eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Doctoral Thesis

Acquisition and processing of new data sources for improved condition monitoring of mechanical systems

Iñaki Bravo Imaz

Supervisors
Alfredo García Arribas
Aitor Arnaiz Irigaray

Leioa, 2018

Abstract

This research is focussed on the development of new ways to perform on-line condition monitoring in mechanical systems. It is focused on technologies that have been quite scarcely investigated in this field, in particular, the use of lubricant oil viscosity monitoring and motor current signature analysis technologies for improving the condition monitoring in gearboxes. On the one hand, a new solution based on magnetoelastic materials is presented for the monitorization of lubricant oil viscosity. On the other hand, motor current signature analysis (MCSA) is presented as a counterpart to traditional accelerometers for the monitorization of mechanical anomalies.

Broadly speaking, nowadays the most common maintenance practice for gearboxes is preventive maintenance. Lubricants and different mechanical parts such as gears and bearings are replaced on a periodic basis regardless of their health state. Steadily, condition based maintenance is gaining importance, as it offers several advantages over preventive maintenance. Still, one of the drawbacks of condition-based maintenance is the need to assess the health state of the component, either by adding sensors (on-line) or by making punctual measurements by qualified staff (periodic inspections). As the price of sensor technologies tends to go down and more electronics is incorporated to the machinery, the trend is to take advantage of the existing on-line monitoring options.

Lubricant oil is similar to the blood that flows inside us. Problems in our body have their reflection in blood, and in an analogous manner, problems in the gearbox will be manifested in the lubricant oil. That is the reason why lubricant oil must be monitored., and among the main lubricant properties, viscosity is the most important one. If oil viscosity does not remain in the right interval, lubrication won't accomplish its purpose, risking even a catastrophic failure of the mechanical component. However, viscosity measurements are still made via off-line methods, using slow and costly equipment systems. Therefore, in this Thesis the concept of on-line lubricant oil viscosity condition measurement is examined.

In particular, a magnetoelastic kinematic viscosity sensor for on-line or in-line measurements is designed, built and tested. The main advantage of the sensor prototype proposed is its capability to measure oil viscosity in a wide range of values (from 32 cSt up to 320 cSt), which is not known to any other sensor commercially available.

Magnetoelastic materials exhibit an intimate coupling between their magnetic and elastic properties, in such a way that a magnetic excitation produces an elastic response of the material and vice versa. The principles

governing the magnetoelastic effects and their beneficial use for the design of the on-line magnetoelastic viscosity sensor are explained in the Thesis.

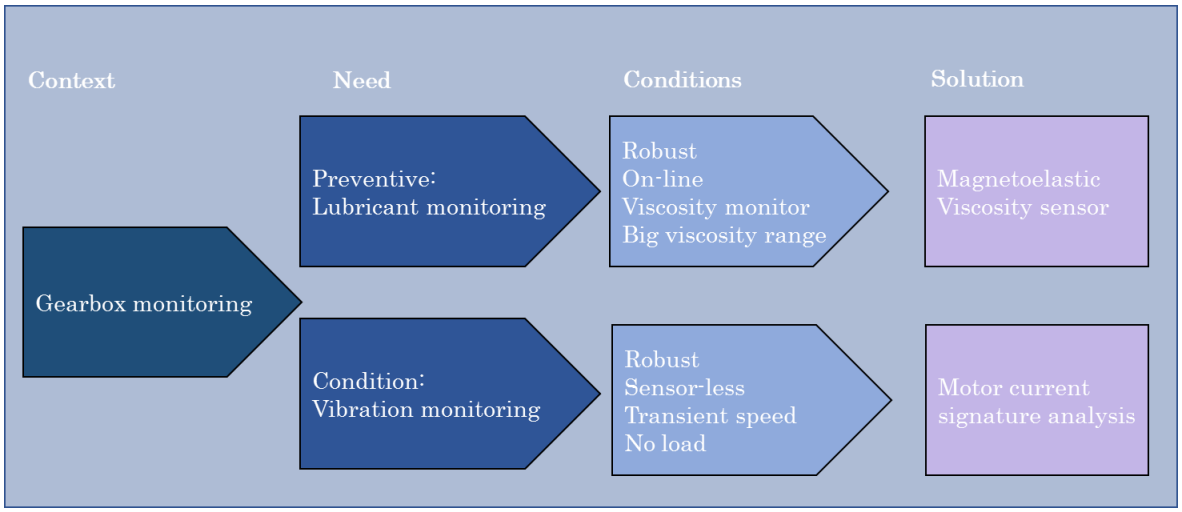
Two different prototypes are presented. The first one is intended as a proof-of-concept of the technology, and the second as a prototype of a practical device for the measurement of the viscosity of different oils. The signal processing to correlate the magnetoelastic response with the viscosity of the oils is described, including the use of a new phenomenological model. Likewise, the relationship between the temperature and the measurements has been studied.

