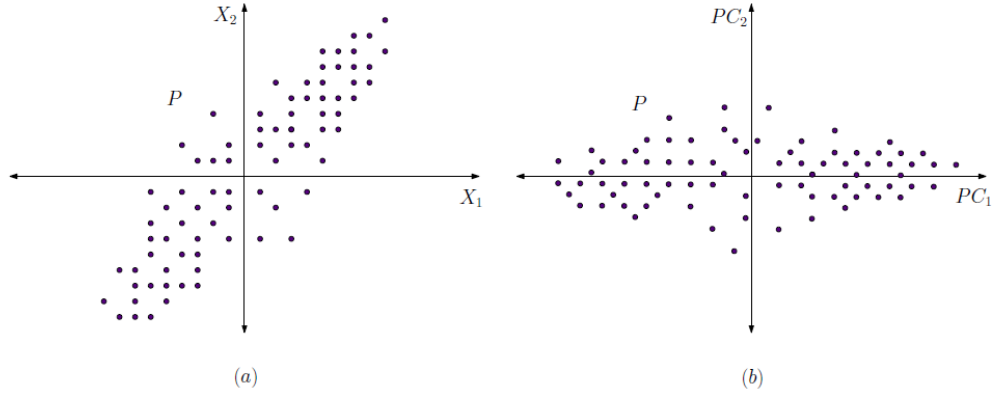


Figure 1.14. (a) Plot of 80 observations on two variables  $X_1$  and  $X_2$ . (b) Plot of the same 80 observations from (a) with respect to their principal components  $PC_1$  and  $PC_2$ .

number  $n$  of  $m$ -dimensional input vectors  $\mathbf{x}_i$ , where  $i = 1, \dots, n$  and  $m < n$ , into new vectors  $\mathbf{s}_i$  through the transformation:

$$\mathbf{s}_i = \mathbf{V}^T \mathbf{x}_i,$$

where  $\mathbf{V}$  is a  $m \times m$  orthogonal matrix where on each column  $\mathbf{v}_j$  is present the eigenvector of the sample covariance matrix

$$\mathbf{C} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T.$$

This means that PCA solves the eigenvalue problem

$$\lambda_j \mathbf{v}_j = \mathbf{C} \mathbf{v}_j, \quad j = 1, \dots, m$$

with  $\lambda_j$  being the eigenvalues of  $\mathbf{C}$  and  $\mathbf{v}_j$  the corresponding eigenvector.

When the orthogonal transformation of the input vector  $\mathbf{x}_i$  is computed, the various components of  $\mathbf{s}_i$ , which are called principal components, are obtained:

$$\mathbf{s}_i(j) = \mathbf{v}_j^T \mathbf{x}_i, \quad j = 1, \dots, m.$$

If the eigenvalues are sorted in descending order, it is possible to select only the first among them in order to reduce the number of principal components in  $\mathbf{s}_i$ . In this stands the PCA dimensional reduction ability. Moreover, the principal components of PCA present some interesting characteristics since  $\mathbf{s}_i(j)$  are uncorrelated and they show sequentially maximum variances. In addition, the mean squared approximation error in the representation of the original inputs through the first several principal components is minimal [Bis06].

On the other side, ICA is relatively recent method that could be considered as generalisation of PCA. This method, in fact, is able to find a linear transform for the observed data using a set of basis functions where the components are not only decorrelated but also mutual independent in complete statistical sense. As stated in [WY07b], ICA could be used

to extract independent components from the mixed signals. The concept of “independence” is related to the fact that information carried by one component can not be inferred from the others. From a statistical point of view, it could be said that joint probability of independent quantities is obtained as the product of the probability of each of them. Given a measured variables data matrix  $\mathbf{x}$ , the ICA model is

$$\mathbf{x} = \mathbf{A}\mathbf{s},$$

where  $\mathbf{A}$  is the unknown so called mixing matrix and  $\mathbf{s}$  is the independent component data matrix.

The goal of this technique is the estimation of the independent component matrix  $\mathbf{s}$  or of the mixing matrix  $\mathbf{A}$  knowing only the measured data matrix  $\mathbf{x}$ . Practically, it is necessary to apply to this former matrix a so called separating matrix  $\mathbf{W}$ :

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x},$$

which guarantees the components of the reconstructed data matrix  $\hat{\mathbf{s}}$  being as much independent as possible.

This last requirement is equivalent to assume that they do not belong to a Gaussian distribution. A way to gain this consists in optimising some measure of the non-Gaussianity of the vector  $\mathbf{W}\mathbf{x}$ .

An interesting evolution of these two techniques in the case of feature selection is the kernel trick. It consists in the projection of the input data into a high-dimensional implicit feature space with a nonlinear mapping, where it is possible to analyse the data. An application of this concept is presented by Widodo and Yang in [WY07b], where they develop an algorithm that combines ICA and the kernel trick to improve the feature extraction process that will be used in condition monitoring and fault diagnosis.

Another study that analyses the feature selection bearing defects quantitatively is proposed by Sugumaran *et al.*, that uses decision trees to select features that provide good performance using a proximal Support Vector Machine [SMR21]. Mahli and Gao develop a proper set of features in time, frequency and wavelet domains through the already cited PCA [MG04]. Also genetic programming (GP) is a powerful method able to generate new features from the original data set without prior knowledge of the probabilistic distribution [GJN05]. In [KNS09], instead, is presented a method based on mutual information which performs feature selection and allows the integration of model based features and data driven features. The power of this approach related to bearing diagnostics is widely expressed in [NK11]. In fact, it is a quantitative statistical method able to compare features for detection that are not dependent on the classifier. Moreover, it offers an ordered optimal feature set with features belonging to different domains providing, all together, a good classification accuracy. Then, the ordering and ranking of the features gives many insights into the system and the operating conditions. This could provide some useful suggestions about which features are



better suited for dealing with fault monitoring and prognostics for the system. Moreover, the author state that mutual information could help in understanding which features could be considered for fault monitoring purposes in their application.

## Chapter 2

# Bearing diagnostics techniques

In Chapter 1 the fundamental concepts related to bearing diagnostics have been presented together with the main steps of a diagnostic process. As already said, once a significant change indicating a potential fault has been detected, it is usually necessary to perform some signal processing techniques to make a diagnosis of the fault itself, which depend greatly on the type of damage expected. The type of analysis to be applied depends on the type of fault, and so it is interesting to investigate how various faults manifest themselves in the vibration signal. As presented in Section 1.2, various approach could be used to evaluate the features that could characterise the system health.

In this Chapter, an overall view of the state of the art of bearing diagnostics techniques related to vibration acquisitions is given (Section 2.1). Then, some of the most used techniques are presented and explained in detail, together with their applications in machinery and bearing diagnostics. The explanation of a time-frequency method, very useful in the processing of non-stationary signals, such as Wavelet Transform (WT), is proposed. Furthermore, an innovative self-adaptive signal processing method that can be applied to non-linear and non-stationary process and named Empirical Mode Decomposition (EMD) is also presented. Another interesting tool to analyse non-stationary signals is Spectral Kurtosis (SK), a statistical tool very powerful in finding the presence of series of transients and their locations in the frequency domain. Finally, a section will be focused on a particular range of analysis related to the variation in machine operative conditions, such as load and speed.

In the next sections, the various techniques are presented explaining their theoretical background and their applications. At present stage, there exists a number of industrial systems for bearing diagnostics which are already available on the market. Appendix B presents a table where their main features and technology content are described.

## 2.1 State of the art

A lot of research work has been published in the last 40 years on the detection and diagnosis of bearing defects through vibration and acoustic methods. One of the earliest papers on bearing diagnostics is written by Balderston of Boeing in 1969 [Bal69]. In his article, he realises how the vibration signals generated by bearing faults is located in the high frequency region of resonances excited by the internal impacts, and he analyses the natural frequencies of bearing rings and rolling elements. In the following years, works related to synchronous averaging come out. In fact, Weichbrodt and Smith uses it to expose local faults in both bearings and gears [WS70]. In particular, Braun made a fundamental analysis of synchronous averaging and applied the technique to bearing signals noticing that bearing signals are not completely periodic but provided of a random variation in period [BD79]. In the 1980s the first works on vibration monitoring techniques are presented: Mathew and Alfredson propose a brief review related to time and frequency domains methods with results on rolling element bearings [MA84], McFadden and Smith talk about the monitoring of rolling element bearings by the high-frequency resonance technique [MS84b], while Kim describes some specific techniques for condition monitoring of rolling bearings [Kim84]. Tandon and Nakra propose, instead, in [TN92b] a detailed review of the different vibration and acoustic methods, such as vibration measurements in time and frequency domains, sound measurements, the shock pulse method and the acoustic emission technique, further updated in [TC99]. Since then, many reviews on bearing diagnostics techniques have been proposed, such as those by Howard [How94], Jardine [JLB06] and Randall [Ran04a], [Ran04b], [RA11].

One of the simplest detection and diagnosis approach is certainly the analysis of the measured vibration signal in time domain, in particular through the evaluation of trending time domain statistical parameters as those presented in Section 1.4.1. Between the various parameters, the kurtosis was firstly proposed by Dyer and Stewart for bearing defect detection [DS78], where the suggestion of its measuring in selected frequency band is given too. Many studies proves the effectiveness of this method in bearing damage detection [MG84], [SCJ84]. However, two are the main drawbacks: kurtosis value tends to come down to the level of an undamaged bearing when the damage is well advanced and it could not detect the incipient damage effectively [MA84], [Kim84]. A further weakness of a time domain technique such as the ratio of peak value to RMS value of acceleration is demonstrated in [MS86], where it is observed as it could be applied with limited success for the detection of localised defects. In more recent years, the time domain parameters are usually evaluated and then used in connection with pattern recognition and classification methods ([HM97], [XSK00], [SABAA04], [SSR08]).

Statistical analysis of vibration signals have been proved to be very useful in detecting faults. As stated in [Col77], the probability density of acceleration of a healthy bearing has a Gaussian distribution, while a faulty one has a non-Gaussian distribution with dominant

tails due to relative increase in the number of high levels of acceleration. A near-Gaussian distribution for some damaged bearings has also been obtained by Mathew and Alfredson [MA84]. Andrade *et al.* suggest in [AEB01] to compare the Cumulative Density Function (CFD)

$$F(x) = \int_{-\infty}^x f(t)dt$$

of a target distribution with that of a reference one.

Then, the likelihood is used to successfully detect the gear tooth fatigue crack. As already said in Section 1.4.1, the study of the probability density curves could be substituted by the statistical moments of the data (1.2) and the relative various moment orders. Another way to express the amplitude characteristics of a vibration signal  $x(t)$  could be the probability density distribution ([DS78], [Ran87])

$$f(x) = \frac{d}{dx}F(x).$$

So, the probability density of a signal could be estimated evaluating the time duration of a signal in a set of amplitude windows. Then, if we consider a typical window at amplitude  $x$  and width  $\Delta x$ , its instantaneous probability  $P(x)$  is

$$P(x \leq x(t) \leq x + \Delta x) = \sum_{i=1}^N \frac{\Delta t_i}{T}.$$

This function varies as the time domain statistics change with the nature of the vibration signal and, so, if a bearing is damaged or not. Moreover, as stated by Dyer in [DS78], the differences in the characteristics of the probability density function can be enhanced considering the integral of the density curve to obtain the probability of exceedence.

On the whole, two approaches could be considered treating time domain statistics. The first one allows the evaluation of the chosen parameter for the whole frequency range of the signal while the other one operates its evaluation after having broken the signal up into discrete frequency bands. Researchers take into account different numbers of bands with differing frequency ranges which seem to have been chosen arbitrarily. As stated in [How94], a possible way to operate is to use at least two frequency bands: one in the base band characterised by the defect frequencies, low order harmonics and sidebands (< 5 kHz), and one in the pass band dominated by structural resonances (5-12.5 kHz). Obviously, the chosen frequencies depend on shaft rotational speed and the system itself. An example is proposed by the author in [How94], where the behaviour of the parameters is observed in four different bands.

Referring to these aspects, some bandpass filtering techniques could also be proposed in the time domain. They are based on the fact that structural resonances lie in the high frequency zone due to impulsive loading caused, as instance, from spalling of the races or rolling elements. These resonances could be detected by a transducer whose resonant

frequency is tuned to it. A method which works on this principle is the shock pulse method, which envisage the use of a piezoelectric transducer having a resonant frequency based at 32 kHz [TC99]. The impacts in the bearings initiate damped oscillations in the transducer at its resonant frequency and cause the shock pulses. Then, an indication of the condition of rolling bearings could be given by the measurement of the maximum value of the damped transient. This shock pulse method has gained wide industrial acceptance and has been reported to be successful in the detection of rolling element bearing defects as demonstrated in [MA84], [Kim84], [But73]. However, some researches demonstrate how this method could not effectively detect defects at low speeds ([TN92a], [Smi82]).

Further time domain parameters could be, as defined in [How94], the clearance factor (CF), the impulse factor (IF) and the shape factor (SF), defined in discrete form respectively as

$$\text{CF} = \frac{\text{peak}}{\frac{1}{N} \left( \sum_{i=1}^N \sqrt{|x(i)|} \right)^2} \quad \text{IF} = \frac{\text{peak}}{\frac{1}{N} \sum_{i=1}^N \sqrt{|x(i)|}} \quad \text{SF} = \frac{\text{RMS}}{\frac{1}{N} \sum_{i=1}^N \sqrt{|x(i)|}}.$$

Between these three parameters, the clearance factor it is said to be the most sensitive and robust in detecting incipient fatigue spalling.

Another way that has been used to define statistical moments is called beta distribution, characterised by the following probability density function [MIS92]:

$$p(x) = \frac{1}{B(b-a)^{\alpha+\beta-1}} (x-a)^{\alpha-1} (b-x)^{\beta-1},$$

where  $a < x < b$ .

In the case of signal analysis,  $a$  and  $b$  indicate the minimum and maximum amplitudes in time domain respectively. On the whole, according to the results obtained by Martin *et al.*, the moments based on this beta distribution have less sensitivity to noise than those based on the Gaussian one [MIS92].

The application of diagnostic techniques in the domain of frequency has been well developed since the 1960s, when the Fast Fourier Transform (FFT) and the cepstrum have been firstly presented as said in Section 1.4.2. However, as said in [How94], other methods could be used to estimate the power spectrum of a discrete time signal, such as the maximum likelihood method, the Auto Regressive (AR) estimation and the Moving Average (MA) technique. These methods starts to be increasing popular in signal processing and diagnostics at the end of 1980s. A typical use of spectrum illustrated by [Ran85] is related to the detection of some variations in the behaviour of a system through spectrum comparison and trending techniques. When a bearing is in good condition, it is supposed to produce a certain spectrum which is chosen as baseline. Variations that could be observed comparing this spectra with others are signs of changes in system condition. Usually, in bearing analysis, these variations are localised in higher frequency ranges. On the wall, these spectrum comparisons could be

performed on a logarithmic amplitude scale or dB. Problems arises when small fluctuations in rotating speeds are present, but they could be overcome using logarithmic frequency axis, as stated in [Kja90].

Various spectral parameters have been defined so far to trend vibration spectra or to trend simple spectral amplitudes of significant components. Some examples could be the matched filter RMS (Mfrms) and the RMS of the spectral difference (Rdo), defined respectively as [MM91]:

$$\text{Mfrms} = 10 \log \left\{ \frac{1}{N} \sum_{i=1}^N \left( \frac{A_i}{A_i(\text{ref})} \right)^2 \right\} \quad \text{Rdo} = \left\{ \frac{1}{N} \sum_{i=1}^N (L_{ci} - L_{oi})^2 \right\}^{\frac{1}{2}},$$

where  $A_i(\text{ref})$  is the amplitude of the  $i$ th spectral line in the reference spectrum,  $A_i$  is the amplitude of the  $i$ th spectral line in the current spectrum,  $L_{ci}$  is the dB-level of the current  $i$ th spectral line and  $L_{oi}$  is the dB-level of the reference spectrum  $i$ th spectral line.

One of the most used vibration signal processing techniques in the past years for roller element bearing detection and diagnosis is the envelope analysis. Since its development in 1970s by Mechanical Technology Inc., it has been known with many names. Firstly, it was called high frequency resonance technique (HFRT), then it was named amplitude demodulation, demodulated resonance analysis, narrow band envelope analysis before being known by its actual name [How94]. The target of the original HFRT was to shift the frequency analysis from the very high range of resonant carrier frequencies to the much lower range of the fault frequencies, so that they could be analysed with good resolution [DBH74]. The authors operate the frequency shifting through analogue rectifiers.

Between the most interesting and innovative works in the bearing diagnostics field stands the McFadden and Smith's classic paper on the modelling of bearing fault signals [MS84a]. Here, they treat the fault pulses as periodic and they explain as the series of broadband bursts excited by the shocks are further modulated in amplitude by two factors. In fact, the strength of the bursts depends on the load borne by the rolling element, and this is normally modulated by the rate at which the fault is passing through the load zone. Furthermore, where the fault is moving, the transfer function of the transmission path varies with respect to the fixed positions of response transducers. So, on the whole, the concept behind envelope analysis is the presence of impulse in vibration when a localised defect in a rolling element bearing comes into contact with another surface in the bearing itself. The impulses occurrence frequency are known as the characteristics bearing defect frequency already defined in Equations (1.4), (1.5), (1.6), (1.7). A problem that arises treating envelope analysis is the determination of the best frequency band to envelope. Various trend parameters and post-processing procedures have been studied since 1980s, which rely on theoretical predictions of typical spectral patterns that could appear for certain modes of failure. Some examples could be the use of median filtering of envelope spectrum followed by clipping, comb filtering and autocorrelation, the evaluation of energy in selected frequency band or of envelope kurtosis

([Ste83], [CG86], [How94]).

As stated by Randall *et al.* in [RAC01], for many years the envelope analysis has been one of the most successful techniques for diagnosing faults in rolling element bearings. They observe as the reasons for the advantage of this approach have been explained qualitatively, but not quantitatively. Thus, they want to give a quantitative evaluation of the degree to which information about bearing faults can be extracted by frequency analysing the raw signal. Moreover, they want to demonstrate the relationship between the classical envelope analysis and spectral correlation analysis, which is one of the tools used to characterise cyclostationary signals. From the definition given in [Ant09], “a signal is said to exhibit cyclostationarity if there exists a cascade of linear and non-linear transformations that produces a periodic component” so it could be seen as a signal that exhibits some hidden periodicity of its energy flow. The development of the cyclostationarity theory is closely linked to that of modern signal processing: the first works on this topic starts in the 1960s, but it is truly since the 1980s that cyclostationarity has become a subject of active research. Gardner gives a great tribute establishing many of the theoretical foundations, the terminology and foreseeing many applications, recognising its great potentiality in applications such as gears, belts, chains, shaft, bearings [Gar94]. However, the first application of the cyclostationarity principles to bearing signals is that by McCormick and Nandi [MN98] followed by many other uses [AG01], [LQ03], [AR05].

Another signal processing technique firstly appeared in 1960s is the cepstrum analysis [BHT63], that was initially used to seismological data and started to be applied to vibration condition monitoring in 1980s ([Ran80], [Cou85], [Mat87]). The principal use of cepstrum for bearing fault detection refers to find the periodicity in the frequency spectrum which corresponds to bearing frequency harmonic and associated sidebands pattern [How94]. So, this method represents a sensitive measure of growth of harmonics/sidebands family and it is characterised by insensitivity to measurement point and loading. Moreover, it can operate on specific section of spectrum, can be used for separation of different families and it is not sensitive to local or distributed defects.

A method that was developed in the early 1980s initially for signal processing in the context of neural network modelling is the Blind Source Separation (BSS). Then, during the last two decades, various studies have demonstrated its possible applications in various fields such as mechanical problems too (rotating machine and bearing diagnostics, online condition monitoring, noise analysis of diesel engines) [AB05]. The success of this method comes certainly from the fact that any BSS technique is able to reveal the underlying structure of a set of observed phenomena such as random variables, measurements or signals. So, this would solve the problem, found in many applications, of recovering initial (and unobservable) signals from measured data. Furthermore, only a small number of assumptions is required about the signals. This fact lies in the term “blind”, that refers to the little a priori knowledge had about the nature of the rough data and about the physical phenomenon of

interest.

The BSS generative model is defined as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad (2.1)$$

where  $\mathbf{x}(t)$  denotes the measured signals, that are essentially  $m$  observed data,  $\mathbf{s}(t)$  indicates the  $n$  desired signals, which are named *sources* or *components* of the system and the matrix  $\mathbf{A}$  is called the *mixing matrix*. Since the existence of  $n$  sources  $\{s_1(t), \dots, s_n(t)\}$  is supposed and the observation of as many mixtures  $\{x_1(t), \dots, x_m(t)\}$ , with  $m = n$ , (2.1) could also be written in the subscript notation as

$$x_i(t) = \sum_{j=1}^n a_{ij}s_j(t) \quad i = 1, \dots, m. \quad (2.2)$$

The signals  $\mathbf{x}(t)$  is known, so the BSS problem consists in estimating the sources  $\mathbf{s}$ , that are unknown, as the mixing coefficients  $a_{ij}$  too. Moreover, noise may corrupt the data too and, in this case, the model (2.2) becomes

$$x_i(t) = \sum_{j=1}^n a_{ij}s_j(t) + v_i(t) \quad i = 1, \dots, m, \quad (2.3)$$

which in matrix is

$$\mathbf{x}(t) = \mathbf{y}(t) + \mathbf{v}_{noise}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t), \quad (2.4)$$

where  $\mathbf{v}_{noise}(t)$  is the noise vector corrupting the data.

These previous equations refers to vectors of observed signals and components at a discrete time instant  $t$ . Supposing to consider a more general matrix  $\mathbf{X} \in \mathbb{R}^{m \times N}$  containing all the observations, where  $N$  is the number of available samples, Equation (2.4) becomes

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{V}, \quad (2.5)$$

where  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{V} \in \mathbb{R}^{m \times N}$  and  $\mathbf{S} \in \mathbb{R}^{n \times N}$ .

Since it is physically realistic to assume that the different signals originate from different physical processes, it is plain that they have to be unrelated and, mathematically, it is expressed assuming the statistical independence of sources. This consideration is important to the definition of the separation process and it is joined to the assumption of a squared mixing matrix, that allows to have the same number of sensors and sources ( $n_x = n_s$ ). Although some decompositions or matrix factorisations provide an exact reconstruction data (i.e.,  $\mathbf{X} = \mathbf{A}\mathbf{S}$ ), different cost functions and imposed constraints may lead to different types of matrix factorisation. Various source separation algorithms are available and their principles can be summarised by the four fundamental approaches presented in [CCPL05].

The power of this technique in facing mechanical problems is shown in a special issue published by Mechanical System and Signal Processing in 2005 [AB05]. It contains a wide



range of BSS technique applications in various fields, such as rotating machine and bearing diagnostics, online condition monitoring, noise analysis of diesel engines, with also the presentation of conceptual, methodological, theoretical and algorithmic aspects of the subject. Moreover, an interesting analysis about BSS application in vibrating signals is proposed by Antoni in [Ant05].

Many time-frequency domain techniques have been developed having great potential in detecting and diagnosing bearing problems in rotating machine, especially when the signal to noise ratio is low and a large number of frequency components is present (Section 1.4.3). Lots of this techniques were born at the beginning of the XIX century, but gain a great development thanks to the advent of fast computers. For example, Short Time Fourier Transform (STFT) has been applied firstly by Gabor in 1946 to speech communication [Gab46]. It reached its great diffusion when computers started to be fast enough to compute the Fourier transform of a discrete data set within a reasonable time frame, so at the end of 1970s [How94]. The visualisation of the STFT allows to see the transient nature of impacts occurring at various times and the nature of the excitation of different resonance. However, a difficulty lies in the inherent relationship between time and frequency resolution since a finer time resolution could be achieved at expenses of frequency resolution and vice-versa. This aspect led to the development of other alternative techniques to time-frequency analysis as, for example, the Wigner Ville Distribution (WVD).

WVD allows, in fact, to achieve a better resolution being able to overcome the spectrogram limitation since it is not based on signal segmentation. The distribution name itself include the names of the two main researchers which contribute to the creation and development of this technique: Wigner defined the Wigner distribution in 1932 [Wig32], while Ville gives his contribution at the end of 1940s [Vil48]. Since the 1990s, WVD has been widely used to analyse vibration data in order to monitor and diagnose damages in rotating machinery ([RTH90], [MQ91]). In particular, this technique and its derivatives have been found to be useful in the diagnostics of short time non stationary events such as impacts, amplitude and phase modulation and resonance excitations [How94]. However, some difficulties arises with the severe cross terms as indicated by the existence of negative power for some frequency ranges. Moreover, the WVD of discrete time signals suffers from the aliasing problem, which may be overcome by employing various approaches.

A technique similar to STFT and WVD in providing a time-frequency map of the signal is Wavelet Transform (WT). However, some differences are present. The main one lies in the fact that the latter considers high frequency components with a sharper time resolution than the low frequency components. An interesting comparison between these three methods is presented in [How94], where they are applied to the vibrations acquired for two configuration of a gear test rig: in the first bearings are in good condition, while in the other a bearing is characterised by a inner race defect. The results comparison shows that STFT seems to be more suited in indicating structural resonances due to bearing impacts rather than in

indicating the precise nature of the resonant response in time domain. On the contrary, WT is more well suited to indicate the transient nature of the resonant excitation in time domain, but it does not solve the response of the high frequency resonance in a clear way in the frequency domain. The WVD, instead, could get the best from time and frequency domain, but it is more difficult to interpret. In fact, the impulses could be identified in time domain but less precisely as in WT, and the structural resonance are not apparent as in STFT case.

In Section 2.2 the wavelet transform (WT) method and some further development will be presented. As stated in [RP11], however, the main drawback of this approach is the necessity to know a priori the frequencies of the original signal that should be analysed. So, the signal processing is not totally automatic and, then, the computational efforts are proportional to the number of the considered frequencies. A possible technique which could overcome this aspect and that is based intrinsically on data Empirical Mode Decomposition (EMD). Thanks to these aspects, it is particularly eligible for non-stationary and non-periodic signals, as those produced by bearings are. The EMD and the following Hilbert transform (HT) will be widely presented in Section 2.3.

## 2.2 Wavelet transform

The wavelet transform (WT) is a relatively new and powerful tool in the field of signal processing, which overcomes problems that other techniques face, especially in the processing of non-stationary signals. This technique, in fact, provides powerful multi-resolution analysis in both time and frequency domain and thereby becomes a rather useful tool to extract the transitory features of non-stationary vibration signals produced by the faulty bearing. In general, this analysis results in a series of wavelet coefficients indicating how close the signal is to the particular wavelet itself. In order to extract the fault feature of signals more effectively, an appropriate wavelet base function should be selected. It is clear that the power of this method comes from the simultaneous interpretation of the signal in both time and frequency domain that allows local, transient or intermittent components to be exposed.

A wavelet  $\psi(t)$  is a waveform of effectively limited duration that has an average value of zero overtime and has a finite energy, as described by this equation:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0.$$

Wavelet analysis consists in the breaking up of a signal into shifted and scaled versions of the original (or mother) wavelet that, thus, could also be seen as basis functions. This particular function is capable of describing a signal in a localised time (or space) and frequency (or scale) domains. In this way, it is possible to say that a continuous wavelet transform could

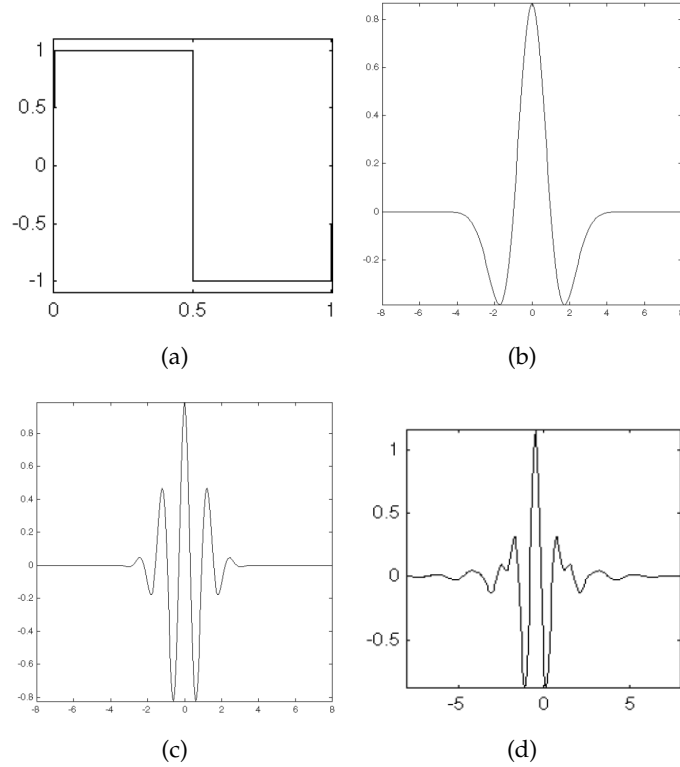


Figure 2.1. Various wavelet functions  $\psi(t)$ : (a) Haar, (b) Mexican hat-shaped, (c) Morlet , (d) Meyer (Figures taken from [Mat02]).

be defined as

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi^* \left( \frac{t-b}{a} \right) dt,$$

where  $x(t)$  is the waveform signal,  $a$  is the scale parameter,  $b$  is the time parameter,  $\psi(\cdot)$  is a wavelet and “\*” denotes complex conjugate.

The choice of the type of the basis function depends on the application as well as the computation efforts required. Thus, the appropriate selection of the mother wavelets is very important, since results are heavily dependent on the chosen wavelet shape. Wavelets of various shapes exist, including the first-ever wavelet, named Haar wavelet (Figure 2.1(a)), and other types such as Mexican hat-shaped (Figure 2.1(b)) and Gaussian-shaped. However, for singularity analysis and detection, wavelets should have an impulse-like shape to capture the sudden change in the signal. Morlet (Figure 2.1(c)) and Gaussian wavelets are found to have an excellent representation of a singularity (discontinuity) as the vibration signal has a harmonic oscillation joined to a singularity representing the fault. On the whole, it is desirable to have a linear phase scaling function that could maintain a constant group delay and the envelope of the original vibration signal. These devices are useful since they will avoid problems during the signal reconstruction process. Furthermore, the use of a wavelet shape resembling that of the targeted fault greatly helps in the wavelet selection decision.

Generally, wavelet analysis allows the use of long time intervals, where more precise low-frequency information is desired, and also permits the use of shorter time intervals if accurate high-frequency information is needed. It is also employed for the accurate extraction of narrow-band frequency signals. The main advantage of wavelets is certainly its ability to perform a local analysis of a signal, or to zoom on any interval of time without losing the spectral information contained therein. Then, it is capable of revealing some hidden aspects of the data that other signal analysis techniques fail to detect. This property is particularly important for damage (crack) or fault detection applications. In fact, wavelet analysis could identify aspects like trends, breakdown points, discontinuities in higher derivatives, and self-similarity. Furthermore, because it affords a different view of data than those of traditional techniques, wavelet analysis can often compress or de-noise a signal without appreciable degradation. One possible drawback, instead, could lie in its frequency resolution, that may be quite poor in the higher frequency region. Hence, the WT still faces difficulties when trying to discriminate signals containing high frequency components, such as impact faults.

Three main families of WT could be considered: the Continuous Wavelet Transform (CWT), the Discrete Wavelet Transform (DWT) and Wavelet Packet transform (WPT). CWT was firstly formulated by Morlet and Grossmann and defined as the sum over all time of the signal multiplied by scaled and shifted versions of the mother wavelet function [PC04]. In this way, the modulus of the CWT for a signal  $f(t)$  could be seen as

$$C_f(a,b) = \int_{-\infty}^{+\infty} x(t)\psi(a,b,t)dt,$$

with  $t$  time,  $a$  scale and  $b$  location or space.

The CWT is certainly one of the best transforms for singularity detection. Impact faults could be detected by finding the singularity in the signal. So, its identification is based on the use of the local maxima lines through finding the abscissa, where the wavelet modulus maxima converge at fine scales. The “continuity” of the method stands in the scales at which it operates, that could be each one possible, and in the shifting, because, during computation, the analysing wavelet is shifted smoothly over the full domain of the analysed function.

Daubechies and Mallat developed the passage from continuous to discrete signal analysis introducing the discretisation of the modulus of the CWT as

$$C_f(a,k) = \sum_n x(n)\psi(a,k,n), \quad n, k = 1, 2, \dots, N$$

where  $t = n\Delta T$  and  $b = k\Delta T$  with  $\Delta T$  sampling interval,  $N$  is the number of samples [PC04]. An efficient way to implement this scheme using filters was developed in 1988 by Mallat, through a classical scheme known in the signal processing as a two-channel subband coder. In particular, it decomposes the signal in its high-scale, low-frequency components, named *approximations* (A) and its low-scale, high-frequency components, which is called *details* (D). The filtering process, with the approximation and detail components, is represented in Figure

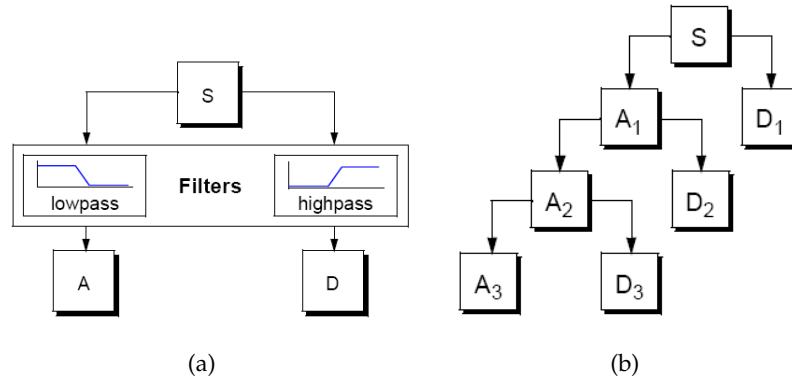


Figure 2.2. Filtering process: (a) the signal (S) is decomposed into approximation (A) and detail (D) components; (b) wavelet decomposition tree (Figures taken from [Mat02]).

2.2(a). This decomposition process can be iterated and the successive approximations  $A_i$  are decomposed in turn: one signal is broken down into many lower-resolution components, obtaining the so called wavelet decomposition tree (Figure 2.2(b)). As it could be imagined, the analysis process is iterative and so, in theory, it can be continued indefinitely. However, in reality, the decomposition can proceed only until a certain suitable number of levels based on the nature of the signal, or on a suitable criterion such as entropy.

A generalisation of DWT, proposed in 1992 by Coifman, Meyer and Wickerhauser, is the Wavelet Packet method, offering a richer range of possibilities for signal analysis [PC04]. In fact, as said before, in wavelet analysis, a signal is split into a detail  $D$  and an approximation  $A$ , which is split again into a second-level approximation and detail and so on (Figure 2.2(b)). Referring to this Figure, the whole signal  $S$  could be seen as the sum of various elements:

$$S = A_1 + D_1 = A_2 + D_2 + D_1 = A_3 + D_3 + D_2 + D_1$$

The different contribution given by Wavelet Packet analysis stands in the splitting of the details  $D$  too. In this way, the decomposition involves both the approximations  $A$  and the details  $D$ , as shown by Figure 2.3. So, the signal could be expressed as

$$S = A_1 + AAD_3 + DAD_3 + DD_2$$

The choice of the possible encodings must, then, be considered. A way is the one cited previously, that is, the use of an entropy-based criterion to select the most suitable decomposition of a given signal. This means to analyse each node of the decomposition tree and to quantify the information to be gained by performing each split. In particular, WPT allows the extraction of signal features that combine both non-stationary and stationary characteristics and is found to be one of the best analysing tools of vibration signals and fault detection. In general, WPT is faster than CWT as it uses orthogonal and bi-orthogonal bases with a better resolution in the high-frequency region [ABSC11]. As for the previous methods, also

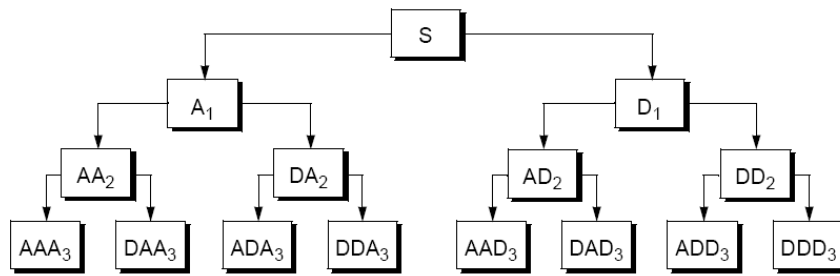


Figure 2.3. Wavelet packet decomposition tree with various levels decomposition (Figures taken from [Mat02]).

for WPT the selection of the mother wavelet function plays a big role for its efficiency. From a simulation point of view, it is found that Daubechies and discrete Meyer (Figure 2.1(d)) are the best wavelets to be used in vibration signal analysis [ABSC11].

In 1990s, WT has been successfully used in signal processing, such as image coding, compressing and edge detection. In the field of mechanical fault diagnosis, wavelet analysis has also been used in rolling bearing faults detection and gearbox diagnosis since early years 2000s. It is clear that the power of this method comes from the simultaneous interpretation of the signal in both time and frequency domain that allows local, transient or intermittent components to be exposed. One of the main drawback, however, is its dependence on wavelet basis function choice. For example, in order to detect the low frequency signals, Li *et al.* use Haar wavelets [LQL07], which, under different and particularly large scales, have good low-pass filter characteristics in the frequency domain. This fact allows the use of Haar continuous wavelet transform for such purposes and it is verified through the use of this transform to the diagnosis of various types of machine faults. Zou *et al.*, instead, use CWT with a Daubechies 10 mother wavelet and Hilbert transform to analyse the torsional vibration signal of a cracked rotor [ZCP04]. Though their analysis they are able to detect a deep crack by two peaks in a three-dimensional wavelet spectrum and that of a shallow crack by six peaks. Related to bearings, Junsheng *et al.* apply the CWT in order to diagnose a fault in roller bearings through an impulse response wavelet [JDY07]. They use it to target the characteristics of vibration signals and they find that this technique is superior to the Morlet wavelet in the scale-wavelet power spectrum comparison method.

## Recent applications

In more recent times, many applications have been presented related to the application of wavelet transform to rotating elements and machine diagnostics. For example, Hu *et al.* present a novel method for fault diagnosis of rolling element bearing based on an improved wavelet package transform, a distance evaluation technique and the support vector machines

(SVMs) ensemble [HHZZ07]. Another approach combining wavelet analysis and SVM is the one proposed by Abbasion *et al.* in [ARFI07] where an electric motor which has two rolling bearings is studied. The results they achieve through wavelet analysis and SVM are fully in agreement with empirical results.

The most recent developments regarding the application of wavelet analysis to diagnostic purposes range over various aspects. In 2011 Hong *et al.* propose an algorithm which takes advantage of time-frequency localised features of wavelet packets in order to classify the system condition and pinpoint the fault position [HASC11]. Fisher's linear discriminant is used to classify system conditions and all the wavelet packets are exhaustively searched in order to unravel the optimal base (or packet) that discriminates different system conditions based on the Fisher value. It is the time-frequency localised feature of the wavelet packet that further allows the detection of the fault position and even of the faulty component in the bevel gear. This algorithm is validated using a machine fault simulator that consists of a motor, a bevel gear and a bearing. The bevel gear condition may be varied to simulate normal, cracked, and broken conditions using vibration signal. The proposed algorithm precisely locates fault position as well as identifies system conditions. In the same year, Al-Badour *et al.* present a paper discussing the vibration analysis of rotating machinery using time-frequency analysis and wavelet techniques [ABSC11]. They start from the fact that commonly-used signal analysis techniques, based on spectral approaches, are powerful in diagnosing a variety of vibration-related problems in rotating machinery but they fail to do so in several practical cases involving non-stationary data. This could result either from fast operational conditions or from the presence of a fault causing a discontinuity in the vibration signal being monitored. To successfully resolve fast-changing signals (such as transient signals) resulting from non-stationary environments they consider wavelet transform tools. In fact, they result to be superior to both the fast and short-time Fourier transforms in effectively analysing non-stationary signals. A suitable choice of a mother wavelet function is done and the wavelet tool is applied to a vibration monitoring system to accurately detect and localise faults occurring in this system. Two cause of non-stationarity are considered: stator-to-blade rubbing and fast start-up and coast-down of a rotor. In this paper two wavelet techniques are used for the analysis of the monitored vibration signals: the continuous wavelet and wavelet packet transforms. The proposed approach is interesting because exploits the fact that the WPT, largely undiscovered and untested in vibration analysis, provides a much richer spectral characterisation of a signal than the CWT by decomposing the detail signal along the approximate one at every scale. In addition, in this paper, a novel algorithm combining the two techniques and the idea of windowing a signal into a number of shaft revolutions to localise faults is applied.

One of the paper presented in 2012 describes a novel wavelet decomposition scheme named adaptive morphological gradient lifting wavelet (AMGLW), whose goal is the detection of bearing defects [LZM<sup>+</sup>12]. The adaptability of this new technique allows this scheme

to select between two filters, to mean the average filter and morphological gradient filter and to update the approximation signal based on the local gradient of the analysed signal. Both a simulated signal and vibration signals acquired from bearing are employed to evaluate and compare the proposed AMGLW scheme with the traditional linear wavelet transform (LWT) and another adaptive lifting wavelet (ALW) developed in literature. Experimental results reveal that the AMGLW outperforms the LW and ALW obviously for detecting bearing defects. The strength of this scheme is given by the fact that the impulsive components can be enhanced while the noise can be simultaneously depressed: this allows the characteristic frequencies of a bearing to be clearly identified. Furthermore, the AMGLW gets an advantage over LW in computation efficiency. In this way, it is quite suitable for online condition monitoring of bearings and other rotating machineries. A work related to gearbox diagnosis, instead, is presented by Omar and Gaouda [OG12]. They propose a novel wavelet-based technique for detecting and localising gear tooth defects in a noisy environment. This method uses a dynamic windowing process during the analysis of the gearbox vibration signals in the wavelet domain. The gear vibration signal is processed through a dynamic Kaiser's window with varying parameters. Characteristics such the window size, the shape and the sliding rate are modified to increase the similarity between the non-stationary vibration signal and the selected mother wavelet. The window parameters are also continuously modified so that they could provide maximum wavelet coefficients localised at the defected tooth. This technique is applied on laboratory data corrupted with high noise level and it shows accurate results in detecting and localising gear tooth fracture with different damage severity.

### 2.3 Empirical Mode Decomposition

Empirical Mode Decomposition is a method firstly presented in 1990s by Huang *et al.* which is based on the local characteristic time scales of a signal [HSL<sup>+</sup>98]. This approach could be seen as a self-adaptive signal processing method that can be applied to non-linear and non-stationary process. The strength of EMD is so its self-adaptiveness which contributes in its high efficiency in non-stationary data analysis. In particular, it allows a complex signal function to be decomposed into a number of Intrinsic Mode Functions (IMFs). These IMF represent the natural oscillatory modes embedded in the raw signal and they work as the basis functions which are determined by the signal itself rather than by pre-determined functions. A component is said to be an Intrinsic Mode Function is satisfies the following definition [HSL<sup>+</sup>98]:

- In the entire data set, the number of extrema and the number of zero crossings must either be equal or differ at most by one;
- At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.



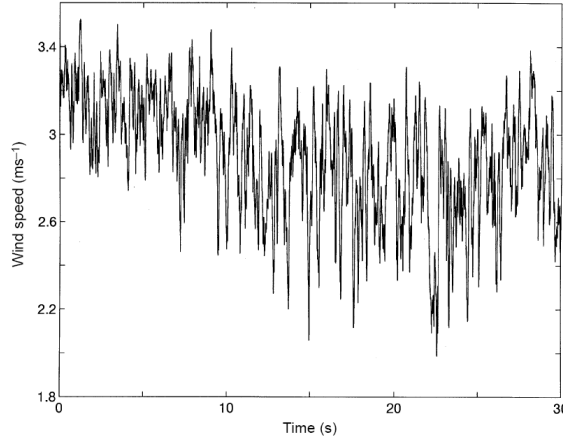


Figure 2.4. The calibrated wind data from a wind-wave tunnel (Figure taken from [HSL<sup>+</sup>98]).

The first point is rather similar to the traditional band requirement for stationary Gaussian processes and means that all maxima must be over the time axes while all minima must be under it. Moreover, it helps in assuring that the signal does not present wave “rising” or “descending”, as could happen in the case of sum of sinusoidal signals. The second aspect is related, instead, one the signal shape, since it must be symmetric, with time mean locally equal to zero in order to avoid unwanted fluctuations induced by asymmetric waveforms. It must be noticed that in the evaluation of the mean of non-stationary signal a local time scale should be introduced, but in this case it is not possible for the signal nature. This problem is overcome evaluating the mean of the two envelopes, as explained in the second point of the IMF definition.

According to [HSL<sup>+</sup>98], a *sifting* process is used in order to extract the IMFs from a given signal  $x(t)$ . It consists of different steps:

- I. Identify all the extrema of the signal, and connect all the local maxima by a cubic spline line as the upper envelope. Repeat the same procedure on the local minima to produce the lower envelope.
- II. Designate the mean of the two envelopes as  $m_1$ , and the difference between the signals  $x(t)$  and  $m_1$  as the first component,  $h_1$ , i.e.

$$x(t) - m_1 = h_1. \quad (2.6)$$

Ideally, if  $h_1$  is an IMF, then take it as the first IMF component of  $x(t)$ . Otherwise, consider  $h_1$  as the original signal and repeat the first two step obtaining

$$h_1 - m_{11} = h_{11}. \quad (2.7)$$

Repeat the sifting process up to  $k$  times when  $h_{1k}$  becomes an IMF, that is

$$h_{1(k-1)} - m_{1k} = h_{1k}. \quad (2.8)$$

The first IMF component is then designated as

$$c_1 = h_{1k}. \quad (2.9)$$

III. Separate  $c_1$  from the original signal  $x(t)$  to obtain the residue  $r_1$ :

$$r_1 = x(t) - c_1. \quad (2.10)$$

IV. Consider  $r_1$  as the original signal and repeat the above process  $n$  times, obtaining the other IMFs  $c_2, c_3, \dots, c_n$  satisfying

$$\begin{aligned} r_1 - c_2 &= r_2 \\ &\vdots \\ r_{n-1} - c_n &= r_n \end{aligned} \quad (2.11)$$

V. Stop the decomposition process when  $r_n$  becomes a monotonic function from which no more IMFs can be extracted. The sum of Eq. (2.10) and Eq. (2.11) gives

$$x(t) = \sum_{i=1}^n c_i + r_n. \quad (2.12)$$

From Eq. (2.12) we can see how the signal  $x(t)$  has been decomposed into  $n$  empirical modes and a residue  $r_n$ , that could be interpreted as the mean trend of the signal. Each IMF  $c_i$  includes different frequency bands ranging from high to low and is stationary.

To illustrate the sifting process, Huang *et al.* use in [HSL<sup>+</sup>98] a set of wind data collected in a laboratory wind-wave tunnel with a high-frequency response Pitot tube located 10 cm above the mean water level. The wind speed is recorded under the condition of the initial onset of water waves from a calm surface. Calibrated wind data are given in Figure 2.4 where it could be noticed how the data are quite complicated and characterised by many local extrema but no zero crossings. The application of the sifting process to this signal produces 9 components (Figures 2.5 (a)-(b)) that enhance the general separation of the data into locally non-overlapping time scale components. They observe as, in some components such as  $c_1$  and  $c_3$ , the signals are intermittent and then the neighbouring components might contain oscillations of the same scale, but signals of the same time scale would never occur at the same locations in two different IMF components.

A possible further step after IMFs evaluation consists in the computation of the Hilbert transform for each component [HSL<sup>+</sup>98]:

$$H[c_i(t)] = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{c_i(\tau)}{t - \tau}$$

that allows the signal to be expressed as

$$z_i(t) = c_i(t) + jH[c_i(t)] = a_i(t) \exp(j\omega_i(t)), \quad (2.13)$$

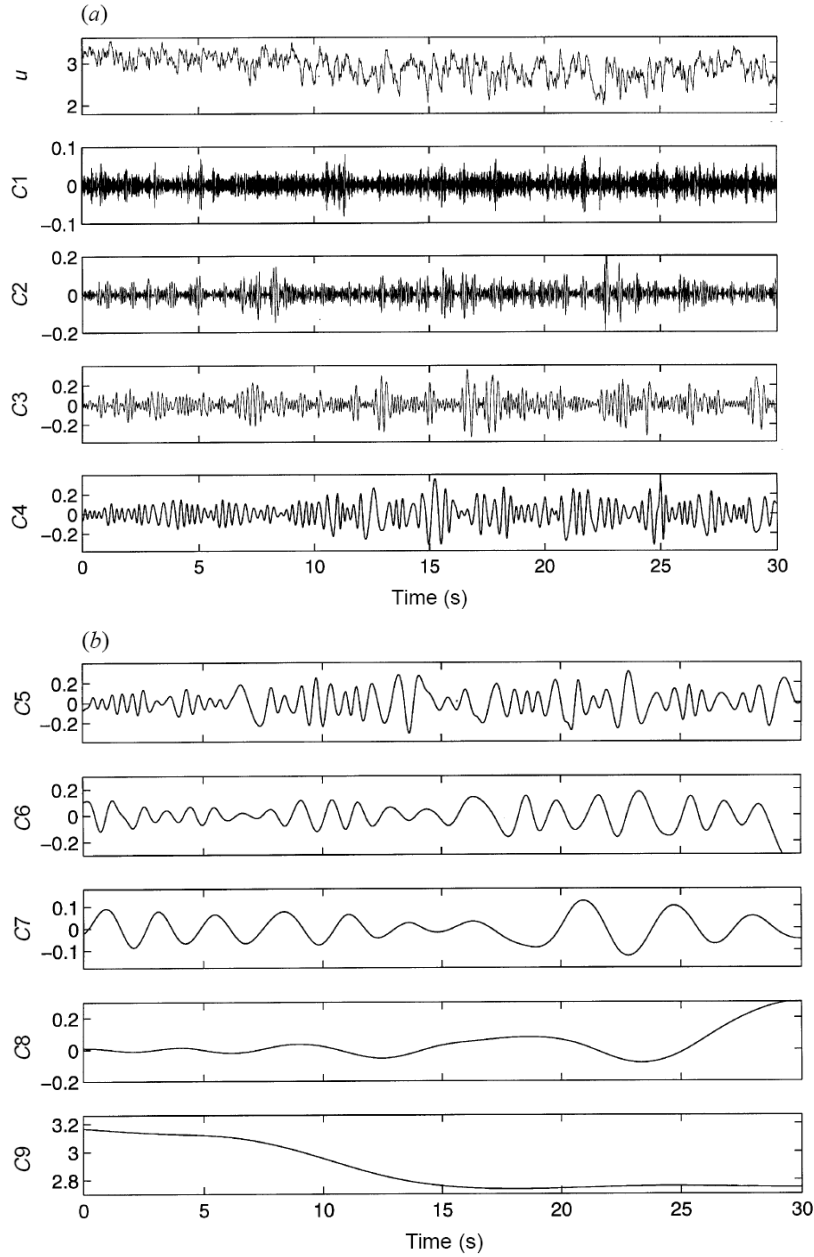


Figure 2.5. Empirical mode decomposition components computed from the wind data: original data ( $u$ ) and the components  $c_1 - c_4$  (a), components  $c_5 - c_9$  (b) where the last component,  $c_9$ , is not an IMF but the trend (Figure taken from [HSL<sup>+</sup>98]).

with amplitude and phase defined, respectively, as

$$a_i(t) = \sqrt{c_i^2(t) + H^2[c_i(t)]} \quad \theta_i(t) = \arctan\left(\frac{H[c_i(t)]}{c_i(t)}\right).$$

So, the instantaneous frequency is defined as

$$\omega_i(t) = \frac{d\theta_i(t)}{dt}$$

and the original data could be decomposed as

$$x(t) = \text{Re} \sum_{i=1}^n a_i(t) \exp(j \int \omega_i(t) dt).$$

A time-frequency distribution, named the Hilbert-Huang spectrum  $H(\omega, t)$ , could be defined referring to 2.13 in order to represent the amplitude and the instantaneous frequency in a three-dimensional plot, being the amplitude the height in the time-frequency plane. From the Hilbert-Huang spectrum

$$H(\omega, t) = \text{Re} \sum_{i=1}^n a_i(t) \exp(j \int \omega_i(t) dt)$$

it is possible to define the marginal spectrum  $h(\omega)$ :

$$h(\omega) = \int_0^T H(\omega, t) dt,$$

with  $T$  total data length.

Given these definitions, the Hilbert spectrum could be seen as a measure of each frequency and time amplitude contribution, while the marginal spectrum offers a measure of the total amplitude (or energy) contribution from each frequency value. Moreover, the marginal spectrum itself represents the cumulated amplitude over the entire data span in a probabilistic sense. Huang *et al.* observe that frequency in  $h(\omega)$  owns a different meaning from the Fourier spectral analysis. While in the Fourier representation, the existence of energy at a frequency  $\omega$  represents a component of a sine or a cosine wave persisting through the time span of the data, in this definition it means simply that, in the whole time span of the data, there is a higher likelihood for such a wave to have appeared locally. So, the frequency in the marginal spectrum indicates only the likelihood that an oscillation with such a frequency exists and the exact occurrence time of that oscillation is given in the full Hilbert spectrum.

An interesting analysis on EMD is that proposed by Flandrin *et al.* in [FRG04], where they study EMD behaviour in stochastic situations involving broadband noise. They report numerical experiments based on fractional Gaussian noise obtaining that EMD acts essentially as a dyadic filter bank resembling those involved in wavelet decompositions. From their results comes out that the filter associated to the mode  $c_1$  is essentially highpass (although it contains a non negligible lowpass part in the lower half-band) while the modes of higher indexes are characterised by a set of overlapping bandpass filters. Moreover, each mode of index  $(k+1)$ ,  $k \geq 2$  occupies a frequency domain which is roughly the upper half-band of that of the previous residual of index  $k$ . Also Gao *et al.* in [GDFM08] consider the filtering features of EMD and compare it with those of wavelet decomposition (Figure 2.6). They observe as the bandwidth of the IMF  $c_i$ ,  $1 \leq i \leq n$ , is determined by the features of the signal itself rather than other factors, such as sampling frequency. In this fact stands the EMD characteristic of

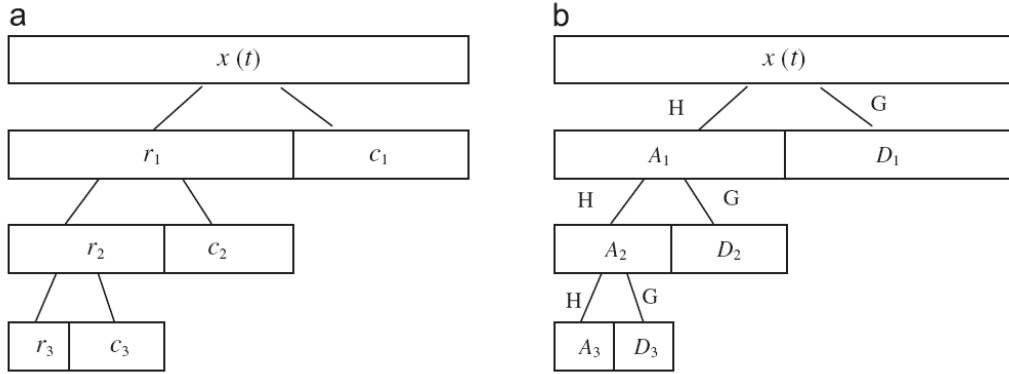


Figure 2.6. The filtering features of Empirical Mode Decomposition (a) and Wavelet Decomposition (b) (Figure taken from [GDFM08]).

being an adaptive band-pass filter bank with the bandwidth self-adaptively determined by the signal decomposed, as shown in Figure 2.6 (a). If it is compared to the filter features of wavelet decomposition (Figure 2.6 (b)) it could be seen that the frequency band of the signal is divided into two equal parts of high frequency and low frequency in every step of wavelet decomposition. In this case, the bandwidth of  $A_j$  and  $D_j$  ( $1 \leq j \leq p$ , where  $p$  is the maximum layer of wavelet decomposition), is decided by the sampling frequency of the signal and the layer  $j$ . Hence, since there is no influence of the signal characteristic in the bandwidth evaluation, wavelet decomposition is not self-adaptive.

Many applications to a wide range of problems have been done so far, from geophysics to structural health monitoring as shown in [HS05]. Lots of authors apply EMD to rotating machines and bearings with diagnostic intents, usually in association with other techniques. An example is [GDFM08] where Gao *et al.* researches show that IMFs sometimes fail to reveal the signal characteristics due to the effect of noises. So, they introduce the concept of combined mode function (CMF): the neighbouring IMFs  $c_i, c_{i+1}, \dots, c_{i+m}$  are combined to obtain  $c_s$ , which represents an oscillation mode depicting signal features more precisely:

$$c_s = c_i + c_{i+1} + \dots + c_{i+m}, \quad 1 \leq i \leq n - m,$$

where  $n$  is the maximal number of IMFs.

They analyse the adaptive filtering features of EMD and CMF and the simulation signals are applied to test their performance. Through these numerical experiments they could prove that CMF can increase the precision of EMD when the signal is over-decomposed. Furthermore, a practical fault signal of a power generator from a thermal-electric plant is analysed to diagnose the fault by their method and results show that it can extract the rotating machine fault characteristics and identify the fault patterns effectively. They find the reason of CMF better performance in the fact that EMD is by nature an adaptive filter bank and their CMF method in fact redesigns the filter bank. Junsheng *et al.* use EMD jointly with an Auto Regressive (AR) model to extract fault features in roller bearings. The AR model, in fact,

contains the characteristics of this system. They state that the roller bearing fault could be identified by the AR model parameters avoiding the construction of the mathematical model and the study of the fault mechanism. A drawback is that, unfortunately, the AR model can only be applied to stationary signals, while the fault vibration signals of a roller bearing are not. Accordingly, in their paper [JDY06] they carry out a pretreatment on fault vibration signals before the creation of the AR model. Firstly, they apply EMD to decomposes the signal into a finite number of IMFs. These IMF components, as explained before, contains the real information of the signal being stationary and an AR model could be defined for each one of them. Feature vectors are then built considering the AR parameters and remnant's variance for different condition or types of faults. Finally the condition of the roller bearing is identified considering the synthesised distance between the feature vectors of the model and those of the analysed signal. The authors test their approach through some experimental analysis which allow them to validate the effective ability in fault recognition. Another diagnostic method proposed by Yu *et al.* is based, instead, on EMD and Artificial Neural Network (ANN) [YDJ06]. Also in this case, the EMD is used as a kind of preprocessing technique and then ANN is applied in order to determine the work condition of the analysed roller bearings. In the presented method, the EMD energy entropy is the parameter that is used to check when the work condition of roller bearing changes. So, the energy of each IMF component is adopted as ANN input features to identify the work condition of the roller bearing. From the experiment results obtained on their test rig, they could state that the combination of these two techniques could successfully identify the work condition and fault patterns of roller bearings. The comparison with results obtained through wavelet packet analysis verify the higher identification ability of their method too. In [YYC08] EMD is combined with another classification technique called Support Vector Machine (SVM). The target of Yang *et al.* is to repair to the problems that usually arises using a conventional envelope analysis method. For example, in this case, it is necessary to select the central frequency of filter in advance, looking for spectral line of fault characteristic frequencies in envelope spectrum. The method they presents tries to overcome this problem through three main steps. At the beginning, EMD is applied to the original modulation signals. Then, they consider some IMFs including the most dominant fault information and they evaluate their envelope spectra. Here, they take into account the ratios of amplitudes at the different fault characteristic frequencies and define them as the characteristic amplitude ratios. Finally, these features, seen as fault characteristic vectors, enters the SVM classifiers in order to identify the various working conditions.

## Recent applications

As already said, EMD has been widely used in rotating machines and bearing diagnosis. However, standard EMD algorithm is designed to process only real-valued data. When

signals are acquired by multiple sensors or from different health conditions, they are processed individually and, so, fault features are extracted from each individual set of IMFs. Two main problems originates if acquisitions and the following extraction is operated in this way. Firstly, since EMD had a local and data-driven nature, it could happen that the decomposition results of signals from different sources do not match and so a multi-scale comparison between different damage conditions could be rather difficult. Furthermore, the joint information between these multiple sensors is not considered since each signal from each sensor is treated individually. To ensure proper decomposition of signals from different sources, standard EMD has recently been extended to the multivariate versions of EMD. Some examples are those suitable for the bivariate [RFGL07], for the trivariate [RM10a] and for multivariate signals [RM10b]. Bivariate EMD estimates the local mean by mapping a bivariate signal to a number of real-valued projected signals and averaging the local mean of these projected signals. For example, it could be employed to facilitate the synchronisation of image fusion and to diagnose a fault of a rotor. Following the same idea, trivariate EMD is proposed for three-dimensional signals. Multivariate EMD generalises this concept for  $n$ -dimensional signals.

Many other recent applications have been developed considering this almost 20-years old technique. For example, in 2011 Ricci and Pennacchi present a method to overcome the problem of IMF selection that are going to be used for the calculation of Hilbert-Huang spectrum [RP11]. This choice, in fact, is usually done on the basis of user's experience. On the contrary, in their paper, a merit index is introduced that allows their automatic selection. The effectiveness of this improvement is proven thanks to some experimental tests performed on a test-rig equipped with a spiral bevel gearbox that is usually characterised by some difficulties in the diagnosis of serious gears damages. Through their merit index, the defective gearbox is always identified, also in the case of transient operating conditions. A further application related to bearing fault detection is that presented by Tsao *et al.* in [TLLP12]. They want to overcome the limit of the traditional envelope analysis which is based on the examination of all the resonant frequency bands. In this paper, EMD is used to choose an appropriate resonant frequency band for characterising feature frequencies of bearing faults. Thanks to EMD band-pass filtering nature, these resonant frequency bands could be allocated in a certain IMF. The method is validated through comparisons with the traditional envelope analysis in the detection of damages on the inner or outer ring of bearings. Their experimental results confirms the efficiency of their approach in diagnosing these faulty bearings. Zhao *et al.* propose another application to rotating machine for the cases of data obtained from multiple sensors and of multiple health conditions [ZPZ12]. Their method consists in extracting fault information through multivariate EMD and full spectrum. Firstly, multivariate EMD is used to extract the IMFs embedded in raw multivariate signals and a criterion based on mutual information is proposed to select a sensitive IMF. Then, a full spectral feature is extracted from the selected fault-sensitive IMF to capture the

joint information between signals measured from two orthogonal directions. This proposed method is both applied to simulated data and to real acquisitions on rotating machinery.

An interesting aspect problem related to EMD is those of mode mixing. A way to answer this problem is named Ensemble Empirical Mode Decomposition (EEMD). Its purpose is the introduction of a controlled amount of white noise to the original EMD. After the addition of this known noise into the raw signal, the signal in the band will have a uniformly distributed reference scale which forces the EEMD to exhaust all possible solutions in the sifting process for minimising mode mixing effect. Even though EEMD has become rather popular, the proper settings for the number of ensemble and the amplitude of white noise that should be added are still not formally prescribed. It must be noticed that the amplitude of added white noise and the number of ensemble are two important parameters. Besides, the performance of EEMD is also affected by the content of the raw signal needed to be analysed. Guo and Tse discuss in [GT10] the influence of various parameters setting on the results in these reducing mode mixing problem. They operate tests both on simulated and on real machine signals. The results they obtain could provide an interesting guideline on setting these parameters properly in order to enhance the ability of EEMD in machine fault diagnosis.

## 2.4 Spectral Kurtosis

Between the various tools to analyse non-stationary signals, Spectral Kurtosis (SK) is a statistical tool which can indicate the presence of series of transients and their locations in the frequency domain. The general idea is that the presence of faults in rotating elements (such as rolling bearings) generates series of impacts producing transient vibration signals. Unluckily these signals are often seriously corrupted by strong levels of background noise, so this problem can be formulated as the detection of transient signals, called  $X(t)$ , in strong additive noise, called  $N(t)$ . The measured signal is then of the form

$$Y(t) = X(t) + N(t)$$

and a model able to describe incipient faults in rolling bearings is the generalised shot noise process.

A possible way to analyse the characterisation of incipient faults is then to consider statistical indicators sensitive to the peakiness of the signal. The kurtosis is a feasible one since, as already explained in Section 1.4.1, it achieves high values in presence of the fault signal  $X(t)$  and an ideal zero value in case of background noise  $N(t)$ . In order to avoid the masking of the signal by the strong vibrations produced from several competing sources which span a large frequency range, the kurtosis could be locally applied in different frequency bands. That is the case of spectral kurtosis presented by Antoni in [Ant06]. Furthermore, in [AR06] the author proposes a SK estimator based on the Short-Time Fourier Transform (STFT) using an analysis window  $w(n)$  whose duration is  $N_w$  that shorter than the mean between two



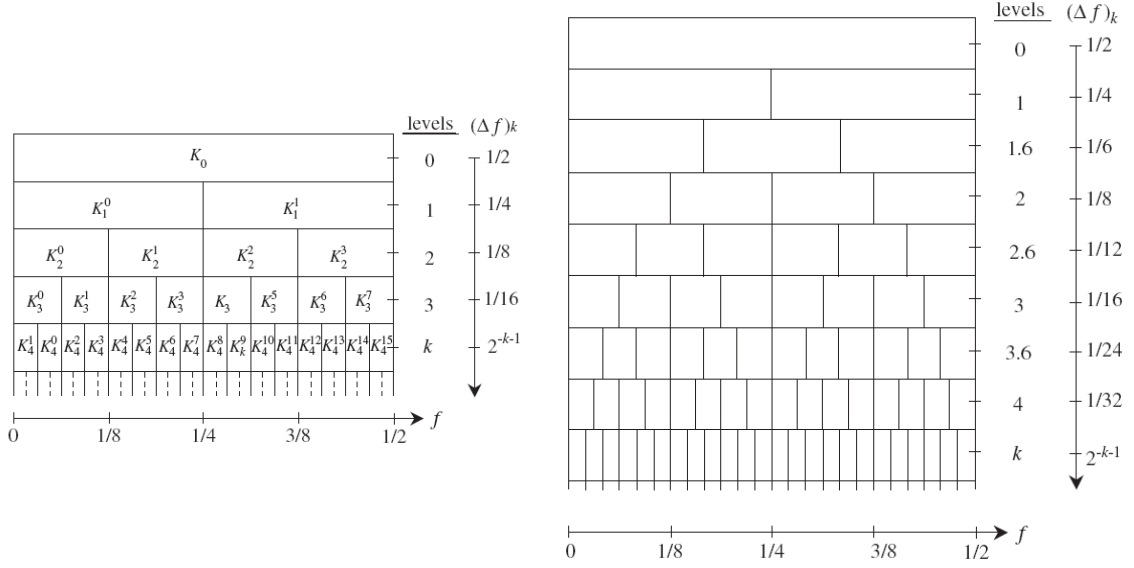


Figure 2.7. Paving of the (frequency/frequency resolution) plane (left) and its extension to the case of the 1/3-binary tree kurtogram estimator (right) (Figure taken from [Ant07]).

consecutive impulses. So, the SK of a vibration signal  $Y(t)$  with sampling frequency  $f_s$  and fault rate of occurrence  $f_d$  becomes approximately

$$K_Y(f) \approx \left( \frac{f_s}{f_d} \gamma_{4w} (\kappa_X + 3) - 2 \right) \frac{1}{[1 + \rho(f)]^2},$$

where  $\gamma_{4w}$  is the time–bandwidth product of the square of the analysis window,  $\kappa_X$  is the intensity of the fluctuations in the impulse amplitudes and  $\rho(f)$  is the noise-to-signal ratio. It is clear that  $K_Y(f)$  increases if the window length  $N_w$  of the STFT decreases and this fact suggests that a shorter  $N_w$  should be preferred.

According to the authors, it is then necessary to create a robust detection filter with an imposed band-pass structure which needs only two parameters to be identified, namely the central frequency  $f_c$  and the bandwidth  $\Delta f$ , or equally the frequency  $f$  and the window length  $N_w$ . The name “kurtogram” is thus coined to indicate the SK map formed as function of  $f$  and  $N_w$  and its maximum values gives the optimal central frequency and bandwidth of the band-pass filter. A further simplification is proposed in [Ant07] and is called *fast kurtogram*. It involves the computation of the kurtosis of the coefficients obtained at the output of filter-banks of quasi-analytic filters with central frequency  $f$  and bandwidth  $\Delta f$ . The fast kurtogram gives a strict method to find automatically the best frequency band able to filter the signal. Moreover, it detects in which frequency band transients take place and it returns the complex envelope for each selected band. Once some transients have been identified thanks to the kurtogram, it is possible to compute the PSD of the filtered signal in order to find out spectral lines related to defect frequencies. Figure 2.7 shows the general paving of the (frequency/frequency resolution) plane and its extension to the case of the

1/3-binary tree kurtogram estimator respectively.

Antoni starts to use Spectral Kurtosis with detection intent in [Ant06], where he proposes a formalisation of the SK itself by means of the Wold-Cramér decomposition of “conditionally non-stationary” processes. This definition then engenders many useful properties enjoyed by the SK. In particular, he establishes to which extent the SK is capable of detecting transients in the presence of strong additive noise by finding a closed-form relationship in terms of the noise-to-signal ratio. He succeeds in proposing a Short-Time Fourier Transform-based estimator of the SK which helps to link theoretical concepts with practical applications. After having demonstrated the high potential of the SK to detect and characterise non-stationary signals, in [AR06] they show how it could be efficiently used in the vibration-based condition monitoring of rotating machines. The first important aspect is that the SK provides a robust way of detecting incipient faults even in the presence of strong masking noise which is of practical importance for monitoring purposes. Moreover, it offers an almost unique way of designing optimal filters for filtering out the mechanical signature of faults, with a great contribution in diagnostics.

They demonstrate these facts through practical industrial cases. The first example of application to bearing surveillance in [AR06] concerns the analysis of rolling element bearing signals originated from a single-stage gearbox at the University of New South Wales where different faults are introduced in the bearings and acceleration signals are measured in each case. Kurtosis values of various signals are actually close to zero in all cases, due to the very strong signals from the gears, which masked everything else. A demonstration of SK ability is given in Figure 2.8, where the difference in dB-spectra corresponding to the analysed signals and their comparison with the estimated SK are displayed. For previous analysis of the authors, it has come out that a reasonable window length is  $N_w = 32 - 64$ , i.e. a spectral resolution of about 750–1500 Hz per line. The difference in dB-spectra represents the difference between the dB-spectrum under normal (non-faulty) condition and the one for a faulty signal. Usually, this should be able to recognise the frequency band where the fault are located and, so, it is widely used in vibration monitoring if some historical data are available [AR06]. These plots show that the SK and the difference in dB-spectra have very similar shapes and, thus, they observe as this parameter could validate the effectiveness in detecting incipient faults. A further confirmation could be found in Figure 2.9, where the same analysis is done for a ball fault. Thanks to these observations, Antoni and Randall state that SK could substitute the difference in dB-spectra when no historical data are available.

Another example is related to the use of the SK in a surveillance programme on a single-stage gearbox submitted to an accelerated fatigue test. Through the great amount of data acquired, they could test the ability of the SK in detecting incipient spall faults. In Figure 2.10 the SK computed for all measurements in chronological order (from number 1 to 160) in the frequency range from 0 to 40 kHz is showed. In this case the spectral resolution of the estimates is 78 Hz per line which must be constant during the whole test. They observe as

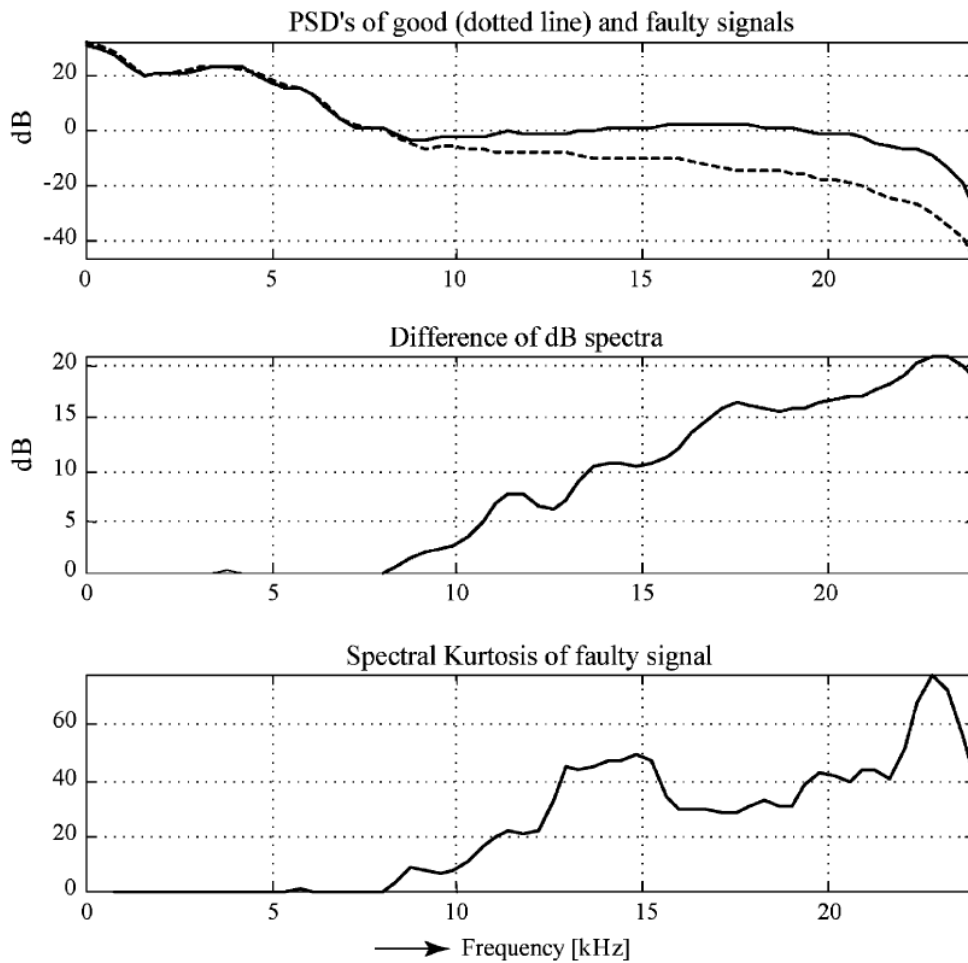


Figure 2.8. PSD, difference in dB-spectra and SK in the case of the outer race fault with spectral resolution 750 Hz per line (Figure taken from [AR06]).

SK takes temporary high values over the whole frequency range on days 1, 6, and 8. Since at the beginning of the test they could be due to the system breaking in, on days 6 and 8 they could arise because of the appearance of incipient spall defects on the gear teeth. It must be noticed that, from three hours before the end of the test, SK starts to have abnormal high values certainly due to excessive spalling. The authors compare these results with those obtained with Power Spectral Density and they could state that it only starts to indicate the appearance of excessive spalling one hour before the test stopping.

A further example of SK in machine diagnostic is presented in [BR09], where Barszcz and Randall apply it to detect a tooth crack in the planetary gear of a wind turbine. The work starts from a real case of catastrophic gear failure on a wind turbine, not noticed by the applied methods. The method they apply is based on SK and shows its ability compared to the various diagnostic techniques that could be applied. In fact, while these detect the

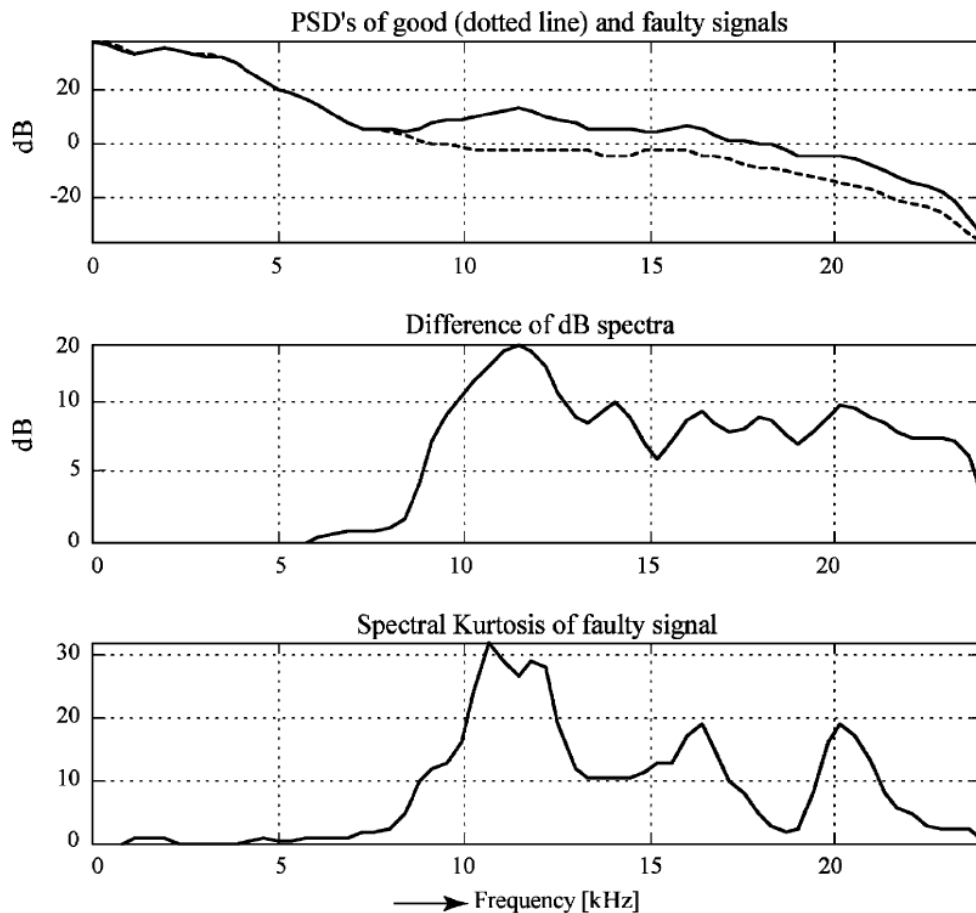


Figure 2.9. PSD, difference in dB-spectra and SK in the case of the ball fault with spectral resolution 750 Hz per line (Figure taken from [AR06]).

failure only few minutes before its appearance, SK is able to identify the existence of the tooth crack several weeks before the gear failure. In [BCIR08], instead, the authors want to find a detection procedure of generalised roughness that could affect roller bearings. Once the bandwidth where the effect of the fault is stronger is identified through SK and kurtogram, the energy of the signal in this bandwidth is used as a diagnostic index. Experiments confirm the effectiveness and reliability of the proposed approach applied to vibration monitoring, with results consistent at various damage levels and rotation speeds. Combet and Gelman propose in [CL09] a methodology for the enhancement of small transients in gear vibration signals in order to detect local tooth faults, such as pitting, at an early stage of damage. They consider an optimal denoising Wiener filter based on the SK estimating. Then, the authors apply it to the gear residual signal which is classically obtained after removing the mesh harmonics with TSA. Various improvements are obtained thanks to this approach over the direct estimation from the raw vibration signal: the signal-to-noise ratio improves, there

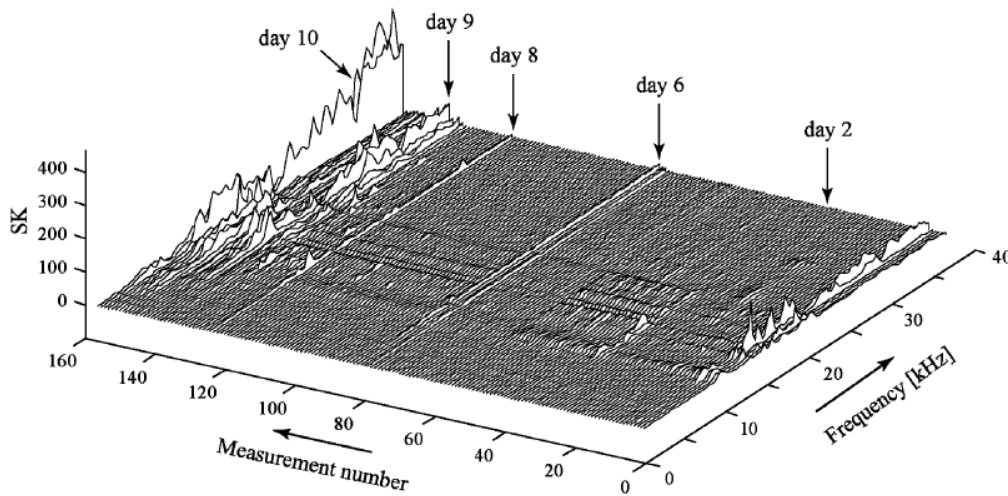


Figure 2.10. SK of measurements on a gearbox submitted to an accelerated fatigue test. Abnormally high values were detected on days 2, 6, 8 and 10. The test was stopped on day 12 due to excessive spall damage (Figure taken from [AR06]).

are reduced interferences from other stages and excited structural resonance(s) within the frequency range of the mesh harmonic components appears to be easier to be detected. The application of this method to an industrial case shows how it is able to detect relatively small tooth surface pitting (less than 10%) in a two-stage helical reduction gearbox.

### Recent applications

According to Lei *et al.*, the concept of kurtogram developed by Antoni in [Ant07] suffers from limited accuracy improvement in extracting transient characteristics from a noisy signal and in identifying machinery fault. In [LLHZ11] more precise filters to be incorporated into the kurtogram method are proposed in order to ameliorate its accuracy in discovering characteristics and detecting faults. In particular, they state that a filter based on WPT is able to filter out noise and match the fault characteristics of noisy signals. So, in this paper they introduce an improved kurtogram method adopting WPT as the filter of kurtogram to have greater performances. The ability of this improved method is verified through the application to vibration signals collected from two kinds of rolling element bearings. The comparison with the original kurtogram shows the effectiveness of their method in extracting fault characteristics and diagnosing faults of rolling element bearings. Another conjoint application of WP and SK is the one developed by Taiyong and Jinzhou in [TJ11]. Their goal is the effective identification of the weak fault characteristic frequencies of roller bearings in presence of strong background noise. The research is based on kurtosis coefficients and the segmentation thresholds noise reduction method by autocorrelation analysis of WPD coefficients. This could improve the signal-to-noise ratio and the composition of high

frequency resonance signal components. Moreover, combined with the spectral kurtosis theory, it determines the parameters of band pass filter. According to the authors, this method can effectively determine the parameters of band-pass filter and allow the extraction of the fault signal characteristic frequency. They propose also some experimental analysis for bearings weak fault diagnosis, which proves that this method is effective and practical. Also Chen *et al.* apply SK and WP together in [CZHJ12]. Since SK has been proved to be a powerful tool to detect and characterise non-stationary signals, they want to compensate for the filter limitation of the SK and improve the resolution of spectral analysis as well as optimally match fault features. They try to do this introducing Adaptive Redundant Multiwavelet Packet (ARMP) into SK. They apply their method to the rolling element bearing and gear fault detection to validate its reliability and effectiveness. The comparison with the conventional frequency spectrum, envelope spectrum, original SK and some single wavelet methods allows them to affirm that their technique could improve the accuracy of frequency-band selection and enhance the ability of rotating machinery fault detection.

Barszcz and Jabłoński state some other limits of SK, arising during their analysis on a signal from a jet engine. In fact, although the fast kurtogram does not yield any significant band, the investigation on the machine revealed severe bearing race defects. They think this behaviour could be due to the low value of signal-to-noise ratio, to the variety of peaks contained into the signal (exciting resonances probably) and to the large characteristic sought frequency. On the whole, the narrowband amplitude demodulation of a vibration signal enables the extraction of components carrying information about rotating machine faults. However, the quality of the demodulated signal depends on the frequency band selected for the demodulation. As demonstrated in [AR06], SK is a very efficient method for detection of such faults, including defective rolling element bearings and gears. According to the author, problems holds, for example, in the presence of a relatively strong, non-Gaussian noise containing high peaks or for a relatively high repetition rate of fault impulses. So, in [BJ11], the selection of the optimal frequency band is done through novel method that attempts to overcome the previous drawbacks. Moreover, Barszcz and Jabłoński introduce a new tool for the presentation of the results that they call *Protrugram*. This method requires some a priori knowledge of the monitored machine kinematics added to visual postprocessing that could reject discrete tones. However, its strengths are its superior detection ability of modulating signals in presence of higher noise and invulnerability to random impulse responses present in the signal. The main difference between the *Protrugram* and the fast kurtogram lies in the different values that are optimised: while the former considers the kurtosis of envelope spectra amplitudes as a function of the centre frequency, the latter uses kurtosis of the time signal filtered in different bands. This new method is validated both on simulated and on real data such as a test rig, a simulated signal and a jet engine vibration signal.

In 2011 Wang and Liang propose, instead, an Adaptive Spectral Kurtosis (ASK) technique for the fault detection of rolling element bearings [WL11]. This method is characterised by

the adaptive determination of the bandwidth and centre frequency which is implemented through successive attempts to right-expand a given window along the frequency axis by merging it with its subsequent neighbouring windows. The main parameters that has to be evaluated in practical application are the initial window function, bandwidth and window overlap on the merged windows. Thanks to simulated experiments, they found that this technique can further enhance the SK-based method as compared to the kurtogram approach. Moreover, experimental signal could assess the effectiveness of the proposed method in fault detection of the rolling element bearings. The same authors present also, in 2012, a multi-fault detection method based on the same ASK but for the analysis of the vibration signal from single sensor [WL12]. They start developing a theoretical model of multiple bearing faults. Then, the comparison with the kurtogram and Protrugram techniques shows that it could extract better the signatures of multiple bearing faults even in the presence of strong background noise. Simulation data and acquisition of a bearing with multiple faults and two faulty bearings helps in validating the performance of the proposed method.

Eftekharnjad *et al.*, instead, propose the application of SK to Acoustic Emission (AE) signature from bearings [EHAM12]. According to the authors, the state that the majority of research works in that field has involved the use of conventional time domain analysis for processing the AE signals from the machines and there has been little attention given to application of more advanced signal processing methods. So, their research presents a study in which some advanced signal processing techniques such as Wavelet and SK have been applied to offer improved diagnosis for bearing defect detection. Other recent works join SK with other signal analysis technique, such as EMD. In particular, Guo *et al.* present a hybrid method based on SK and EEMD for faulty bearing signal recovery from large noise [GTD12]. In fact, in presence of large noise, it may be difficult to recover impulses. So, in their paper, they first filter the raw vibration signal through an optimal band-pass filter based on SK and, then, they apply EEMD method to decompose the filtered signal. Various bearing signals are used to validate the efficiency of the proposed method and the results demonstrate how this processing method can effectively recover the impulses generated by bearing faults from the raw signal also if large noise is present.

## 2.5 Speed and load variations

A rather recently challenging topic is related to the analysis of vibration in case of varying load and speed. In 2009, Bartelmus and Zimroz show that for condition monitoring of planetary gearboxes it is important to identify the external varying load condition. In their paper [BZ09], they analyse the influence that many factors have on the vibration signals generated by a system containing a planetary gearbox. Processing real measured vibration signals obtained in the industrial environment they found that a planetary gearbox in bad condition is more susceptible to load than a gearbox in good condition. They observe the estimated load

time traces obtained through a demodulation process of the vibration acceleration signal for the gearbox in both good and bad conditions. In this paper it is clearly showed how two dominant low-frequency causes could influence vibration signal modulation and how these causes are the varying load itself. In order to better understand these aspects and to verify the correlation between simulation and experimental results, Bartelmus *et al.* develop two models of gearboxes operating under varying load conditions. In fact, in [BCZH10] a fixed-axis two-stage gearbox and a planetary gearbox, based on two mechanical systems used in the mining industry, are studied. They propose a new transmission error function that could be able to detect changes in technical conditions and under load variation. The diagnostic features are energy based parameters such as RMS over the signal or the arithmetic sum of the amplitudes of spectral gearmesh components. The results obtained in various simulation demonstrate how there is a strong correlation between load values, changes in condition and the features themselves. Zimroz and Bartelmus work on complex mining mechanical systems in [ZB12] too. Here, they study the application of an adaptive filtering that could recover weak impulsive signal in order to detect bearings local damages in the case of varying load. This filter is applied as a pre-processor to find impulsive cyclic weak signal that could be present in that raw vibration signals. They state that these periodic and cyclic impulses are related to local faults that produces effects such as impulse, concentric forces and stresses in the kinematic pairs. The problem in their case is that these elements lied to local faults are masked by other signal sources. The authors propose an adaptive system and apply, for the first time, the Normalised Least Mean Square (NLMS) to signals obtained from machines working in industrial condition. Another type of analysis on gearboxes subjected to fluctuating loads and operating speeds is that presented by Heyns *et al.* in [HGdP12]. Starting from the fact that the non-intrusive nature of vibration data enables the monitoring of enclosed systems such as gearboxes, they analyse vibration data generated under fluctuating operating conditions. Moreover, they want to face the problem due to the relatively little prior knowledge that usually analysts have on a specific gearbox. They investigate how to remove the non-fault related components in the structural response of a gear assembly through an adaptive time series model based on Bayesian model selection. In that way, they could obtain a residual signal robust enough when fluctuating operating conditions are present. They generate this residual signal evaluating the difference between the output from a time series regression model and the observed waveform. The regression model considered is based on the weighted average from an assembly of AR models and is iteratively updated through a Bayesian model selection framework. In their analysis, the author obtain the final residual signal through the merging of two signals by point wise selecting each time the smaller of the two. These two have been defined running the time series regression both forwards and backwards. Thus, Heyns *et al.* are able to state that this final residual signal is much more capable of localising fault induced signal irregularities. The authors propose also a statistical framework that could be used to detect the gear state



condition through the analysis of the residual signal itself. Two sets of simulated data are used to verify the proposed methodology. Firstly, a simple dynamic gear model would approximately simulate the structural response of the gear casing under various loading and gear health conditions. Moreover, a single stage gearbox is run until destruction under constant speed and loading and data are acquired. In both cases, the results obtained by the authors provide an intuitive understanding of the nature of the gear damage. In particular, from the simulated data, they could say that the proposed methodology is able to recognise the gear damage both in stable and in under significantly fluctuating operating conditions.