Concerning motor current signature analysis (MCSA), the objective of the Thesis is to advance in the design of a system that can monitor a gearbox in normal operation. In this sense, the work is oriented towards the analysis of transients in speed, maintaining the load fixed. A gearbox test bench is used to reproduce different faults and acquire data in different operating conditions. With respect to the analysis of the current signals from the motor moving the gearbox, wavelet analysis is selected as the most convenient technique for the analysis of transients in speed.

The Thesis describes the experimental part of the work, including the test bench, the design of experiments, the type of gearbox faults monitored, and the organization of the data pool. The procedures for data reduction, preprocessing and analysis are presented, which produce different sets of features describing the health state of the system. The performance of these features is assessed using different classification algorithms and the results are discussed and compared including the comparison with other pre-processing techniques, such as dual level time synchronous averaging

The techniques developed have been applied to both gears and bearings inside the gearbox.

The investigation performed demonstrates that the combination of using transient information from the feeding current and the use of wavelets to analyze the data, maximize the value of motor current signal for condition monitoring in a gearbox, and enables its widespread deployment in maintenance procedures.


Context
Need
Conditions
Solution
Gearbox monitoring
Preventive:
Lubricant monitoring
Condition:
Vibration monitoring
Robust
On-line
Viscosity monitor
Big viscosity range
Robust
Sensor-less
Transient speed
No load
Magnetoelastic
Viscosity sensor
Motor current
signature analysis

Resumen

Este trabajo está centrado en el desarrollo de nuevas maneras de realizar la monitorización en línea del estado de salud de sistemas mecánicos mediante tecnologías poco utilizadas hasta ahora en este campo. En particular, se han investigado el uso de la monitorización de la viscosidad del aceite lubricante y la tecnología de análisis de las características de la corriente que alimenta el motor para mejorar la monitorización del estado de las cajas de engranajes. Por un lado, se presenta una nueva solución basada en materiales magnetoelásticos para la monitorización de la viscosidad del aceite lubricante. Por el otro, el análisis de la corriente alimentación del motor (MCSA por sus siglas en inglés) se presenta como alternativa de los acelerómetros tradicionales para la monitorización de anomalías mecánicas.

En general, la práctica de mantenimiento más habitual hoy en día, para las cajas de engranajes es el mantenimiento preventivo. Los lubricantes y las diferentes partes mecánicas, como los cojinetes, se reemplazan periódicamente, independientemente de su estado de salud. El mantenimiento basado en el estado de salud (*condition monitoring* en inglés) está ganando importancia progresivamente, ya que ofrece varias ventajas sobre el mantenimiento preventivo. Aun así, uno de los inconvenientes del *condition monitoring* es la necesidad de evaluar el estado de salud del componente, ya sea agregando sensores (en línea) o haciendo mediciones regulares por personal calificado (inspecciones periódicas). A medida que el precio de las tecnologías de monitorización con sensores tiende a disminuir y se incorporan más dispositivos electrónicos a la maquinaria, la tendencia es sacar provecho las opciones de monitorización en línea existentes.

El aceite lubricante es similar a la sangre que fluye dentro de nosotros. Los problemas en nuestro cuerpo tienen su reflejo en la sangre, y de manera análoga, los problemas en las cajas de engranajes se manifiestan en el aceite lubricante. Esa es la razón por la cual se debe monitorizar el aceite lubricante, y entre sus principales propiedades, la viscosidad es la más importante. Si la viscosidad del aceite no está en el intervalo correcto, no logrará su finalidad, incluso pudiendo crear un fallo catastrófico del componente. Sin embargo, la práctica habitual es realizar las mediciones través de sistemas de equipos fuera de línea, lentos y costosos. Por lo tanto, en esta Tesis se ha investigado el concepto de medición de la viscosidad del aceite lubricante en línea.