A further application presented by many research group is related to wind turbines. Is it well known as they has become widely used for the electric power generation and, due to their wide diffusion, it has become very important and interesting their condition monitoring. Moreover, the complexity of the structure itself with a multitude of gears and bearings gives more importance to the diagnostic aspects. In fact, the different signals need usually to be separated from each other in order to locate the source of a problem. Moreover, these structures differ from many other gearboxes since they operate with a widely varying load over time periods corresponding to individual analysis records. This aspect is rather important because gives particular problems with gear vibrations, where the effects of load have to be separated from those of condition. Moreover, the most efficient turbines operate with widely varying speed as well, and this has to be taken into account in the analysis procedures too. In [ZBBU11] the authors concentrate on the wind turbine main bearing diagnosis, being an example of machine working under non-stationary load. In fact, as already state, vibration signals acquired from bearings working under varying load and speed condition make the diagnosis difficult. In general, problems originates from the variation of diagnostic features due to these variation, to the low energy of sought features and to the high noise levels. In this paper, Zimroz *et al.* propose a diagnostic approach for main rotor bearings used in wind turbines. In particular, they considers two vibration parameters such as the peak-to-peak vibration acceleration and the generator power that is related to the operating conditions. Data covering several months, including the replacement of the bearing because of its failure and the installation of the new one, could be analysed by the authors. They observe as the data are rather variable and so a decision making regarding the condition of bearings is pretty difficult. So, they consider an approached based on the analysis of healthy and damaged data in a 2-dimension space, where the dimensions are feature are load. Zimroz *et al.* are, then, able to demonstrate the ability of their approach in successfully classify data according to the belonging to the damaged and undamaged bearings group. Randall *et al.* discuss a number of ways of compensating for speed and load variation in [RSM12], through signals obtained from a round robin study of naturally occurring faults in a wind turbine transmission. Since the gearbox is mounted on a test rig for the measurements, the speed variation is lower compared to that measured in practise. However, they could analyse larger speed and load variations using signals from a simpler gearbox on another test rig and through another

wind turbine. Their analysis starts with the order tracking, based on the pseudo-encoder signals, and proceeds with the extraction of the synchronous averaging of the gear signals through the whole gear train. Moreover, they compare also the spectra in the high frequency range and the corresponding cepstra, for faulty and normal condition. Interesting results are obtain in order to extract speed and load information from the acceleration signal.

As could be imagined, the speed is also a condition that influences vibrations, so it is could be a challenging task the creation of diagnostic algorithms able to proper detect damage. In fact, when the speed of rotational speed of a motor changes continuously, it is rather difficult, for example, to do proper diagnostics of the ball bearings. As known, unfortunately, the large part of algorithms proposed in the literature need a constant rotation frequency of the motor to identify fault frequencies in the spectrum. Some solutions are proposed by Cocconcelli *et al.* in various papers. In [CR11], for instance, they present a review of some techniques for the diagnostics of ball bearing based on the experience of the authors on an industrial application. They treat the case of direct-drive motors, whose advantage stands in the removal of cams or gearboxes afterwards the motor, with a consequent reduction of economic and maintaining costs. However, the bearing diagnostics in these cases is rather difficult. In fact, as already said, most of the literature techniques are based on the search of fault-characteristic frequencies in the vibration spectrum with a linear dependence on the rotational frequency of the shaft. So, the speed must be supposed to be constant. Unfortunately, in this kind of motors it changes continuously and the idea of fault frequencies is meaningless. The authors make these considerations because they treat a specific industrial case of bearing diagnostics in a packaging machine working under variable speed condition. Between the various techniques they present, an improved version of the order tracking, the use of correlation function and Wavelet Analysis are included. Moreover, also two supervised learning approaches such as ANN and SVM are considered. The former of these two techniques is used in [CRZB11]. Cocconcelli *et al.*, in fact, propose a method for the diagnostics of ball bearings in direct-drive motors by means of ANN after having proved to be a powerful tool to diagnose the health of ball bearing even in variable-speed applications. In [CZRB12b] direct-drive motors are considered too, in this case the authors focus on signal segmentation using the position profile of the shaft as trigger. The position is directly measured by an encoder. Each machine cycle is then analysed in time domain and, using the encoder, the signal machine contribution is deleted. STFT is then used to extract the features in order to detect the damage, considering the averaging in time-frequency domain of the STFT for each cycle to enhance fault signature. The damage indicator is so given by the sum of STFT coefficients. In [CZRB12a], instead, Cocconcelli *et al.* suggest to combine the information gathered by the already known SK and by the energy distribution to obtain an automatic selection of a filtering band. In this band, should be, then, possible to extract from the STFT map the most informative component in time domain, reducing the complexity of the output to a mono-dimensional vector. A possible fault indicator is then given by the

scalar value obtained if the output exceed a given threshold.

An innovative application related to the speed variation is that presented in [BPRR12]. The authors start from the fact that the transmission path given by the excitation to the measured vibration on the surface of a mechanical system could introduces a distortion both in amplitude and in phase. Moreover, in case of variable speed conditions, the transfer function of the mechanical system contribute to their distortion with a variation in time. In this way, the order tracking usually operated with a tachometer has reduced effectiveness and, consequently, a discrete-random separation performed by a synchronous averaging is not reliable any more. The innovation presented in this paper stands in the identification, both in the time domain and in the order spectrum of the signal, of extent of the distortion. The important consequences related to rotating machinery diagnostics are then highlight. The main domain of their application is gears. Along their work, they express quantitatively the effect of the transfer function of the mechanical system on the spectrum of the order tracked vibration. Moreover, they demonstrate how the identification of bands could be computed in closed form. This is rather important in bearing and gears fault detection, since a good separation of deterministic and random components is fundamental. In fact, those bands are an important indication for the user because they helps in localising the biases introduced by the transmission path on the spectra of both the deterministic part and residual [BPRR12]. An experimental example allows to demonstrate that these bands can be efficiently used in gear diagnostics when the goal is to extract the phase shift information, where apply then Phase Domain Averaging or Improved Synchronous Averaging. In general, the innovation is also for the programmer of an automatic algorithm for gear or bearing diagnostics. In fact, it is not needed any more the user intervention in filters and parameters definition for the signal processing.

## Chapter 3

# Anomaly detection

The concept of “anomaly detection” refers to the identification of patterns in a given data set that does not conform to an established normal behaviour. These non-conforming patterns are often referred to as anomalies, outliers, discordant observations or exceptions. In particular, “anomalies and “outliers” are the two terms most commonly used in the context of anomaly detection and, sometimes, they are also interchangeably. The detection of outliers or anomalies in data has been studied in the statistics community as early as the XX century. Over time, a variety of anomaly detection techniques has been developed in several areas, including statistics, machine learning and data mining. Many of these methods have been specifically developed for certain application domains, while others are more generic. Some examples of applications can be fraud detection, ecosystem disturbances in natural world, intrusion detection in computer networks, test analysis in medicine.

### 3.1 Anomaly detection

As said previously, anomaly detection refers to the detection of patterns in data that do not conform to a well defined notion of normal behaviour. An example of anomaly is shown in Figure 3.1 where a 2-dimensional data set is presented. The two data set, named as  $N_1$  and  $N_2$ , are considered regions of normal behaviour since most observations lie in these two regions. Points  $o_1, o_2$ , and region  $O_3$  too, are labelled as anomalies because they lie sufficiently far away from the regions  $N_1$  and  $N_2$ .

Causes of anomalies might be induced in the data for a variety of reasons. The main underlying causes are often unknown as the several sources that give rise to anomaly too. For example, if an object belongs to a different type or class respect to the totality of data, it can be considered as anomalous. This fact is stated in the often-quoted definition of an outlier given by the statistician Douglas Hawkins [Haw80]: “an outlier is an observation that differs so much from the other observations as to arouse suspicion that it was generated by a different mechanism”. Some examples are fraud, intrusion and abnormal test results. These

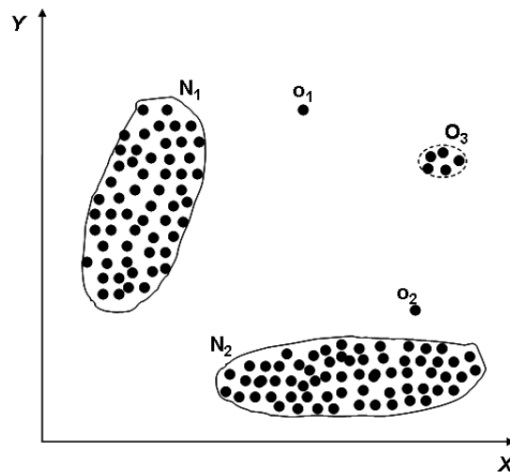


Figure 3.1. Example of anomalies in a 2D data set (Figure taken from [CBK09]).

anomalies are usually the focus of anomaly detection in the field of data mining. Sometimes, errors can occur during data collection and measurement. This is another source of anomaly, due to human error, to the presence of noise or to problem with the measuring device. In this case, the quality of data is reduced as well as the subsequent data analysis. A type of data processing, named data cleaning, deals with the removal of these anomalies. An interesting type of anomaly is represented by extremes or unlikely variations. This is the case, for example, of a tall person in a group: he is anomalous in one of his characteristics (height) respect all others objects in the same class. Formally speaking, we speak about natural variation because most of the objects lie near a centre and the likelihood that an element is different from this average object is small.

The aspects presented here can help in understanding that the definition of the representative normal region is a key challenge. Firstly, the boundary between normal and outlying behaviour is often not so precise. In addition, in many domains normal behaviour keeps evolving so that a current notion of “normal behaviour” might not be sufficiently representative in the future. Moreover, the exact notion of an outlier is different for each application domains and, as said before, data might contain noise. Another key aspect of any anomaly detection technique is the nature of the input data. In fact, each input data instance can be described using a set of attributes (also referred to as variable, characteristic, feature) that can be of various kind, such as binary, categorical or continuous. If the data instance analysed consists of only one attribute it is said univariate, otherwise multivariate. The nature of attributes determines the applicability of anomaly detection techniques. Furthermore, input data can also be categorised considering the relationship present among data instances.

Another important aspect of anomaly is its nature. Three general categories could be defined related to this feature: point, contextual and collective anomalies. The first case

refers to an individual data instance anomalous with respect to the rest of data. This is the simplest type of anomaly and it is the focus of majority of researches on anomaly detection. If a data instance is instead anomalous in a specific context but not otherwise, it is named as a contextual anomaly. The last one is the case of a collection of related data instances anomalous with respect to the entire data set. Moreover, different definitions and approaches to anomaly can be provided from various anomaly detection techniques. Three general groups can be considered: model-, proximity- and density-based techniques. In the first case the anomaly detection technique provide a model of the data and the anomalies are those objects that do not fit the model very well. For example, if we consider a set of clusters, the anomaly is an element that do not belong to any of them. If we see anomalous and normal objects as two different classes, classification techniques can be used to build models of the classes. If it is possible to define a proximity measure between objects a proximity-based technique can be considered. In this case, anomalies are represented by those objects distant from most of the other elements. The last approach refers to the estimate of the density of objects in a data set. It is a rather easy computation and it allows to identify anomalies as objects belonging to region of low density and relatively distant from neighbours.

An important aspect for any anomaly detection technique is the manner in which the anomalies are reported as *output*. This can be done typically considering scores or labels. The scoring techniques assign an anomaly score to each instance in the test data depending on the degree to which that instance is considered an anomaly. In this way we have a ranked list of anomalies as output. In the other case a label (normal or anomalous) is assigned to each test instance. The concept of label is also tied to the different approaches that can be held toward anomaly detection and, in particular, to three broad categories of anomaly detection techniques that exist.

## 3.2 Classification based techniques

Classification based anomaly detection techniques operate as a more general classification: after a first step when a model is learnt from a set of labelled data object, a new instance is classified into one of the classes defined from the model. These two phases are named training and testing respectively. If we talk about anomaly detection techniques we use the available labelled training data to built the classifier that would allow us to state if a new test instance is normal or anomalous [TSK05].

A way to denote if an instance is normal or anomalous is to assign it a label. In general, it is quite expensive to obtaining labelled data both accurate and representative of all types of behaviour. The labelling operation is often done manually by a human expert and hence requires substantial effort to obtain the labelled training data set. It is also rather difficult to get a proper labelled set of anomalous data which could cover all possible type of anomalous behaviour. In addition, new types of anomalies might arise, for which there

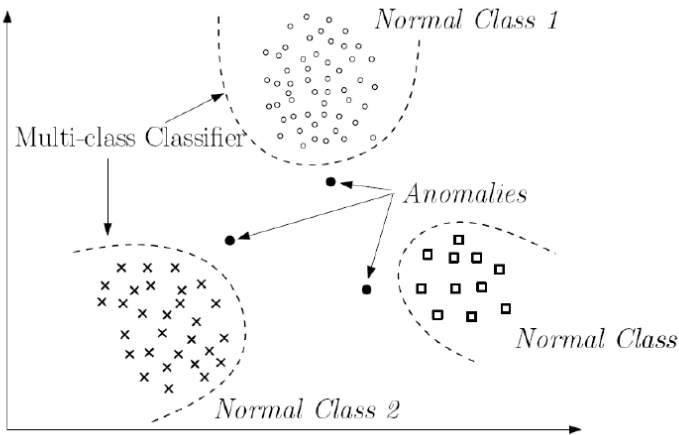


Figure 3.2. Example of multi-class classification (Figure taken from [CBK09]).

is no labelled training data. Three broad categories of anomaly detection techniques can be treated relatively to the extent to which the labels are available: unsupervised, supervised and semi-supervised.

*Unsupervised anomaly detection* refers to lots of practical situation in which class labels are not available. Techniques that operate in this mode do not require training data, so they are most widely applicable. The objective is to assign a label to each instance that reflects the degree to which this is anomalous. It must be noted that the implicit assumption that normal instances are far more frequent than anomalies in the test data is done. In the case that this assumption is not true such techniques suffer from a rather high false alarm rate. If there are many anomalies that are similar to each other they could be labelled as normal. This means that anomalies must be distinct from one another, as well as normal instances, in order to have a successful detection.

In the case that training data contain labelled normal data but no information about the anomalous instances is available, we talk about *semi-supervised anomaly detection*. The goal is to find an anomaly label for a set of given objects by using information obtained from labelled normal data. As opposed to the previous case, if many related outliers are present in the set that has to be labelled no impact on their evaluation is expected. The typical approach used in such techniques is to build a model for the class corresponding to normal behaviour and use this model to identify anomalies in the test data. It must be noticed that lots of semi-supervised techniques can be adapted to operate in an unsupervised mode by using a sample of the unlabelled data set as training data. However, such adaptation assumes that the test data contains very few anomalies and the model learnt during training is robust to these few anomalies. In the case of this kind of technique we can see that while models can be easily understood and the normal behaviour can be accurately learnt, a high false alarm

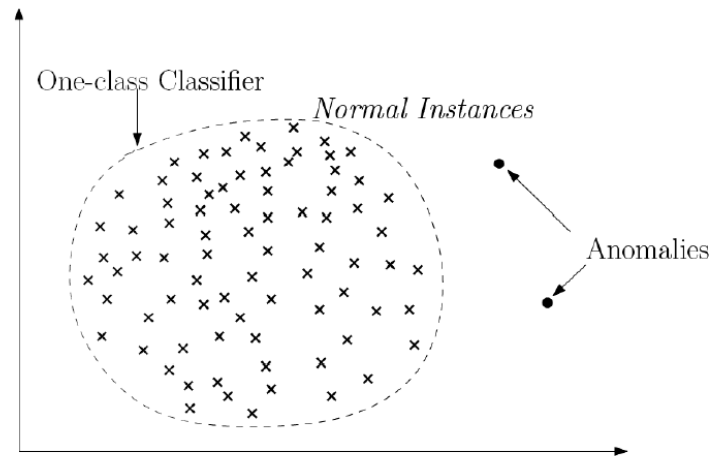


Figure 3.3. Example of one-class classification (Figure taken from [CBK09]).

rate is possible. In fact, previously unseen data records may be recognised as anomalies.

In the previous cases we dealt with situation where class labels are available only for a class of data or, at worst, not available at all. If, instead, a labelled training set containing both anomalous and normal objects is present, we talk about *supervised anomaly detection*. In this case a classifier is trained, that is, a predictive model for normal and anomaly classes is built. When any unseen data instance comes out, it is compared against the model to determine which class it belongs to. Two issues arise in particular in supervised anomaly detection. A first aspect regards the fact that anomalous instances are far fewer compared to the normal instances in the training data. This problem is widely been addressed in the data mining and machine learning literature. Moreover, another usually challenging aspect is the attainment of accurate and representative labels, especially for the anomaly class. If we want to state some advantage of this technique we can say that this problem is similar to building predictive models that can be easily understood with a rather high accuracy in detecting many kinds of known anomalies. On the contrary, instead, both normal and anomaly class are required and unknown and emerging anomalies can not be detected.

As we have seen, classification is based on the labelling of data, being them both training and testing. Considering labels available for training data instances, classification based anomaly detection techniques can be grouped into two broad categories: multi-class and one-class anomaly detection techniques. The main difference between these two approaches is the number of classes that could be labelled during the training. In fact, *multi-class classification* based anomaly detection techniques assume that the training data contains labelled instances belonging to multiple normal classes. In this way, a classifier is learnt to distinguish between each normal class against the rest of the classes. It means that a new test instance



is considered anomalous if it is not classified as normal by any of the classifiers (Figure 3.2). In some cases a confidence score is associated together with the prediction made by the classifiers and if none of them is confident in classifying the test instance as normal, this one is anomalous labelled. If, instead, all the training instances belong to a single class, we talk about *one-class classification* based anomaly detection techniques. The core of most of this kind of techniques learn a discriminative boundary around the training normal instances using a one-class classification algorithm. In the case of any new test object, if it does not fall within the learnt boundary, then it is labelled as anomalous. An example of this procedure is presented in Figure 3.3.

Classification based techniques, and especially the multi-class techniques, have the great advantages that powerful algorithms able to distinguish between instances belonging to different classes could be used. Moreover, the testing phase is a fast procedure because each test object needs to be compared against the pre-computed model. Unfortunately, some drawbacks must be considered when we use some classification based techniques. In fact, if we talk about multi-class techniques, we must remember that they rely on the availability of accurate labels for various normal classes and often this is not possible.

### 3.2.1 Neural Network based

The term “neural network” has its origins in attempts to find mathematical representations of information processing in biological neural systems ([RM86], [Bis06]). Indeed, it has been used very broadly to cover a wide range of different models and, from the perspective of practical applications of pattern recognition it could be seen as an efficient models for statistical pattern recognition, in particular the *multilayer perceptron*.

The basic neural network model could be described as a series of functional transformations. First we construct  $M$  linear combinations of the input variables  $x_1, \dots, x_D$  in the form

$$a_j = \sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^{(1)} \quad j = 1, \dots, M$$

where the superscript <sup>(1)</sup> indicates that the corresponding parameters are in the first “layer” of the network.

The parameters  $w_{ji}^{(1)}$  are usually known as *weights*, the parameters  $w_{j0}^{(1)}$  as *biases* and the quantities  $a_j$  are known as *activations*. Each of them is then transformed using a differentiable, nonlinear activation function  $h(\cdot)$ , usually sigmoidal functions such as the logistic sigmoid, to give

$$z_j = h(a_j),$$

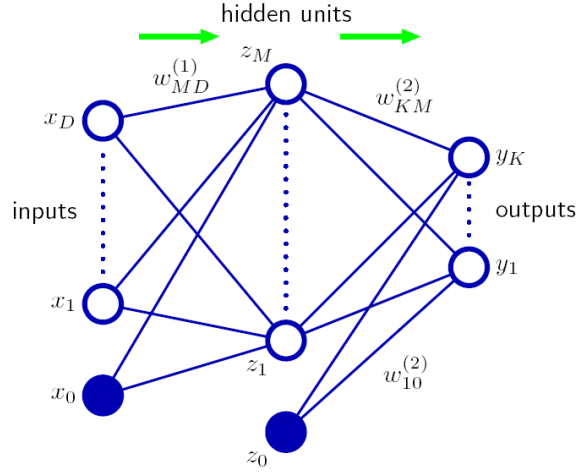


Figure 3.4. Network diagram for the two-layer neural network corresponding to (3.1) (Figure taken from [Bis06]).

which is named *hidden unit*.

A further linear combined of this values gives *output unit activations*

$$a_k = \sum_{j=1}^M w_{kj}^{(2)} z_j + w_{k0}^{(2)} \quad k = 1, \dots, K$$

with  $K$  total number of outputs.

The latter transformation corresponds to the second layer of the network, and again the  $w_{k0}^{(2)}$  are bias parameters. Finally, the output unit activations are transformed using an appropriate activation function to give a set of network outputs  $y_k$ . The critical point of the choice of activation function is determined by the nature of the data and the assumed distribution of target variables. If we consider a multiple binary classification problems, each output unit activation is transformed using a logistic sigmoid function so that

$$y_k = \sigma(a_k), \quad \text{where } \sigma(a) = \frac{1}{1 + e^{(-a)}}.$$

The combination of all this stages produces the overall network function:

$$y_k(\mathbf{x}, \mathbf{w}) = \sigma \left( \sum_{j=1}^M w_{kj}^{(2)} h \left( \sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^{(1)} \right) + w_{k0}^{(2)} \right), \quad (3.1)$$

where the vector  $\mathbf{w}$  contains the set of all weight and bias parameters.

In this way, the neural network model is simply a nonlinear function from a set of input variables  $x_i$  to a set of output variables  $y_k$  controlled by a vector  $\mathbf{w}$  of adjustable parameters. This function can be represented in the form of a network diagram as shown in Figure 3.4.

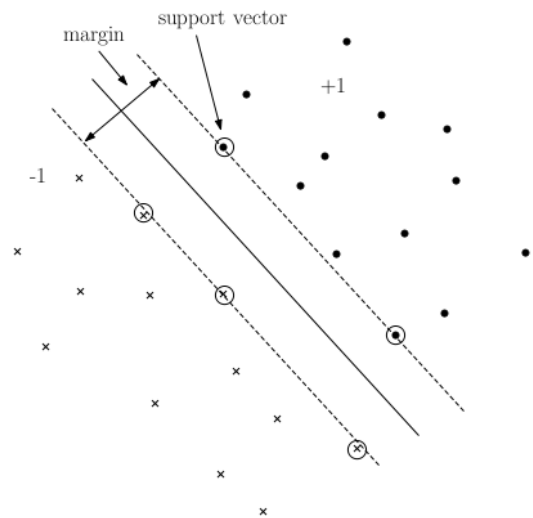


Figure 3.5. SVM representation of the two classes (labelled “+1” and “-1”), the maximum margin and the support vectors.

Here, the input, hidden, and output variables are represented by nodes, and the weight parameters are represented by links between the nodes, in which the bias parameters are denoted by links coming from additional input and hidden variables  $x_0$  and  $z_0$ . The arrows denote the direction of information flow through the network during forward propagation. In fact, the process of evaluating (3.1) could be seen as a forward propagation of information through the network.

The neural network approach can be applied to anomaly detection in multi-class as well as one-class setting. In the case of multi-class two steps are provided: after that a neural network is trained on the normal training data to learn the different normal classes, each test instance is provided as an input to the neural network. At this point two results are possible: the network accepts the test input, so it is considered normal, or it rejects it, so it is an anomaly. Chandola *et al.* [CBK09] present a table with some examples of classification based anomaly detection techniques using neural networks going from multi-layered perceptrons to radial basis function based algorithm.

### 3.2.2 Support Vector Machines based

Support vector machine (SVM) is a computational learning method developed by Vapnik in the 80s and based on the statistical learning theory, rather suitable in case of classification and regression of a large set of data [Vap82]. An input data  $\mathbf{x}$  in an  $n$ -dimensional space is given and it is made up of a number  $M$  of samples, each of these belonging to a class, namely a positive or a negative one. The goal of SVM is the construction of a hyperplane that has to separate the two classes trying to place a boundary between them. Moreover, this boundary has to be oriented such that its distance from the nearest data points in each class is maximal.

In this way, an optimal separating hyperplane is created, namely the *maximum margin* (Figure 3.5). Moreover, the points in both classes nearest to this margin are called *support vectors* and, once selected, they contain all the information necessary to define the classifier.

If we consider a data input  $\mathbf{x}_i$ , with  $i = 1, \dots, M$  and  $M$  number of samples, these samples are assumed belonging to the positive class, labelled  $y_i = 1$ , or to the negative class, labelled  $y_i = -1$ . Thus, in case of linear data, the hyperplane that separates the data belonging to the two classes is expressed as

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b = 0,$$

where the  $M$ -dimensional vector  $\mathbf{w}$  and the scalar  $b$  are the parameters that define exactly its position.

So, the decision function, necessary to create the separating hyperplane between the positive and the negative class, is made using  $\text{sign}f(x)$ . Moreover, the separating hyperplane should be satisfy the constraints

$$\begin{aligned} f(x_i) &= 1 & \text{if } y_i &= 1, \\ f(x_i) &= -1 & \text{if } y_i &= -1, \end{aligned}$$

or, as a complete equation

$$y_i f(\mathbf{x}_i) = y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad i = 1, \dots, M.$$

The nearest data points lie instead on the planes  $f(\mathbf{x}) = \pm 1$  and so the margin between them is  $2\|\mathbf{w}\|^{-1}$ . The goal is then to maximise this one and it can be done through the following optimisation problem:

$$\begin{aligned} &\text{minimise} && \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^M \xi_i \\ &\text{subject to} && \begin{cases} y_i(\mathbf{w} \cdot \mathbf{x} + b) \geq 1 - \xi_i & i = 1, \dots, M \\ \xi_i \geq 0 & i = 1, \dots, M \end{cases} \end{aligned}$$

where  $\xi_i$  (named *slack variables*) measure the distance between the margin and the points  $x_i$  lying on the wrong side of the margin itself,  $C$  is the error penalty and  $y_i \in -1, +1$  are the negative and positive class labels respectively.

It is possible to simplify this problem using the Kuhn-Tucker condition that convert it into the equivalent Lagrangian dual problem [Bis06]

$$\text{minimise} \quad L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^M \alpha_i y_i(\mathbf{w} \cdot \mathbf{x}_i + b) + \sum_{i=1}^M \alpha_i. \quad (3.2)$$

So, the goal is the minimisation of Equation (3.2) with respect to  $\mathbf{w}$  and  $b$ , while requiring the derivatives of  $L$  to  $\alpha$  to vanish. At optimal point, we have the following saddle-point

equations [WY07a]:

$$\frac{\partial L}{\partial \mathbf{w}} = 0, \quad \frac{\partial L}{\partial b} = 0,$$

which is replaced by

$$\mathbf{w} = \sum_{i=1}^M \alpha_i y_i \mathbf{w}_i, \quad \sum_{i=1}^M \alpha_i y_i = 0. \quad (3.3)$$

Equation (3.3) means that  $\mathbf{w}$  is contained in the subspace spanned by the  $\mathbf{x}_i$  and if it is substitute in Equation (3.2), the following dual quadratic optimisation problem is obtained:

$$\begin{aligned} \text{maximise } L(\alpha) &= \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j=0}^M \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \\ \text{subject to } \alpha_i &\geq 0, \quad i = 1, \dots, M, \quad \sum_{i=1}^M \alpha_i y_i = 0 \end{aligned}$$

where  $\alpha_i$  are the Lagrangian multipliers, necessary to define the following decision function

$$f(\mathbf{x}) = \text{sign} \left( \sum_{i,j=1}^M \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{x}_j) + b \right). \quad (3.4)$$

In case of non-linear classification, SVM can also be applied mapping the data to be classified onto a high-dimensional feature space, where it is then possible to apply the linear classification. A non-linear vector function such as  $\Phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_l(\mathbf{x}))$  could solve the problem, so that the decision function (3.4) becomes

$$f(\mathbf{x}) = \text{sign} \left( \sum_{i,j=1}^M \alpha_i y_i (\Phi^T(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)) + b \right). \quad (3.5)$$

Furthermore, if a kernel function  $K(\mathbf{x}_i, \mathbf{x}_j) = (\Phi^T(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j))$  is applied, it is not necessary to evaluate  $\Phi(\mathbf{x})$  explicitly in the feature space, reducing computational problems. So, in the decision function (3.5) the term  $y_i (\Phi^T(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j))$  could be replaced by  $K(\mathbf{x}_i, \mathbf{x}_j)$ , which is easier to evaluate. There are various possible choices of kernel function:

- linear  $K(\mathbf{x}, \mathbf{x}_j) = (\mathbf{x}^T \mathbf{x}_j)^d$
- polynomial  $K(\mathbf{x}, \mathbf{x}_j) = (\gamma \mathbf{x}^T \mathbf{x}_j + r)^d, \gamma > 0$
- Gaussian RBF,  $K(\mathbf{x}, \mathbf{x}_j) = \exp(-\|\mathbf{x} - \mathbf{x}_j\|^2 / 2\gamma^2)$

This fact is rather outstanding because it enables SVM to be used in case of very large feature spaces because the dimension of classified vectors does not influence directly the SVM performance.

The Support Vector Machine as presented is fundamentally a two-class classifier. However, often it happens that problems involves more than two classes. Two possible approaches

could be considered to combine multiple two-class SVMs in order to build a multiclass classifier: One-against-all (OAA) and One-against-one (OAO). The first one is the earliest implementation and it constructs  $K$  separate SVMs, where  $K$  is the number of various classees. Then, the  $i$ -th SVM model is trained using the data from one class as the positive examples and the data from the remaining  $K - 1$  classes as the negative examples. In this way, given  $l$  training data  $(x_1, y_1), \dots, (x_l, y_l)$ , where  $x_i \in \mathbb{R}^n, i = 1, \dots, l$  and  $y_i \in \{1, \dots, K\}$  is the class of  $x_i$ , the  $i$ th SVM solve the following problem [WY07a]:

$$\begin{aligned} & \text{minimise} \quad \frac{1}{2} \|\mathbf{w}^i\|^2 + C \sum_{i=1}^M \xi_j^i (\mathbf{w}^i)^T \\ & \text{subject to} \quad \mathbf{w}^i \cdot \phi(\mathbf{x}_j) + b^i \geq 1 - \xi_j^i \quad \text{if } y = i, \\ & \quad \quad \quad \mathbf{w}^i \cdot \phi(\mathbf{x}_j) + b^i \leq -1 + \xi_j^i \quad \text{if } y \neq i, \\ & \quad \quad \quad \xi_j^i \geq 0, \quad j = 1, \dots, l, \end{aligned}$$

where the training data  $\mathbf{x}_i$  is mapped to a higher-dimensional space by function  $\phi$  and  $C$ , is the penalty parameter.

The other approach, OAO, proposes the construction and the training of  $K(K - 1)/2$  different 2-class SVMs on all possible pairs of classes. To train data from the  $i$ th and the  $j$ th classes, the following binary classification problem has to be solved:

$$\begin{aligned} & \text{minimise} \quad \frac{1}{2} \|\mathbf{w}^{ij}\|^2 + C \sum_t \xi_t^{ij} (\mathbf{w}^{ij})^T \\ & \text{subject to} \quad \mathbf{w}^{ij} \cdot \phi(\mathbf{x}_t) + b^{ij} \geq 1 - \xi_t^{ij} \quad \text{if } y_t = i, \\ & \quad \quad \quad \mathbf{w}^{ij} \cdot \phi(\mathbf{x}_t) + b^{ij} \leq -1 + \xi_t^{ij} \quad \text{if } y_t = j, \\ & \quad \quad \quad \xi_t^{ij} \geq 0, \quad j = 1, \dots, l. \end{aligned}$$

Then, various approach could be considered to decide which class a data instance belongs to. An intuitive approach to get the global classification result is to consider outputs of the binary classifiers as votes and select the class that gets most votes. Another method linearly couples the outputs of the classifiers with the mixture matrix created in the training phase to minimise the error between the correct class decision and the linear combination of outputs of the binary classifiers. This approach weights the votes given by classifiers so that the classification is more likely correct compared to the previous one.

In particular, SVM techniques have been applied to anomaly detection in the one-class setting: a region that contains the training data instances is learnt in order to create a boundary. This can be done using kernels, such as polynomial or radial basis function (RBF),

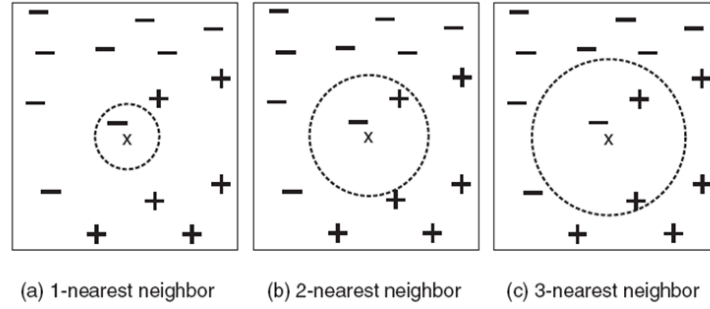


Figure 3.6. The 1-, 2- and 3-nearest neighbours of an instance (Figure taken from [TSK05]).

in case of more complex regions. For each test object, if it falls within the learnt region it is declared as normal, otherwise as anomalous. These aspects will be developed better in Chapter 5.

### 3.3 Nearest Neighbour based techniques

In classification, the training set of data contains both attributes  $\mathbf{x}$  and a class label  $c$ . When a new dataset of training data points is given, the goal is to return the correct class  $c(\mathbf{x})$  of a novel  $\mathbf{x}$ . The concept behind Nearest Neighbour is that elements similar to  $\mathbf{x}$  (in  $\mathbf{x}$ -space) should have the same class label and so, given a text example, its proximity to the rest of data points in the training set must be evaluated. The key point is then the choice of the measure of similarity (or dissimilarity). This is usually done through a *dissimilarity function* between the data points  $\mathbf{x}$  and  $\mathbf{y}$ , or also *distance*, which could be computed in different ways [TSK05]. One of the most used measure of dissimilarity is based on the Euclidean distance between the vector representation of data points:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2},$$

where  $n$  is the number of dimensions,  $x_k$  and  $y_k$  are the  $k^{\text{th}}$  components of  $\mathbf{x}$  and  $\mathbf{y}$  respectively.

The generalisation of this measure is the Minkowski distance metric:

$$d(\mathbf{x}, \mathbf{y}) = \left( \sum_{k=1}^n |x_k - y_k|^r \right)^{1/r},$$

with  $r$  parameter.

The three most common examples of Minkowski distance are the case of  $r = 1$ , also known as city block distance,  $r = 2$ , the already cited Euclidean distance, and  $r = \infty$ , named

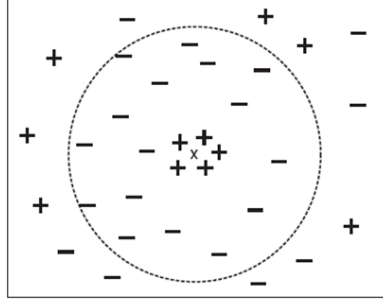


Figure 3.7.  $k$ -nearest neighbours with large  $k$  (Figure taken from [TSK05]).

supremum distance and defined as

$$d(\mathbf{x}, \mathbf{y}) = \lim_{r \rightarrow \infty} \left( \sum_{k=1}^n |x_k - y_k|^r \right)^{1/r}.$$

Having defined as  $d(\mathbf{x}, \mathbf{y})$  the distance between two points  $\mathbf{x}$  and  $\mathbf{y}$ , some properties holds:

- Positivity:  $d(\mathbf{x}, \mathbf{y}) \geq 0 \quad \forall \mathbf{x}, \mathbf{y}, \quad d(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}$
- Symmetry:  $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$
- Triangle inequality:  $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}) \quad \forall \mathbf{x}, \mathbf{y}, \mathbf{z}.$

When a measure satisfies all three properties it is known as metric.

The concept of nearest neighbour analysis has been used in several anomaly detection techniques. In this context, it must be considered the assumption that normal data instances occur in dense neighbourhoods, while anomalies occur far from their closest neighbours. Nearest neighbour based anomaly detection techniques can be broadly grouped into two categories: techniques that use the distance of a data instance to its  $k^{\text{th}}$  nearest neighbour as the anomaly score and techniques that compute the relative density of each data instance to compute its anomaly score. The first classification procedure could be understood looking at Figure 3.6, where the 1-, 2- and 3-nearest neighbours of a data point located at the centre of each circle are wanted. The point is classified considering the class of its neighbours and if they have more than one label, the point is assigned to their majority class. Following this consideration, the data point in Figure 3.6(a) is assigned to the negative class, while in Figure 3.6(c) is classified as positive. When a situation as Figure 3.6(b) occurs, that is, there is a tie between the classes, the final choice is taken through a random choice. Thus, it is very important to choose a proper value for  $k$  because there is the risk of overfitting due to noise, if  $k$  is too small, and of misclassification, if  $k$  is too large.

Nearest neighbour based techniques present various consistent advantages. The first point is that they are unsupervised in nature and do not make any assumptions regarding



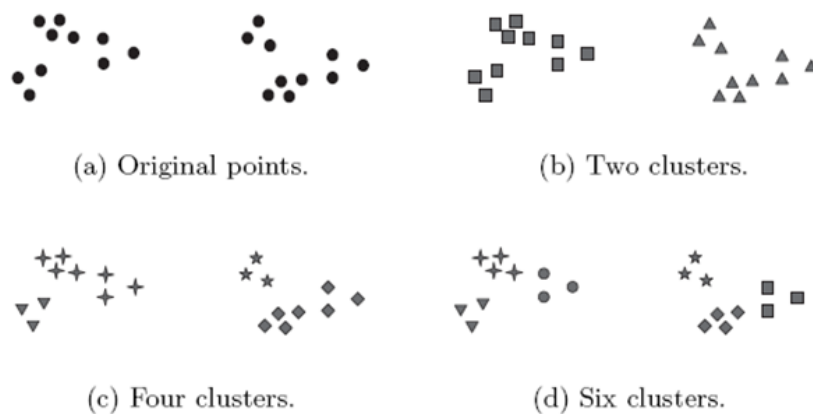


Figure 3.8. Different way of clustering the same set of points (Figure taken from [TSK05]).

the generative distribution for the data being purely data driven. Moreover, semi-supervised techniques perform better than unsupervised techniques in terms of missed anomalies, since the likelihood of an anomaly to form a close neighbourhood in the training data set is very low. In general, adapting this techniques to a different data type is straightforward, and primarily requires defining an appropriate distance measure for the given data. Some disadvantages are present too, because for unsupervised techniques if the data has normal instances that do not have enough close neighbours or if the data has anomalies that have enough close neighbours, the technique fails to label them correctly, resulting in missed anomalies. For semi-supervised techniques, instead, if the normal instances in test data do not have enough similar normal instances in the training data, the false positive rate for such techniques is high. Moreover, the computational complexity of the testing phase is also a significant challenge. In fact, it involves computing the distance of each test instance with all instances belonging to either the test data itself, or to the training data, to compute the nearest neighbours. Finally, the performance of this technique greatly relies on a distance measure, defined between a pair of data instances, that can effectively distinguish between normal and anomalous instances. Defining distance measures between instances can be challenging when the data is complex, e.g. graphs, sequences, etc.

### 3.4 Clustering based techniques

Cluster analysis groups data objects based only on information found in the data able to describe objects and their relationships. A cluster is formed when elements within a group are similar to one another and are different from the object in other groups. The better cluster distinction arises when there is greater similarity within a group and greater difference

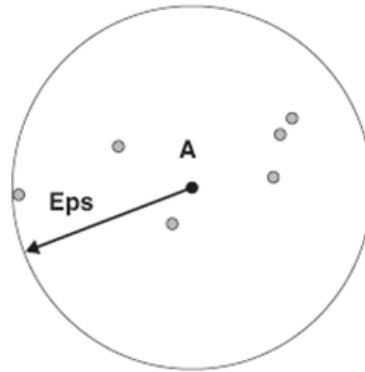


Figure 3.9. Center-based density (Figure taken from [TSK05]).

between the various group. To show the difficulty in the clustering formation, Figure 3.8 shows a set of twenty points and three different ways of dividing them into clusters. It could be seen as the definition of a cluster is rather imprecise and the best definition depends on the nature of data and the desired results.

Clustering analysis could be mainly grouped into three techniques: K-means, agglomerative hierarchical clustering and DBSCAN [TSK05]. The first one attempts to find a user-specified number of clusters ( $K$ ) represented by their centroids. It starts from the choice of these  $K$  centroids, where  $K$  is, obviously, the number of desired clusters. Then, each point is assigned to the closest centroid and the clusters are formed. So, the various centroids are updated for each cluster evaluating, usually, the mean of the points belonging to the cluster itself. These two steps, assignment and update, are repeated until the centroids remain the same. The assignment of a point to the closest centroid must be done through a proximity measure and, usually, the Euclidean distance is used. The agglomerative hierarchical clustering approach starts with each point being seen as a singleton cluster and then repeatedly merging the two closest clusters until a single, all-encompassing cluster remains. Two basic approach could be used to generate a hierarchical clustering: agglomerative or divisive. In the first case, the points are seen, at the beginning, as individual clusters and, at each step, the closest pair is merged. A notion of cluster proximity is requested. In the latter case, an all-inclusive cluster is considered and it is split at each step until only one-point clusters remain. For this approach, the point is the choice of the cluster to split and how to do the splitting. DBSCAN, instead, is a density-based clustering algorithm that produces a partitional clustering with the clusters number automatically determined by the algorithm itself. It must be noticed that this technique omits the points in low-density regions since they are classified as noise. In the centre-based approach, density is estimated for a particular point in the data set by counting the number of points within a specified radius,  $Eps$ , of that

point, including the point itself (Figure 3.9). As could be imagined, the density of any point depends on the specified radius, so if it too large the density will be equal to the number of dataset points and if it is too small, the points will have density equal to 1.

Clustering based anomaly detection techniques can be grouped into three categories. The first one rely on the assumption that normal data instances belong to a cluster in the data, while anomalies either do not belong to any cluster. Techniques based on the above assumption apply a known clustering based algorithm to the data set and declare any data instance that does not belong to any cluster as anomalous. A disadvantage of such techniques is that they are not optimised to find anomalies, since the main aim of the underlying clustering algorithm is to find clusters. The second category assumes that normal data instances lie close to their closest cluster centroid, while anomalies are far away from their closest cluster centroid. It must be observed that if the anomalies in the data form clusters by themselves, the two previous techniques will not be able to detect such anomalies. To address this issue the third category of clustering based techniques rely, instead, on the assumption that normal data instances belong to large and dense clusters, while anomalies either belong to small or sparse clusters.

Several clustering based techniques require distance computation between a pair of instances. Thus, in that respect, they are similar to nearest neighbour based techniques. The choice of the distance measure is critical to the performance of the technique and so the discussion regarding the distance measures hold for clustering based techniques also. The key difference between the two techniques, however, is that clustering based techniques evaluate each instance with respect to the cluster it belongs to, while nearest neighbour based techniques analyse each instance with respect to its local neighbourhood.

Clustering based techniques can operate in an unsupervised mode, and it is a great advantage, and such techniques can often be adapted to other complex data types by simply plugging in a clustering algorithm that can handle the particular data type. Furthermore, the testing phase for clustering based techniques is fast since the number of clusters against which every test instance needs to be compared is a small constant. Unluckily the performance of clustering based techniques is highly dependent on the effectiveness of clustering algorithm in capturing the cluster structure of normal instances. Moreover, many techniques detect anomalies as a by-product of clustering, and hence are not optimised for anomaly detection. On the whole, several clustering algorithms force every instance to be assigned to some cluster. This might result in anomalies getting assigned to a large cluster, thereby being considered as normal instances by techniques that operate under the assumption that anomalies do not belong to any cluster. It must be also observed that several clustering based techniques are effective only when the anomalies do not form significant clusters among themselves. Finally, the computational complexity for clustering the data is often a bottleneck.

### 3.5 Examples in machine diagnostics

The topic presented so far have been widely applied to rolling bearings condition monitoring and fault diagnosis. The first step is the selection of proper features able in representing the condition of the machine. After this, a suitable method must be able to detect damages and to prevent further escalation and losses. On the whole, normal data refers to components without defects and are available. Since a labelling of these instances could be done, we recall semi-supervised techniques. An anomalous behaviour must be identified as soon as possible in order to prevent bigger damages and breakdowns. In [CBK09] some anomaly detection techniques used in this domain are listed.