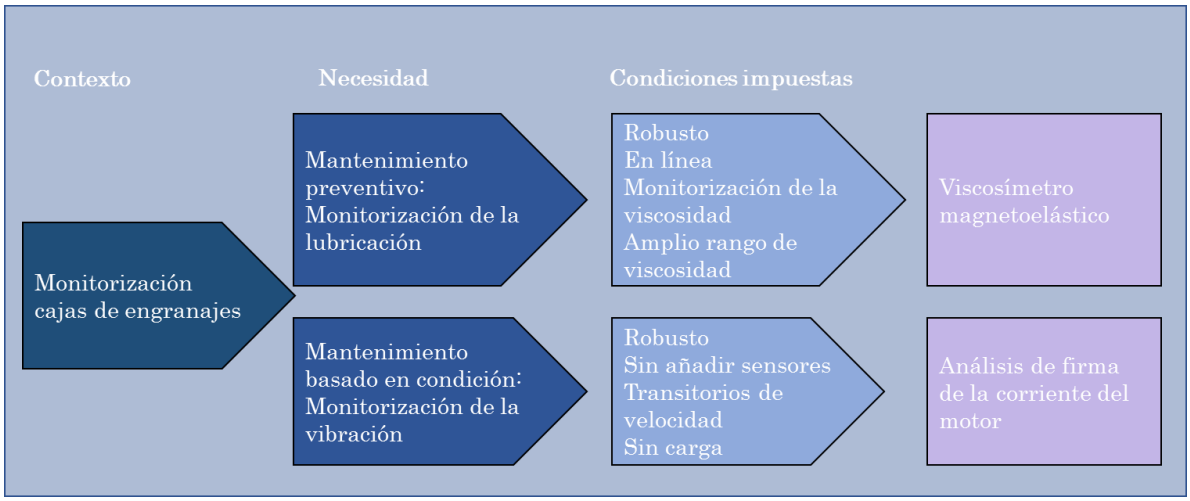
En particular, se ha desarrollado un sensor magnetoelástico de viscosidad cinemática para mediciones en línea. La principal ventaja del sensor propuesto es su capacidad de medir en una amplia gama de valores de viscosidad (desde 32 cSt hasta 320 cSt). No se conoce ningún otro sensor equivalente comercialmente disponible con un rango similar.

Los materiales magnetoelásticos presentan un acoplamiento entre sus propiedades magnéticas y elásticas, de modo que una perturbación magnética da lugar a una respuesta elástica y viceversa. En la Tesis se explican los principios que rigen los materiales magnetoelásticos y su uso de manera eficiente para el diseño de un sensor de viscosidad magnetoelástica en línea.

Se presenta el diseño de dos prototipos. El primero se concibe como una prueba de concepto de la tecnología, mientras que el segundo consiste en una implementación práctica para la medición de la viscosidad de diferentes aceites. Se describe así mismo el procesado de la señal obtenida para su correlación con la viscosidad de los aceites, incluyendo el uso de un nuevo modelo fenomenológico. Se analiza también los resultados de la medida de la viscosidad a diferentes temperaturas.

Con respecto al análisis de las características de la corriente de alimentación del motor (MCSA), el objetivo de la Tesis es poder diseñar un sistema para monitorizar una caja de engranajes en funcionamiento normal. En este sentido, se ha abordado el análisis de transitorios de velocidad, manteniendo la carga fija. Se ha utilizado un banco de pruebas de cajas de engranajes para reproducir diferentes fallos y adquirir datos en diferentes condiciones de operación. Con respecto al análisis de la señal de corriente del motor que mueve la caja de engranajes, se ha seleccionado el análisis *wavelet* como técnica adecuada para el análisis de los transitorios de velocidad. En la Tesis se presenta la parte experimental del trabajo, describiendo primero el banco de pruebas, el diseño de experimentos, el tipo de fallos de la caja de engranajes monitorizadas y la organización del conjunto de datos. A continuación se presenta el proceso de reducción de datos, pre-procesamiento y análisis, lo que proporciona como resultado un conjunto de características que describen el estado del sistema. El rendimiento de estas características se evalúa utilizando diferentes algoritmos de clasificación, comparando los resultados con otras técnicas de pre-procesamiento existentes, como el promediado síncrono en el tiempo de doble nivel. Las técnicas desarrolladas se han aplicado a la monitorización tanto de engranajes como de rodamientos en el interior de las cajas de engranajes.

Los resultados de esta investigación demuestran que la combinación del análisis de la corriente de alimentación del motor en régimen transitorio y el uso del pre-procesamiento basado en wavelets maximizará el valor de la señal de corriente del motor para determinar el estado de salud en una caja de engranajes, estableciendo la posibilidad de su despliegue generalizado en procedimientos de mantenimiento.


Contexto
Necesidad
Condiciones impuestas
Monitorización cajas de engranajes
Mantenimiento preventivo: Monitorización de la lubricación
Mantenimiento basado en condición: Monitorización de la vibración
Robusto
En línea
Monitorización de la viscosidad
Amplio rango de viscosidad
Robusto
Sin añadir sensores
Transitorios de velocidad
Sin carga
Viscosímetro magnetoelástico
Análisis de firma de la corriente del motor

Acknowledgments

It has been hard to arrive to be able to write these lines, passing over times where I thought it would simply be impossible to finish the document, just wanting to run away combined with other more positive moments. Overall and looking back the good times prevail over not so good ones. Anyway, I feel that I have progressed a lot since I started the Thesis both in the professional field and in the personal one.

The first person that I would like to thank is my supervisor Dr. Alfredo García-Arribas for his guidance, hard work and for being available every time I needed him. Another person that I must specially thank is my co-supervisor Dr. Aitor Arnaiz, again for the guidance and arduous work. It would not have been possible to be here without their dedication, supervision and wide knowledge on the subjects covered in this Thesis.

To IK4-Tekniker that gave me the opportunity of doing this Thesis, without its support it would have been impossible.

My colleagues from IK4-Tekniker deserve to be mentioned also. Specially the people with whom I have worked closely Eneko Gorritxategi, Egoitz Konde, Dr. Susana Ferreiro, Dr. Santiago Fernandez, Eduardo Gilabert, Kerman López de Calle, Dr. Alberto Villar and Deitze Otaduy. They have been and are a big support and helped in the confection of the Thesis. I cannot forget that they have also been a bit irritating asking when I was going to finish the Thesis. In this category, my colleague Dr. Francesco Pagano is the head with his emails asking when I was going to finally finish the Thesis, and trying to compete with whom would finish before (unfortunately he won that one but I hope to have a chance to return it). Also, Borja Pozo, David Cantero and Patricia Lopez for sharing those Saturdays of Thesis.

The people with whom I had the pleasure to work in my stage at the University of Cincinnati. Specially Dr. Jay Lee who accepted me in his group. My colleagues in the stage Dr. Hossein Davari Ardakani, Dr. David Siegel, Dr. Edzel Lapira, Dr. Wenyu Zhao, Zongchang Liu, Dr. Behrad Bagheri and Michael Lyons. They have been of significant help.

Dr. Matteo Corbetta for his help in this last moments of the Thesis.

Ms. Thu-Hien Pham form the DLR with whom I had the pleasure to work in the Actuation 2015 project.

The people of the Electricity and Electronic department of the Science and Technology faculty in the UPV/EHU that shared their time with me, Jon

Gutierrez, Iñaki Orue, Manu Barandiarán and Valentin Toyos. Prof. Dimitris Kouzoudis with whom I interchanged knowledge and material.

To my family Pedro Bravo and Ramoni Imaz for their support and understanding. To my girlfriend Leyre Iglesias for her help lift my spirit when I needed it.

To my friends (others than my colleagues in IK4-Tekniker which I also consider my friends) for the support elevating the mood by inviting to some beers. I wanted to highlight Carlos Bueno and María Vitores.

Thank you all from the bottom of my heart.

1

Introduction

The advances in different fields of technology and their economic success deeply rely in the correct operation of the equipment and machinery, without unscheduled interruptions or unexpected failures. In this context, the growing interest in safety, higher quality, better sustainability and in the preservation of goods, together with the pressure to increase efficiency, boost the importance of system maintenance to guarantee the faultless performance of the assets during the maximum useful life. This situation has paved the way towards the emergence of more advanced maintenance strategies.

The implementation of these advanced maintenance strategies involves moving from a *preventive* concept to a *condition-based* one [1.1]. In the first case, maintenance is done in scheduled and safe intervals regardless of the actual condition of the asset. This strategy cannot avoid unforeseen failures, since these time intervals are chosen according with the normal operation of the system, during which it is expected to function flawlessly. In contrast, condition based-maintenance takes into account the condition of the system by means of periodic inspections. Therefore, an unforeseen degradation of the asset or a malfunction can be detected and corrected by performing the maintenance operations, thus avoiding sudden failures.

The supervision of the state of the system is usually called *condition monitoring (CM)*[1] [1.2] and can be implemented in several ways. It can be performed manually, off-line in procedures also denominated non-destructive inspections[2], carried out periodically. During the inspections, an operator with enough expertise needs to access the equipment inspected and makes the required measurements. The analysis and interpretation of the data is also manually driven, and may take from seconds to days depending on several factors.

On the other hand, *on-line* is applied to condition monitoring scenarios where the sampling or measurement automatically generates a set of data. It can be sent to a processing system physically separated from the item monitored (remote analysis), or can also be locally analysed through embedded software systems. Finally, on-line information generated can be

[1] There are other terms that are closely related to Condition Monitoring and sometimes used equally. Health Monitoring (HM) has a similar meaning, often connected to aerospace applications and prognostic scenarios.

[2] Usually referred also as Non-Destructive Testing (NDT). NDT differs from other inspections where part of the component or system being tested is destructed in the inspections process - e.g. to cut and inspect a bearing with electronic microscope to find out root cause failure (e.g. changes in material mechanical properties).

used as support for human decision, as well as being part of an automated control system (e.g. halting mechanism depending on asset condition).

As the demand for continuous asset management increases, it also pushes the search of reliable and cost-effective ways of on-line monitoring [1.3]. It is therefore understood that, in the implementation of more favourable maintenance strategies, the automation is key, limiting as much as possible the manual inspection processes, to minimize costs and assure the efficient maintenance and operation of the assets. Hence, it is mandatory to acquire, process and communicate in an adequate way all the potentially useful signals that can be extracted from the system [1.4]. Reciprocally, the change of maintenance strategies has been fostered in the last years by the increase in options and technologies at the various stages of health monitoring (acquisition, transmission and processing of information), which are rendering a decrease in cost, thus making such policies applicable in new fields.