Many applications of neural network can be found in order to monitor conditions of structures, of machines and of bearing in particular. In the paper [LCTH00] by Li *et al.* an approach for motor rolling bearing fault diagnosis using neural networks and time/frequency-domain bearing vibration analysis is presented. They use vibration simulations to assist in the design of various motor rolling bearing fault diagnosis strategies. Both simulation and real-world testing results obtained indicate that neural networks can be effective agents in the diagnosis of various motor bearing faults through the measurements and interpretation of motor bearing vibration signatures. Samanta *et al.* propose a study in [SABAA04] to compare the performance of bearing fault detection using three types of artificial neural networks, namely, multilayer perceptron (MLP), radial basis function (RBF) network, and probabilistic neural network (PNN). The time domain vibration signals of a rotating machine with normal and defective bearings are processed for features extraction. These features are then used as inputs to the three ANN classifiers for two-class (normal or fault) recognition. The results show the relative effectiveness of the classifiers in detection of the bearing condition.

Regarding SVM, Shin *et al.* present in [SEK05] a test technique for machine fault detection and classification in electro-mechanical machinery from vibration measurements using a one-class support vector machines approach. In particular, they examine the performance of the proposed method by comparing it with that of multilayer perception, one of the artificial neural network techniques, based on real benchmarking data. Experiments performed on artificial and real data set show that the performance of one-class method is mostly superior to that of multilayer perception. Yang *et al.* propose their study in [YTTW05]. They present a fault diagnosis scheme based on multi-class SVMs for a rotating machinery that is trained and tested on data obtained from a fault simulator. In particular, they propose a comparison between two kinds of algorithms, the SVMs and ANNs and the great characteristics of SVMs are demonstrated. Widodo *et al.* present a very complete survey of machine condition monitoring and fault diagnosis using support vector machines (SVMs) [WY07a]. They attempt to summarise and review the recent developments of SVM in machine condition monitoring and diagnosis. They state that SVM has excellent performance in generalisation so it can produce high accuracy in classification for machine condition monitoring and diagnosis.

In [JLB06] are presented various research and developments in diagnostics and prognostics of mechanical systems implementing condition based maintenance (CBM). They state that a commonly used algorithm in machine fault classification is the nearest neighbour algorithm that fuses two closest groups into a new group and calculates distance between two groups as the distance of the nearest neighbour in the two separate groups and they cite some applications in bearing diagnostics. An application of K-nearest neighbour is that shown in [ZJN05] where genetic programming is used for feature generation, and then the K-nearest neighbour is applied as a classifier on roller bearing data too. The paper by Wei *et al.* deals with the issue of incipient fault diagnosis for rolling element bearing too [WGZ<sup>+</sup>11] and proposes an envelope demodulation technique based on wavelet packet transform and energy operator application to extract the fault feature of vibration signal. Then, the relative spectral entropy of envelope spectrum and the gravity frequency are combined to construct two-dimensional features vector that characterises each fault pattern. At this point, K-nearest neighbours is used to perform faults identification automatically. Their experimental results prove that the method could avoid inaccurate diagnosis which only depends on the recognition of characteristic frequency and that it is effective in the automatic fault diagnosis of bearing.

In their review on machinery diagnostics and prognostics implementing CBM [JLB06], Jardine *et al.* talks about cluster analysis too. In fact, as a multivariate statistical analysis method, it is a statistical classification approach that groups signals into different fault categories on the basis of the similarity of the characteristics or features they possess and some application in machinery fault diagnosis. In their review, Jardine *et al.* cite the application of the most commonly used distance measures (Euclidean distance, Mahalanobis distance, Kullback–Leibler distance and Bayesian distance) and also some new distance metric. For example, in [LHZC08] the authors present a fault diagnosis method of rotating machinery based on a new clustering algorithm using a compensation distance evaluation technique (CDET). Their new clustering algorithm is described and applied to incipient fault and compound fault diagnosis of locomotive roller bearings. The results obtained show that it is able to reliably recognise not only different fault categories and entities but also the compound faults, and demonstrates the superior effectiveness and practicability of the algorithm. The agglomerative hierarchical algorithms using the Ward's method is applied in [AKT09]. The authors consider a Self-Organizing Maps (SOM) and they investigate its performance in induction motor rotor fault detection and severity evaluation. In particular, the agglomerative hierarchical algorithms is used to automatically dividing the map into interesting interpretable groups of map units that correspond to clusters in the input data. They observe that the results obtained with the presented approach allows the detection of a rotor bar fault just directly from the visualisation results and also the estimation of the extent of rotor faults.

## Chapter 4

# Experimental tests

In this Chapter the procedure lied to the whole process of data acquisition is presented. A bearing test rig has been assembled in the laboratory of the Department of Mechanical and Aerospace Engineering (DIMEAS) of Politecnico di Torino by our group, Dynamics & Identification Research Group (DIRG). The goal is the study of the vibrations due to the bearings in a controlled situation, with the reduction of spurious signals coming from other rolling and vibrating elements. Moreover, it is possible to analyse various damages types and to change the load and the rotational speed applied. Variations of the oil and of bearing temperature could also be checked.

In Section 4.1, the whole test rig is presented, starting from its project to its final construction. Then, the elements composing the acquisition system (accelerometers, Oros OR38) is described in Section 4.2. Our tests relies on the analysis of vibrations of various kind of faults on bearings, that are presented in Section 4.3. Finally, the two setups that are provided for the data acquisition in different kinds of tests are showed.

The development and the design of the test rig and the consequent acquisitions campaign are related to a collaboration with Avio S.p.A. industry. Avio is a world leader in the design, development and manufacturing of aerospace propulsion components and systems for both for civil and military aircraft. Our group works with Avio within a Piemonte region project named “Piattaforme per l’Aerospazio – Progetto GREAT 2020” with the topic “Progetto McLab: Unità prognostica per trasmissioni meccaniche di potenza (Gearbox)”. The goal of our collaboration is the definition of some algorithms that could detect the damage on a gearbox with a certain confidence level. Almost all the tests that we operate on our rig are done in order to assess and verify the algorithms to work in various situation thanks to the fact that here we can work in controlled conditions.

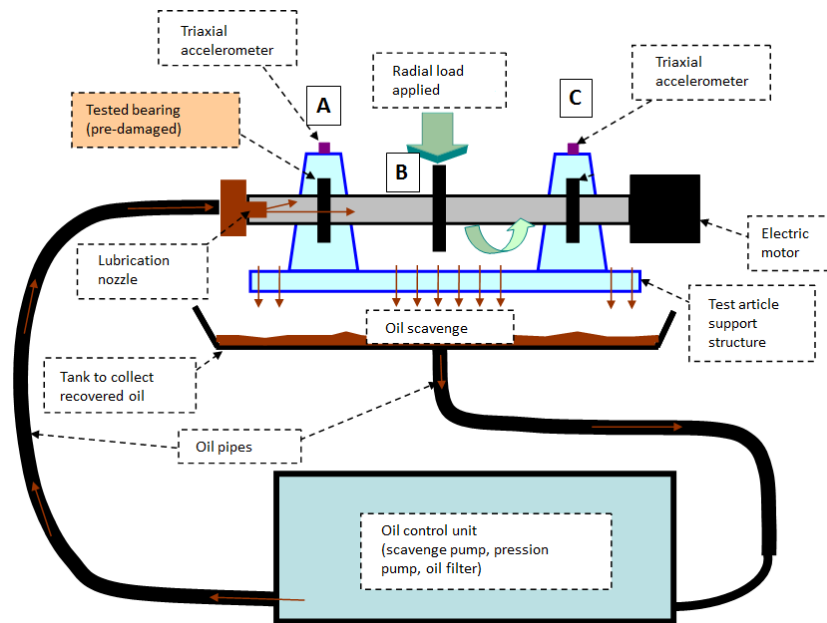


Figure 4.1. Test rig project and its components; point A, B and C indicates the places where are present the bearings.

## 4.1 Bearing test rig

As outlined in the introduction, the data we are going to analyse have been acquired on a test rig assembled by DIRG at DIMEAS laboratory. It has been designed with Avio S.p.A. collaboration in order to reproduce, in controlled laboratory conditions, real gearbox bearings behaviour. With this target, some components of the original gearbox have been taken and adapted on our test rig. The principal goal is the reduction of spurious signals coming from the mechanical sounds of other bearings, rotating shafts, gear wheels meshing and other vibrating elements, as usually happens in real applications. Moreover, it is designed to perform accurate testing of bearings with different levels of damage in a controlled environment.

Figure 4.1 shows the project of the test rig and all the important components it is made of. Three bearings are located on a shaft which rotates thanks to an electric motor (point A, B and C in Figure 4.1). It must be said that this shaft is the same used on the real industrial aeronautic gearbox EJ200. On the central bearing (Point B) a load is applied, while the one in Point A could be changed in order to analyse various damage types. In the Figure is also presented the oil path: after having lubricated the bearing through a nozzle, it is collected in a tank. Then, a control unit allows it to continue circulating in some pipes and return to the nozzle.

The next step in the test rig realisation is the design of the project, which is showed in Figure 4.2 and 4.3 in two SolidWorks® rendering. The three bearings (Point A, B, C in the

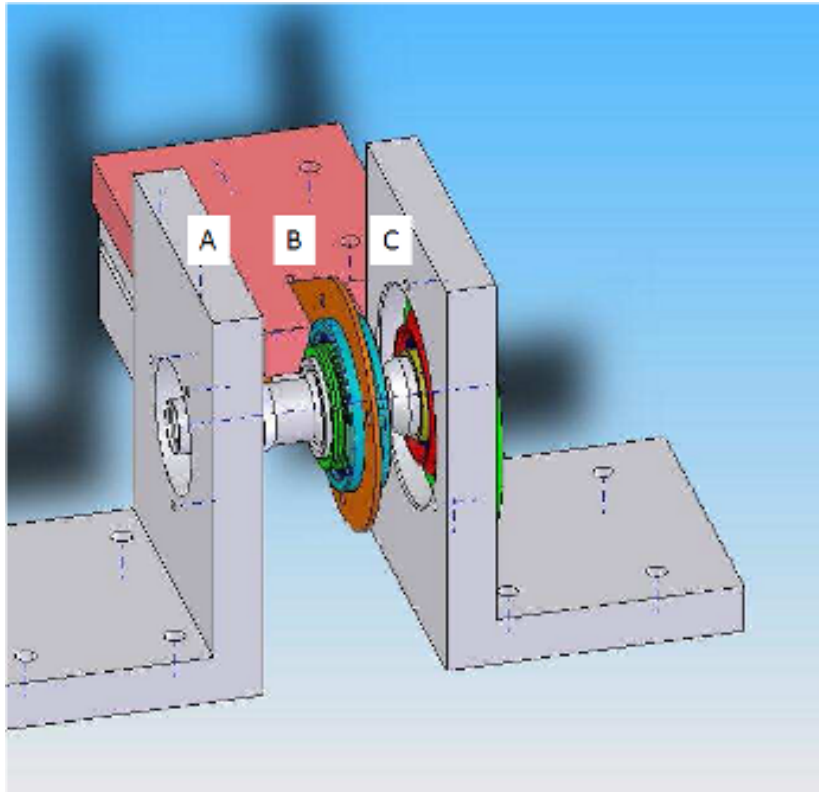


Figure 4.2. Test rig in a Solid Works rendering: bearings are placed in Point A, B and C.

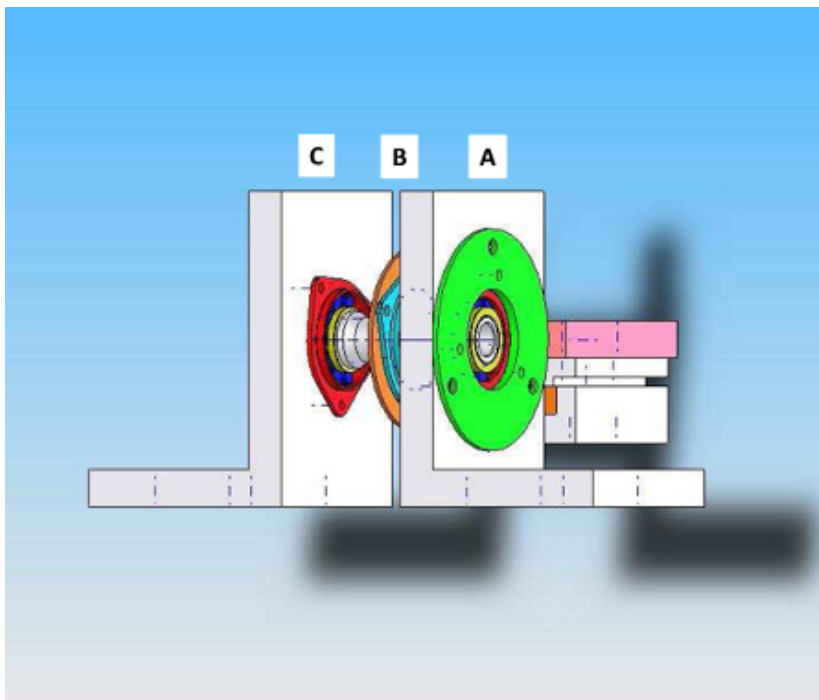


Figure 4.3. Test rig axonometric view in a Solid Works rendering: bearings are placed in Point A, B and C.



ISO Gradation	Viscosity		Density kg/l @ 15°C	Pour point °C	Flash point °C
	cSt @ 40°C	cSt @ 100°C			
15	15	3.72	0.859	-42	188

Table 4.1. Main technical characteristics of the Mobil DTE 11M oil that is used to lubricate the test rig.

Figures) rotate on the shaft as expected by the project. Two of them, in particular those in Point A and C, are taken directly from the original gearbox and they have an internal diameter of 25 mm and 10 rolling elements. The other one (in Point B) is 40 mm diameter and has a greater number of rolling elements and comes from another gear of the same gearbox. Its goal is the transmission to the shaft of the same radial load applied to the gearbox shaft. Moreover, this bearing is able to face, under stress, a load two times heavier than the other two bearings and its oversizing is provided in order to avoid early damages and wears.

Each one of these bearings is lubricated through a proper oil jet circulating thanks to the lubrication system described previously. The oil used is the Mobil DTE 11M, whose main features are an optimal fluidity at low temperatures, a great shear stress resistance and a high efficiency also when viscosity decreases. Moreover, thanks to its high viscosity index, good performances are assured in a wide temperatures range, from the lowest at the system switch to the highest during the system operating condition. This oil has also been chosen thanks to its rather safe and healthy properties. The main technical characteristics of the oil used in our test rig are showed in Table 4.1.

Figure 4.4 shows a picture of the real test rig. Here it is possible to see the tank that collects all the oil (Point G) and allows its circulation through the lubrication system. If necessary, in the second part of the tests, a resistance has been provided in order to heat the oil itself. The goal is to reach faster equilibrium conditions and to operate various oil temperatures analysis. In the same Figure could also be seen the shaft and the bearings located as described before (Point A, B and C). Point D shows, instead, the electric spindle which allows the shaft to rotate from 6000 to 30000 RPM, with also its supply and the refrigerating system (coloured cables). The radial load is applied to the central bearing, as requested, through two springs (Point E in the Figure) and could be measured through a load cell (Point F). This one, has 0.5 mV/N sensibility and its calibration characteristics are showed in the plot in Figure 4.5(a).

## 4.2 Sensors and acquisition system

To acquire the vibration signals, triaxial accelerometers such as the one in Figure 4.5(b) are placed in various positions on the test rig case. Technical characteristics of the accelerometers are presented in Table 4.2. As could be seen in Figure 4.5(b), they are strictly fixed to the bearing support thanks to a set screw. Moreover, the  $x$ ,  $y$  and  $z$  axis corresponds to the axial,

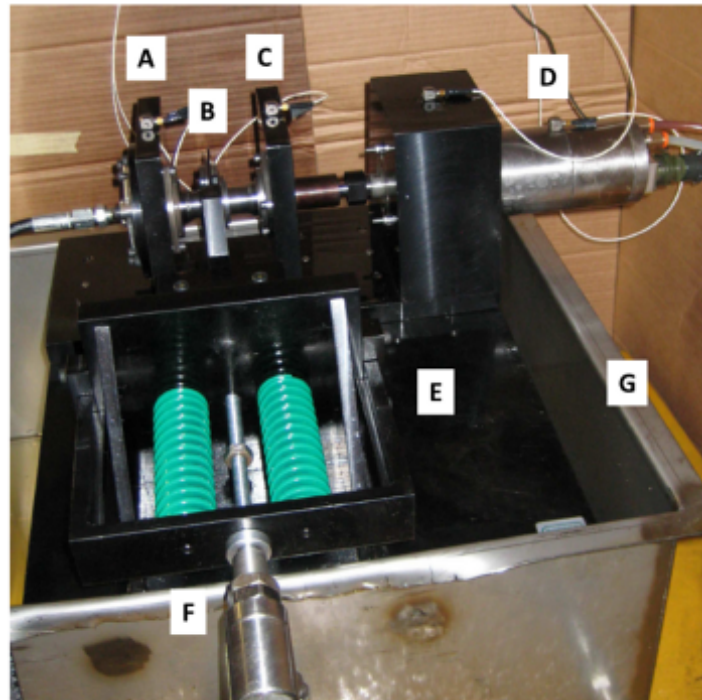
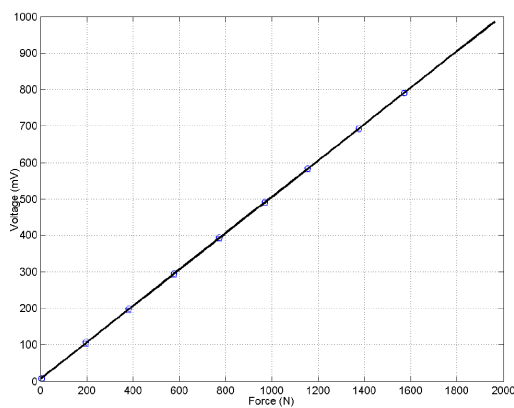


Figure 4.4. Test rig view: in Point A, B and C are placed the bearings, Point D indicate the electric spindle, while in Point E could be noticed the springs used to apply the load; in Point F is placed the load cell to measure the load itself and Point G shows the tank containing the oil.



(a)



(b)

Figure 4.5. (a) Cell load characteristic and (b) triaxial accelerometer used for signal acquisition.

radial and tangential direction respectively.

All the accelerometers channels are connected to an *Oros OR38* data acquisition system, which is shown in Figure 4.6(b). It is able to sample up to 24 channels at the same time with a maximum sampling rate of 102400 sample per second with a passing band of 40 kHz, taking into account the anti-aliasing filters. Each channel could be amplified or weakened

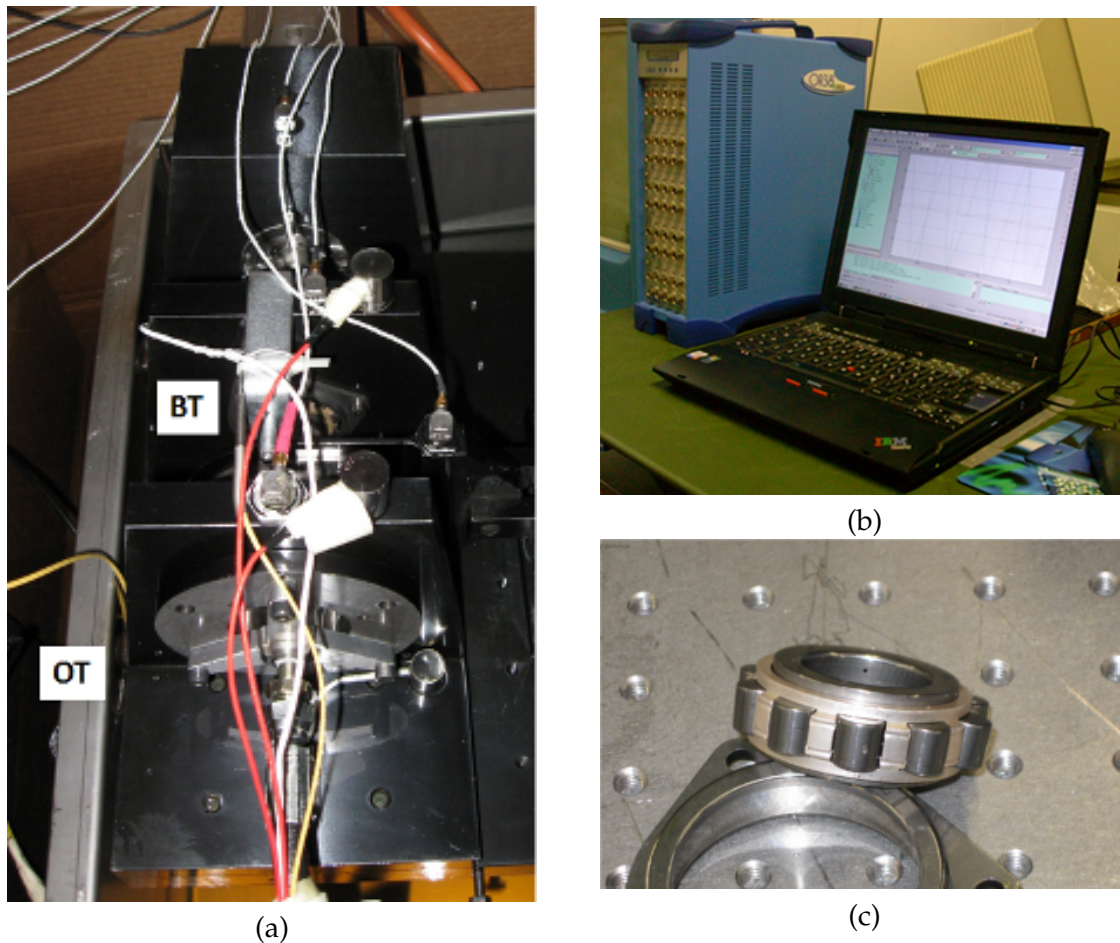


Figure 4.6. (a) Test rig view, where point BT and OT are bearing and oil temperature acquisition points respectively; (b) Oros OR38 acquisition system with its PC; (c) type of bearing used during the test (in this case a  $450\ \mu\text{m}$  indentation on a rolling element).

Company	Model	Type	Full scale g	Nominal sensibility mV/g	Resonance frequency Hz
Kistler	8763A500	triaxial	500	10	55

Table 4.2. Technical characteristics of the accelerometers used for signals acquisitions .

independently, considering a scale from 3.16 mV to 20 V, to use at best the 24 bit characterising the analog-to-digital converter resolution. The system could be set to measure a certain number of seconds of time for each channel without any signal conditioning before the acquisition. In this way, raw data are always available and it is possible to analyse them *a posteriori*. The acquisition system Oros OR38 could register in a very little time the measures thanks to the internal disc. Moreover, a ethernet crossover cable connects it to its control system, which is installed on a portable PC. Then, data saved in the memory could be easily exported on other PC in various format such as ASCII and .mat files (Matlab<sup>®</sup> format).

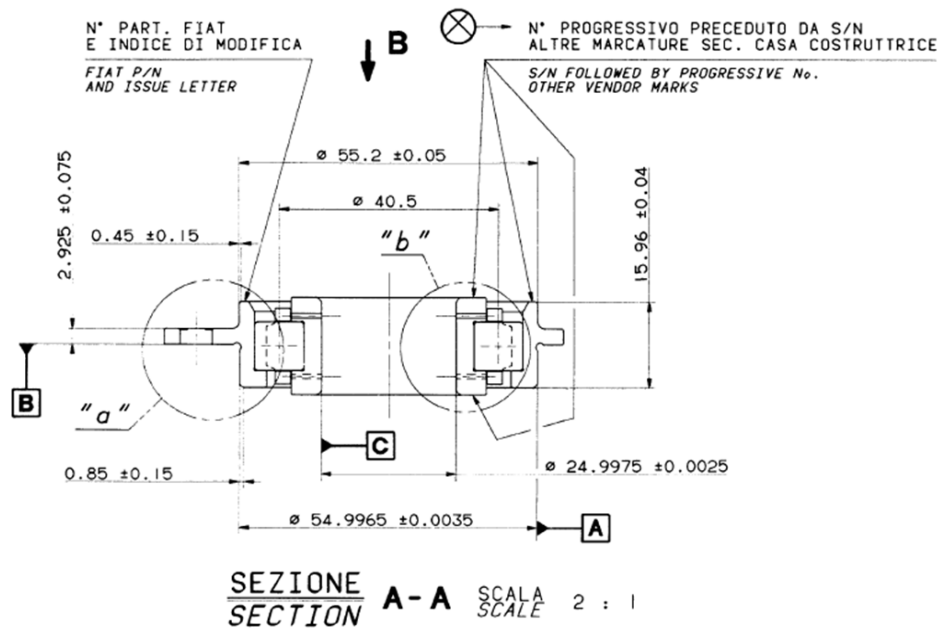


Figure 4.7. Section and measures of the bearing that is assembled in its case in Point A (Figure 4.4) and that could be changed during the various measurements.

A further parameter that could be measured is the temperature, in particular that reached by the bearing and the oil. To monitor them, two thermocouples are placed in the system: the former is in proximity of the outer ring bearing (Point BT in Figure 4.6(a)) and the latter lies into the oil in the tank (approximately in Point OT in Figure 4.6(a)). In this way it is possible to measure and check the temperature near the bearing and the one reached by the oil. Moreover, as said before, it is possible to heat the oil in the tank with a resistance in order to change its temperature.

### 4.3 Bearings

The bearings used during the acquisitions and that could be changed in Point A in the test rig (Figure 4.4) are supplied by SKF by keeping the specifications requested by Avio. They belong to the rolling element bearings type, where the rotating elements are cylinders with length slightly greater than diameter. SKF deals also with the creation of the various damages in their laboratories. All bearing belong to the same kind, whom section and measures could be seen in Figure 4.7. A picture of this bearing with also its external structure is in Figure 4.6(c), where the particular case of a damaged roller is showed. Thanks to the structure of the test rig, it is possible to assemble and disassemble on it several bearings affected by different damages. The various types of faults have been furnished are presented in Table 4.3, where their denomination, the different damages and related entity are defined. Moreover, some examples of these damages could be seen in Figure 4.8. For instance, Figure 4.8(a) shows the

Denomination	Type	Entity
0A	healthy	-
1A	inner ring indentation	450 $\mu\text{m}$
2A	inner ring indentation	250 $\mu\text{m}$
3A	inner ring indentation	150 $\mu\text{m}$
4A	rolling element indentation	450 $\mu\text{m}$
5A	rolling element indentation	250 $\mu\text{m}$
6A	rolling element indentation	125 $\mu\text{m}$
7A	inner ring sandblast	Ra 0.73 $\mu\text{m}$
15A	electroerosion inner ring	110 $\mu\text{m}$
16A	notch along roller axis	50 $\mu\text{m}$

Table 4.3. Various type of damages affecting the different bearings with their denomination and entity.

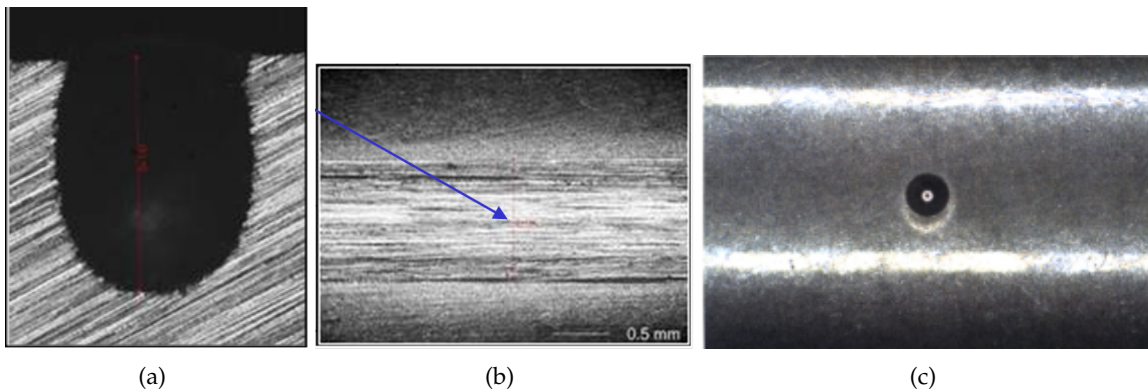


Figure 4.8. Three damages characterising three analysed bearings: (a) 110  $\mu\text{m}$  electroerosion on the inner ring, (b) 50  $\mu\text{m}$  notch along roller axis and (c) 450  $\mu\text{m}$  damage on a roller (15A, 16A and 4A in Table 4.3 respectively).

110  $\mu\text{m}$  electroerosion on the inner ring, while Figure 4.8(b) displays the 50  $\mu\text{m}$  notch along roller axis. In Figure 4.8(c), instead, the fault on the rolling element of 450  $\mu\text{m}$ , which has been produced through a Rockwell type indentation hedge, could be seen. The case 16A, that is the notch along the roller axis, provides a low number of acquisitions since it generate too high vibrations and noise.

#### 4.4 Test setup

The test rig described in Section 4.1 is used to acquire a large set of data. The whole test campaign has lasted about two years, has concerned the two slightly setups described in the following and has resulted in about 250 GB of data. The main goal has been the creation of a consisted data base which could allow the evaluation of different situations such as damage identification and evolution and temperature variation. An acquisition consists

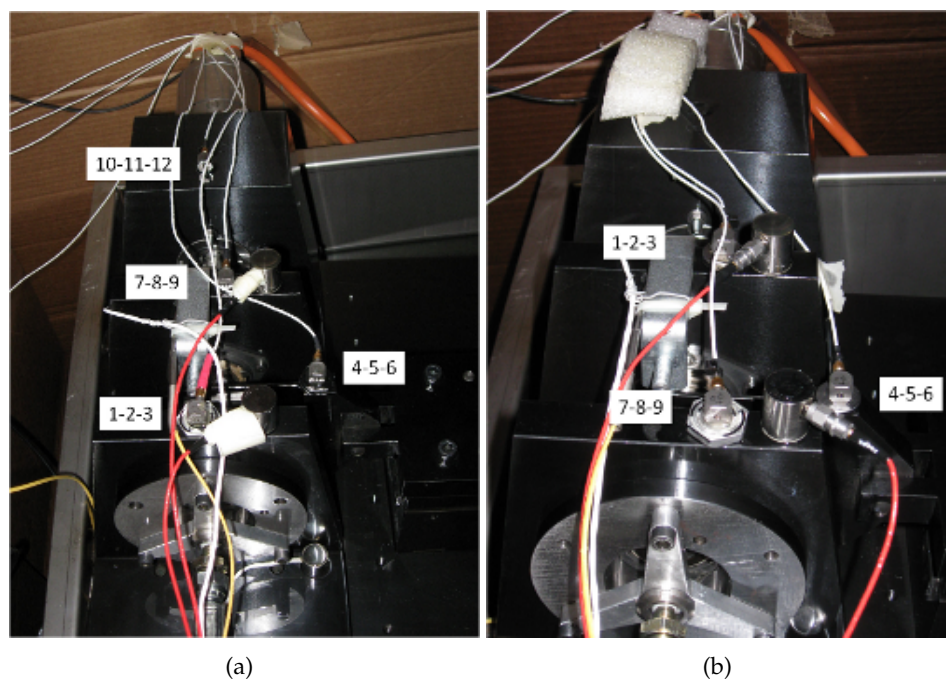


Figure 4.9. The two setups used for the data base acquisitions. The main differences in the two cases stands in the position and in the different number of accelerometers (4 in Setup I in (a) and 3 in Setup II in (b) ) and in the introduction of a resistance to heat the oil in Setup II.

Direction	Acc 1	Acc 2	Acc 3	Acc 4
Axial	1	4	7	10
Radial	2	5	8	11
Tangential	3	6	9	12

(a)

Direction	Acc 1	Acc 2	Acc 3
Axial	7	4	1
Radial	8	5	2
Tangential	9	6	3

(b)

Table 4.4. Axis direction of various channels for the accelerometers in Setup I (Table (a)) and Setup II (Table (b)). The number of the accelerometers (Acc 1, Acc 2, etc.) refers to their position in Figures 4.9(a) and 4.9(b) going from the foreground to the background.

usually in 8 seconds of accelerations registered by the three channels of various triaxial accelerometers at a sampling frequency of 102400 Hz. For each acquisition, also oil and bearing casing temperatures (see Figure 4.6(a)) are checked through two thermocouples and then registered. As described previously, this test rig has been assembled so that it could be possible to change and monitor conditions such as the speed, the load and the temperature. Moreover, it is possible to assemble and disassemble the various bearings characterised by the different damages (Table 4.3).

#### 4.4.1 Setup I

The test rig has been firstly built in 2009 and in this setup four accelerometers are considered for the vibrations acquisitions, as shown by Figure 4.9(a). Three of them, that are those

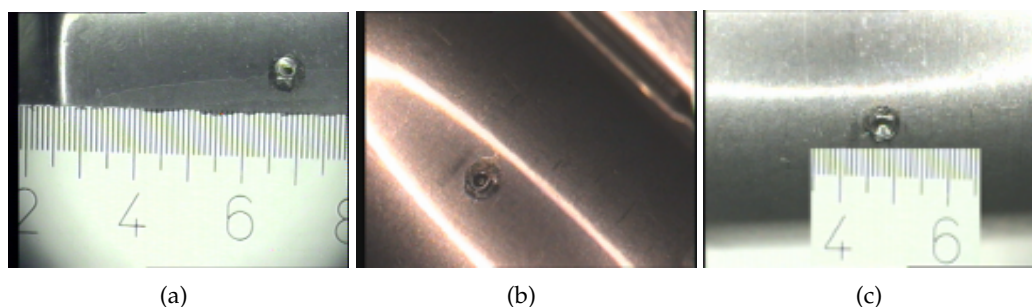


Figure 4.10. (a) Initial condition of roller fault (Picture taken on 9th July 2009), (b) damage appearance after 50 hours of running (Picture taken on 15th October 2009), (c) damage when other 50 hours have passed (Picture taken on 27th October 2009). The scales are  $10^{-1}$  mm.

producing channels 1-2-3, 4-5-6 and 7-8-9, are placed on the bearings in order to acquire the vibrations produced directly. They corresponds to Point A, B and C in Figure 4.4. The last accelerometer, giving channels 10-11-12, is placed on the engine support in order to acquire the mechanical noisy components given by the engine itself. Various channels directions are showed in Table 4.4. Moreover, two thermocouples are also present (Figure 4.6(a)) and placed one in contact with the casing of the bearing and one in the oil tank.

In the following sections, the various tests and acquisitions handled with this setup are described.

#### 4.4.1.1 Different type of damage

These measurements regards the acquisition of accelerations produced by the nine different bearings presented in Table 4.3 that are set up in various days on the test rig. The test consists in crossing five rotating speeds (6000, 12000, 18000, 24000 and 30000 RPM) with three possible values of applied load (1000, 1400 and 1800 N). Also the no load condition is acquired. It must be observed that the acquisitions for the crossing conditions are carried out in close succession, avoiding the whole system to reach stabilisation. This means that oil and bearing temperatures are not equal during all the acquisitions.

Each acquisition registers 8 seconds of acceleration signal with sampling frequency of 102400 Hz. In order to have a more consistent data base, during our analysis each signal will be divided in 8 parts, each of them standing for 1 second of measure.

#### 4.4.1.2 Endurance

An endurance test is a typical procedure in mechanics in order to evaluate the evolution of a certain type of damage. In this case it regards the bearing characterised by the  $450 \mu\text{m}$  fault on the roller element (4A) which is subject to a running procedure of about 200 hours. The speed of the shaft is fixed at 18000 RPM while a load of 1800 N is applied on the system.

Two different acquisition standard are used (Figure 4.11). During the first 50 operating

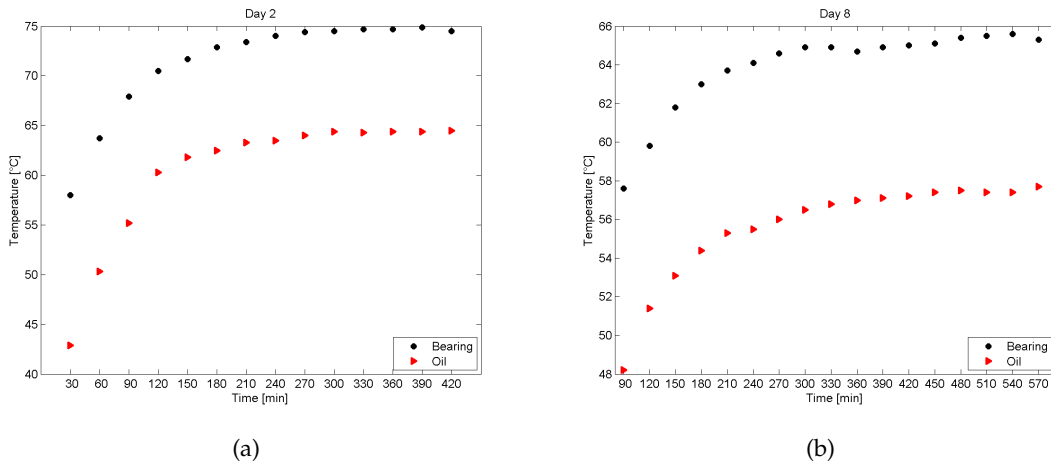


Figure 4.11. Two acquisition standards in the endurance test: measures are taken every 30 minutes after that (a) 30 minutes or (b) 90 minutes have passed from the system switch on.

hours, the system is switch on and then, after half an hour, the measures are taken every 30 minutes. This means that the first measurements of each day of acquisition are taken also if the whole system has not reached the stability. The bearing is then taken apart in order to check the indentation evolution. However, any particular difference could be noticed comparing the two cases. In next part of the test, the system is switched on and the first acquisition is then taken after 90 minutes, so that the stability is almost reached. Then, accelerations are registered every 30 minutes. When other 50 hours of rotation have passed, the bearing is again disassembled but also in this case no significant difference is observed. However, any particular difference could be noticed comparing the two cases. The final part of the endurance test lasts about 100 hours and also in this case the first measure is registered after 90 minutes from the switch on.

#### 4.4.2 Setup II

The four accelerometers of Setup I have been taken away in 2010, while the test rig has not been subjected to any variations. A new accelerometers installation, instead, has been displaced. The test rig is that shown in Figure 4.9(b), where only 3 accelerometers are used. Note that channels numbers for each accelerometers are changed compared to Setup I. In fact, channels 1-2-3 correspond to channels 7-8-9 in Setup I while channels 7-8-9 are similar to channel 1-2-3 in the other setup. The accelerometer that was previously placed on the engine support has been considered not interesting any more. Unfortunately, the exact accelerometers position of Setup I can not be reproduced, so this means that it is not possible to compare properly the data between the two acquisition campaigns. However, they are placed, as before, in Point A, B and C in Figure 4.4.

The two thermocouples, instead, stand in the same previous positions. Moreover, a



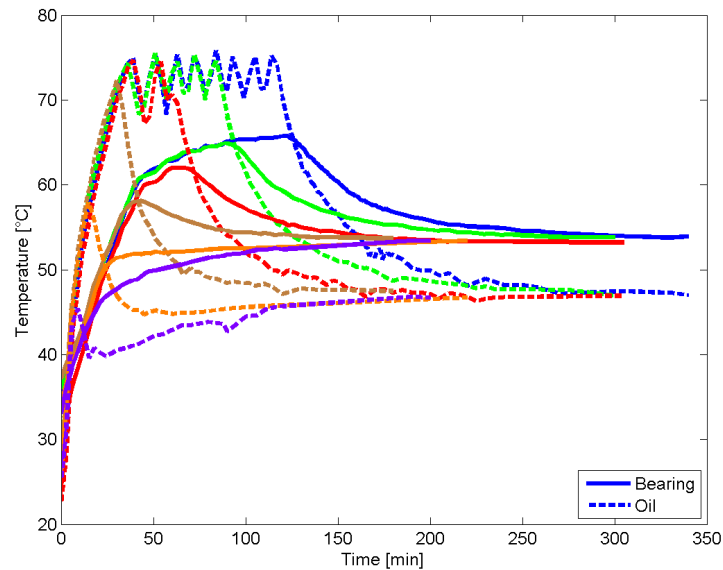


Figure 4.12. Results obtained evaluating bearing and oil temperatures to assess the use of the resistance; each colour stands for a different test related to different intervals of switching on and off. The goal is the evaluation of the proper period of oil heating that allows a faster stable condition reaching. The rotational speed is 12000 RPM and the load applied is 1800 N.

resistance is introduced in the oil tank in order to heat the oil itself. This component allows the system to reach the stability in a lower time so that it is possible to operate the acquisitions with the certainty of being always at the same stable condition. Figure 4.12 shows the results of some tests operated on the system in order to evaluate the proper period of oil heating that allows a faster stable condition reaching. The plot shows that it is not necessary to heat the oil too much (blue, green, red and brown lines). In fact, in those cases, the bearing, that represents the system itself, reaches a too high temperature and it takes some time to arrive to the effectively stable one. This is noticed in the plot: in these four tests the bearing temperature shows a decreasing trend. The optimal condition is the orange one: the oil reaches about 60°C and then the resistor is switched off. In this way, the bearing arrives at an optimal condition (around 50°C) in approximately 30 minutes. This is a very useful and interesting way to exploit the resistance since it helps in reducing the stability reaching times.

#### 4.4.2.1 Temperature variations

The goal of this test is the analysis of the system behaviour under various speed and temperature conditions. The bearing considered is the healthy one (0A) with a 1800 N load applied on the test rig. The shaft rotates at three different speeds (12000, 1800 and 24000 RPM) and four oil temperatures are reached for each one of them (45, 60, 75 and 85 °C, within a 7% margin). The stability of temperature condition is achieved switching on and off the

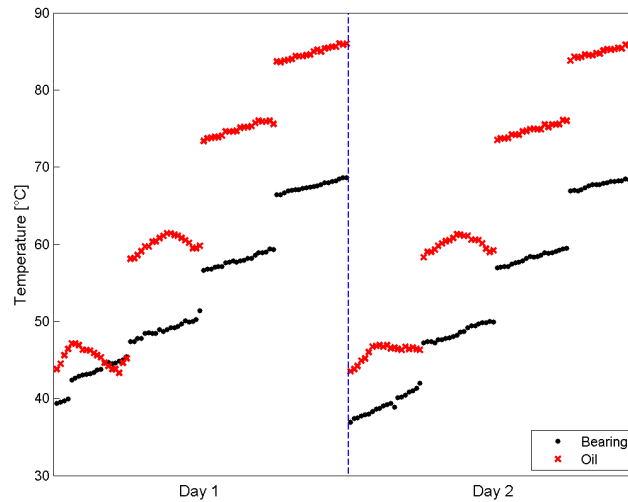


Figure 4.13. Bearing and oil temperatures in the case of 12000 RPM shaft speed rotation in the two days of acquisitions. It could be observed that the temperatures are almost equals in the various “plateau”.

resistance according to the values reported by the thermocouple. For each temperature about twenty measures are acquired and repeated in two different days to verify the repeatability of the acquisitions. Figure 4.13 shows the four oil and bearing temperatures “plateau” for a certain speed (12000 RPM) in the two days of acquisitions. It could be observed that the temperature are the same in both days.

#### 4.4.2.2 Speed and load small variations

This test wants to compare the behaviour of the system for two conditions in various cases of speed and load. The healthy bearing (0A) and the one characterised by  $450 \mu\text{m}$  roller element damage (4A) are considered. The shaft is supposed to rotate at three different speeds (9000, 10500 and 12000 RPM) and three different loads (1400, 1600 and 1800 N) are applied on the bearing in Point B in Figure 4.4. The goal is to acquire data with small condition variations (about 15%), in particular of speed and load. The acquisition consists in crossing the three values of rotating speed with the three values of loading. Forty measures are taken for each configuration in two different days in order to check their repeatability.

In this test the resistance shows its usefulness since, as explained previously, it reduces the time needed to reach the system temperature stabilisation. Some aspects related to the temperature could be noticed, as showed in Figure 4.14. Firstly, the values comparison of different loads at the fixed rotational speed of 12000 RPM points out as they are not particularly influenced by this factor (Figure 4.14(a)). If, instead, the load is the same (1800 N) and the speed changes, a variation of about  $2^\circ\text{C}$  could be observed (Figure 4.14(c)).