The development of new technologies continuously increases the level of automation in maintenance, but there is still considerable room for improvement, most of all in some specific domains. This is the case of gearboxes, a mechanical element that is essential and widely used in different fields such as manufacturing, energy, or transportation, to cite a few. The automation of the health monitoring process in gearboxes is, at least, challenging. As an example, the extraction of lubricant oil samples is not usually an easy task, and the installation of sensors to monitor vibrations may not be easy, not to mention the cost of the sensors. This affects the two areas of condition monitoring that are the focus of this thesis.

Regarding the lubricant oil, nowadays the most common way to monitor its health state in a machine part as a gearbox, is to take samples manually. The process is quite demanding as it involves a large degree of manual work, because, for example, the sampling point may not be easily accessible. Besides the laboratory analysis of the oil sample can take several days, during which the risk of a fault happening increases.

The vibration analysis can be more easily automated, but it still presents serious problems. The installation of sensors to monitor vibration may be cumbersome, since it may most probably involve shutting down the asset, while the suitable sensing points may not be easily reachable. In any case, a high mechanical knowledge of the system is needed for both the installation of the sensors and for the analysis of the gathered information. An operator with large expertise is needed for the analysis of the vibration signal no matter if it is performed manually or automatically, because the parameters

of the automated analysis must be set (*i.e.* initial conditions and limits). Another limiting factor, that in some cases prevents the deployment of vibration measuring sensors is the fact that their installation may reduce the reliability of the asset as a whole. Finally, the cost of the sensors, although steadily decreasing, constitutes sometimes a major drawback.

In this work, we are proposing two strategies for the on-line health monitoring, and the proposal is to apply them in one of the most typical mechanical components that can be found in machinery, *i.e.* gearboxes. Both techniques can be combined or used separately. On the one hand, a proof of concept for an on-line magnetoelastic sensor to determine the oil viscosity is proposed. This address the part of the condition monitoring dealing with the state of the lubricant oil in gearboxes, since a viscosity in the correct range is one of its most important properties to correctly perform its intended task. On the other hand, the health state of the gears inside a gearbox in a specially designed test stand is analysed through the current signal feeding the motor that drives the system, focusing specially in non-stationary conditions. This technique avoids the addition of additional sensors, as the feeding current is usually readily available in the control panel of the motor. Advanced signal processing techniques, such as wavelet analysis, have been applied to extract information of transients during speed transitions.

The organization of this thesis reflects the above structure. Following the introduction, there is a chapter dealing with the magnetoelastic viscosity sensor and another describing the motor current signature analysis. Both are quite independent from each other but share the effort to establish and on-line monitoring of the health state of gearboxes to implement a condition based maintenance scheme. In the rest of this introductory chapter some key concepts are explained which help establishing the context of the research work presented in this Thesis.

Content

1 Overview of industrial maintenance strategies

The importance of maintenance is such that it represents an average 4 % of the fixed assets costs in various industries, varying from 2-6 % for oil industry to 6/8% for steel and discrete manufacturing industries [1.5]. There is an enormous potential to moderate this cost since it is estimated that the amount can be reduced up to 50 % while a decrease in unforeseen accidents of 55 % can be achieved by applying the appropriate maintenance strategies.

Several attributes must be considered when selecting the type of maintenance [1.6]. The choice involves different aspects such as the investment required, possible safety and environmental problems, approximate failure costs, the reliability of the maintenance policy, the measured or estimated mean time between failures or the mean time to repair the facility in case of shutdown among others. Certainly, some of these factors are not easy to evaluate because of their intangible and complex nature. In fact, selecting the best maintenance strategy is usually a big challenge in itself.

There are three main different maintenance strategies [1.7] :

Corrective maintenance (CM): A faulty component is repaired (or adjusted, or changed) after the fault happened. It is still a very common strategy and, in fact, it can be very adequate in cases where failure is not frequent and does not have a great impact on the component operation (e.g. light bulbs within a factory).

Preventive maintenance (PM): A component is repaired in fixed time intervals regardless of the health state, before it reaches its useful life (which is defined as the period of time for which the asset will be economically feasible for use in a business, *i.e.* the asset is accomplishing the mission for which it was designed). This strategy is mainly implemented dealing with components that are under direct wear as well as on those subjected to fatigue, corrosion, oxidation or chemical instability, for instance. (*e.g.* the maintenance performed in cars, with interventions at fixed time or distance covered).

Condition based maintenance (CBM): The health state of the component is periodically (or even continuously) monitored so that it is repaired in the opportune moment, upon perceiving indications of a malfunction, but before the fault has happened. For instance, certain components produce measurable signals of the beginning of the deterioration process such as the increase of vibration, noise or heat, which indicate a potential fault that must be prevented. In this strategy, the interval between measurements of the

health state and the time for the adequate maintenance action must be smaller than the time to failure.

It is to be noted that, according to the International Standard (Figure 1) [1.8], condition based maintenance is officially reported as a particular type of preventive maintenance, where the interval between maintenance actions is driven by the condition instead of scheduled by 'time' or other inputs (preventive predetermined maintenance). However, for the sake of simplicity, during this Thesis we will distinguish between the three maintenance strategies (corrective, preventive, condition-based) as described above.

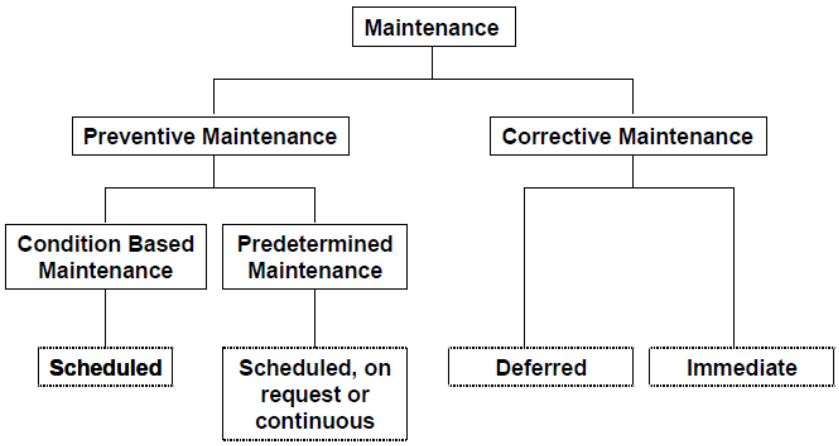


Figure 1. Overall view of the different maintenance strategies [1.8].

Additionally, some authors identify a fourth maintenance strategy denominated *proactive* maintenance [1.9]. It is defined as the set of actions aimed to identify, detect and correct the root causes that can potentially lead to a failure. The maintenance action is then executed to amend the conditions that could cause the deterioration of the system. Proactive maintenance strategies study the function and the degradation of the asset to discover incipient faults and correct them, before the performance is affected. Beneficially, proactive maintenance enlarges the life-time of the asset, due to the early detection of the root causes of the failures, and to the concomitant maintenance actions carried out. For instance, in the case of lubricant oil, if detected to be degraded before it causes a failure, and changed before the lubricated system suffer any damage, the expected life of the asset is enlarged. In definitive, proactive maintenance can be seen as an evolution of CBM, based in similar signal information brought about by the monitoring of the asset condition.

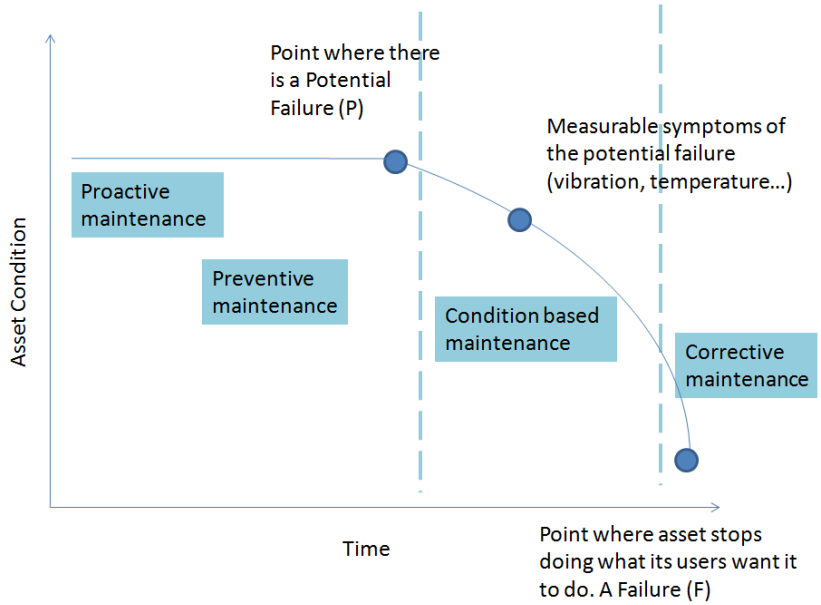


Figure 2. So-called P-F curve showing the time intervals were typically each type of maintenance strategy takes place. P indicates the moment where a potential failure starts developing and first symptoms appear. F indicates the moment where the failure develops and the asset is no longer able to perform the tasks in a normal manner.