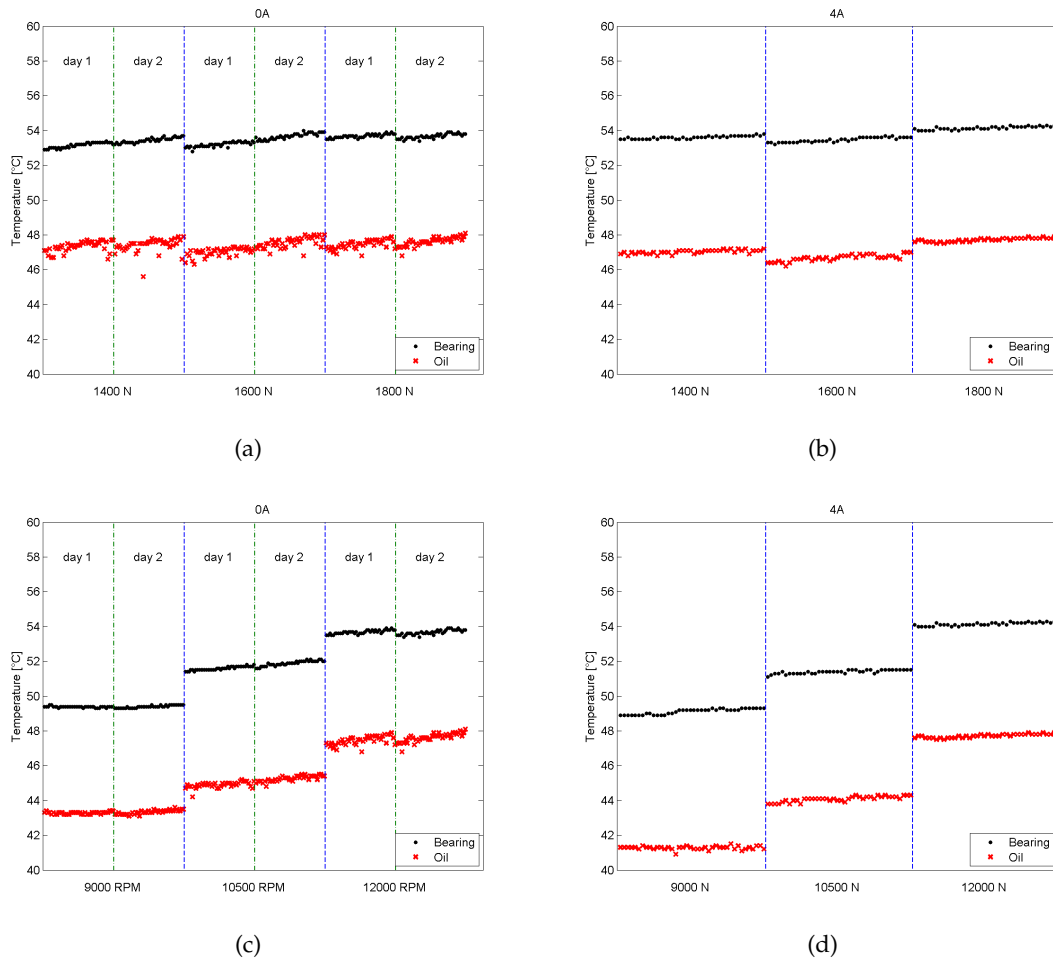


Figure 4.14. Bearing and oil temperatures for the healthy bearing and for the  $450\ \mu\text{m}$  roller element damage. Figures (a) and (b) are related to the 12000 RPM shaft speed, while Figures (c) and (d) refers to the fixed 1800 N load. In the case of the healthy bearing the acquisitions develop in two different days.

Furthermore, the values do not seem to be so different in the two days of acquisition for the case 0A, as shown by Figure 4.14(a) and 4.14(c). Finally, from the comparison of Figure 4.14(a) with 4.14(b) and Figure 4.14(c) with 4.14(d) it could be said that the temperatures of the system are rather not so influenced by the state (healthy or damaged) of the system. All these observations lead us to affirm that there are some configurations in which we are sure to analyse data without the influence of temperature.

# Chapter 5

## Test results

In this chapter, the techniques presented in Chapter 2 are applied to real data acquired on the test rig assembled by our research group and presented in Chapter 4. The results of this work have been published in some articles such as [FMP<sup>+</sup>10], [PFGM12] and [PMFG12].

In the first part of this Chapter, the Spectral Kurtosis technique, discussed in Section 2.4, is used to analyse the acquisition of Setup I (presented in Section 4.4.1). Results obtained through the application of the fast kurtogram on our data are showed. Moreover, the development of a diagnostic method based on this technique is presented in Section 5.1.5. This one has been implemented, in conjunction with that based on PCA and discussed in [PGT<sup>+</sup>11], in the context of the project developed with Avio S.p.A. They have been tested on two industrial aircraft gearboxes revealing a good detection ability.

In the other part of the Chapter, the work aims in the removal of effects produced in vibrations by the variation of speed and load. Empirical Mode Decomposition is applied to the data acquired on Setup II (presented in Section 4.4.2) to extract the features contained in the signal. Then, a classifier such as One Class SVM is trained to check the ability of the features in effectively identify the damage independently from the operating conditions. Moreover, the method is applied to data acquired for various temperature to check if this factor could also influence the recognition ability. A comparison with wavelet decomposition is also presented.

### 5.1 Damage detection through Spectral Kurtosis

The Spectral Kurtosis is an interesting statistical tool to analyse non-stationary signals, as those produced by bearings, because it is able to indicate the presence of series of transients and their locations in the frequency domain (Section 2.4).

### 5.1.1 The model

The signal  $Y(t)$  acquired in the case of a fault could be seen as the sum of the original damaged signal  $X(t)$  with a strong additive noise  $N(t)$ :

$$Y(t) = X(t) + N(t)$$

where  $X(t)$  is transitory and  $N(t)$  is stationary.

Moreover, the signal  $X(t)$  could be properly modelled as a generalised shot noise process [Ant06]:

$$X(t) = \sum_k \mathbf{X}_k h(t - \tau_k),$$

with  $\{\mathbf{X}_k\}_{k \in \mathbb{Z}}$  and  $\{\tau_k\}_{k \in \mathbb{Z}}$  being sequences of random variables accounting for possibly random amplitudes and occurrences of impacts, respectively, and  $h(t)$  being the impulse response resulting from a single impact.

Usually, the characterisation of incipient faults is achieved by means of statistical indicators which are sensitive to the peakiness of the signal, such as the kurtosis and the crest factor (Formulas in (1.1)). RMS value, instead, is only able to characterise the signal intensity and not to provide information on the temporal variation of its envelope. As clearly explained by Antoni and Randall in [AR06], two are the main advantages of the kurtosis indicator:

- I. it takes high values in the presence of the fault signal  $X(t)$ , that is true when the impulse responses generated by the impacts are sufficiently well separated and when the signal-to-noise ratio is sufficiently high
- II. it is ideally zero when only background noise  $N(t)$  is present, that is verified whenever it can be assumed Gaussian as a result of the superposition of a large number of independent (stationary) sources.

However, they notice that, in practise, these conditions are rarely satisfied and the most serious discrepancy concerns the concept of “sufficiently high” signal-to-noise ratio. In fact, the background noise often embodies strong vibrations from several competing sources (e.g. harmonics of rotating parts, random impulses from friction and contact forces, flow noise, etc.) which span a large frequency range and seriously mask the signal of interest. As consequence, the kurtosis is unable to capture the peakiness of the fault signal and does not differ from the zero value. The authors propose then its evaluation in various frequency bands rather than in the whole signal, defining the parameter Spectral Kurtosis (SK) as the fourth-order normalised cumulant (Section 2.4):

$$K_X(f) = \frac{\langle |H(t,f)|^4 \rangle}{\langle |H(t,f)|^2 \rangle^2} - 2, \quad (5.1)$$

where  $|H(t,f)|$  represents the complex envelope of the signal  $x(t)$  at frequency  $f$ ,  $\langle f(t) \rangle = \lim_{N \rightarrow \infty} N^{-1} \sum_N f(t)$  stands for the temporal average of function  $f(t)$  and -2 substitutes the -3

of the classical kurtosis since  $H(t, f)$  is complex.

As observed in [Ant07], the most important properties of the so-defined SK are that:

- I. SK of a stationary process is a constant function of frequency;
- II. SK of a stationary Gaussian process is identically zero;
- III. in the presence of stationary additive noise  $N(t)$ , the SK of a non-stationary process  $X(t)$  is

$$K_Y(f) = \frac{K_X(f)}{[1 + \rho(f)]^2}. \quad (5.2)$$

In the previous equation,  $\rho(f)$  represents the noise-to-signal ratio that could be expressed as the ratio of the noise Power Spectral Density (PSD) over the signal PSD:  $\rho(f) = S_N(f)/S_X(f)$ . From Equation (5.2) it could be noticed that  $K_Y(f)$  tends to  $K_X(f)$  in the frequency with low noise-to-signal ratio while it would be almost zero where the noise effect is stronger. These fact explain SK ability in finding the frequency bands where the fault signal is easier to be detect.

In [AR06], the authors explains how the SK could be evaluated through a Short-Time Fourier Transform (STFT) estimator, considering an analysis window  $w(n)$  with duration  $N_w$  shorter than the mean spacing between two consecutive impulses. Given a vibration signal  $Y(t)$ , acquired at sampling frequency  $f_s$ , with the repetitive fault rate of occurrence  $f_d$ , its approximate SK is

$$K_Y(f) \approx \left( \frac{f_s}{f_d} \frac{\gamma_{4w}}{N_w} (\kappa_X + 3) - 2 \right) \frac{1}{[1 + \rho(f)]^2}, \quad (5.3)$$

where  $\gamma_{4w}$  is the time–bandwidth product of the square of the analysis window and  $\kappa_X$  is the intensity of the fluctuations in the impulse amplitudes.

From Equation (5.3) the following remarks could be given [AR06]:

- I. SK is higher if is higher the sampling rate  $f_s$  of the signal;
- II. SK is lower when the rate of repetition  $f_d$  of the impulses, i.e. with the rotation speed of the machine, increases;
- III. SK is lower when the window length  $N_w$  of the STFT is smaller;
- IV. SK is higher if is higher the intensity of the fluctuations in the impulse amplitudes  $\kappa_X$ .

The remark related to  $N_w$  could suggest that a shorter  $N_w$  is convenient to reach high SK values, but, in the meanwhile, if it is too short, the SK would have poor spectral resolution. Thus, the authors propose is “that several window lengths  $N_w$  should be tried before selecting that which best emphasises the transient characteristics of the fault signal in the SK estimate; this should be paralleled with a similar procedure on the PSD estimate where the control of spectral resolution is made easier” [AR06].

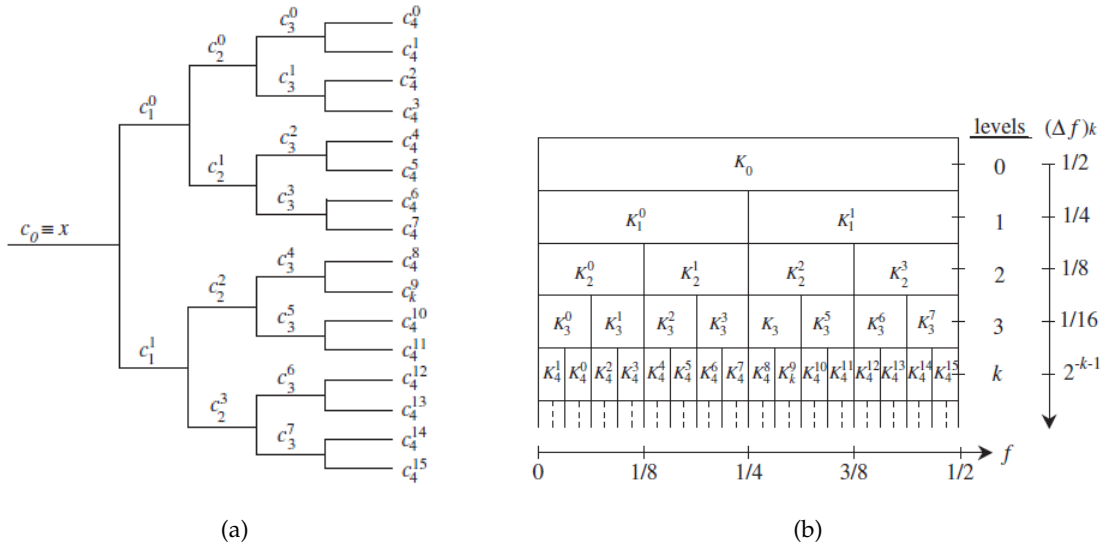


Figure 5.1. (a) Fast computation of the kurtogram by means of an arborescent filterbank structure and (b) fast kurtogram paving of the (frequency/frequency resolution) plane (Figure taken from [Ant07]).

### 5.1.2 Fast Kurtogram

The next step, as presented by [Ant07], is the realisation of some filters that could extract the fault signal through SK. A detection filter  $g(t)$  with a defined band-pass structure and only two parameters to set up, such as the central frequency  $f_c$  and the bandwidth  $B_f$  of the filter which maximises the kurtosis, in the envelope of the filtered signal. This is equivalent to find the frequency  $f$  and the window length  $N_w$  maximising the kurtosis itself on all the possible choice. In [Ant06], it is coined the name “kurtogram” to define the map formed by the STFT-based SK as function of  $f$  and  $N_w$  and so “the optimal central frequency  $f_c$  and bandwidth  $B_f$  of the band-pass filter are found as those values which jointly maximise the kurtogram.” The problem is then the way to select the best frequency resolution of the SK of an arbitrary signal, at a given frequency, or better, the optimal combination of frequency  $f$  and a frequency resolution  $\Delta f$  which maximises the SK. In particular, it originates the concept of (frequency/frequency resolution) dyad and the kurtogram said before is this representation of the SK in the  $(f, \Delta f)$  plane.

The whole exploration of the plane is a rather slow operation especially if the goal is on line industrial damage detection. However, Antoni demonstrates that it is sufficient to evaluate the kurtosis of the coefficients obtained after the evaluation of the quasi-analytic filter banks with central frequency  $f$  and bandwidth  $\Delta f$ . This is the so called “fast kurtogram” [Ant07]. A possible way to represent the filters construction is showed in Figure 5.1(a), where the coefficients  $c_k^i(n)$  could be interpreted as the complex envelope of signal  $x(t)$  as positioned at the central frequency  $f_i$  with bandwidth (frequency resolution)  $(\Delta f)_k$  defined, respectively,

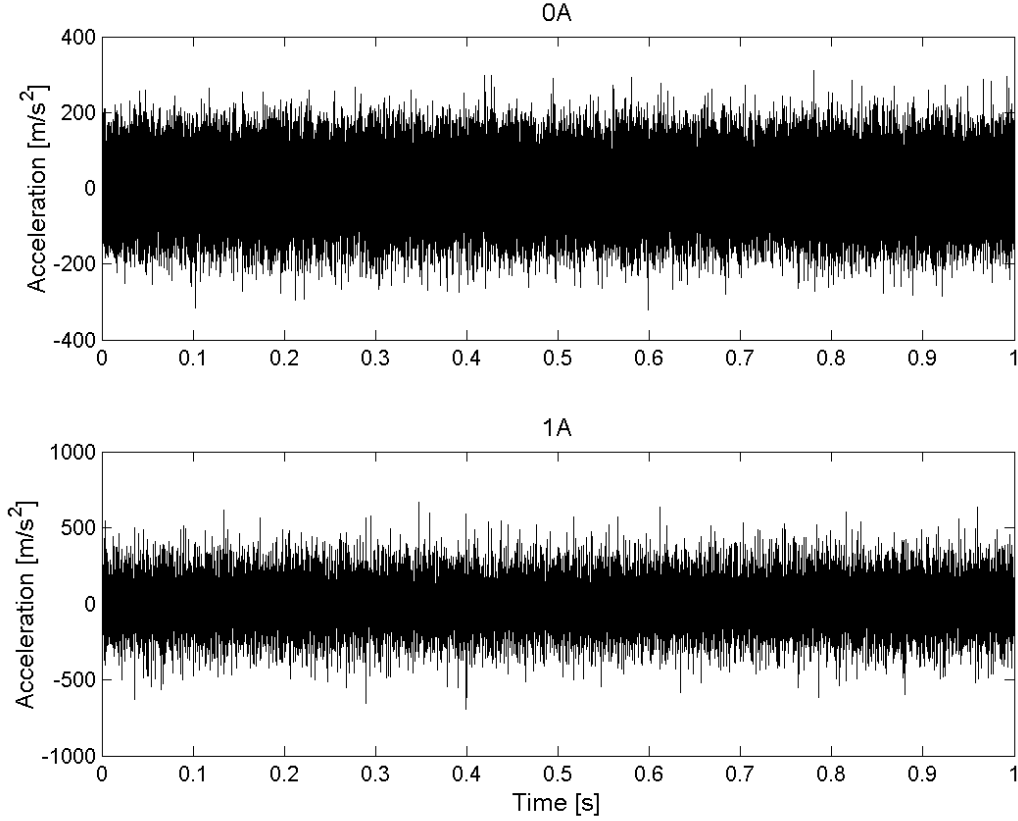


Figure 5.2. Acceleration signals (1 second) for the healthy bearing (top) and the 450  $\mu\text{m}$  damage on the inner ring (bottom), with the shaft rotating at 18000 RPM and 1800 N of load applied. The channel considered in the axial one for Accelerometer 1.

as

$$f_i = (i + 2^{-1}) 2^{-k-1} \quad (\Delta f)_k = 2^{-k-1},$$

where  $k$  represents the level of the decomposition.

The next step to produce the kurtogram is, then, the kurtosis evaluation for each sequence, according to Equation (5.1):

$$K_k^i = \frac{\langle |c_k^i(n)|^4 \rangle}{\langle |c_k^i(n)|^2 \rangle^2} - 2 \quad i = 0, \dots, 2^k - 1, \quad k = 0, \dots, K - 1.$$

Figure 5.1(b) shows the kurtogram representation at nodes  $\{f_i; (\Delta f)_k\}$  of the  $(f, \Delta f)$  plane.

On the whole, the power of fast kurtogram stands in its ability in finding automatically the best frequency band where it is possible to filter and modulate the signal. Moreover, three are the most interesting points:

- I. it reveals if there is any transitory in the signal;
- II. it shows in which frequency band they are localised;



III. it produces the complex envelope  $c_k^i(n)$  in each selected frequency band.

### 5.1.3 Application to bearings data

The procedure explained in Section 5.1.2 brings Antoni to develop an algorithm named `Fast_Kurtogram.m` which is freely accessible on the Internet<sup>1</sup>. Here, a signal and a certain level of decomposition are given as input in order to obtain and visualise the paving of the (frequency/frequency resolution) plane through a pyramidal decomposition of the signal into that number of levels. The kurtogram shows the kurtosis at each level and at each frequency on a colour image. A high value of the kurtogram indicates a frequency band where the signal is impulsive. In this way, the best frequency band where it is possible to filter and modulate the signal could be found. The algorithm could also bandpass filter the signal in a user-specified frequency band and provide a graphical visualisation of both the filtered signal and its envelope spectrum.

To demonstrate how this algorithm works it is applied on the data acquired on the Setup I (Section 4.4.1), in particular on those referred to various faulty bearings described in Section 4.4.1.1. Figure 5.2 shows 1 second of acceleration signals acquired on the test rig in the case of the healthy bearing and when a 450  $\mu\text{m}$  damage is present on the inner ring. This signal is the Channel 1 registered by the accelerometer producing channels 1-2-3 as showed in Figure 4.9(a), which is the axial component. Both signals in Figure 5.2 are characterised by the shaft rotating at 18000 RPM and a 1800 N load being applied in Point B in Figure 4.4. These two signals and those acquired for the other bearings in the same operating conditions are passed through the kurtogram algorithm. Figure 5.3 shows the graphical results for 1 second of acceleration for the healthy case, for the 450  $\mu\text{m}$  inner ring damage, for the 450  $\mu\text{m}$  rolling element damage, for the Ra 0.73  $\mu\text{m}$  sandblast on the inner ring and for the 125  $\mu\text{m}$  inner ring electroerosion. In addition, Table 5.1 propose the values obtained when this analysis is done for the other entities of inner ring and roller element damage too. It is clear how the maximum kurtosis is lower when the bearing is healthy, while the highest value is reached with the bigger damage on the rolling element. The smaller entities (150  $\mu\text{m}$  and 250  $\mu\text{m}$ ) for both the damages on the inner ring and on the roller show, instead, lower kurtosis values. A further aspect that could be analysed regards the position of the obtained  $f_c$  in the frequency domain. It could be noticed that for 4A and 5A, roller element fault of 450 and 250  $\mu\text{m}$  respectively, it is rather high (35200) and concentrated in a not so wide bandwidth (6400). This is due to the nature of the damage and reveals the ability of the technique in the detection and localisation of transient in the analysed signal.

<sup>1</sup><http://www.utc.fr/antoni/programm.htm>

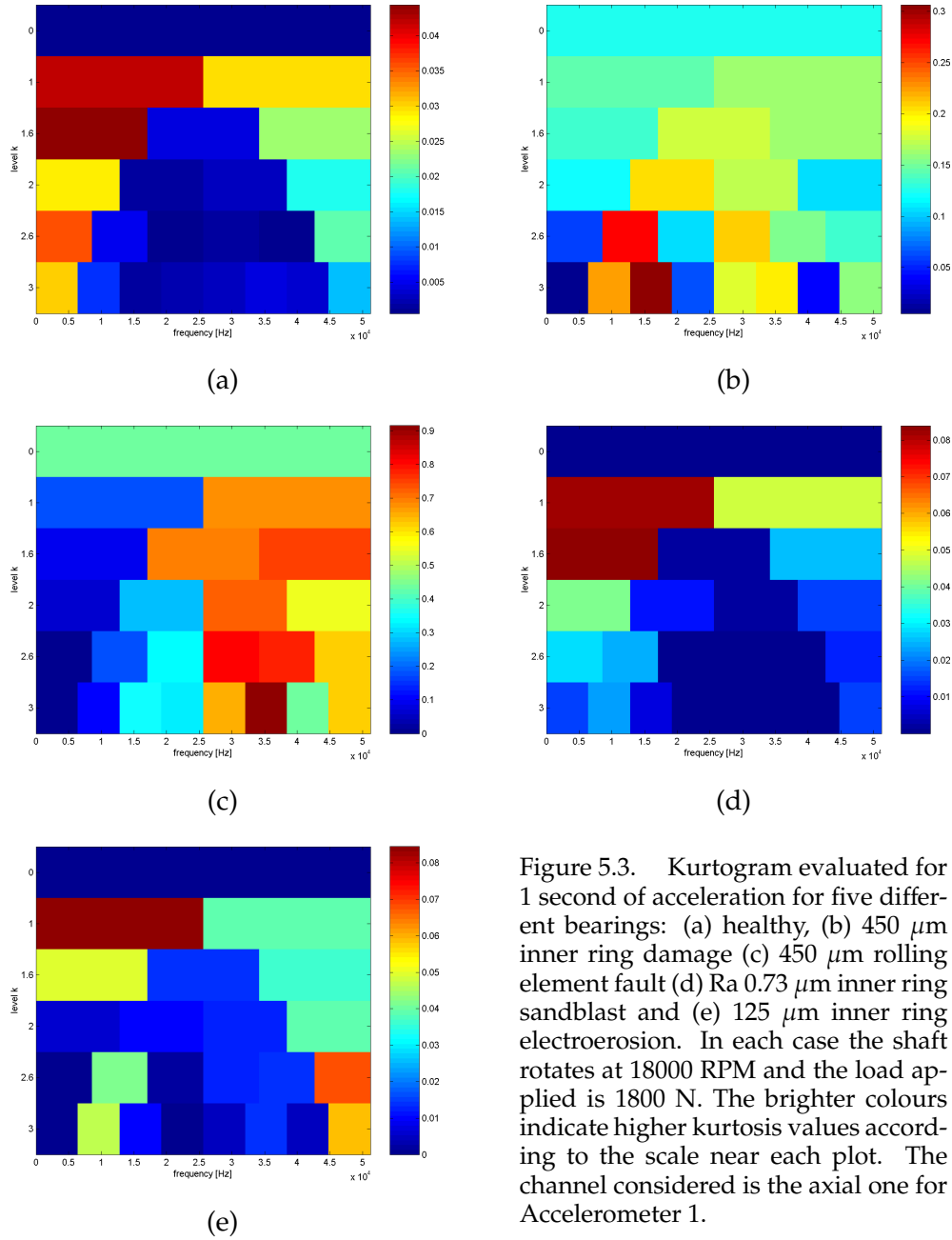


Figure 5.3. Kurtogram evaluated for 1 second of acceleration for five different bearings: (a) healthy, (b) 450  $\mu\text{m}$  inner ring damage (c) 450  $\mu\text{m}$  rolling element fault (d) Ra 0.73  $\mu\text{m}$  inner ring sandblast and (e) 125  $\mu\text{m}$  inner ring electroerosion. In each case the shaft rotates at 18000 RPM and the load applied is 1800 N. The brighter colours indicate higher kurtosis values according to the scale near each plot. The channel considered is the axial one for Accelerometer 1.

### 5.1.4 Methodology

Since now, the main features characterising the algorithm have been presented, together with the results that it produces. In fact, it has demonstrated its ability in the filtering of a signal in a particular informative band. We want to exploit this interesting aspects in order to develop a bearing fault diagnosis method. The next step that we could consider is the visualisation, and analysis, of the envelope and the Power Spectral Density (PSD) of the signal filtered according to the central frequency  $f_c$  and the bandwidth  $B_w$  obtained as output from the

Bearing type	max kurtosis	lev	$B_f$	$f_c$
0A	0.044	1.5	17067	8533
1A	0.306	3	6400	16000
2A	0.074	1	25600	12800
3A	0.153	2.5	6400	16000
4A	0.916	3	6400	35200
5A	0.467	3	6400	35200
6A	0.078	1	25600	12800
7A	0.084	1.5	17067	8533
15A	0.084	1	25600	12800

Table 5.1. Values obtained through the fast kurtogram algorithm: “max kurtosis” is the maximum value of kurtosis parameter in the frequency band found, whose amplitude is  $B_f$ , carrier frequency is  $f_c$  and is placed at level  $lev$  in the kurtogram plane. Refer to Table 4.3 for the various damage types, while the speed is 18000 RPM with an applied load of 1800 N. The channel considered is the axial one for Accelerometer 1.

kurtogram. Figure 5.5 shows the spectra of the previous signals of Figure 5.2, for the healthy and the inner ring damaged bearings, after being filtered in the frequency domain. While in the first case any particular peak could be identified, in the other one the spectrum reveals a lot of spectral lines exceeding the mean level of the remaining PSD. These lines could be associated to the fault presence.

From the inspection of the spectra, two indicators of fault presence could be defined. Our analysis, that brings to the definition of the parameters  $par_1$  and  $par_2$ , could be formalised through the following steps:

- I. the signal is passed through the Fast\_Kurtogram algorithm;
- II. the spectrum of the filtered signal is evaluated;
- III. the first 25 peaks are taken into account, avoiding the shaft rotation frequency multiples;
- IV. the parameters are evaluated as

$$par_1 = \frac{kurt_{max} \sum_{pk}}{mean_{sp}} \quad par_2 = kurt_{max} \quad (5.4)$$

where  $kurt_{max}$  is the maximum value of kurtosis given by the algorithm,  $\sum_{pk}$  is the sum of the considered peaks and  $mean_{sp}$  represents the mean value of the spectrum.

This procedure is operated on all kinds of bearings and for each speed and load. The healthy bearing is considered as reference element because of its nature and the mean  $\mu_{0A}$  and standard deviation  $\sigma_{0A}$  of values acquired in this case for a certain speed and load are used to define a threshold:

$$thrs = \mu_{0A} + \alpha \sigma_{0A}. \quad (5.5)$$

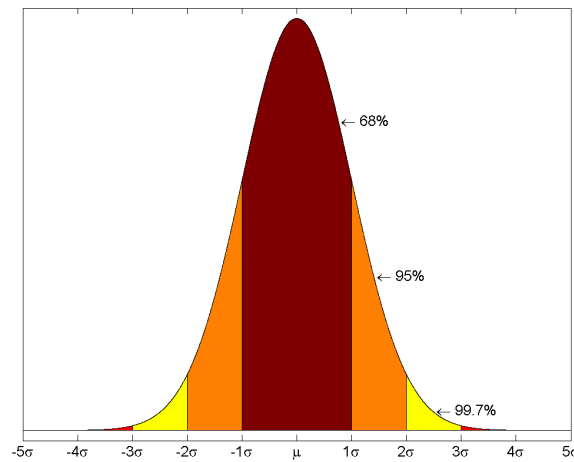


Figure 5.4. A plot of a normal distribution with the various 1-standard deviation widths. Different colours represent the values at various  $\sigma$ -distance from the mean  $\mu$ .

Once defined the threshold, the same parameters are acquired for the damaged bearing and compared with this value and this reasoning is taken into account:

```
if par_d > thrs
then damaged
else healthy
```

It is clear that is rather important the assign a proper value to the parameter  $\alpha$ , on whom, then, depends the threshold according to Equation (5.5).

In order to assign a proper value to this parameter, it is interesting to remember some notion about the normal distribution and about the so called “empirical rule”. As stated in statistics, in fact, if we consider a normal distribution, nearly all values lie within three standard deviations of the mean [CB02]. From Figure 5.4 it is possible to understand better the concept behind this rule. In fact, about 68% of values of the distribution lie within one standard deviation  $\sigma$  from the mean  $\mu$ , as showed by the brown part under the bell-shaped distribution. The orange area, instead, sum with the previous one, contains about 95% of values, that means a two  $\sigma$  distance from  $\mu$ . Finally, the yellow part, added to the others, arrives to include nearly all (99.7%) the distribution. In this last case, it is usually said that the values lie within three  $\sigma$  far from the mean. The ideas presented here could be an interesting hint in order to assign a proper value to the parameter  $\alpha$ .

Some other features that must be taken into account in the definition of a suitable threshold are lied to the concepts of False Alarms (FA) and Missed Alarms (MA). In fact, they are two important aspects in diagnostics in order to furnish a reliable diagnostic technique. The FA rate is correlated to the number of elements that, although being healthy, are instead classified as damaged. The MA refers, as said the name, to the faulty elements that are not

properly recognised. In general, between the two parameters, the FA rate has to be as lower as possible (~1-2%) because a damage warning could cause the machine stopping to check the state of the system. If this happens but no fault could be find, it means that a FA has verified. As could be imagined, this is an expensive and useless procedure that must be avoided as much as possible. Referring to (5.5), it is clear that lower is threshold  $thrs$ , higher is the FA rate, while higher is the threshold, lower is the MA rate. The conjunction of these last aspects with those related to the empirical rule allows us to think that a proper value that could be assigned to the parameter  $\alpha$  lies in the interval [2,3].

The next step is then related to the choice of a criteria that could join the results obtained in the various channels. A reasonable way to operate could be summarised as follows:

- I. consider a certain acquisition and evaluate the parameter;
- II. check for each chosen channel if the parameter is over the threshold evaluated for the channel itself;
- III. sum the results for each acquisition;
- IV. check if a certain voting criteria is confirmed.

The “voting criteria” could be seen as a way to classify the data according to their behaviour in the considered channels. In our case, we choose a 4/6 criteria: if at least 4 channels over the 6 considered detect a certain acquisition as damage, it is labelled as “damaged”, otherwise it is seen as “healthy”. In terms of FAs and MAs, we have a FA if at least 4 channels identify a 0A acquisition as “damaged”, while a MA originates when a faulty bearing data is instead classified as “healthy” by at most 3 channels.

### 5.1.5 Damage identification

As example of application of the method presented in the previous Section, we can consider the two signals showed in Figure 5.2, whose spectrum are presented in Figure 5.5. Since now, the step I and II of the procedure have been applied. The next stage is the identification of the 25 higher peaks: in Figure 5.6 are those signed with the red lines. The evaluation of  $par_1$  in Equation (5.4) in these two cases furnishes:

Bearing type	$kurt_{max}$	$sum_{pk}$	$mean_{sp}$	$par_1$
0A	0.044	432.036	3.375	5.682
1A	0.306	1815.539	3.829	145.297

From the results it is clear the big difference between the two bearings and it is very interesting in order to diagnose damages.

If this operation is executed for all the cases considered in Table 4.3, for each second of acquisition and for various channel, we obtain the plots in Figure 5.7. We select six channels,

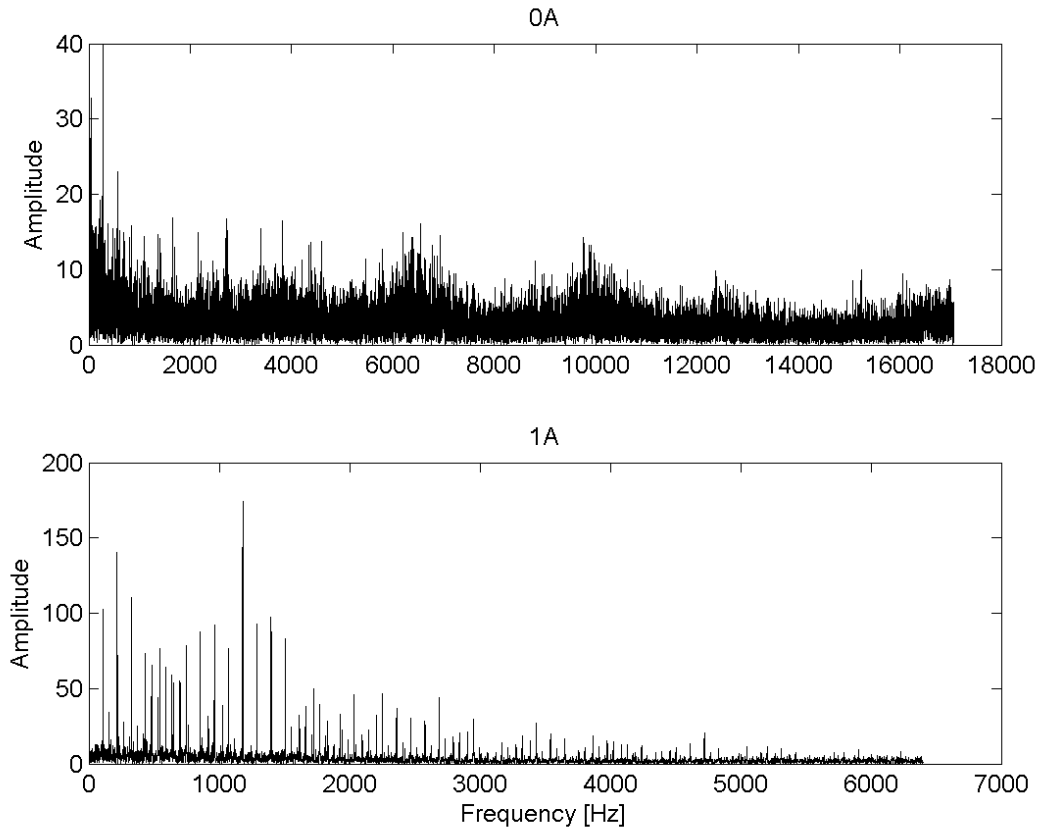


Figure 5.5. PSD of the two signals of Figure 5.2 as they comes out filtered by Fast\_Kurtogram algorithm: healthy bearing (top) and 450  $\mu\text{m}$  damage on the inner ring (bottom).

that are 1-2-3 and 7-8-9 in Figure 4.9(a). The simple observation of values obtained for each channel demonstrate that this method could be very interesting for damage detection. For example, if we consider Channel 1, it is clear that values in the healthy case (0) are rather different from those of bigger entity of damages (450  $\mu\text{m}$ ) on the inner ring (1) and on the rolling element (4). If, instead, we analyse a channel more distant from the damaged bearing such as Channel 7, we can notice that other types of damages localise farther. In fact, in this case, the sandblast and the electroerosion on the inner ring (7 and 15 respectively) acquire values bigger than the threshold  $thrs$ . A further observation regards the entity of damages, since, when it is smaller,  $par_1$  holds lower values, as could be intuitively imagined.

The results obtained for the parameter  $par_1$  (Equation (5.4)) evaluated for the data at 18000 RPM and 1800 N load could be seen in Figure 5.8(a). The voting criteria of 4/6, described in Section 5.1.4, is used and we could say that no FA are present. In fact, all the healthy data are correctly seen as “healthy” (green points). Instead, some MAs characterise the case 2A and 5A since they are erroneously labelled as “healthy” (green points). However, it is clear that a change in the voting criteria would modify the results and so this is a rather important

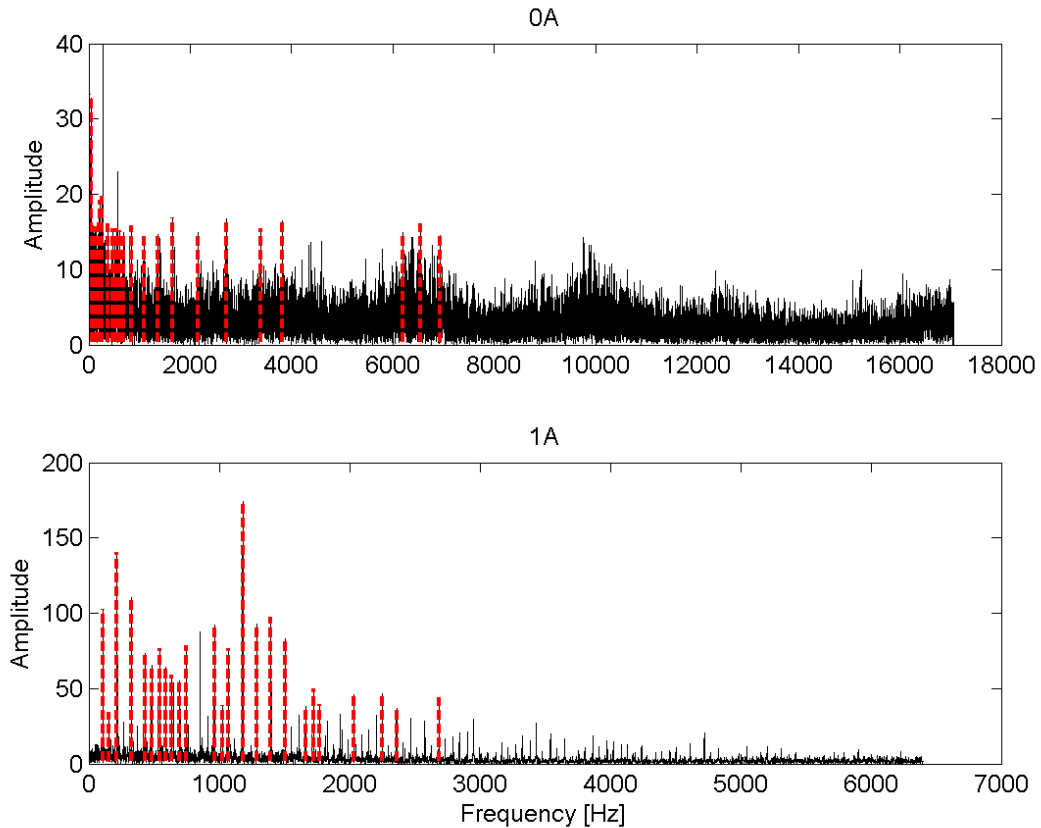


Figure 5.6. Power Spectral Density of the healthy bearing (top) and the one with  $450 \mu\text{m}$  damage on the inner ring (bottom). The red lines highlights the peaks that are taken into account in the parameter evaluation.

choice. In this case, in fact, if the voting criteria decreases to 3/6 these MAs would disappear. If, instead, we operate the same evaluation using  $par_2$  (Equation (5.4)) we notice that no MAs are present with the 4/6 voting (Figure 5.8(b)).

An interesting comparison could be done taking into account the value of the Root Mean Square parameter (Equation 1.1) computed for each second of acquisition. The same procedure of threshold evaluation based on the healthy elements is operated and also in this case the 4/6 voting criteria is applied. The results are presented in Figure 5.8(b). As for  $par_1$ , any FA is obtained, while many more MAs could be observed. In particular, we could say that this parameter is not able to detect properly the lower entities of inner ring and rolling element damage (2A, 3A, 5A, 6A). However, differently from the  $par_2$ , in this case it is not sufficient to simply vary the voting criteria since for 2A and 3A all the channels furnish MAs. This comparison could help us in understand the potential of the parameter obtained through the Fast\_Kurtogram algorithm, since it is able to recognise also small damages.

In Table 5.2 the % of acquisitions labelled as “damaged” by the algorithm adopting a criteria of 4/6 are presented for all the speeds and loads. Both parameters  $par_1$  and  $par_2$  are

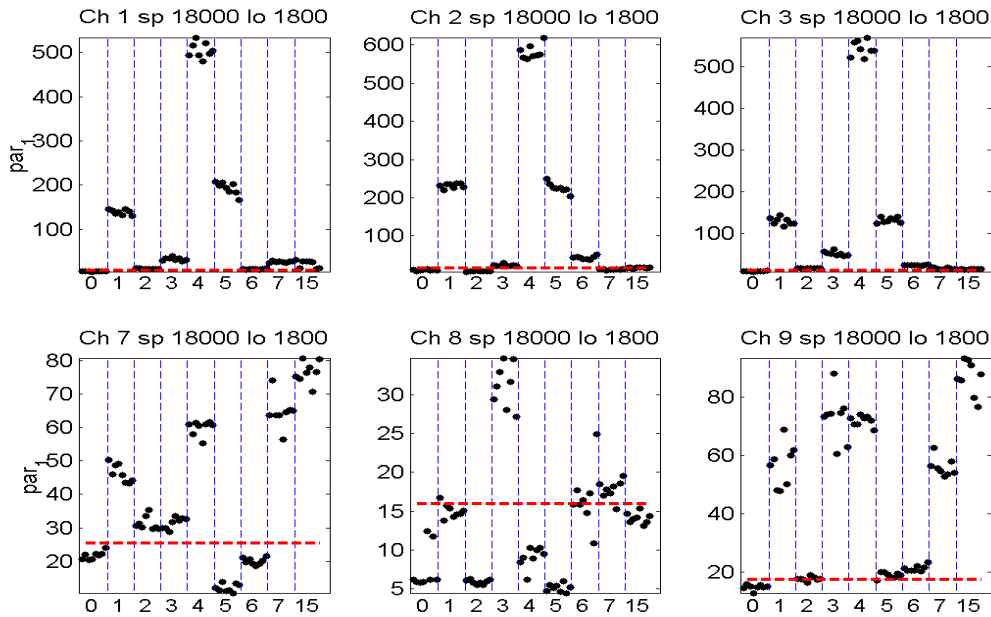


Figure 5.7. Parameter  $par_1$  evaluated for the six selected channels signals in the cases of various types of bearings (at 18000 RPM speed and 1800 N load). The damages codes are in Table 4.3. The red line represents the threshold  $thrs$  obtained according to Equation (5.5) assuming  $\alpha = 3$ .

taken into account. The first aspect that could be noticed is the absence of FAs in all the cases, that is, no data in healthy case are recognised as “damaged”. This is a good point, but it must be said that the main cause stands in the definition of the threshold  $thrs$ . The bigger damages on the inner ring and on the rolling elements are almost always recognised (1A and 4A). The sandblast on the inner ring (7A), instead, is well recognised for all loads but only for the speeds greater than 12000 RPM. Another particular behaviour is noticed for the electroerosion on the inner ring (15A): the fault is correctly labelled only at the speeds 12000 and 18000 RPM. These two facts could be explained thinking that these two defects rises their effects at these particular shaft speeds. Comparing the % for each speed, it seems that the performances are better especially when the load is absent or maximum (1800 N). The comparison between the two parameters  $par_1$  and  $par_2$ , instead, does not show a particular predominance of one of them in damage detection.