Although no modern asset has been designed to last forever, in normal working conditions and with the adequate maintenance, it may last longer than the nominal lifetime. Figure 2 describes how each type of maintenance strategy is related to the condition of the asset and the time at which the maintenance is performed. The condition of the asset, represented by a relevant health index, is at its maximum value at the beginning, but start decreasing with time and decline rapidly towards a catastrophic failure. In Figure 2, the point P symbolizes a potential failure, representing a condition in which a functional failure does not affect the health of the system yet, but starts emitting signals (symptoms) anticipating the failure. The preventive maintenance must be scheduled before this P point. The potential failure is not detectable until the system presents measurable symptoms. This is when condition based maintenance should be performed. The assets can still continue working when a failure has begun, but eventually it will evolve to a functional failure (F) if the convenient actions have not been taken, causing the system to fail. Corrective maintenance is then necessary. It is worthy to highlight that, in real systems, combinations of these maintenance strategies are also implemented.

2 Gearboxes and traditional condition monitoring approaches

Gearboxes are the basic mechanical systems for controlled power transmission. They are conceived to transmit torque from an axis to another, by means of toothed wheels (gears). Gearboxes have been known to mankind since ancient times[3][1.10]. Today, they are widely used in transportation, impulsion, elevation, movement and, in general, in any application that implies the transmission of movement and torque.

A gearbox is basically composed by gears separated in different axis (shafts) that engage between them. The gears have teeth designed to ensure that the pitch circles of the engaging gears roll on each other without slipping, thus providing a smooth transmission of the rotary movement from one axis to the next. For its correct operation, a gearbox also includes other necessary components such as

- Shaft: connecting the gearbox to the external mechanical systems (e.g. engines) and to support internally the different gears.
- Bearings: required to support the shafts to the housing and to reduce friction.
- Lubricant: mandatory to reduce friction and to refrigerate the engaged parts.
- Seals: impeding the leaking of lubricant out of the gearbox.
- Housing: to enclose and robustly support the mechanical parts and the lubricant.

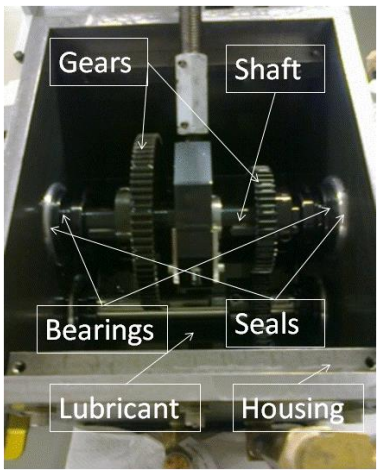


Figure 3. Main elements of gearboxes. GPS test bench.

[3] The earliest known geared artifact is the so-called "south pointing chariot". Dating from ca. 2600 BC, it is a chariot with a quite complex differential gear train that supposedly was designed to point south no matter the direction in which it was moving.

The design of the gearbox is determined by the application, the environment, and the physical constraints of the system in which it is to be used [1.11] Until the onset of the industrial revolution the requirements weren´t very demanding, since the power transmitted was not large, and the speeds were low. Gear design was then based on testing and specialized knowledge obtained by accumulated practical experience. The increasingly demanding needs of the industrial revolution fostered the design and fabrication of more complex gears with increased performance, while also providing improved tools for their manufacture. These developments have continued until present days where gears are commonplace.

Essentially, the configuration of the gearbox is defined by four main parameters:

- Power to be transmitted.
- Gear ratio required for reduction or multiplication.
- Speeds.
- Arrangement of shafting.