### 5.1.6 Various disassemblies

An interesting evaluation that could be done analysing the data acquired during the endurance test presented in Section 4.4.1.2 regards the influence of disassemblies on the system. As explained in Section 4.4.1.2, the bearing, characterised by a  $450 \mu\text{m}$  on the rolling element, is subjected to about 200 hours of running in order to analyse the damage evolution. During



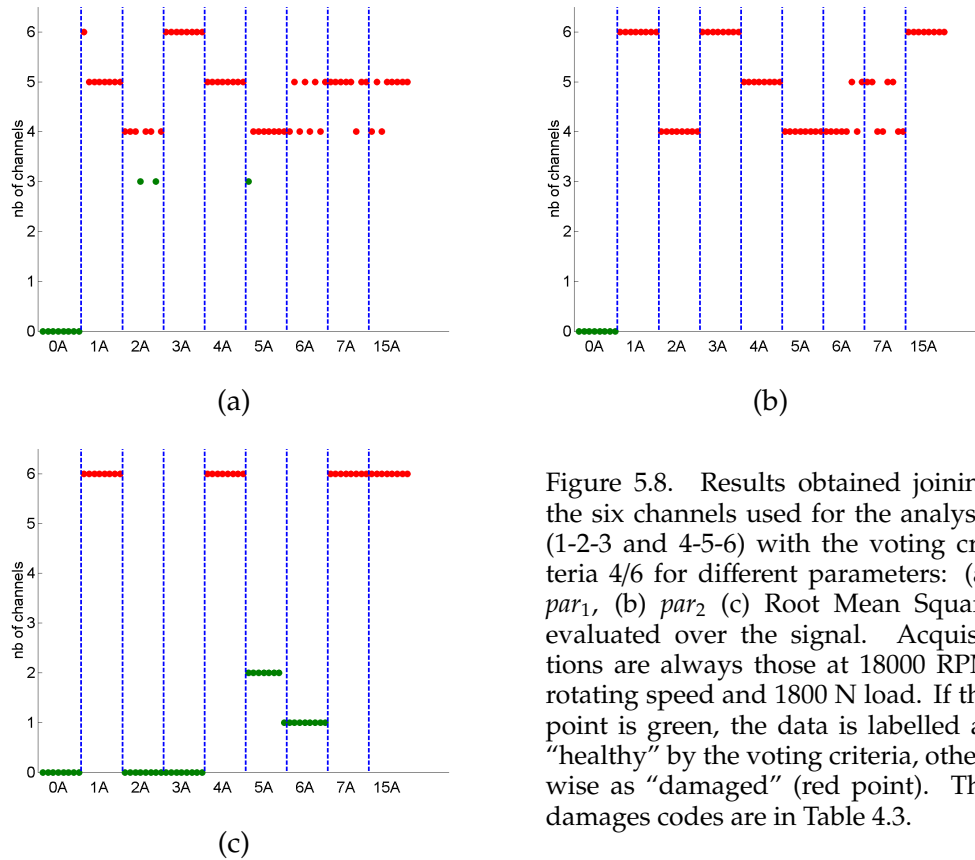


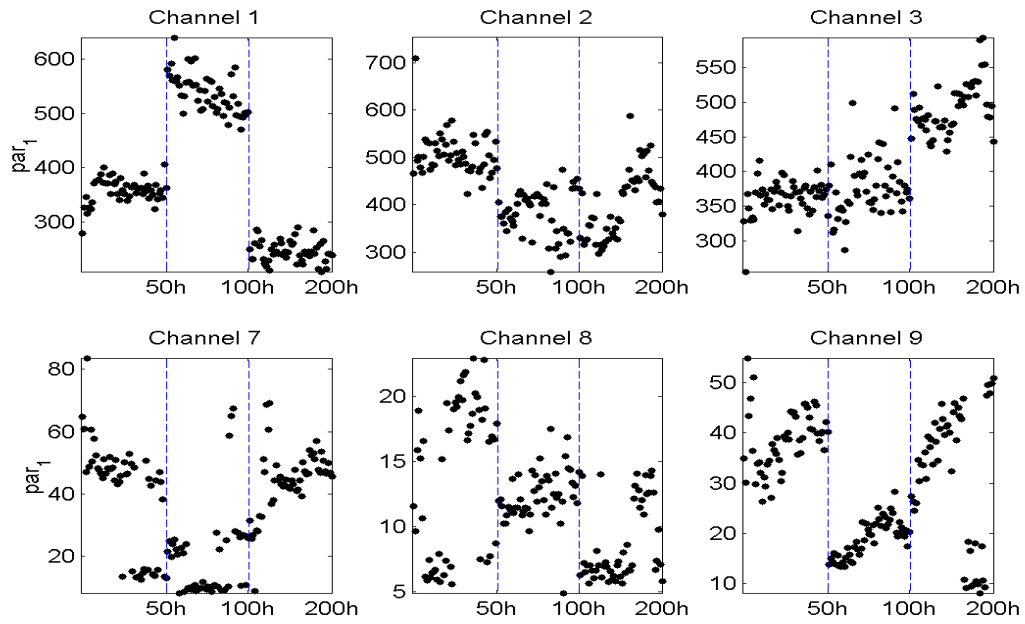
Figure 5.8. Results obtained joining the six channels used for the analysis (1-2-3 and 4-5-6) with the voting criteria 4/6 for different parameters: (a)  $par_1$ , (b)  $par_2$  (c) Root Mean Square evaluated over the signal. Acquisitions are always those at 18000 RPM rotating speed and 1800 N load. If the point is green, the data is labelled as "healthy" by the voting criteria, otherwise as "damaged" (red point). The damages codes are in Table 4.3.

this time, it is taken apart twice in order to check damage evolution: after 50 hours and after other 50 hours of running. Correlated to these events, it would be interesting to check if the parameters we evaluate remain almost unvaried before and after these disassemblies. With this target,  $par_1$  and  $par_2$  are computed and showed in Figure 5.9(a) and 5.9(b) respectively. Only the acquisitions related to a stable condition reached along a day are considered, that is, when the bearing has come at almost constant temperature. It has been noticed that this usually happens after about three hours after the switch on.

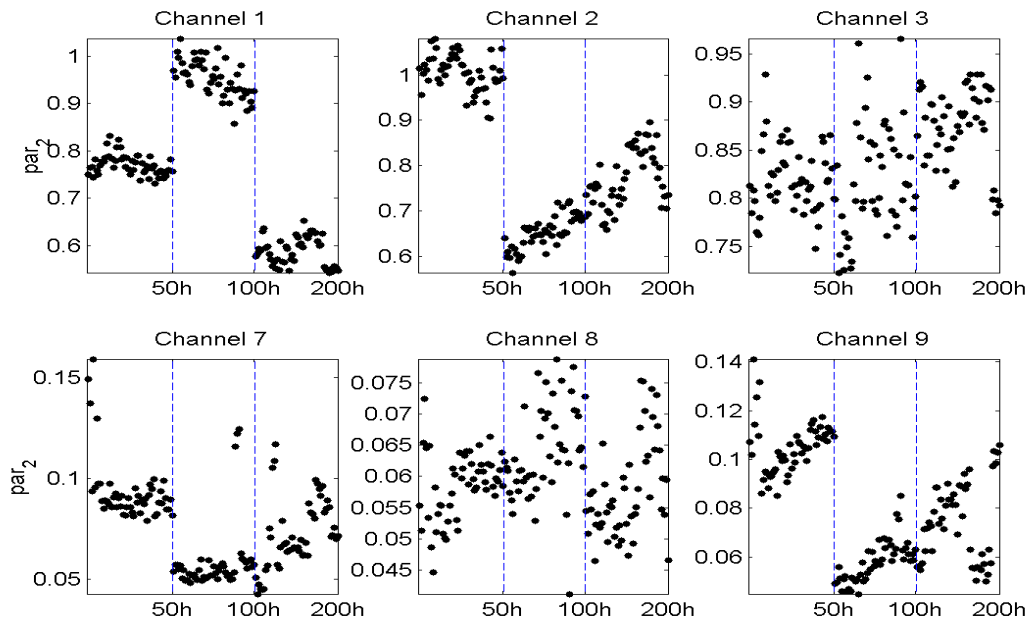
The first aspect to notice in the Figures is that each channel has a different behaviour, probably due to the different direction (axial, radial or tangential) and position (1-2-3 are on the damaged bearing). For example, Channel 1, which registers the axial direction in the bearing proximity, is the one showing the most consistent variations ( $\approx 30\%$ ) between the various phases of the running, both for  $par_1$  and  $par_2$ . It could be said that the first bearing inspection (after 50 hours of running) has brought a general changing in all the channels analysed. Probably, when the test rig has been prepared, the shaft has been fastened to the structure in a slightly different way compared to the previous assembly. However, some channels, such as the tangential one near the bearing position (Channel 3) and the radial of the other accelerometer (Channel 8), seems to be less influenced by this operation.

Speed	Load	Parameter	0A	1A	2A	3A	4A	5A	6A	7A	15A
6000	/	<i>par</i> <sub>1</sub>	0	100	100	100	100	100	12.5	100	100
		<i>par</i> <sub>2</sub>	0	100	100	100	100	100	87.5	100	100
6000	1000	<i>par</i> <sub>1</sub>	0	100	62.5	0	100	100	0	0	25
		<i>par</i> <sub>2</sub>	0	100	100	12.5	100	100	12.5	37.5	100
6000	1400	<i>par</i> <sub>1</sub>	0	100	100	87.5	100	100	75	100	100
		<i>par</i> <sub>2</sub>	0	100	100	100	100	100	100	62.5	100
6000	1800	<i>par</i> <sub>1</sub>	0	100	0	100	100	62.5	0	25	62.5
		<i>par</i> <sub>2</sub>	0	100	0	100	100	62.5	0	0	87.5
12000	/	<i>par</i> <sub>1</sub>	0	100	100	100	100	75	100	100	100
		<i>par</i> <sub>2</sub>	0	100	0	100	100	100	100	100	100
12000	1000	<i>par</i> <sub>1</sub>	0	100	37.5	37.5	100	100	0	100	100
		<i>par</i> <sub>2</sub>	0	100	100	75	87.5	87.5	0	100	100
12000	1400	<i>par</i> <sub>1</sub>	0	100	87.5	50	100	100	0	100	100
		<i>par</i> <sub>2</sub>	0	100	0	0	75	100	0	100	100
12000	1800	<i>par</i> <sub>1</sub>	0	100	100	100	100	100	50	100	100
		<i>par</i> <sub>2</sub>	0	100	100	100	100	100	100	100	100
18000	/	<i>par</i> <sub>1</sub>	0	100	0	0	100	75	50	100	100
		<i>par</i> <sub>2</sub>	0	100	0	62.5	100	75	100	100	100
18000	1000	<i>par</i> <sub>1</sub>	0	87.5	50	0	87.5	0	100	100	100
		<i>par</i> <sub>2</sub>	0	100	0	0	62.5	0	87.5	100	100
18000	1400	<i>par</i> <sub>1</sub>	0	87.5	0	0	100	0	0	100	100
		<i>par</i> <sub>2</sub>	0	100	25	62.5	100	0	0	100	100
18000	1800	<i>par</i> <sub>1</sub>	0	100	75	100	100	87.5	100	100	100
		<i>par</i> <sub>2</sub>	0	100	100	100	100	100	100	100	100
24000	/	<i>par</i> <sub>1</sub>	0	100	100	100	100	0	100	100	0
		<i>par</i> <sub>2</sub>	0	100	100	100	100	0	0	100	100
24000	1000	<i>par</i> <sub>1</sub>	0	100	0	0	0	100	0	100	25
		<i>par</i> <sub>2</sub>	0	100	0	0	62.5	100	0	100	0
24000	1400	<i>par</i> <sub>1</sub>	0	100	87.5	0	100	100	0	100	100
		<i>par</i> <sub>2</sub>	0	100	0	0	100	100	0	100	0
24000	1800	<i>par</i> <sub>1</sub>	0	100	100	100	100	100	0	100	75
		<i>par</i> <sub>2</sub>	0	100	100	100	100	100	37.5	100	100
30000	/	<i>par</i> <sub>1</sub>	0	0	0	0	0	0	0	0	0
		<i>par</i> <sub>2</sub>	0	100	62.5	0	100	0	0	75	87.5
30000	1000	<i>par</i> <sub>1</sub>	0	100	0	62.5	100	100	0	100	0
		<i>par</i> <sub>2</sub>	0	100	0	62.5	100	100	0	100	100
30000	1400	<i>par</i> <sub>1</sub>	0	100	0	100	100	100	0	100	100
		<i>par</i> <sub>2</sub>	0	100	0	100	100	100	0	100	100
30000	1800	<i>par</i> <sub>1</sub>	0	100	0	62.5	100	100	0	100	100
		<i>par</i> <sub>2</sub>	0	100	0	100	100	100	0	100	100

Table 5.2. Results obtained for the two parameters *par*<sub>1</sub> and *par*<sub>2</sub> considering the voting criteria 4/6, considering all the speeds and loads for the various bearing types. The values are the % of acquisitions labelled as “damaged” by the algorithm.



(a)



(b)

Figure 5.9. Parameters  $par_1$  and  $par_2$ , (a) and (b) respectively, evaluated for six selected channels in the endurance test only at a stable condition of the system for each day of acquisition. The blue vertical lines indicates when the bearing is taken apart for inspection.

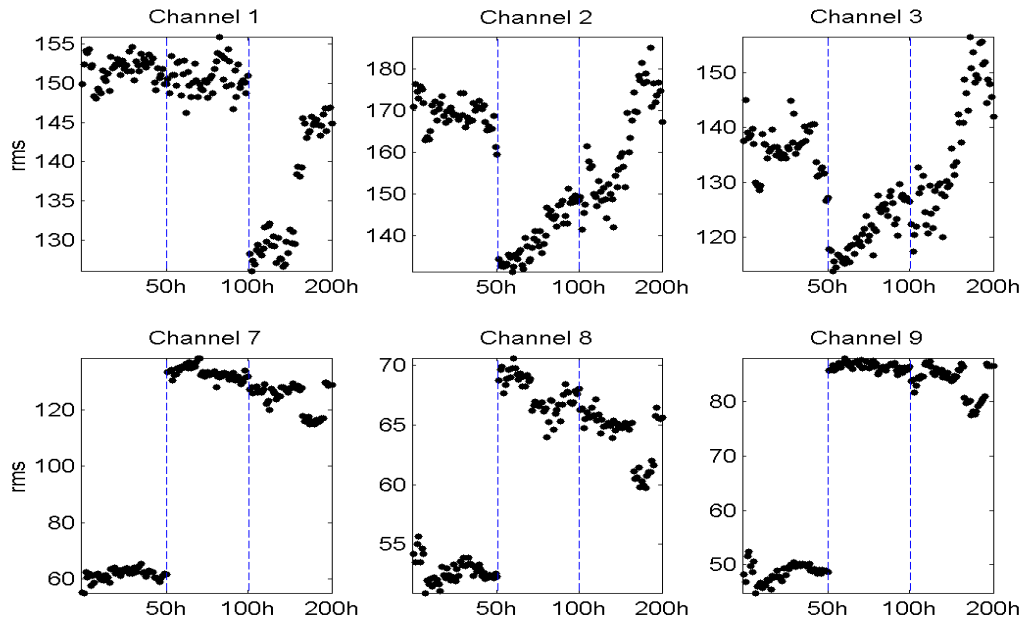


Figure 5.10. RMS parameter evaluated for six selected channels in the endurance test only at a stable condition of the system for each day of acquisition. The blue vertical lines indicates when the bearing is taken apart for inspection.

When the state of the bearing is checked the second time, after 100 hours of running, variations are again noticed mainly in the already cited Channel 1. All the other channels shows, in general, approximately similar values (Channel 2, 8 for  $par_1$  and Channel 3 and 8 for  $par_2$ ) or an overall increasing trend that is not modified by the disassembly (Channel 3, 7 and 9 for  $par_1$  and Channel 2, 7 and 9 for  $par_2$ ). This last aspect is rather important since it means that this events does not modify the general increasing trend of the parameter due to the evolution of the damage.

An interesting comparison could be done evaluating the Root Mean Square parameter for the same acquisition (Figure 5.10), since its values indicate the system overall vibration. Also in this case, Channel 1 is the only one component showing a consistent variation after the 100 hours inspection. All the other channels, instead, seems to be more influenced by the other inspection, operated after 50 hours of running. The observation of the values registered after this disassembly demonstrates a different behaviour in the two accelerometers. In fact, while Channels 2 and 3 reveal an increasing trend, Channels 7, 8 and 9 are characterised by decreasing values.

In the previous section, the ability of the presented method in damage detecting has been proved. It could be interesting to try to apply this also to the data used in this section. The threshold  $thrs$  is defined as explained previously, but, since any healthy data are present, we consider a section of the endurance test as the training part. All the other acquisitions are

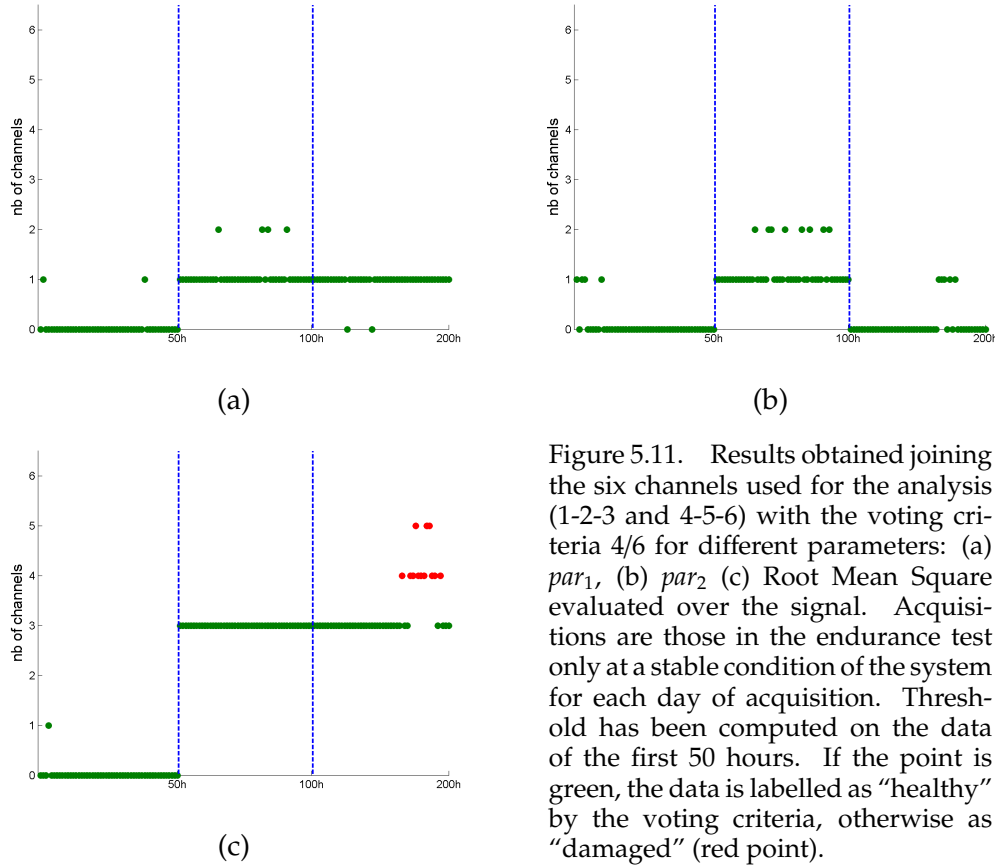


Figure 5.11. Results obtained joining the six channels used for the analysis (1-2-3 and 4-5-6) with the voting criteria 4/6 for different parameters: (a)  $par_1$ , (b)  $par_2$  (c) Root Mean Square evaluated over the signal. Acquisitions are those in the endurance test only at a stable condition of the system for each day of acquisition. Threshold has been computed on the data of the first 50 hours. If the point is green, the data is labelled as “healthy” by the voting criteria, otherwise as “damaged” (red point).

Parameter	Train			Test			Train			Test		
	P1	P2	P3	P2	P1	P3	P3	P1	P2			
$par_1$	0	0	0	0	0	0	0	0	0			
$par_2$	0	0	0	0	0	0	0	0	0			

Table 5.3. Results obtained for the two parameters  $par_1$  and  $par_2$  considering the voting criteria 4/6. “P1” stands for first part of the test, “P2” for the second and “P3” for the third, according to the various disassemblies. The values are the % of acquisitions labelled as “damaged” by the algorithm.

tested in order to see where they place respect to  $thrs$ . This is done for the six channels as before and the 4/6 voting criteria is taken into account. The results obtained allows us to evaluate if the method proposed is not influenced by the changing of the system. Figure 5.11 shows a graphical visualisation of this idea for the first 50 hours of acquisitions considered at “healthy” and the others are checked. It is clear that for both parameters  $par_1$  and  $par_2$  the disassembly do not seem to influence too much the results, since less that 3 channels of the 6 have a value over the threshold. The comparison with the RMS shows, instead, as in the last part of the test, this parameter is seen as “damaged” is some case. This would mean that RMS is more sensible to the system disassembly than the other two parameters. Table 5.3

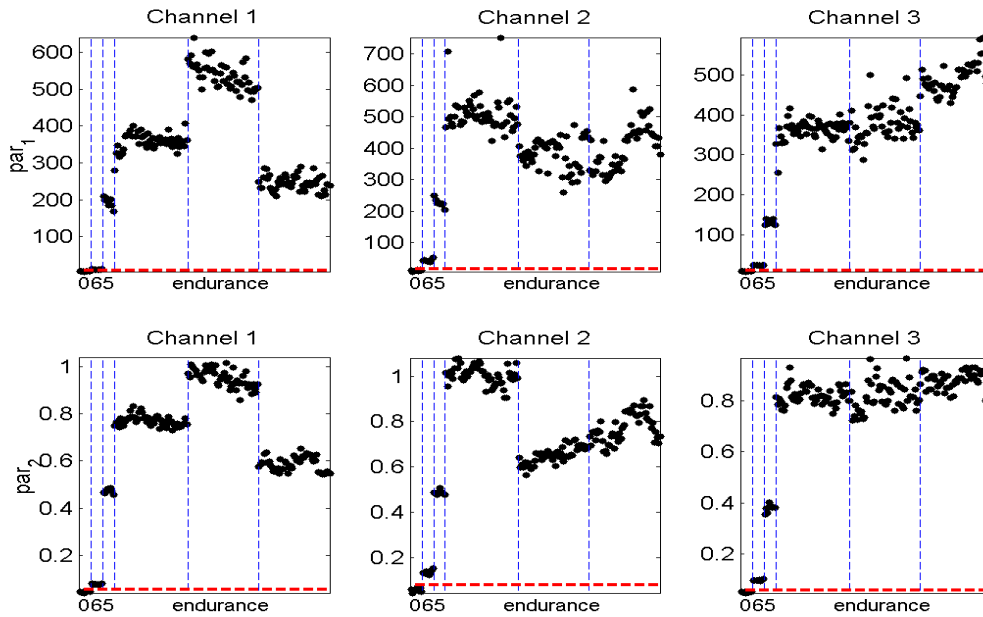


Figure 5.12. Parameters  $par_1$  and  $par_2$  evaluated for the healthy bearing (0), for the  $150\ \mu\text{m}$  and  $250\ \mu\text{m}$  damage on the rolling element, (6) and (5) respectively, and for the endurance test on the  $450\ \mu\text{m}$  damage on the rolling element. For this last case only the acquisitions at stable conditions are considered. The rotating shaft is 18000 RPM and the load applied is 1800 N. The red line represents the threshold  $thrs$  obtained according to Equation (5.5) assuming  $\alpha = 3$ . The blue lines in the endurance part are located where the bearing inspections hold.

demonstrates clearly that each time a different part of the test is considered to evaluate the threshold  $thrs$ , it does not happen that at least 4 channels of the 6 are over  $thrs$ . So, we could say that these parameter does not seem to be influenced by the system variations due to the fact that the bearing has been taken apart.

### 5.1.7 Damage evolution

A particular analysis that could be led considering data acquired on the Setup I described in Section 4.4.1, concerns the evolution of the damage on the rolling element. In fact, we could compare the parameter  $par_1$  and  $par_2$  evaluated for the healthy bearing, for the  $150\ \mu\text{m}$  and  $250\ \mu\text{m}$  damage on the rolling element for and the endurance test run on the  $450\ \mu\text{m}$  damage on the same component. Since in this last case the shaft rotates at 18000 RPM and a load of 1800 N is applied (Section 4.4.1.2), data acquired for the other entity of fault at the same operating conditions are considered too. Results are presented in Figure 5.12, where the channels of accelerometer placed directly on the bearing that is changed are considered (Channels 1, 2 and 3). It is clear from the pictures how the distances from the threshold defined by healthy data is generally proportional to the damage entity. In fact, when the fault is lower (6), the values are close to the healthy ones, while when it is increased (5) the

parameters are bigger. Finally, all along the endurance test, the greatest distance from the threshold is observed.

This particular analysis, related to the various distances from *thrs*, could gain some interest because it could be used in order to quantify the entity of a damage. However, we could see that it holds mainly, according to the data acquired, for the rolling element indentation. In fact, we can observe the results presented in Figure 5.7 for the inner ring indentation. Referring to 1, 2 and 3 (450  $\mu\text{m}$ , 250  $\mu\text{m}$  and 150  $\mu\text{m}$  respectively) we could see that the previous reasoning is no more consistent. The bigger entity (1) parameters locate, in general, rather far from the threshold, as could be expected, but this do not hold for the other two cases. The medium entity fault (2) takes, in fact, bigger values compared to the smallest one (3), unlike the expected behaviour. This happens for all the channels considered during the analysis.

## 5.2 External condition removal in bearing diagnostics through EMD and One-Class SVM

A rather worth of interest aspect is the search for methods able to remove effects produced in vibrations by various factors, such as environmental temperature or varying speed and load. Some works have been presented related to this topic. Some examples are presented in [PGT<sup>+</sup>11] and in [MLBGA11], where the multi-variate statistical technique named Principal Component Analysis (PCA) is used successfully in bearing fault detection and rotating shaft. Other factors influencing vibrations, in the case of rotating elements, are varying load and speed. As already explained in Chapter 2, a variation in these factors produces some difficulties in recognising the presence of fault in a signal.

The influence of elements such as temperature, speed, load in the performances of a diagnostic technique came out during our work and our diagnostics tests on some industrial aircraft test rigs. The need to deepen this aspects, takes us to project many acquisitions on our DIRG laboratory test rig. Here, it is possible to monitor and vary these parameters in a controlled environment. The acquisitions realised on the Setup II described in Section 4.4.2 have been mainly planned in order to answer to these problems. In fact, the intent of our work is to find a parameter able to remove the influence of various external conditions in order to detect properly a damage in a roller bearing.

### 5.2.1 One-Class Support Vector Machine

The Support Vector Machine (SVM) method has been presented in Chapter 3. As said, it is well suited for classification, because given some data points which belong to a certain class it is able to state the class a new data point would be in. If we consider an  $n$ -dimensional input data made up of a number of samples belonging to two classes, namely positive or negative, SVM constructs a hyperplane that separates them. Moreover, this boundary would satisfy the condition that the distance from the nearest data points in each class is maximal. In this way, an optimal separating hyperplane is created, namely the maximum margin. The points in both classes nearest to this margin are called support vectors and, once selected, they contain all the information necessary to define the classifier.

Every time a new element appears, it could be classified according to where it places respect to the separating hyperplane. SVM could also be applied in case of non-linear classification using a function  $\phi(x)$  that maps the data onto a high-dimensional feature space, where the linear classification is then possible. Furthermore, if a kernel function  $K(x_i, x_j) = (\phi^T(x_i) \cdot \phi(x_j))$  is applied, it is not necessary to evaluate explicitly  $\phi(x_i)$  in the feature space. Various kernel function could be used, such as linear, polynomial or Gaussian RBF. This property enables SVM to be used in case of very large feature spaces because the dimension of classified vectors does not influence directly the SVM performance.



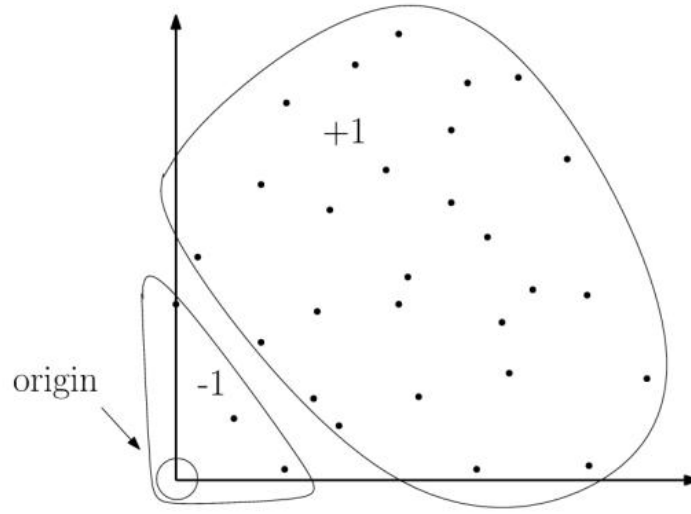


Figure 5.13. One-Class SVM classifier where the origin is the only member of one class.

It is clear that in the previous cases, two or more classes of data are given since the beginning of the analysis. In more general diagnostic applications, instead, only one type of data objects is usually acquired: the healthy one. This could be seen as the detection of patterns in data that do not conform to a well defined notion of normal behaviour, so we could refer to anomaly detection. One-Class SVM is the application of the SVM approach to the general concept of anomaly detection, as presented by Schölkopf *et al.* in [SWS<sup>+</sup>00]. In their method they construct a hyper-plane around the data, such that this is maximally distant from the origin and can separate the regions that contain no data. They propose to use a binary function that returns “+1” in region containing the data and “-1” elsewhere. For a hyper-plane  $w$  which separates the data  $x_i$  from the origin with maximal margin  $\rho$ , the following quadratic program has to be solved:

$$\min_{w \in F, \xi \in \mathbb{R}^n, \rho \in \mathbb{R}} \frac{1}{2} \|w\|^2 + \frac{1}{vn} \sum_i \xi_i - \rho \quad (5.6)$$

$$\text{subject to } (w \cdot \Phi(x_i)) \geq \rho - \xi_i, \quad \xi_i \geq 0 \quad (5.7)$$

where  $\xi$  represents the slack variable and  $v$  is a variable taking values between 0 and 1 that monitors the effect of outliers (hardness and softness of the boundary around data).

If  $w$  and  $\rho$  solve the minimisation problem presented in Equations (5.6) - (5.7), the decision function

$$f(x) = \text{sign}((w \cdot \Phi(x_i)) - \rho) \quad (5.8)$$

is positive for most instances representing the majority of data.

Figure 5.13 shows graphically the idea presented here, with only few points around the origin that are negatively labelled.

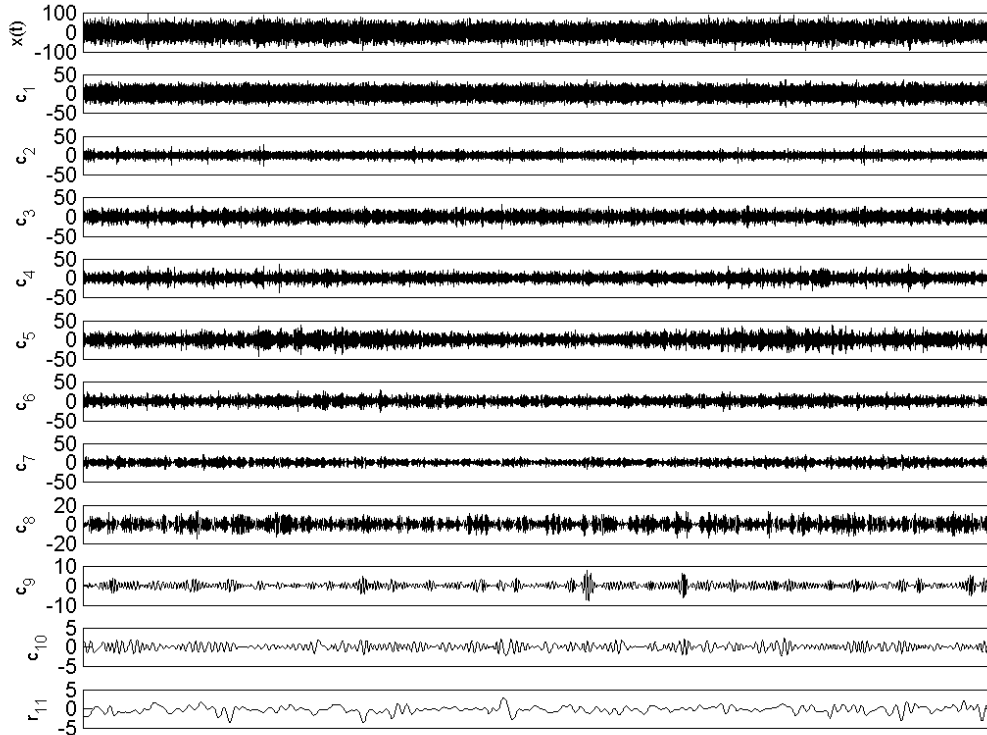


Figure 5.14. 1 second of acceleration signal  $x(t)$  and its decomposition to the 10th IMF  $c_i$  for the healthy bearing (0A).  $r_{11}$  represent the residual of the signal.

## 5.2.2 Empirical Mode Decomposition application

The innovative technique in the time-frequency domain called the Empirical Mode Decomposition (EMD) has already been presented in Section 2.3. Its main ability is the decomposition of any complicated signal into a collection of Intrinsic Mode Functions (IMFs) based on the local characteristic time scale of the signal. As already said, it is self-adaptive because the IMFs, working as the basis functions, are determined by the signal itself rather than being pre-determined. Hence, EMD is highly efficient in non-stationary data analysis.

An example of how a signal results after the decomposition through this technique is presented in Figures 5.14 and 5.15. Here, 1 second of the original accelerations acquired for the healthy case and for the 450  $\mu\text{m}$  rolling element fault respectively are considered ( $x(t)$ ) and decomposed with EMD. Each  $c_i$  represents an Intrinsic Mode Function according to the definition and the method explained in Section 2.3. In this case and in the following, we decide to stop the decomposition at the 10th level, since in these components the most dominant fault information is contained. Figure 5.16(a) shows the energy content evaluated at each IMF for the signals presented in Figures 5.14 and 5.15. Although there are some

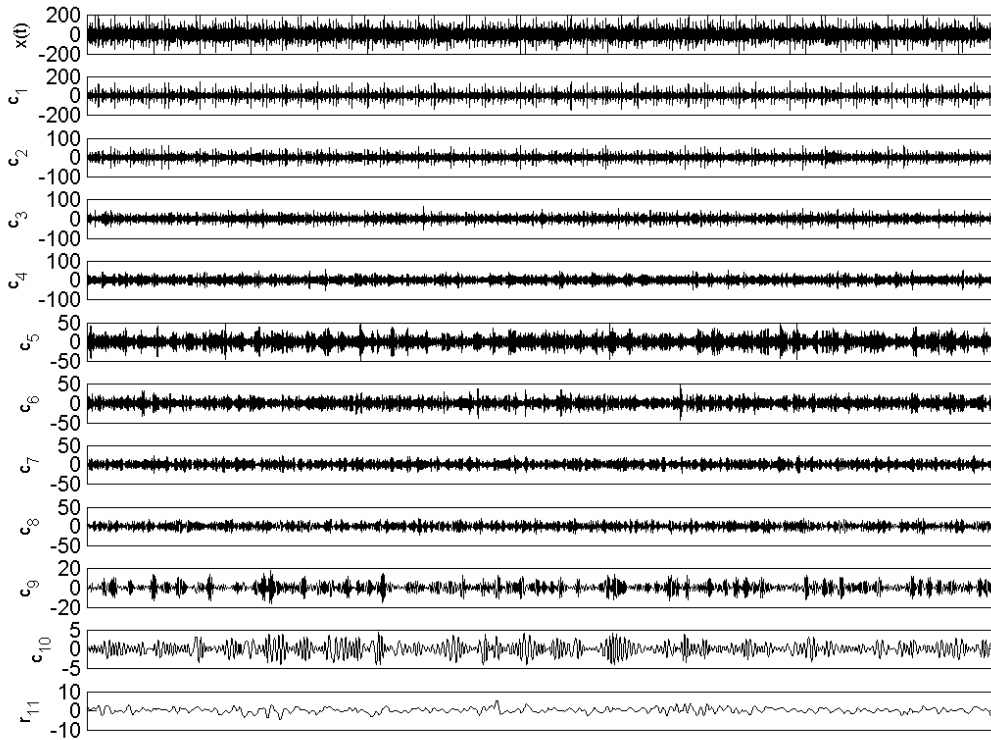


Figure 5.15. 1 second of acceleration signal  $x(t)$  and its decomposition to the 10th IMF  $c_i$  for the  $450 \mu\text{m}$  rolling element indentation on the bearing (4A).  $r_{11}$  represent the residual of the signal.

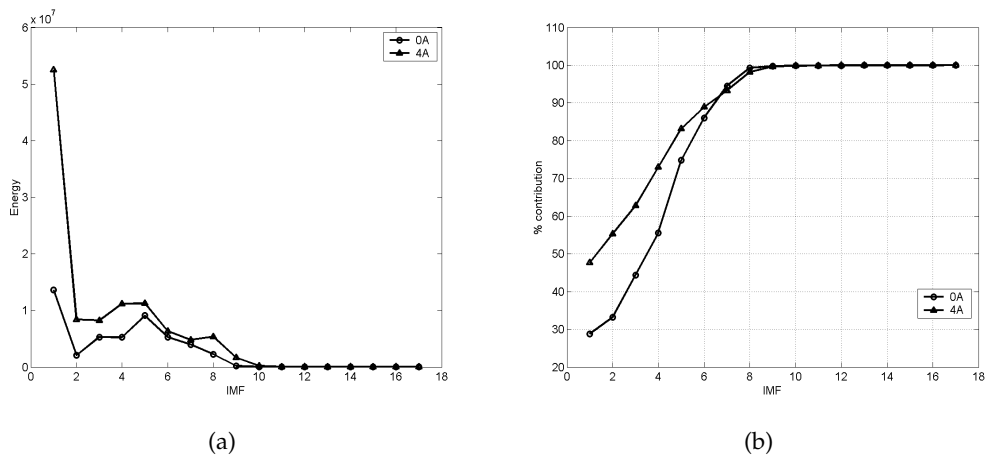


Figure 5.16. (a) Energy content of each IMF for the healthy bearing (0A) and for the  $450 \mu\text{m}$  rolling element damaged bearing (4A) of Figures 5.14 and 5.15; (b) overall percentage in the energy contribution of each IMF for the two cases.

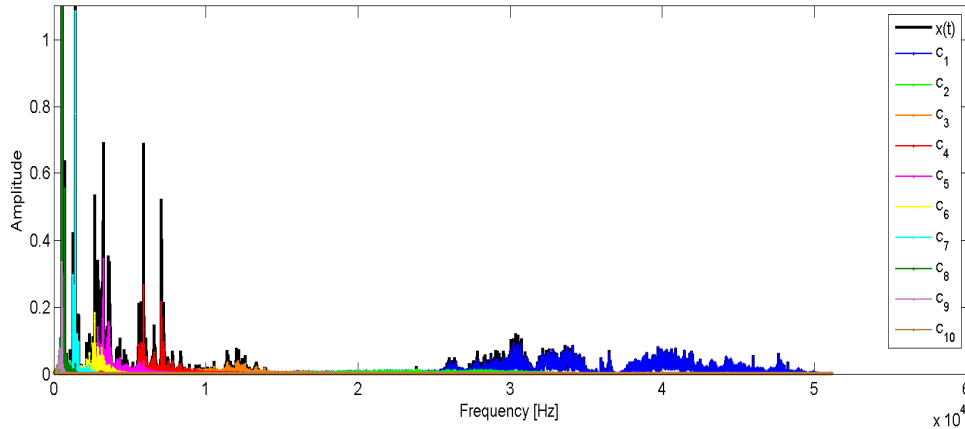


Figure 5.17. Power Spectral Density for the signal and for each IMF for the damaged case presented in Figure 5.15. Each colour represents a  $c_i$  component.

difference between the two cases in the first components, it could be seen that from the 10th IMF no more contribution to the energy evaluation seems to be added. This is confirmed in the plot in Figure 5.16(b), where the contribution percentage for each IMF is computed. It is clear how between the 9th and the 10th  $c_i$  almost the 100% of energy of the signal is reached for both the bearings. So this strengthens the idea that for our analysis it would be sufficient to stop the decomposition at the 10th IMF.

An interesting observation could be given referring to the analysis presented by Flandrin *et al.* in [FRG04] and cited in Section 2.3. As already said, they assert, with the aid of numerical experiments, that EMD acts essentially as a dyadic filter bank. In particular, the filter associated to the mode  $c_1$  is a highpass one while the modes of higher indexes are characterised by a set of overlapping bandpass filters. Moreover, for the modes of index  $(k + 1), k \geq 2$  a frequency domain which is roughly the upper half-band of that of the previous residual of index  $k$  is defined. It is possible to validate these affirmations analysing the elements showed in Figure 5.15: signal and IMFs in the roller fault. The Power Spectral Density for all these elements could be observed in Figure 5.17. The original signal spectrum is plotted with the black line while the frequency contents of the other components could be identified with different colours. It is clear how each IMF locates in a different position in the frequency scale, starting from the highest with  $c_1$  to the lowest with  $c_8, c_9$  and  $c_{10}$ . This reflects the results obtained by Flandrin *et al.*: in fact  $c_1$  contains the signal as if it has been passed through a high pass filter, while the frequency content for the other modes is represented by smaller and smaller bands.

The evaluation of IMFs in this section is done through the algorithm developed by Rilling, Frandrin and Gonçalves, which is available as freeware code on the Internet<sup>1</sup>. In this algorithm, they propose and use an improvement of the usual approach in the sifting process

<sup>1</sup><http://perso.ens-lyon.fr/patrick.flandrin/emd.html>

stopping criteria [RFG03]. They base it on two thresholds  $\theta_1$  and  $\theta_2$  whose goal is to assure globally small fluctuations and, at the same time, to take into account locally large excursions. So, they introduce the mode amplitude

$$a(t) = \frac{e_{max}(t) - e_{min}(t)}{2},$$

together with an evaluation function

$$\sigma(t) = \left| \frac{m(t)}{a(t)} \right|.$$

The sifting process is so iterated in two times: until  $\sigma(t) < \theta_1$  for a  $(1-\alpha)$  part of the total duration and until  $\sigma(t) < \theta_2$  for the other one. The default values considered for the parameters are  $\theta_1 = 0.05$ ,  $\theta_2 = 0.5$  and  $\alpha = 0.05$ .

### 5.2.3 Methodology I

The background and the theoretical aspects of the two methods that now we want to use jointly have been exposed in the previous sections. Our goal is the search for a method able to identify a damage in a rotating element of a roller bearing by removing the effect of external conditions influencing vibrations. Our diagnosis method consists of different steps:

- I. Collect vibration signals under various condition of speed and radial load applied, both for a healthy and a damaged bearing;
- II. Apply EMD and decompose the original signal into IMFs; then choose the first  $n$  to extract the features used during the analysis;
- III. Evaluate the total energy for the  $n$  selected IMFs:

$$E_j = \int_{-\infty}^{+\infty} |c_j(t)|^2 dt \quad j = 1, \dots, n; \quad (5.9)$$

- IV. Create a feature vector with the energies of the  $n$  selected IMFs:

$$F = [E_1, \dots, E_n]; \quad (5.10)$$

- V. Normalise the feature vector dividing  $F$  for this value:

$$E_N = \sqrt{\sum_{j=1}^n |E_j|^2}; \quad (5.11)$$

- VI. Obtain the  $n$ -dimension normalised feature vector:

$$F' = [E_1/E_N, \dots, E_n/E_N]; \quad (5.12)$$

- VII. Consider 75% of healthy data as training and the remaining 25% as test together with damaged data. All loads and speeds are analysed together;
- VIII. Train the one-class SVM classifier on training data and evaluate the label assigned by the classifier to test data. The real class is known so mistakes in labelling could be computed;
- IX. Repeat point VII. and VIII. 30 times permuting healthy data order to give statistical significance to the analysis and evaluate the error percentage in labels assignment.

#### 5.2.4 Application to bearings data

Several conditions can influence data during acquisitions, as we discovered through the analysis on the industrial aircraft test rig, such as speed, external load, temperature variations. So, we decide to conduct some experiments on our test rig (Setup II described in Section 4.4.1) since it has been designed to perform accurate testing of bearings in a controlled laboratory conditions. As already said, the minimisation of spurious signals coming from other vibrating elements, assures that the only variations in accelerations are given by speed and load, that can be properly changed and monitored. Thus, detecting and removing the effects of such factors is important in order to avoid any bias during the application of diagnostic techniques. In fact, a small variation in speed or in the temperature of the oil circulating in a system produces deviations that a diagnostic algorithm may erroneously detect as a damage, thus providing a False Alarm. Our target is the development of a method able to identify a damage in a rotating element of a roller bearing by removing the effect of speed and external load.

The signals considered are those acquired on the Setup II as described in Section 4.4.2.2 for the axial direction named Channel 7. It belongs to the accelerometer placed directly on the bearing that could be changed (7-8-9 in Figure 4.9(b)). Since the behaviour for the data in a certain condition is very similar all along the various acquisitions, we analyse only the first 10 elements. As already explained, we consider three different speed values (9000, 10500 and 12000 RPM) and three radial loads (1400, 1600 and 1800 N) and 1 second of acquisitions is taken into account for each combination. The variation of the speeds and the loads is around 15%, so it is not so wide. Acquisitions in the nine cases are done both for the healthy bearing and for 450  $\mu\text{m}$  indentation on a rolling element. Notice that the temperature of the oil circulating is almost constant between the different acquisitions, so we are certain that the only variations detected through vibrations are caused by load and speed changing. In Figure 5.18 Root Mean Square values for the 10 acquisitions in each condition are evaluated. This plot shows how this parameter is not so much dependent on the load: in fact no variations could be notice when it changes, both for the healthy and for the damaged case. The speed, instead, influences RMS: it is clear from the Figure that it increases going from

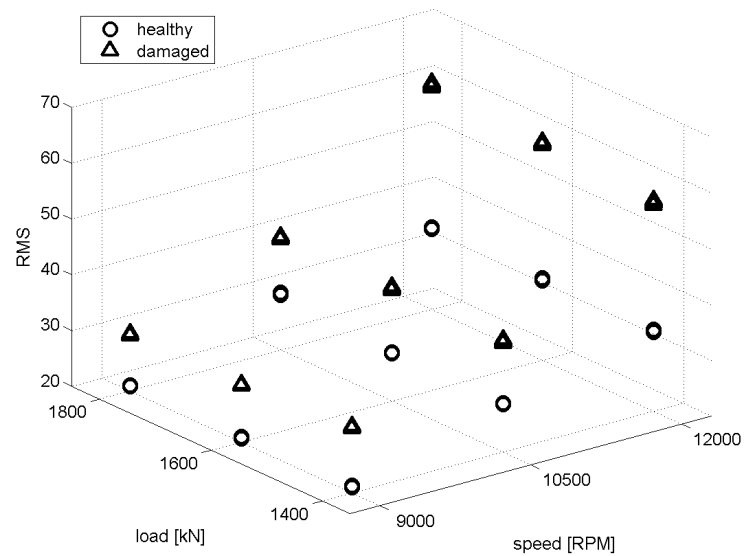


Figure 5.18. RMS values for the healthy and the faulty bearing in the nine combinations of speed and load considered.

9000 to 12000 RPM for the two bearings. Moreover, it can be noticed that, in low speed case, this parameter for damaged bearing is almost near to the healthy one when it reaches the highest speed. For example, RMS value for a damaged bearing at 9000 RPM for the three loads is around 30. If we consider the healthy case at the highest speed evaluated (12000 RPM) RMS is around 34. Thus, the undamaged bearing at the highest speed has a parameter value bigger than the faulty one at the lowest speed. It means that if we consider the RMS parameter taking into account all nine conditions together, the difference between healthy and faulty bearings may be strongly biased. This observation leads to the need of a parameter that could avoid such kind of problems.

According to the results shown in Figures 5.16(a) and 5.16(b), we decide to take into account for our analysis the first 10 IMFs, since they include the most dominant fault information. So, the feature vector  $F'$  that has to be defined following the methodology in Section 5.2.3 is a 10-dimension one. The analysis through OCSVM is done starting from the first two dimensions of the feature vector. Then we add a new dimension each time until the whole normalised feature vector  $F'$  is used. We choose to include the features from the beginning according to the fact that EMD operates in form of collection of filters organised in a filter bank structure, as observed in Figure 5.17. In such way, taking the features starting from the beginning of the vector, we move from higher frequency contents to lower ones.

As stated in Section 5.2.3, the 75% of healthy data are used to train the classifier, while the 25% of them are added to damaged data as testing instances. Since the exact belonging is known, it is interesting to evaluate the errors in labelling made by the OCSVM classifier.

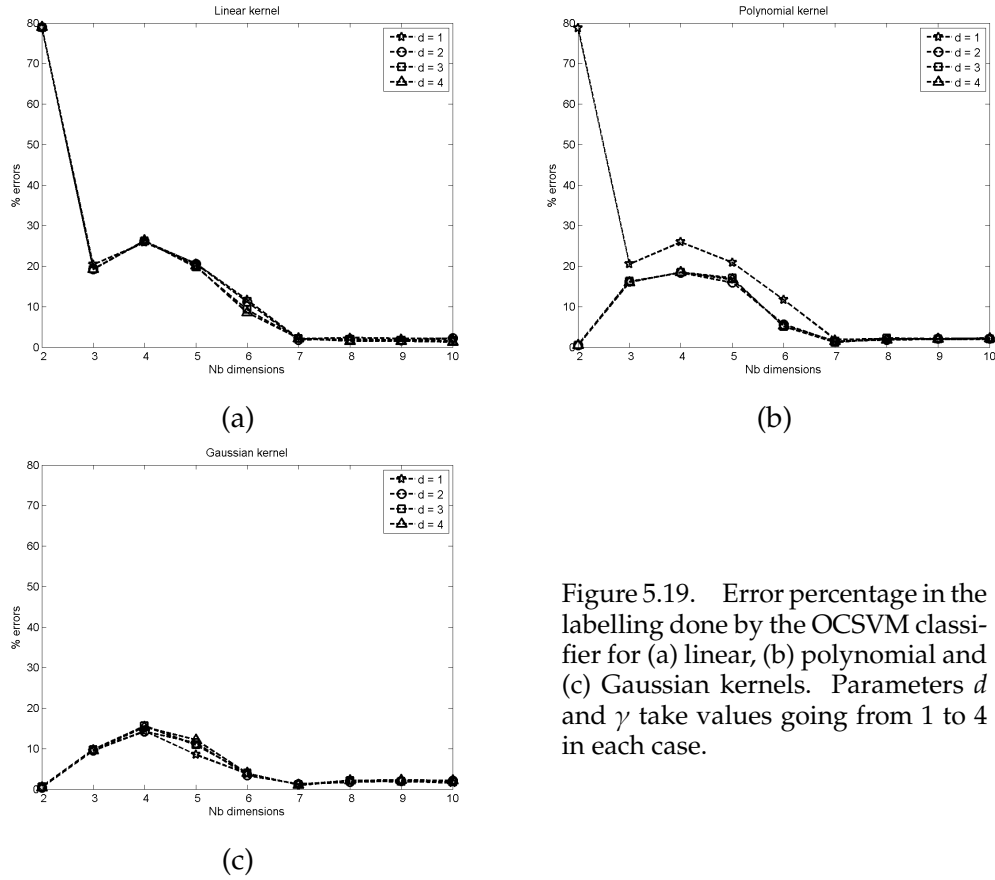


Figure 5.19. Error percentage in the labelling done by the OCSVM classifier for (a) linear, (b) polynomial and (c) Gaussian kernels. Parameters  $d$  and  $\gamma$  take values going from 1 to 4 in each case.

In this way, an evaluation of the relation between the number of dimensions and a proper identification procedure could be done. Moreover, three different SVM kernels are compared through the application to the acquired data:

- linear:  $K(x_i, x_j) = (x_i^T x_j)^d$
- polynomial:  $K(x_i, x_j) = (x_i^T x_j + 1)^d$
- Gaussian:  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$

For each kernel, parameters  $d$  and  $\gamma$  take values going from 1 to 4 and labelling mistakes are evaluated in percentage.

Figures 5.19 presents the different behaviours of the three kernels when the number of features and the parameters values increase. The error percentage for the linear kernel tends to decrease when the dimensions go from 1 to 10 (Figure 5.19(a)). Hence, in order to provide a good detection ability a greater number of features should be considered. The same behaviour is observed for polynomial kernel when  $d = 1$ , while for the other values of the parameter  $d$  less errors are present for 2, 7, 8, 9 and 10 dimensions (Figure 5.19(b)). The error trend in the case of a gaussian kernel does not seem to be conditioned by parameter  $\gamma$ , while the minimum number of labelling errors are found when the feature vector has 2



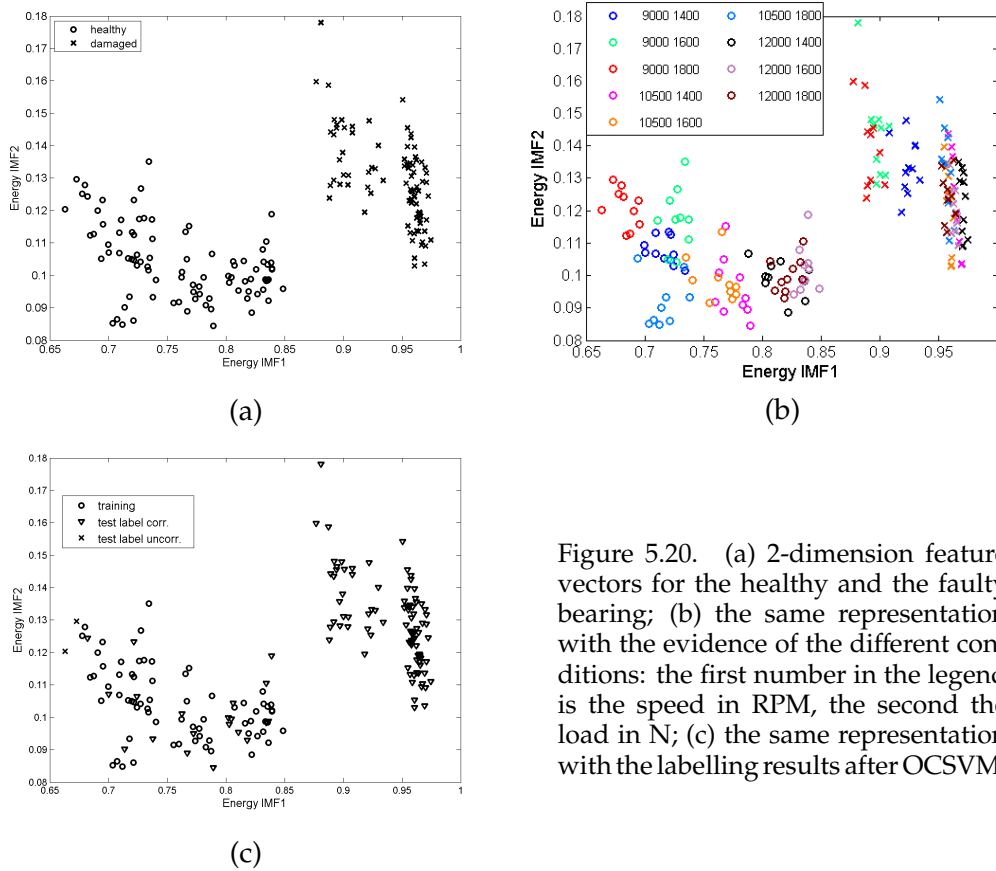


Figure 5.20. (a) 2-dimension feature vectors for the healthy and the faulty bearing; (b) the same representation with the evidence of the different conditions: the first number in the legend is the speed in RPM, the second the load in N; (c) the same representation with the labelling results after OCSVM.

or more the 7 dimensions (Figure 5.19(c)). On the whole, a gaussian kernel or a polynomial one with parameter  $d > 1$  give successful results in detecting the damage regardless of speed and load influence.

To emphasise this fact, we concentrate on the 2-dimensions feature vector  $F'$ , since it gives interesting results and because it is easier to visualise. If we consider the totality of 180 values (10 for each one of the nine conditions for both bearings) computed using our methodology, we obtain the plot in Figure 5.20(a). In this picture, it is clear how data divide into two groups according to their state rather than depending on their condition of load and speed. This explains the great efficiency of the classifier in damage identification, due to the perfect distinction between the two classes of data.

Furthermore, any dependence on different loads and speeds seems to be removed as pointed out in Figure 5.20(b). The nine symbols represent the various conditions for the undamaged and damaged bearing and no particular division based on the rotational speed or on the load applied is noticed. According to the results presented in Figure 5.19, the use of a Gaussian kernel with the parameter  $\gamma = 1$  to the OCSVM furnishes a small error percentage. We can visualise this in Figure 5.20(c): testing data are, in fact, well classified (triangles) and only two elements belonging to the faulty class are instead labelled as healthy (cross)

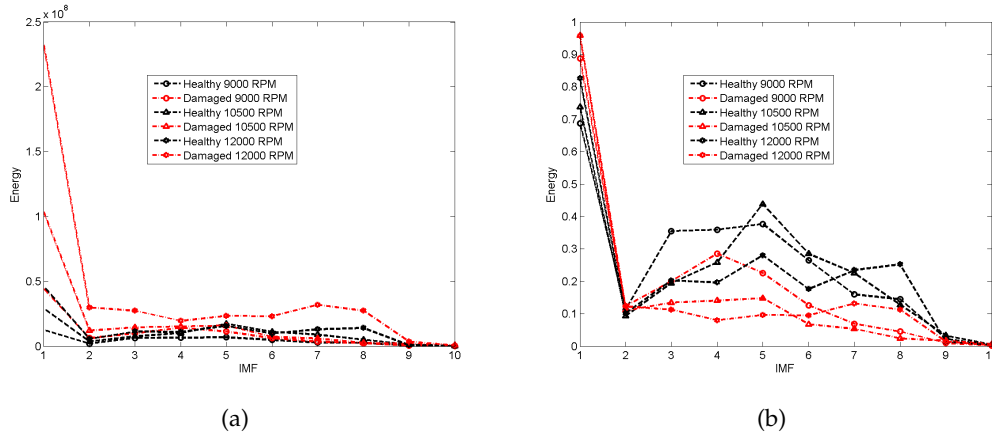


Figure 5.21. Feature vectors for the healthy and damaged bearing at the three different speeds (9000, 10500 and 12000 RPM) and the same vectors normalised over the various components.

originating about 1.79% errors. This value is obtained in the considered case. However, following point IX in Section 5.2.3, we obtain that 0.54% of label are wrongly assigned by the OCSVM classifier, so we are treating a proper identification technique.

The comparison between Figures 5.21(a) and 5.21(b) gives an explanation about the ability of the method in the speed and load influence removal. Values of the acquisition feature vector are plotted for each of the speeds considered, both for the healthy and for the damaged bearing. In Figure 5.21(a) these vectors are plotted, while in Figure 5.21(b) the same vectors, but after the normalisation, are considered. The vector normalisation adopted at step V and VI in Section 5.2.3 helps to remove the contribution of highest energies and, so, to mitigate the various conditions influence on the features.

Moreover, as it could be noticed in the Figure, this aspect is particular clear for the frequency content represented by the component  $c_2$ . The normalised values of the energies here, in fact, tend to be very similar independently of the speed considered, giving a great contribution in the removal of this parameter influence. The same happens for the 9th and 10th IMF, as could be observed in Figure 5.21(b). If we plot the first two dimensions of the data avoiding the normalisation step, we obtain Figure 5.22(a). It is clear how the division between the condition healthy/damaged is still present, but too many differences between the various speeds and loads arise. Figure 5.22(b) helps in identifying how the data are clustered. Here, it is clear as the elements tends to group according to their speed for both bearing conditions rather than according to their load. Moreover, a clear overlapping between the healthy case at 12000 RPM and the faulty one at 9000 RPM could be noticed. This gives rise to a high percentage of labelling error for the OCSVM: in fact, in this case, the error stabilises at 19.64%.

Since now only Channel 7 has been considered, that belongs to the accelerometer placed directly on the bearing that could be changed (7-8-9 in Figure 4.9(b)). As already said, Channel

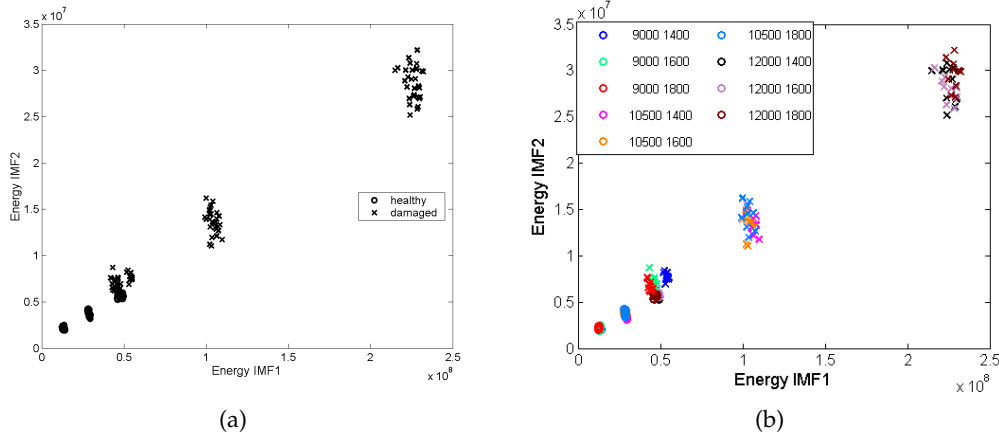


Figure 5.22. (a) 2-dimension feature vectors for the healthy and the faulty bearing if we consider the data avoiding the normalisation step; (b) the same representation with the evidence of the different conditions: the first number in the legend is the speed in RPM, the second the load in N.

Ch	Kernel	par	Number of IMFs								
			2	3	4	5	6	7	8	9	10
7	linear	1	78.81	20.33	26.19	20.27	11.90	2.02	2.32	2.11	2.41
	polynom	2	0.54	16.07	18.30	15.09	5.95	0.89	2.05	2.08	2.11
	gauss	1	0.54	9.85	13.90	10.51	3.18	1.04	1.90	1.93	1.73
8	linear	1	3.70	7.11	0.86	0.63	0.86	1.22	1.01	1.37	0.95
	polynom	2	1.43	4.55	0.92	0.89	0.98	0.95	0.77	1.46	1.04
	gauss	1	0.98	4.20	0.80	1.34	0.83	1.13	1.13	1.04	1.13
9	linear	1	15.33	6.37	7.05	5.42	2.88	2.62	2.41	2.35	2.68
	polynom	2	9.26	6.67	6.37	5.18	2.68	2.41	2.35	2.38	2.26
	gauss	1	7.17	6.64	6.10	5.36	3.21	2.53	2.35	2.29	2.17
1	linear	1	63.90	63.01	63.27	61.64	56.73	78.45	78.07	78.04	78.01
	polynom	2	77.41	63.13	63.24	59.29	57.56	77.95	77.86	78.18	78.10
	gauss	1	80.30	72.92	62.05	57.11	58.45	77.71	78.07	78.09	77.71
2	linear	1	73.90	33.99	31.55	17.95	17.71	16.70	16.75	17.26	16.82
	polynom	2	74.17	33.84	31.07	18.78	16.82	16.96	16.79	16.49	17.65
	gauss	1	74.55	35.89	30.92	18.57	17.14	16.85	16.10	16.70	16.58
3	linear	1	55.09	65.42	61.25	61.61	67.23	66.61	65.21	65.92	66.49
	polynom	2	56.64	67.68	60.71	60.98	65.86	66.43	67.38	65.92	66.19
	gauss	1	62.14	68.63	60.80	60.33	66.22	65.80	65.42	65.54	65.65

Table 5.4. Error labelling percentage done by OCSVM for three different kernels considering the Channels 7-8-9, referring to the accelerometer placed directly on the bearing that could be changed, and Channels 1-2-3, related to the other accelerometer (Figure 4.9(b)). Feature vectors, obtained through the Methodology presented in Section 5.2.3, are evaluated starting from a 2-dimension one and adding each time a new component.

7 registers the vibration in the axial direction, while Channel 8 and Channel 9 account for the radial and tangential one, respectively. In Table 5.4 the results obtained for also these

channels are presented. For each of them, the various kernels cited before are taken into account, with only one value for the parameter  $d$  or  $\gamma$ . The feature vector starts from 2 dimension and then, at each step, the next component is added to reach a 10-dimensional case.

The radial dimension seems to be the one with an overall low error rate for the different kernels and for all the various types of feature vector. Channel 7, instead, seems to be more efficient when only 2 components are considered, for the polynomial and Gaussian kernel, and when the 7th one is added, for all the kernels. For the tangential acquisitions, we could observe as the error percentage has a general decreasing trend when the number of dimensions increases. In the same Table, results obtained for the channels belonging to the accelerometer placed far from the damage are presented. In these cases, more labelling errors are obtained, usually with a  $> 50\%$  rate. An exception is observed for the radial direction (Channel 8), when the feature vector has more than 5 dimensions. This means that, far from the damage, the detection ability decreases due, probably, to this fact.

### 5.2.5 Methodology II

In the previous sections, the problem of finding a feature vector able to remove the influence of speed and load has been faced considering a single channel each time. Another possible way to operate involves, instead, three channels of a certain accelerometer together to form the feature vector. Thus, the performance of the classifier are checked for each IMF in a 3-dimensional space composed by the axial, the radial and the tangential components of one point. In this way, it is possible to do some considerations related to the frequency components that are less influenced by the speed and load variation in the damage recognition.

The diagnostic method following this approach consists of different steps:

- I. Collect vibration signals under various condition of speed and radial load applied, both for a healthy and a damaged bearing;
- II. Apply EMD and decompose the original signal into IMFs; then choose the first  $n$  to extract the features used during the analysis;
- III. Evaluate the total energy for the  $n$  selected IMFs:

$$E_j = \int_{-\infty}^{+\infty} |c_j(t)|^2 dt \quad j = 1, \dots, n; \quad (5.13)$$

- IV. Create a vector with the energies of the  $n$  selected IMFs:

$$F = [E_1, \dots, E_n]; \quad (5.14)$$

V. Normalise the vector dividing  $F$  for this value:

$$E_N = \sqrt{\sum_{j=1}^n |E_j|^2}; \quad (5.15)$$

VI. Obtain a  $n$ -dimension normalised vector:

$$F' = [E_1/E_N, \dots, E_n/E_N]; \quad (5.16)$$

- VII. Repeat points from I. to VI. for each one of the three channels of an accelerometer;
- VIII. Create a 3-dimensional feature vector composed by the energy values of the  $j$ th IMF for each channel:

$$F_v(j) = [F'_{ch1}(j), F'_{ch2}(j), F'_{ch3}(j)] \quad j = 1, \dots, n;$$

- IX. Consider 75% of healthy data as training and the remaining 25% as test together with damaged data. All loads and speeds are analysed together;
- X. Train the one-class SVM classifier on training data and evaluate the label assigned by the classifier to test data. The real class is known so mistakes in labelling could be computed;
- XI. Repeat point IX. and X. 30 times permuting healthy data order to give statistical significance to the analysis and evaluate the error percentage in labels assignment.

### 5.2.6 Application to bearing data

The signals considered are the same used in the previous sections, that is, data acquired on the Setup II described in Section 4.4.2.2. The shaft rotates at three different speeds (9000, 10500 and 12000 RPM), while three radial loads (1400, 1600 and 1800 N) are applied. The healthy bearing and the one characterised by the 450  $\mu\text{m}$  indentation on a rolling element are considered. The methodology presented in Section 5.2.5 is then applied to these data taking into account the accelerometer placed on the bearing that could be changed.

The first evidence of this approach could be observed in Figures 5.23 and 5.24. The 3-dimensional vector containing the three directions of the considered accelerometer are plotted. In each plot, each colour stands for a different speed/load crossing between the nine considered. Then, the circles refers to the healthy case while the crosses to the faulty one. It is clear as data reveal a different behaviour in each IMF. So, some considerations arises related to the influence of the bearing condition (healthy/faulty) and to the division due to the speed and load values.

A great separation is observed between the healthy and faulty states, with the exception

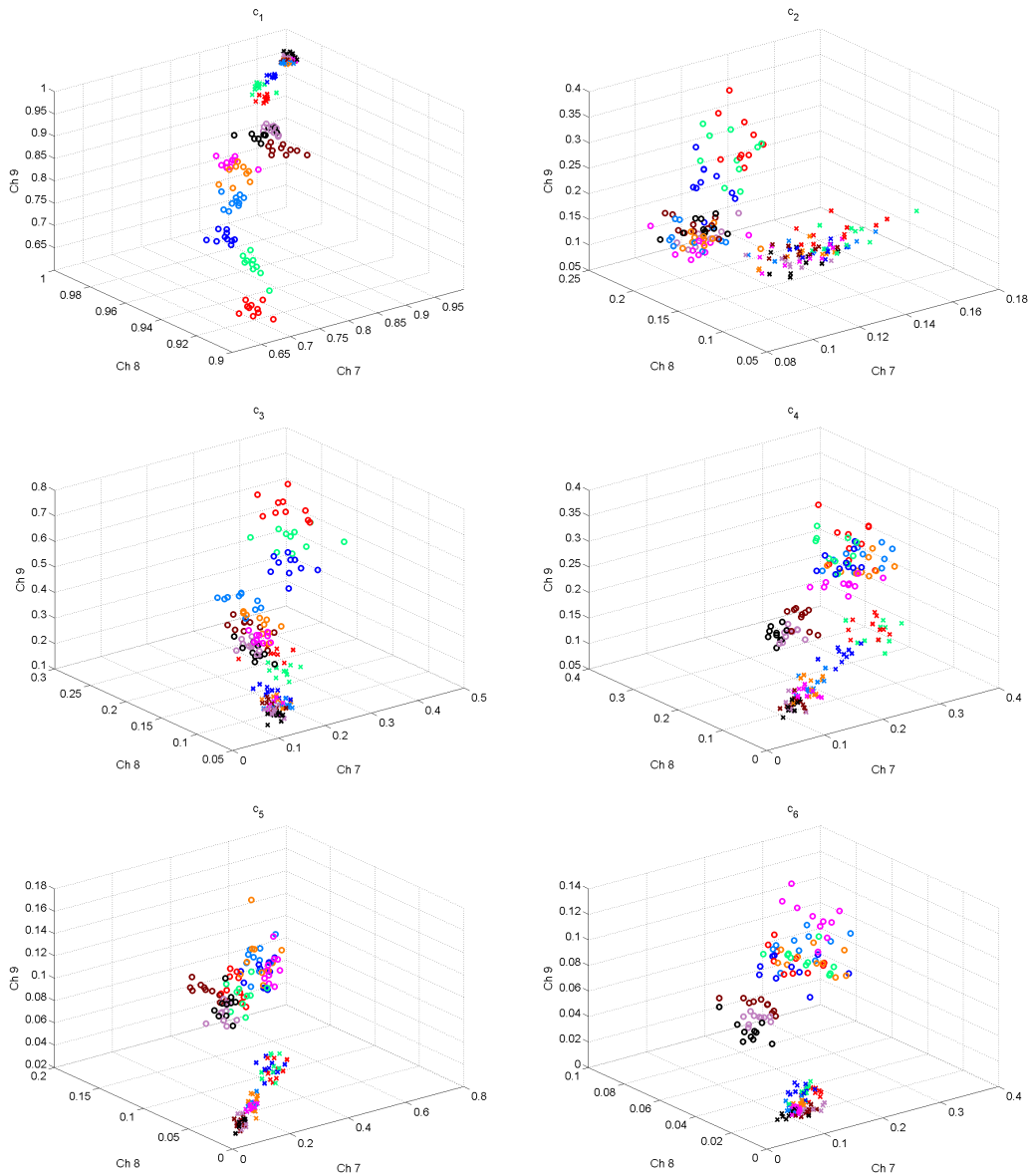


Figure 5.23. 3-dimension feature vectors obtained according to Methodology II for the healthy (circle) and the faulty (cross) bearing for the first six IMFs; different conditions of speed and load are represented by different colours, as showed by the legend in Figure 5.24 .

of the last two components  $c_9$  and  $c_{10}$  (Figure 5.24). Usually, data tends to cluster according to this condition rather than referring to the various speed/load case. A different behaviour is noticed for the first component  $c_1$ , where these characteristics influence the data clustering (Figure 5.23). However, in some components, data join other elements belonging to the same speed independently from the load, as, for example, in  $c_4$ ,  $c_5$  and  $c_6$ .

A possible way to analyse how data places in the 3-dimensional space obtained considering the three accelerometer directions, is to examine if the various clusters are distant from

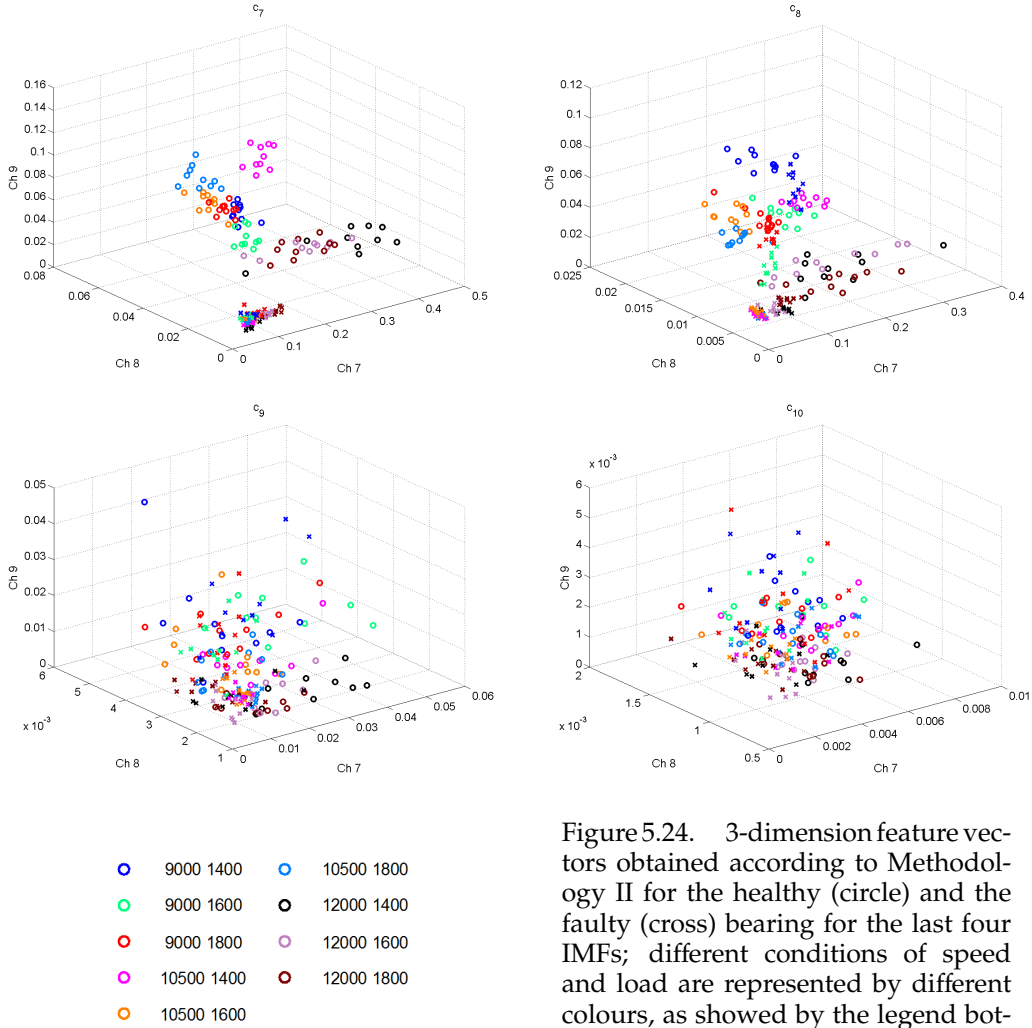


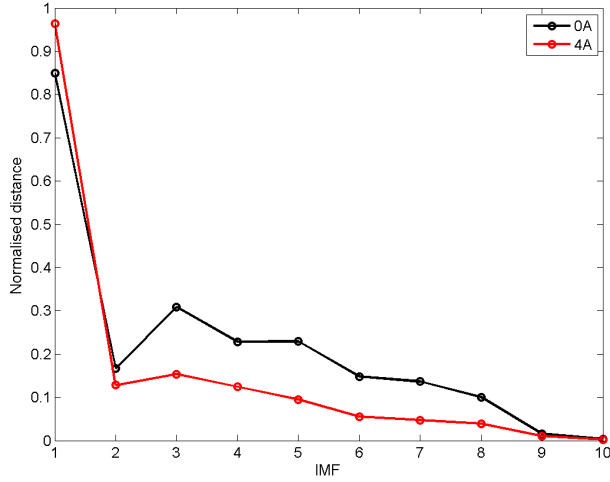
Figure 5.24. 3-dimension feature vectors obtained according to Methodology II for the healthy (circle) and the faulty (cross) bearing for the last four IMFs; different conditions of speed and load are represented by different colours, as showed by the legend bottom left: the first number is the speed in RPM, the second the load in N.

each other for the healthy and faulty case. An overall comparison follows these steps:

- I. Consider a IMF component;
- II. Take into account the 3-dimensional elements for each speed/load condition for the healthy bearing;
- III. Evaluate the centre of each cluster;
- IV. Compute the Euclidean distance between the combinations of all the nine conditions:

$$d_{1,2} = \sum_{i=1}^N |ctr_1^2 + ctr_2^2|^{1/2},$$

where  $ctr_1$  and  $ctr_2$  are the centres of the first and the second clusters respectively;



(a)

IMF	Frequency range kHz
$c_1$	$\approx[25.6-51.2]$
$c_2$	$\approx[12-27]$
$c_3$	$\approx[6-14]$
$c_4$	$\approx[4-9]$
$c_5$	$\approx[2.5-5]$
$c_6$	$\approx[1.5-3.5]$
$c_7$	$\approx[1-2]$
$c_8$	$\approx[0.4-1.2]$
$c_9$	$\approx[0.15-0.7]$
$c_{10}$	$\approx[0.1-0.3]$

(b)

Figure 5.25. (a) Mean distances between all the combinations of various clusters for the healthy (0A) and 450  $\mu\text{m}$  indentation on a rolling element bearing (4A); feature vectors are evaluated with data acquired at Channels 7, 8 and 9. (b) Table showing, approximately, the frequency ranges defined by the various IMFs according to the plot in Figure 5.17 and to [FRG04].

Kernel	par	IMF									
		1	2	3	4	5	6	7	8	9	10
linear	1	12.56	15.30	49.29	53.54	25.39	12.20	80.27	39.58	77.95	65.57
polynom	2	0.33	12.02	7.68	9.79	3.39	6.04	2.38	10.00	73.01	61.70
gauss	1	0.92	11.49	8.63	10.45	3.67	5.95	2.29	9.88	72.59	60.30

Table 5.5. Error labelling percentage done by OCSVM for three different kernels considering the feature vector made up of Channels 7, 8 and 9. Feature vectors, obtained through the Methodology presented in Section 5.2.5, are evaluated for each IMF.

V. Consider the mean value between all the distances;

VI. Do Points II.-VI. for the damaged case;

VII. Do all the operations for every IMF component.

A graphical way to visualise the results obtained after these steps is Figure 5.25(a). Here, component  $c_1$  shows clusters that are rather dependent on the various speed/load crossings for both the bearing states, as it has been observed in Figure 5.23. A general decreasing trend in the clusters separation originates from the 4th and 5th IMF.

The following aspect that has to be considered is the labelling ability of the classifier for each component. So, the error rate obtained for the various decomposition levels is evaluated according to the procedure described in Section 5.2.5. Table 5.5 shows the results for the accelerometer placed directly on the bearing that could be changed for three different



Kernel	par	IMF									
		1	2	3	4	5	6	7	8	9	10
linear	1	79.91	56.55	73.24	79.58	70.74	80.80	80.95	77.53	76.10	78.39
polynom	2	80.89	29.88	53.39	55.00	43.75	75.63	79.61	79.85	80.24	79.85
gauss	1	81.04	31.43	53.30	55.39	43.57	75.33	79.91	79.94	79.85	79.85

Table 5.6. Error labelling percentage done by OCSVM for three different kernels considering the feature vector made up of Channels 1, 2 and 3. Feature vectors, obtained through the Methodology presented in Section 5.2.5, are evaluated for each IMF.

kernels. So, the 3-dimension feature vector is composed by the values evaluated for Channels 7, 8 and 9. The lower labelling errors are obtained for the polynomial and the Gaussian cases. The first IMF is the one producing a percentage  $< 1$  in both cases, so it means that in this frequency range the detection ability is higher. Also for  $c_5$  and  $c_7$  the elements that are wrongly labelled are less than 5%, so it means that, in these bands, the influence of speed and load is not so consistent. An interesting comparison with values arising from the analysis of the other accelerometer, located more distant from this bearing, could be done. Values in Table 5.6 demonstrate a general bigger error labelling rate for the various channels and for each IMF.

If we join the results showed in Figure 5.25(a) with those presented in Table 5.6 for the Gaussian kernel, some interesting considerations arise. In fact, it is possible to compare the various components according to their error rate and their clusters separation. For example, the first IMF has the best performance in the labelling even though each cluster locate far from each other. The components  $c_9$  and  $c_{10}$ , on the contrary, present 3-dimensional points strictly grouped together that are, then, more difficult to label. A right arrangement is reached, for example, at  $c_7$ . In this case, the clusters for the various speed/load combinations are not so separated but, in the meanwhile, the classifier is able to label data in a proper way with a low error percentage. So, this frequency range furnishes a good separation between a healthy and a damage condition being not so influenced by speed and load variations. Table 5.25(b) says that we are considering rather low frequencies.

### 5.2.7 Comparison with Wavelet Decomposition

A signal processing method that we consider and compare with EMD is the Wavelet Decomposition (WD), presented in Section 2.2. This idea comes natural since the filter features of WD have some similarities with those of EMD, as showed by Figure 5.26. In the case of WD, the frequency band of the signal is divided into two equal parts of high frequency and low frequency in every step of the decomposition. Here, the bandwidth of  $A_j$  and  $D_j$ , where  $1 \leq j \leq p$ , with  $p$  maximum layer of the decomposition, is decided by the  $j$ th layer itself in conjunction with the sampling frequency. This is the main difference with EMD and the main feature allowing us to say that WD is not a self-adaptive technique.

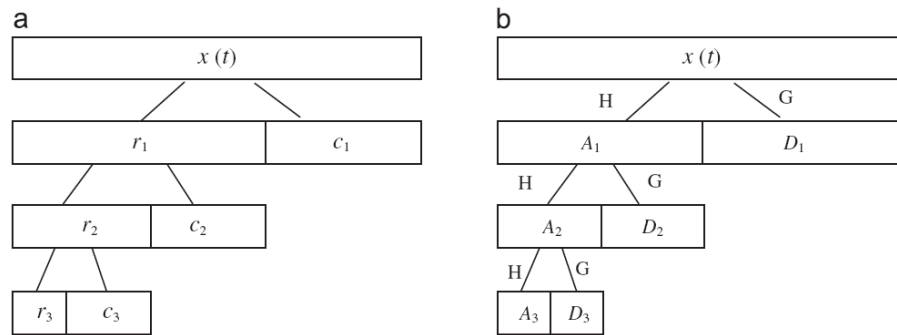


Figure 5.26. The filtering features of Empirical Mode Decomposition (a) and Wavelet Decomposition (b) (Figure taken from [GDFM08]).

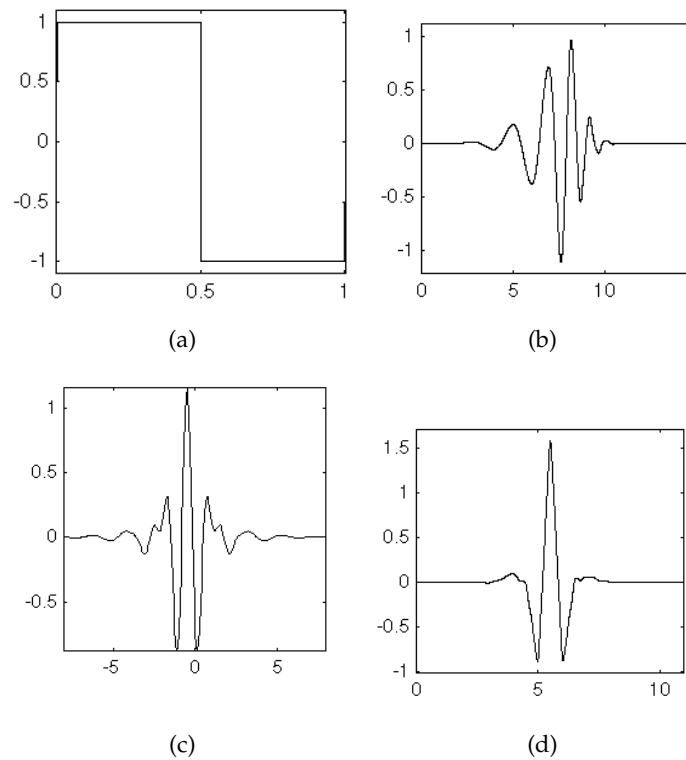


Figure 5.27. Various mother wavelets considered in the analysis: (a) Haar, (b) Daubechies order 8, (c) Coiflets of order 2 and (d) Discrete Meyer (Figures taken from [Mat02]).

The methodology presented in Section 5.2.5 is applied in all its steps, but now the total energies are not evaluated for the IMFs but for the signal filtered by the WD. This decomposition is operated to the 10th layer to have the same number of components as before, that we call  $w_1, w_2$  and so on. A multilevel one-dimensional wavelet analysis is considered, using the wavelets showed in Figure 5.27. If we compute the Power Spectral Density for the various level filtered by the WD, Figure 5.28 is obtained. This result is very similar to the one of Figure 5.17, which refers to EMD. So, it is consistent to do a comparison between the

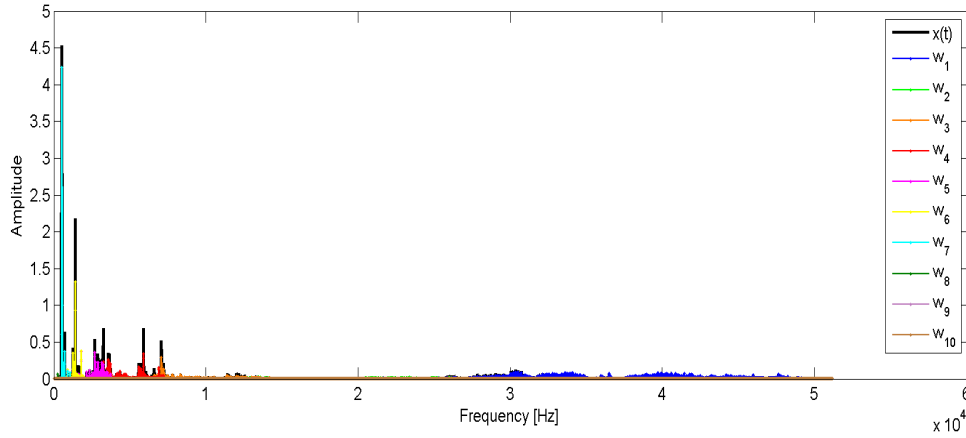


Figure 5.28. Power Spectral Density for the signal and for each level of the wavelet decomposition for the damaged analogue to Figure 5.17. Each colour represents a  $w_i$  component.

Wavelet	Level									
	1	2	3	4	5	6	7	8	9	10
Haar	5.95	0.60	6.76	10.48	0.42	0.57	9.26	55.60	16.88	31.28
Daubechies8	0.39	0.51	5.60	24.94	0.80	0.98	11.34	80.83	69.05	76.64
Coiflets2	0.74	0.57	5.98	19.76	0.60	0.68	11.19	79.43	55.30	73.90
Discrete Meyer	0.77	0.89	1.16	27.32	0.48	0.65	17.68	71.01	9.94	35.18

Table 5.7. Error labelling percentage done by OCSVM for the various wavelet considering the feature vector made up of Channels 7, 8 and 9. Feature vectors, obtained through the Methodology presented in Section 5.2.5, are evaluated for each decomposition level.

results obtained for EMD and WD. The main difference is that the filtering feature of the WD is not self-adaptive, so the frequency domain of the  $k + 1$  level is the upper half-band of that of the previous one of index  $k$ .

The results in Table 5.7 show a rather good labelling ability for all the mother wavelets. The lower error percentages (<1%) refer to the components  $w_1$ ,  $w_2$ ,  $w_5$  and  $w_6$ . This means that in these frequency band the two classes (healthy and damaged) are well separated independently from various speed/load conditions.

# Conclusion

In this thesis, we presented our contribution to the topic of bearing diagnostics. The importance of this subject was noticed during the collaboration of our Dynamics & Identification Research Group (DIRG) with Avio S.p.A. In fact, we worked at a project whose goal was the definition of some algorithms enabling damage detection on a gearbox with a certain confidence level. Data analysis on real industrial gearboxes revealed the influence of some external factors, such as speed, load, temperature and system disassemblies. It is clear that a reliable diagnostic technique should be able to produce the right fault warnings independently by these fluctuations.

In order to understand better these aspects, we acquired a consisted data base on a test rig set up in our laboratory. Here, the environment allowed us to have the certainty of controlled parameter variations. In this way, it was possible to assess and to verify our techniques.

Two methods was presented referring to two different non-stationary signals processing techniques. A first method based on the statistical tool called Spectral Kurtosis, and its application called fast kurtogram, was applied to our data. Various types of bearing faults were considered and in all cases our method gave proper results. Moreover, its ability in acting efficiently also when the bearing is unmounted was observed. A further interesting application revealed that, for the roller element damage, the parameter was able to recognise an increasing trend in the fault entity.


The other diagnostic technique was based on the Empirical Mode Decomposition. The main goal, in this case, was tied to the need of defining a parameter that could identify a damage also in case of small speed and load variations. Two different approaches was proposed based on the energy content of the Intrinsic Mode Functions obtained after the decomposition. In both cases, however, the One Class Support Vector Machine, trained on healthy data for all the various speed and load conditions, was used to label test data. Interesting results demonstrated the ability of this method in recognising the damage properly. Moreover, frequency ranges that seemed to be less influenced by the speed and load in the correct class assignment was found. A comparison with the results obtained applying the Wavelet Decomposition was presented too.

Future works will focus on the potentiality of the method based on EMD. Acquisitions will be operated at the same speed and load conditions for other types of bearing faults too. Thus, the detection ability will be checked in order to verify if it would be effective also for various damage entities. Other conditions such as temperature and different test rig assemblies are going to be analysed too. Moreover, some work will be done on the frequency bands where the detection ability is higher. Somehow, they could be considered as more informative and less influenced by some conditions, so a further and deeper analysis will be dedicated to study this hypothesis.

## **Appendix A**



### **Bearing failures**

In the next pages a table describing the most widespread bearing damages is presented. Some of them have been already explained in Chapter 1. In this Appendix, various failures are listed and described together with factors and events that could cause them and could contribute to their evolution and expansion. For each damage, some countermeasure, usually the most adopted by users and experts, are presented. Finally, an example of how a failure appears to a visual inspection is given through some photos.

Phenomena	Description	Causes	Countermeasures	Picture
Flaking	Bearing surface turns scaly and peels off because of contact load repeatedly received on the raceway and rolling surface during rotation Present when bearing service life is near to the end May be found at early stage and it is necessary to examine the causes	Bearing internal clearance becomes narrower than specified during operation Mistakes during bearing ring assembly (inclination) Flaw created during mounting, or brinelling, nicks, rust, etc. occur on the raceway or rolling surface Inaccurate shape of shaft or housing (imperfect circle, depressions on surface)	Use bearing with heavier rated load Check if abnormal load being generated Lubrication method must ensure better formation of lubricant film, by increasing the viscosity	 Flaking on inner ring of cylindrical roller bearing
Pitting	Small holes (0.1 mm in depth) generated on the raceway surface by rolling fatigue May be found at early stage and it is necessary to examine the causes	Bearing internal clearance becomes narrower than specified during operation Mistakes during bearing ring assembly (inclination) Flaw created during mounting, or brinelling, nicks, rust, etc. occur on the raceway or rolling surface Inaccurate shape of shaft or housing (imperfect circle, depressions on surface)	Lubrication method must ensure better formation of lubricant film, by increasing the viscosity	

*Continued on next page*



Table A.1 – Continued from previous page

Phenomena	Causes	Countermeasures	Picture
Wear	Due to sliding abrasion on roller end face and rib, cage pocket surface, cage, guide surface of the bearing ring	Improper lubricant or lack of lubricant Contamination by foreign matter(s)	
	If due to contamination by foreign matter and corrosion occurs on sliding surface and also on the rolling surface		Wear on outer ring raceway surface of double-row cylindrical roller bearing
Fretting	Occurs if slight sliding is repeatedly caused on the contact surface	Vibration load	Investigation and countermeasures for the source of vibration
	Fretting corrosion occurs on the fitting surface with a rust like powder generation	Slight vibration on fitting surface caused by load	
Cracks	Includes slight cracks, splitting and fracture	Heavy load	Investigation followed by countermeasures for excessively heavy load Removal of thermal impact
		Excessively heavy internal load due to improper installation	
	Excessive interference at fitting, or shaft and housing of improper shape		
	Instantaneous heat generation due to sudden sliding at rolling surface, sliding surface or fitting surface		
	Abnormal heat generated by lack of lubricant		
			Crack and chip in a spherical roller bearing

Continued on next page





Table A.1 – Continued from previous page

Phenomena	Causes	Countermeasures	Picture
Chips	Occurs at a certain part of a bearing ring rib or corner of a roller	Abnormally heavy axial load or impact load Partial impact of hammer or other tool during bearing assembly or disassembly	
		Improvement of assembly and disassembly procedures Improvement of handling method Investigation followed by countermeasures if excessively heavy load	Chip in outer ring rib of cylindrical roller bearing
Brinelling	Depressions created on the part of the raceway surface which comes into contact with the rolling element Due to plastic deformation or to contamination by solid foreign matters	Extremely heavy load (static or impact load) applied to bearing Solid foreign matter caught in bearing parts	
		Investigation followed by countermeasures for excessively heavy load or impact Enhancement of sealing capability Careful washing of shaft and housing to remove foreign matter Oil filtering Investigation of flaking in target bearing together with other bearings	Brinelling on inner ring raceway surface of tapered roller bearing
Nicks	Flaw due to direct impact caused by a hammer or other solid too hitting bearings	Assembly and disassembly of faulty bearing Bearing mis-handling	Bearing assembly and disassembly improvement Bearing handling improvement



Continued on next page

Table A.1 – Continued from previous page

Phenomena	Causes	Countermeasures	Picture
Scratches	Relatively shallow flaw caused by sliding contact	Flaw generated during assembly If on raceway surface and rolling surface due to improper rotation of rolling element, improper lubrication or foreign matter intrusion	
	More serious scratches are regarded as scuffing	Improvement of assembly and disassembly procedures Review and improvement of lubricant and lubrication method Sealing capability enhancement Sufficient cleaning of shaft and housing	Scratch on roller rolling surface of cylindrical roller bearing
Scuffing	Caused by high contact pressure and heat on the rolling surface	Flaw generated during assembly Scuffing on roller end face and rib face due to improper lubrication at contact face, excessive preload, foreign matter intrusion, abnormal axial load	
		If on raceway surface and rolling surface due to improper rotation of rolling element, improper lubrication or foreign matter intrusion	Improvement of assembly and disassembly procedures Review and improvement of lubricant and lubrication method Inspection and countermeasures for abnormal load Sealing capability enhancement Sufficient cleaning of shaft and housing



Continued on next page

Table A.1 – Continued from previous page

Phenomena	Causes	Countermeasures	Picture
Rust	Film of oxide, hydroxide, or carbonate produced on a metallic surface by chemical action	Occurs when humidity in the housing turns into drops of water after equipment stopping and temperature decreasing to dew point, water drops often contaminate the lubricant and, so, rust is generated on the bearing surface If bearings are stored in a humid place for a long time, rust generates on the raceway surface at intervals equal to the rolling elements spacing	 <p>Rust on roller rolling surface of cylindrical roller bearing</p>
Corrosion	Oxidation or dissolution occurring on the surface due to chemical action with acid or alkali	Due to decomposition, under high temperature, of sulfur or chlorine compound contained in lubricant additives Water getting inside bearings	<p>Enhancement of sealing capability Periodic inspection of lubricant. Provision for adequate rust prevention during storage of bearings</p>
Pear skin	Small depressions on the rolling surface obtained entirely as a result of many foreign matters being caught between parts Rolling surface suffering from pear skin appears dim and is rough in texture In the worst case, the surface is discoloured by heat	Contamination by foreign matter or lack of lubricant must be inspected carefully Water getting inside bearings	 <p>Pear skin on inner ring raceway surface of double-row cylindrical roller bearing</p> <p>Shaft and housing careful washing Enhancement of sealing capability Oil filtering Lubricant and lubrication method review</p>



Continued on next page

Table A.1 – Continued from previous page

Phenomena	Causes	Countermeasures	Picture
Discoloration	Bearing surface is discoloured by staining or heat generated during operation	Caused by deterioration of the lubricant or adhesion of coloured substances to the bearing surface, that could be removed by scrubbing or wiping with a solvent	
	Can be classified in: staining, electric pitting, rust, corrosion, and temper colour	Brown discoloration of the rolling or sliding surface due to adhesion of acidic powders generated by abrasion during operation which adhere uniformly to the bearing circumference	
Smearing	Minute seizure is concentrated on the rolling surface	Occurs if the oil film disappears as rolling elements stop rotating due to inappropriate use or improper lubrication and start to slide on the raceway surface	
	Surface is partially melted by heat of high temperature generated by friction	In ball bearings is caused by sliding or spinning of balls	
	On some parts, the surface damaged could become significantly rough	In roller bearings tends to occur when the roller enters into on from the load zone	Review followed by countermeasures to improve the formation of oil film Provision for extreme-pressure lubricant Countermeasures to prevent sliding adoption
			Discoloration on raceway surface, roller rolling surface of cylindrical roller bearing
			Smearing on outer ring raceway surface of cylindrical roller bearing

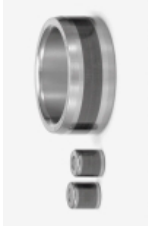

Continued on next page

Table A.1 – Continued from previous page

Phenomena	Causes	Countermeasures	Picture
Creep	Displacement during operation of a bearing ring, relative to the shaft or housing	Interference is too small in relation to the heat or load generated during operation	 <p>Creep of deep groove ball bearing inner ring</p>
Creep	Displacement during operation of a bearing ring, relative to the shaft or housing	Interference is too small in relation to the heat or load generated during operation	 <p>Creep of deep groove ball bearing inner ring</p>

*Continued on next page*

Table A.1 – Continued from previous page

Phenomena	Causes	Countermeasures	Picture
Electric pitting	<p>Bearing surface is partially melted by sparks generated when electric current enters the bearing and passes through an extremely thin oil film at the rolling contact point</p> <p>Classified into pitting or ridge marks</p> <p>Depressions can be observed when pitting is magnified where the surface has been melted by sparks</p> <p>Significant electric pitting causes flaking</p> <p>If pitting is observed by normal visual inspection, the bearing cannot be re-used</p>	<p>Bearing surface is partially melted by electric current passing through the bearing</p> <p>Grounding or grounding maintenance improvement</p> <p>Provision of insulation for bearings or for the section near bearings</p>	 <p>Electric pitting on cylindrical roller bearing</p>
Seizure	<p>Bearing parts are melted and adhere to one another due to abnormal heat or the rolling surface becoming rough so that the bearing can no longer rotate</p> <p>Not possible to use again bearing: the hardness has deteriorated and smooth rotation is impossible on the rough surface</p>	<p>Abnormal heat generated by improper lubrication, excessive preload, or improper contact of rolling elements with the raceway surface (not compensated by the cooling method or lubrication employed)</p> <p>Causes to be investigated</p> <p>Appropriate countermeasures corresponding to the results should be taken</p>	 <p>Seizure of cylindrical roller bearing with rib</p>

## **Appendix B**

### **Market analysis**

In the next pages, some of the bearings diagnostic industrial systems present at this time on the market are described. The company names that develops and produces the considered system is pointed out. The sales pitch that is used to advertise the system under consideration is cited. Then, the main features characterising it are listed. They includes both the methods exploited for the diagnostics and the measures required to obtain the verdict on the machine condition. If known, also the technology content characterising the system is reported. Finally, the main fields of application of the monitoring system itself are marked.

Company	Sales pitch	Features	Technology content	Application
GasTOPS	<b>MetalScan - Oil Debris Sensor</b> Situational Awareness for Critical Equipment	On-line full-flow inductive sensor fitted in lube oil line Detects spall initiation, progression, rate Detects 100% of particles above minimum particle size Measures number, size, mass of ferrous & non-ferrous debris Can be used to quantify damage severity and remaining useful life 15 years of condition indicators verified by military, government & OEMs Verification of parameters including critical mass loss & mass rate Validated by engine tear-down and x-ray debris analysis		F22 Raptor: F119 engine Joint Strike Fighter: F135 engine & lift Eurofighter Typhoon: EJ200 engine SeaKing Helicopter: engine and gbx Pilatus PC12: PWC PT6A engine

*Continued on next page*



Table B.1 – Continued from previous page

Company	Sales pitch	Features	Technology content	Application
Impact Technologies	<b>Bearing Health Monitor (BHM)</b> Real-time Data + Intelligent Processing = Actionable Information	specific software and hardware components can be added to satisfy your specific needs, stand alone or distributed processing operation, built in webserver. support of most common sensors including thermocouples, accelerometers, voltage and current inputs. Can also output voltages and currents to further integration possibilities real-time operating system, redundant power supplies, Watchdog timers, and other features lead to superb reliability can be expanded to accommodate more sensors by simply adding more input modules to chassis		Aerospace Power and Industrial Marine Systems Electronic Systems Maintenance Management Design and Systems Engineering Commercial Systems Integration Ground Vehicles
	<b>BearingLife™ with ImpactEnergy™</b> Early Bearing Fault Detection		Reusable software components (feature extraction, automated detection/tracking and troubleshooting) work in conjunction with commercial off the shelf (COTS) sensor hardware to produce a low-cost, state-of-the-art solution for your bearing problems takes advantage of a variation of the demodulation process associated with the High Frequency Resonance (HFR) techniques to discover mechanical fault frequencies that serve as early warning indicators for impending failure	

Continued on next page

Table B.1 – Continued from previous page

Company	Sales pitch	Features	Technology content	Application
SPM INSTRUMENT	<b>Bearing checker</b> bearing condition at your fingertips	proactive approach to maintenance economically feasible useful and economical complement to maintenance toolbox without specialised training, potential problem sources detected well before damage is detectable by heat or vibration measures shock pulses with a built-in probe and machine surface temperature with an infrared sensor	Stethoscope Shock pulse measurement Temperature measurement	Maritime Industries Power Production Industries Windpower Industries Mining Industries Automotive Industries
	<b>Leonova™ Infinity – SPM Spectrum</b> monitoring for profitable maintenance	Vibration Analysis SPM Spectrum Orbit Analysis Structural Resonances		
Synchrony Dresser-Rand	<b>Magnetic Bearing Monitoring System</b> Spotting Trouble Before It Starts	Monitor trends and changes in operating conditions Predict and schedule pre-emptive maintenance shutdowns Maximise up time and minimise component damage Ensure maximum efficiency	Radial and axial forces Eccentricity value Peak to peak eccentricity Radial expansion Individual coil temperatures Integrated power electronics temperatures Individual coil current Pair coil voltage Velocity & Acceleration	Aerospace & Defence Renewable & Recoverable Energy General Industry

Continued on next page

Table B.1 – Continued from previous page

Company	Sales pitch	Features	Technology content	Application
Timken	<b>StatusCheck™ Series</b> Wireless condition monitoring system for machines	Wireless configuration Flexible software with adjustable alarm thresholds Dual mounting (magnetic or threaded) Suitable for industrial environments Multiple connectivity options Open database format Simple interpretation (no vibration analysis training is required) Ideal for early detection of problems on steady-state equipment Ideal for hard-to-reach applications and environments	Contact temperature probe Two-axis vibration detection Measures acceleration and velocity	fans pumps motors gearboxes

*Continued on next page*

Table B.1 – Continued from previous page

Company	Sales pitch	Features	Technology content	Application
Schaeffler Technologies AG	<b>FAG DTECT X1 s Continuous Monitoring of Plant and Machinery</b> Reliable machine protection through vibration analysis	Reliable machinery protection by means of vibration diagnosis Space saving due to compact construction Suitable for harsh conditions (-20°C to +70°C) Diverse monitoring tasks due to larger number of measurement channels Increased reliability through combination of various process parameters Versatile communication interfaces and connection options Increased operational security due to breakdown protection Flexible and simple installation at the location due to standardised connection systems	signal broken down into its frequency components by means of FFT RMS peak value peak-to-peak value steady component crest factor speed, torque temperature, pressure connection of all common acceleration, velocity and displacement sensors.	Steel and non-ferrous metals Hydroelectricity Raw material extraction and processing Pulp and Paper Shipbuilding
SKF	<b>SKF Machine Condition Advisor CMAS 100-SL</b> Machine monitoring, made simple	Quick and easy to set up and use Lightweight, compact, and ergonomically designed Alert and Danger prompts provide increased diagnostic confidence. Measuring velocity, enveloped acceleration and temperature simultaneously saves time Flexible enough to work with standard 100 mV/g constant current accelerometers	overall "velocity" vibration reading that measures vibration signals from the machine and automatic comparison to pre-programmed (ISO) guidelines temperature with an infrared sensor Velocity Enveloped acceleration	Medium-sized machines and electrical machines Large machinery and electrical machines

Continued on next page

Table B.1 – Continued from previous page

Company	Sales pitch	Features	Technology content	Application
	<b>SKF Inspector 400 Ultrasonic Probe CMIN 400-K</b> Complements your condition monitoring program	Detect pressure and vacuum leaks, including compressed air Check steam traps and valves quickly and accurately Detect arcing, tracking and corona in electric apparatus can detect the early stage of a machine's mechanical malfunction	senses high frequency sounds and electronically translates these signals by a heterodyning process, making them audible	Test bearings, pumps, motors, compressors, etc. Gear/Gearbox inspection

Continued on next page

Table B.1 – Continued from previous page

Company	Sales pitch	Features	Technology content	Application
Kongsberg	<p><b>Bearing Wear Condition Monitoring (BWCM)</b></p> <p>Necessary means to avoid open up inspection on large two-stroke diesel engines</p>	<p>predicting bearing wear in large two-stroke diesel engines before it becomes critical</p> <p>early warning if any of the three crank-train bearings</p> <p>experience unexpected problems during ship operation</p>	<p>high-frequency alternating current flows through a coil cast in the sensor housing. The electromagnetic coil field induces eddy currents in the conductive target which alters the AC resistance of the coil. This change in impedance produces a linear electrical signal proportional to the distance of the target from the sensor. The measurements take place every time the crosshead guide shoe passes Bottom Dead Centre (BDC) during engine running. The sensor is sampling at a speed of 150 kHz. When the crank passes the BDC the sensor measures and detects the minimum distance between the sensor and the crosshead guide shoe. The minimum value is transmitted to the BWCM system.</p> <p>PS-11 sensor based on the eddy current principle (used for measurements against electrically conductive, ferromagnetic materials)</p>	ships

Continued on next page

Table B.1 – Continued from previous page

Company	Sales pitch	Features	Technology content	Application
Amot	Bearing Condition Monitor - Model XTS-W	<p>Combined measurement and display for each cylinder of crosshead bearing, crank-pin bearing, main bearing</p> <p>Monitors condition of crank train bearings in large 2-stroke diesel engines</p> <p>Monitors Bearing Wear (BWM)</p> <p>Monitors Water in Oil (WIO)</p> <p>Monitors Shaft Earth Device (SEDM)</p>	<p>proximity sensing post-processing</p> <p>electronic sensors, which use a low intensity of electromagnetic field to detect the position</p> <p>the guide shoes at bottom dead centre (BDC)</p>	ships

# Bibliography

- [AB05] J. Antoni and S.G. Braun. Special issue on blind source separation: editorial. *Mechanical System and Signal Processing*, 19:1166–1180, 2005.
- [ABSC11] F. Al-Badour, M. Sunar, and L. Cheded. Vibration analysis of rotating machinery using time-frequency analysis and wavelet techniques. *Mechanical Systems and Signal Processing*, 25:2083–2101, 2011.
- [AEB01] F.A. Andrade, I. Esat, and M.N.M. Badi. Gear condition monitoring by a new application of the kolmogorov-smirnov test. *Journal of Sound and Vibration*, 240:1091–1107, 2001.
- [AG01] I. Antoniadis and G. Glossiotis. Cyclostationary analysis of rolling-element bearing vibration signals. *Journal of Sound and Vibration*, 248 (5):829–845, 2001.
- [AKT09] T. Aroui, Y. Koubaa, and A. Toumi. Clustering of the self-organizing map based approach in induction machine rotor faults diagnostics. *Leonardo Journal of Sciences*, 15:1–14, 2009.
- [Ant05] J. Antoni. Blind separation of vibration components: principles and demonstrations. *Mechanical System and Signal Processing*, 19:1166–1180, 2005.
- [Ant06] J. Antoni. The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mechanical System and Signal Processing*, 20:282–307, 2006.
- [Ant07] J. Antoni. Fast computation of the kurtogram for the detection of transient faults. *Mechanical System and Signal Processing*, 21:108–124, 2007.
- [Ant09] J. Antoni. Cyclostationary by examples. *Mechanical System and Signal Processing*, 23:987–1036, 2009.
- [AR05] J. Antoni and R.B. Randall. On the use of the cyclic power spectrum in rolling element bearing diagnostics. *Journal of Sound and Vibration*, 281 (1-2):463–468, 2005.
- [AR06] J. Antoni and R.B. Randall. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machine. *Mechanical System and Signal Processing*, 20 (2):308–331, 2006.
- [ARFI07] S. Abbasion, A. Rafsanjani, A. Farshidianfar, and N. Irani. Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mechanical Systems and Signal Processing*, 21:2933–2945, 2007.



- [Bal69] H.L. Balderston. The detection of incipient failure in bearings. *Material Evaluation*, 27:121–128, 1969.
- [BCIR08] A. Bellini, M. Cocconcelli, F. Immovilli, and R. Rubini. Diagnosis of mechanical faults by spectral kurtosis energy. In *34th Annual Conference of IEEE*, pages 3079 – 3083, 2008.
- [BCZH10] W. Bartelmus, F. Chaari, R. Zimroz, and M. Haddar. Modelling of gearbox dynamics under time-varying nonstationary load for distributed fault detection and diagnosis. *European Journal of Mechanics - A/Solids*, 29(4):637–646, 2010.
- [BD79] S. Braun and B. Datner. Analysis of roller/ball bearings. *ASME Transactions, Journal of Design*, 101 (1):118–128, 1979.
- [BHT63] B. Bogert, M. Healy, and J. Tukey. The quefrency alalysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In *In Proceedings Symposium on Time Series Analysis*, 1963.
- [Bis06] C.M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. 2006.
- [BJ11] T. Barszcz and A.j Jabłoński. A novel method for the optimal band selection for vibration signal demodulation and comparison with the kurtogram. *Mechanical Systems and Signal Processing*, 25 (1):431–451, 2011.
- [BM96] D.C. Baille and J. Mathew. A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings. *Mechanical System and Signal Processing*, 10 (1):1–17, 1996.
- [BPRR12] P. Borghesani, P. Pennacchi, R.B. Randall, and Ricci R. Order tracking for discrete-random separation in variable speed condition. *Mechanical Systems and Signal Processing*, 30:1–22, 2012.
- [BR09] T. Barszcz and R.B. Randall. Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine. *Mechanical System and Signal Processing*, 23:1352–1365, 2009.
- [But73] D.E. Butler. The shock pulse method for the detection of damaged rolling bearings. *NDT Int*, pages 92–95, 1973.
- [BZ09] W. Bartelmus and R. Zimroz. Vibration condition monitoring of planetary gearbox under varying external load. *Mechanical Systems and Signal Processing*, 23:246–257, 2009.
- [CB02] G. Casella and R.L. Berger. *Statistical inference*. Duxbury, 2002.
- [CBK09] V Chandola, A. Banerjee, and V. Kumar. Anomaly detection : A survey. *ACM Computing Surveys*, 41 (3), 2009.
- [CCPL05] S. Choi, A. Cichocki, H. Park, and S.Y. Lee. Blind source separation and independent component analysis: A review. *Neural Information Processing*, 6(1):1–57, 2005.
- [CG86] P. Chivers and P. Gadd. Monitoring of rolling element bearings using vibration

- analysis techniques. Technical report, NGASt 6638 Feasibility Study, 1986.
- [CHY12] Y. Chen, Z. He, and S. Yang. Research on on-line automatic diagnostic technology for scratch defect of rolling element bearings. *International Journal of Precision Engineering and Manufacturing*, 13(3):357–362, 2012.
- [CL09] F. Combet and Gelman L. Optimal filtering of gear signals for early damage detection based on the spectral kurtosis. *Mechanical Systems and Signal Processing*, 23 (3):652–668, 2009.
- [Coh95] L. Cohen. *Time-frequency Analysis*. Prentice-Hall, NJ, 1995.
- [Col77] R.A. Collacott. *Mechanical fault diagnosis*. 1977.
- [Cou85] J. Courrech. New techniques for fault diagnosis in rolling element bearings. In *Proceedings of the 40th Mechanical Failure Prevention Group, Maryland*, 1985.
- [CR11] M. Cocconcelli and R. Rubini. Overview on condition monitoring for bearings in variable speed conditions: the example of a packaging machine. In *Proceedings of XX Congresso Italiano di Meccanica Teorica e Applicata (AIMETA), Bologna, Italy*, 2011.
- [CRZB11] M. Cocconcelli, R. Rubini, R. Zimroz, and W. Bartelmus. Diagnostics of ball bearings in varying-speed motors by means of artificial neural networks. In *Proceeding International Conference on Condition Monitoring, Cardiff, UK*, 2011.
- [CT65] J.W. Cooley and J.W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19:297–301, 1965.
- [CZHJ12] J. Chen, Y. Zi, Z. He, and Yuan J. Improved spectral kurtosis with adaptive redundant multiwavelet packet and its applications for rotating machinery fault detection. *Measurement Science and Technology*, 23:15 pp, 2012.
- [CZRB12a] M. Cocconcelli, R. Zimroz, R. Rubini, and W. Bartelmus. Kurtosis over energy distribution approach for stft enhancement in ball bearing diagnostics. In *Proceedings of the 2nd International Conference on Condition Monitoring of Machinery in Non-Stationary Operations, Hammamet, Tunisia*, 2012.
- [CZRB12b] M. Cocconcelli, R. Zimroz, R. Rubini, and W. Bartelmus. Stft based approach for ball bearing fault detection in a varying speed motor. In *Proceedings of the 2nd International Conference on Condition Monitoring of Machinery in Non-Stationary Operations, Hammamet, Tunisia*, 2012.
- [DBH74] M.S. Darlow, R.H. Badgley, and G.W. Hogg. Application of high frequency resonance techniques for bearing diagnostics in helicopter gearboxes. Technical report, US Army Air Mobility Research and Development Laboratory, 1974.
- [DRR00] G. Dalpiaz, A. Rivola, and R. Rubini. Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears. *Mechanical System and Signal Processing*, 14:387–412, 2000.
- [DS78] D. Dyer and R.M. Stewart. Detection of rolling element bearing damage by statistical vibration analysis. *Trans ASME, J Mech Design*, 100 (2):229–235, 1978.

- [EHAM12] B. Eftekharijad, M. Hamel, A. Addali, and D. Mba. Application of spectral kurtosis to acoustic emission signatures from bearings. *Condition Monitoring of Machinery in Non-Stationary Operations*, Part 4:425–437, 2012.
- [FMP<sup>+</sup>10] A. Fasana, S. Marchesiello, M. Pirra, L. Garibaldi, and A. Torri. Spectral kurtosis against svm for best frequency selection in bearing diagnostics. *Mécanique & Industries*, 11 (6):489–494, 2010.
- [FRG04] P. Flandrin, G. Rilling, and P. Gonçalves. Empirical mode decomposition as a filter bank. *IEEE Signal Processing Letters*, 11 (2):112–114, 2004.
- [Gab46] D. Gabor. Theory of communication. *Journal IEEE*, 93:429–457, 1946.
- [Gar94] W. Gardner. *Cyclostationarity in Communications and Signal Processing*. IEEE Press, New York, 1994.
- [GDFM08] Q. Gao, C. Duan, H. Fan, and Q. Meng. Rotating machine fault diagnosis using empirical mode decomposition. *Mechanical System and Signal Processing*, 22:1072–1081, 2008.
- [GJN05] H. Guo, L.B. Jack, and A.K. Nandi. Feature generation using genetic programming with application to fault classification. *IEEE Transactions on Systems Man and Cybernetics*, 35(1):89–99, 2005.
- [GT10] W. Guo and P. W. Tse. Enhancing the ability of ensemble empirical mode decomposition in machine fault diagnosis. In *Proceedings of Prognostics and Health Management Conference*, 2010.
- [GTD12] W. Guo, P.W. Tse, and A. Djordjevich. Faulty bearing signal recovery from large noise using a hybrid method based on spectral kurtosis and ensemble empirical mode decomposition. *Measurement*, 45(5):1308–1322, 2012.
- [GWMF09] L. Garibaldi, G. Wang, S. Marchesiello, and A. Fasana. Diagnosis of rolling elements bearing based on inverse autoregressive filter. *Key Engineering Materials*, 413-414:635–642, 2009.
- [HASC11] Y.S. Hong, S. H. Ahn, C. K. Song, and Y.M. Cho. Component-level fault diagnostics of a bevel gear using a wavelet packet transform. *Proceedings of the Institution of Mechanical Engineers Part E-journal of Process Mechanical Engineering*, 225(E1):1–12, 2011.
- [Haw80] D. M. Hawkins. *Identification of Outliers*. Monograph on Applied Probability and Statistics, 1980.
- [HGdP12] T. Heyns, S.J. Godsill, J.P. deVilliers, and Heyns P.S. Statistical gear health analysis which is robust to fluctuating loads and operating speeds. *Mechanical Systems and Signal Processing*, 27:651–666, 2012.
- [HHZZ07] Q. Hu, Z. He, Z. Zhang, and Y. Zi. Fault diagnosis of rotating machinery based on improved wavelet package transform and svms ensemble. *Mechanical Systems and Signal Processing*, 21:688–705, 2007.
- [HJL<sup>+</sup>98] M. Huang, L. Jiang, P. K. Liaw, C. R. Brooks, R. Seeley, and D. L. Klarstrom.

- Using acoustic emission in fatigue and fracture materials research. *JOM*, 50 (11), 1998.
- [HM97] F. Honarvar and H.R. Martin. New statistical moments for diagnostics of rolling element bearings. *Journal of Manufacturing Science and Engineering*, 119 (3):425–432, 1997.
- [How94] I.M. Howard. A review of rolling element bearing vibration: detection, diagnosis and prognosis. Technical report, DSTO Research report 0013, 1994.
- [HR00] D. Ho and R.B. Randall. Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals. *Mechanical System and Signal Processing*, 14:763–788, 2000.
- [HS05] N. E. Huang and S. Shen. *Hilbert-Huang Transform and Its Applications*. World Scientific, Singapore, 2005.
- [HSL<sup>+</sup>98] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and nonstationary time series analysis. In *Proceedings of the Royal Society*, pages 903–995, 1998.
- [JDY06] C. Junsheng, Y. Deije, and Y. Yu. A fault diagnosis approach for roller bearings based on emd method and ar model. *Mechanical System and Signal Processing*, 20:350–362, 2006.
- [JDY07] C. Junsheng, Y. Deije, and Y. Yu. Application of an impulse response wavelet to fault diagnosis of rolling bearings. *Mechanical System and Signal Processing*, 21:920–929, 2007.
- [JLB06] A.K.S. Jardine, D. Lin, and D. Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20:1483–1510, 2006.
- [Kim84] P.Y. Kim. A review of rolling element bearing health monitoring (ii): preliminary test results on current technologies. In *Proceedings of Machinery Vibration Monitoring and Analysis Meeting*, Vibration Institute, New Orleans, 1984.
- [Kja90] Bruel & Kjaer. Machine health monitoring using vibration analysis. In *Proceedings of the workshop presented by J. Courrech and R.B. Randall*, Monash University, 1990.
- [KNS09] K. Kappaganthu, C. Nataraj, and B. Samanta. Optimal feature set for detection of inner race defect in rolling element bearings. In *In proceedings of Annual Conference of PHM Society, San Diego*, 2009.
- [LCD<sup>+</sup>12] T. Liu, J. Chen, C. Dong, W. Xiao, and X. Zhou. The fault detection and diagnosis in rolling element bearings using frequency band entropy. In *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 2012.
- [LCTH00] B. Li, M.-Y. Chow, Y. Tipsuwan, and J. C. Hung. Neural-network-based motor

- rolling bearing fault diagnosis. *IEEE Transactions on industrial electronics*, 47 (5):1060–1069, 2000.
- [LHZC08] Y. Leia, Z. Hea, Y. Zia, and X. Chen. New clustering algorithm-based fault diagnosis using compensation distance evaluation technique. *Mechanical Systems and Signal Processing*, 22:419–435, 2008.
- [LLHZ11] Y. Lei, J. Lin, Z. He, and Y. Zi. Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, 25:1738–1749, 2011.
- [LP47] G. Lundberg and A. Palmgren. *Dynamic Capacity of Rolling Bearings*, volume 3 (1). 1947.
- [LQ03] L. Li and L. Qu. Cyclic statistics in rolling bearing diagnosis. *Journal of Sound and Vibration*, 267 (2):253–265, 2003.
- [LQL07] L. Li, L.S. Qu, and X.H. Liao. Haar wavelet for machine fault diagnosis. *Mechanical System and Signal Processing*, 21:1773–1786, 2007.
- [LZM<sup>+</sup>12] B. Li, P. Zhang, S. Mi, R. Hu, and D. Liu. An adaptive morphological gradient lifting wavelet for detecting bearing defects. *Mechanical Systems and Signal Processing*, 29:415–427, 2012.
- [MA84] J. Mathew and R.J. Alfredson. The condition monitoring of rolling element bearings using vibration analysis. *Trans ASME, J Vibr, Acoust*, 106:447–453, 1984.
- [Mat87] J. Mathew. Machine condition monitorin using vibrations analysis. *Journal of the Australian Acoustical Society*, 15 (1):7–13, 1987.
- [Mat02] *MATLAB 6.5, Release 13, 2002, Wavelet Toolbox, The MathWorks, Inc.*, 2002.
- [MG84] L.G. Martins and S.N.Y. Gerges. Comparison between signal analysis for detecting incipient bearing damage. In *Proceedings of the International Condition Monitoring Conference, Swansea, UK*, 1984.
- [MG04] A. Malhi and R.X Gao. Pca-based feature selection scheme for machine defect classification. *IEEE Transactions on Instrumentation and Measurement*, 53:1517–1525, 2004.
- [MIS92] H.R. Martin, F. Ismail, and A Sakuta. Algorithms for statistical moment evaluation for machine health monitoring. *Mechanical System and Signal Processing*, 6 (4):317–327, 1992.
- [MLBGA11] J.M. Machorro-López, A. Bellino, L. Garibaldi, and D.E. Adams. Pca-based techniques for detecting cracked rotating shafts including the effects of temperature variations. In *Proceeding 6th International Conference on Acoustical and Vibratory Surveillance Methods and Diagnostic Techniques, Compiègne, France*, 2011.
- [MM91] C.K. Mechefske and J. Mathew. Parametric spectral trending and classification for monitoring low speed bearings. In *Proceedings of the Asia-Pacific Vibration Conference, Monash University, Melbourne*, 1991.

- [MN98] A.C. McCormick and A.K. Nandi. Cyclostationary in rotating machine vibrations. *Mechanical System and Signal Processing*, 12:225–242, 1998.
- [MQ91] Q. Meng and L. Qu. Rotating machinery fault diagnosis using wigner distribution. *Mechanical System and Signal Processing*, 5 (3):155–166, 1991.
- [MS84a] P.D. McFadden and J.D. Smith. Model for the vibration produced by a single point defect in a rolling element bearing. *Journal of Sound and Vibration*, 96(1):69–82, 1984.
- [MS84b] P.D. McFadden and J.D. Smith. Vibration monitoring of rolling element bearings by the high frequency resonance technique - a review. *Tribol Int*, 17 (1):3–10, 1984.
- [MS86] T. Miyachi and K. Seki. An investigation of the early detection of defects in ball bearings using vibration monitoring - practical limit of detectability and growth speed of defects. In *Proceedings of the International Conference on Rotordynamics, JSMEIFTOMM, Tokyo*, 1986.
- [NK11] C. Nataraj and K. Kappaganthu. Vibration-based diagnostics of rolling element bearings: state of the art and challenges. In *In Proceedings of 13th World Congress in Mechanism and Machine Science, México*, 2011.
- [NP05] C. Nataraj and R.G. Pietrusko. Dynamic response of rigid rotors supported on rolling element bearings with an outer raceway defect. In *ASME Conference Proceedings*, 2005.
- [OG12] F. K. Omar and A.M. Gaouda. Dynamic wavelet-based tool for gearbox diagnosis. *Mechanical Systems and Signal Processing*, 26:190–204, 2012.
- [PC04] Z. K. Peng and F. L. Chu. Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography. *Mechanical System and Signal Processing*, 18:199–221, 2004.
- [PFGM12] M. Pirra, A. Fasana, L. Garibaldi, and S. Marchesiello. Damage identification and external effects removal for roller bearing diagnostics. In *In proceedings of 1st PHM Conference Europe, Dresden*, 2012.
- [PG13] M. Pirra and L. Garibaldi. Tecniche avanzate per la diagnostica di cuscinetti a rotolamento. accepted for "Tecniche nuove – Organi di trasmissione", 2013.
- [PGT+11] M. Pirra, E. Gandino, A. Torri, L. Garibaldi, and J. M. Machorro-López. Pca algorithm for detection, localisation and evolution of damages in gearbox bearings. *Journal of Physics. Conference series*, 305(1):1–10, 2011.
- [PMFG12] M. Pirra, S. Marchesiello, A. Fasana, and L. Garibaldi. External condition removal in bearing diagnostics through emd and one-class svm. In *In proceedings of XVIIIth symposium Vibrations, Chocs et Bruit VISHNO (Vibrations, shocks & noise), Paris, France*, 2012.
- [RA11] R.B. Randall and J. Antoni. Rolling element bearing diagnostics - a tutorial. *Mechanical System and Signal Processing*, 25:485–520, 2011.

- [RAC01] R.B. Randall, J. Antoni, and S. Chobsaard. The relationship between spectral correlation and envelope analysis for cyclostationary machine signals. application to ball bearing diagnostics. *Mechanical System and Signal Processing*, 15 (5):945–962, 2001.
- [Ran80] R.B. Randall. Advances in the application of cepstrum analysis to gearbox diagnosis. In *Proceedings of the 2nd International Conference on Vibrations in Rotating Machinery*, 1980.
- [Ran85] R.B. Randall. Computer aided vibration spectrum trend analysis for condition monitoring. *Maintenance Management International*, 5:161–167, 1985.
- [Ran87] R.B. Randall. *Frequency analysis*. 1987.
- [Ran04a] R.B. Randall. State of the art in monitoring rotating machinery - part 1. *Sound and Vibration*, 38(3):14–21, 2004.
- [Ran04b] R.B. Randall. State of the art in monitoring rotating machinery - part 2. *Sound and Vibration*, 38(5):10–17, 2004.
- [Ran10] R.B. Randall. Signal decomposition for symptom enhancement : A journey from fourier analysis through time/frequency analysis to blind source separation. In *Keynote lecture at ISMA, Leuven*, 2010.
- [RFG03] G. Rilling, P. Flandrin, and P. Gonçálvés. On empirical mode decomposition and its algorithms. In *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, Grado*, 2003.
- [RFGL07] G. Rilling, P. Flandrin, P. Goncalves, and J.M. Lilly. Bivariate empirical mode decomposition. *IEEE Signal Processing, Letters* 14:936–939, 2007.
- [RM86] D. E. Rumelhart and J. L. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press, 1986.
- [RM10a] N. Rehman and D.P. Mandic. Empirical mode decomposition for trivariate signals. *IEEE Transactions on Signal Processing*, 58:1059–1068, 2010.
- [RM10b] N. Rehman and D.P. Mandic. Multivariate empirical mode decomposition. In *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, volume 466, pages 1291–1302, 2010.
- [RP11] R. Ricci and P. Pennacchi. Diagnostics of gear faults based on emd and automatic selection of intrinsic mode functions. *Mechanical System and Signal Processing*, 25:821–838, 2011.
- [RSM12] R.B. Randall, N. Sawalhi, and Coats M. Separation of gear and bearing fault signals from a wind turbine transmission under varying speed and load. In *Proceedings of the 2nd International Conference on Condition Monitoring of Machinery in Non-Stationary Operations, Hammamet, Tunisia*, 2012.
- [RTH90] P. Rao, F. Taylor, and G. Harrison. Real time monitoring of vibration using the wigner distribution. *Sound and Vibration Magazine*, pages 22–25, 1990.
- [SABAA04] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi. Bearing fault detection using

- artificial neural networks and genetic algorithm. *EURASIP Journal on Applied Signal Processing*, 3:366–377, 2004.
- [SCJ84] A.F. Stronach, C.J. Cudworth, and A.B. Johnston. Condition monitoring of rolling element bearings. In *Proceedings of the International Condition Monitoring Conference, Swansea, UK*, 1984.
- [SEK05] H. J. Shin, D.-H. Eom, and S.-S. Kim. One-class support vector machines - an application in machine fault detection and classification. *Computer & Industrial Engineering*, 48:395–408, 2005.
- [Smi82] J.D.. Smith. Vibration monitoring of bearings at low speeds. *Tribology Int*, 15:139–144, 1982.
- [SMR21] V. Sugumaran, V. Muralidharan, and K.I. Ramachandran. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mechanical System and Signal Processing*, 930-942:2007, 21.
- [SR04] N. Sawalhi and R.B. Randall. The application of spectral kurtosis to bearing diagnostics. In *Proceedings of Acoustic, Gold Coast, Australia*, 2004.
- [SR08] N. Sawalhi and R.B. Randall. Simulating gear and bearing interactions in the presence of faults part i. the combined gear bearing dynamic model and the simulation of localised bearing faults. *Mechanical System and Signal Processing*, 22:1924–1951, 2008.
- [SRE07] N. Sawalhi, R.B. Randall, and H. Endo. The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. *Mechanical System and Signal Processing*, 21:2616–2633, 2007.
- [SSR08] V. Sugumaran, G.R. Sabareesh, and K.I. Ramachandran. Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine. *Mechanical System and Signal Processing*, 34:3090–3098, 2008.
- [Ste83] R.M. Stewart. *Application of signal processing techniques to machinery health monitoring*, chapter 23, pages 607–632. Noise and Vibration, Halsted Press, 1983.
- [SWS<sup>+</sup>00] B. Schölkopf, R. C. Williamson, A. J. Smola, J. S. Taylor, and J. C. Platt. Support vector method for novelty detection. *Advances in Neural Information Processing Systems*, 12:582–586, 2000.
- [Tan94] N. Tandon. A comparison of some vibration parameters for condition monitoring of rolling element bearings. *Measurement*, 12:285–286, 1994.
- [TC99] N. Tandon and A. Choudhury. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology International*, 32:469–480, 1999.
- [TJ11] W. Taiyong and W. Jinzhou. Fault diagnosis of rolling bearings based on wavelet packet and spectral kurtosis. In *Proceedings of the 4th International Conference on*



- Intelligent Computation Technology and Automation*, 2011.
- [TLLP12] W.C. Tsao, Y.F. Li, D.D. Le, and M.C. Pan. An insight concept to select appropriate imfs for envelope analysis of bearing fault diagnosis. *Measurement*, 45(6):1489–1498, 2012.
- [TN92a] N. Tandon and B.C. Nakra. Comparison of vibration and acoustic measurement techniques for the condition monitoring of rolling element bearings. *Tribology Int*, 25 (3):205–212, 1992.
- [TN92b] N. Tandon and B.C. Nakra. Vibration and acoustic monitoring techniques for the detection of defects in rolling element bearings - a review. *Shock Vibr Digest*, 24 (3):3–11, 1992.
- [TSK05] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. 2005.
- [Vap82] V. N. Vapnik. *Estimation of dependences based on empirical data*. 1982.
- [Vil48] J. Ville. Théorie et applications de la notion de signal analytique. *Cables et Transmission*, 2 A:61–74, 1948.
- [WGZ<sup>+</sup>11] Z. Wei, J. Gao, X. Zhong, Z. Jiang, and B. Ma. Incipient fault diagnosis of rolling element bearing based on wavelet packet transform and energy operator. *WSEAS Transactions on systems and control*, 10 (3):81–90, 2011.
- [Wig32] E.P. Wigner. On the quantum correction for thermodynamic equilibrium. *Physical Review*, 40:749–759, 1932.
- [WKS<sup>+</sup>09] A. Widodo, E. Y. Kim, J.-D. Son, B.-S. Yang, A.C.C. Tan, D.-S. Gu, B.-K. Choi, and J. Mathew. Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine. *Expert Systems with Applications*, 36:7252–7261, 2009.
- [WL11] Y. Wang and M. Liang. An adaptive sk technique and its application for fault detection of rolling element bearings. *Mechanical System and Signal Processing*, 25:1750–1764, 2011.
- [WL12] Y. Wang and M. Liang. Identification of multiple transient faults based on the adaptive spectral kurtosis method. *Journal of Sound and Vibration*, 331:470–486, 2012.
- [WS70] B. Weichbrodt and K.A. Smith. Signature analysis - non-intrusive techniques for incipient failure identification. Technical report, General Electric Technical Information Series 70-C-364, 1970.
- [WY07a] A. Widodo and B. Yang. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, 21:2560–2574, 2007.
- [WY07b] A. Widodo and B.S. Yang. Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors. *Mechanical System and Signal Processing*, 33:241–250, 2007.

- [XSK00] F. Xi, Q. Sun, and G. Krishnappa. Bearing diagnostics based on pattern recognition of statistical parameters. *Journal of Vibration and Control*, 6 (3):375–392, 2000.
- [YDJ06] Y. Yu, Y. Deije, and C. Junsheng. A roller bearing fault diagnosis method based on emd energy entropy and ann. *Journal of Sound and Vibration*, 294:269–277, 2006.
- [YTWW05] B. S. Yang, Han T., and Hwang W.-W. Fault diagnosis of rotating machiner based on multi-class support vector machines. *Journal of Mechanical Science and Technology*, 3:846–859, 2005.
- [YYC08] Y. Yang, D. Yu, and J. Cheng. A fault diagnosis approach for roller bearing based on imf envelope spectrum and svm. *Measurement*, 40:943–950, 2008.
- [ZB12] R. Zimroz and W. Bartelmus. Application of adaptive filtering for weak impulsive signal recovery for bearings local damage detection in complex mining mechanical systems working under condition of varying load. *Solid State Phenomena*, 180:250–257, 2012.
- [ZBBU11] R. Zimroz, W. Bartelmus, T. Barszcz, and J. Urbanek. Wind turbine main bearing diagnosis - a proposal of data processing and decision making procedure under non stationary load condition. In *Proceedings of SHM, Cracow*, 2011.
- [ZCP04] J. Zou, J. Chen, and Y.P. Pu. Wavelet time–frequency analysis of torsional vibrations in rotor system with a transverse crack. *Computers & Structures*, 82:1181–1187, 2004.
- [ZJN05] L. Zhang, L. B. Jack, and A. K. Nandi. Extending genetic programming for multi-class classification by combining. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2005.
- [ZPZ12] X. Zhao, T.H. Patel, and M.J. Zuo. Multivariate emd and full spectrum based condition monitoring for rotating machinery. *Mechanical System and Signal Processing*, 27:712–728, 2012.