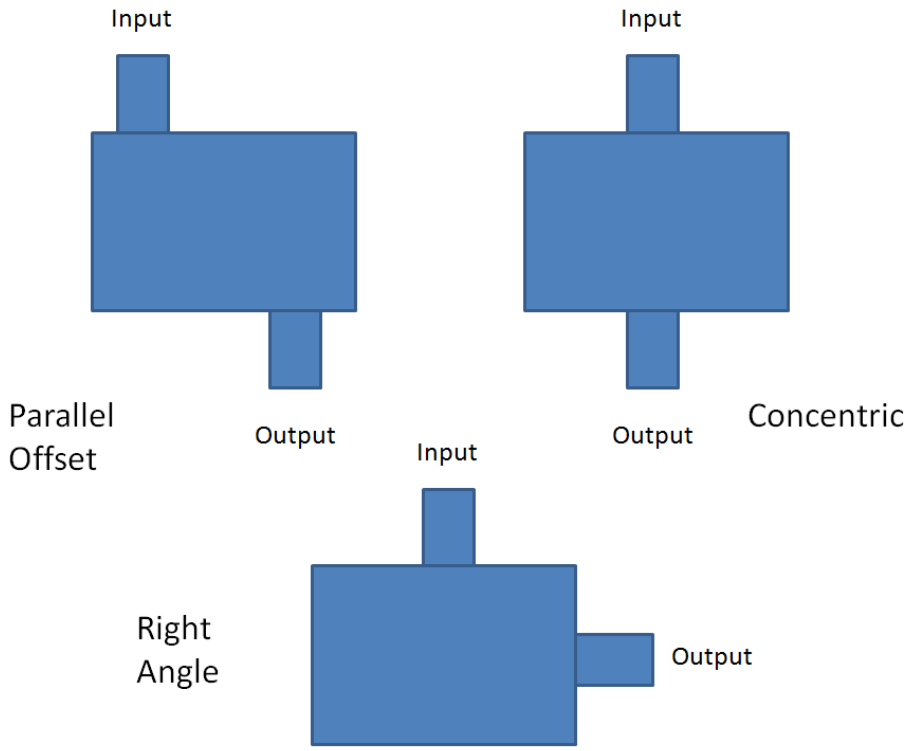


Figure 4. Types of gearbox regarding the input and the output of the shafts [1.11].

The main arrangements of input and output shaft configurations can be described as parallel offset, concentric and right angle, as depicted in Figure 4.

Two diverse types of gears are commonly used: spur and helical gears. Their shape is described in Figure 5.

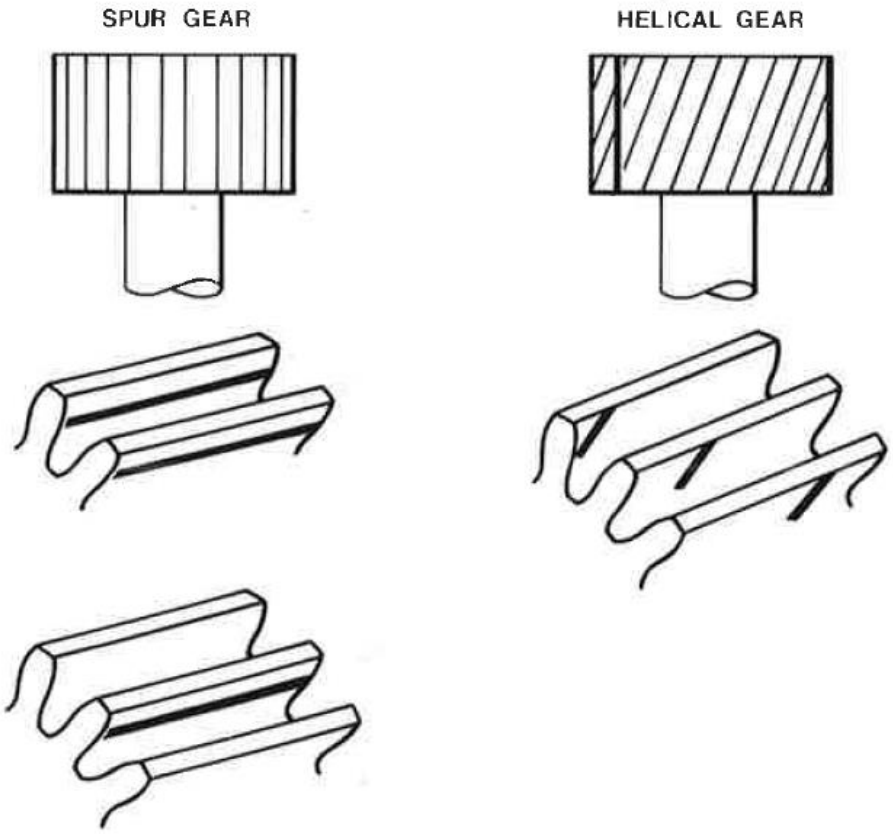


Figure 5. Comparison of spur and helical gear teeth with the contact line depicted for each of the cases [1.11].

Spur gears are straight-toothed gears, having radial teeth. Helical gears have the teeth cut in angle, making the meshing between gears more smooth and gradual than in the case of spur gears, generating less noise and vibration than them. Spur gears transmit the load by one or two gears at any time. They do not generate axial thrust. Helical gears, on the contrary, have an overlap in the axial and transverse planes. The load is shared between a sufficient number of teeth to allow a smooth transfer of load and a constant elastic flexibility. As helical gears generate axial thrust, double helical gears are also used. However, they must be adjusted axially, making their installation more complicated and requiring the use of more sophisticated bearings. Besides, helical gears are more complicated to be machined, making them more expensive. These shortcomings make spur gears highly widespread.

The design of gearboxes must consider not only the transmission requirements, but also its expected endurance and many other aspects such as efficiency, noise generation, space and weight limitation, to cite some.

Besides, the physical environment (*i.e.* dust, humidity, corrosive atmosphere, etc) must be taken into account, as they have implications in the design and maintenance of the gearbox.

2.1 Common problems and failures in gears

Generally, gearboxes are critical components of complex machinery and take up a major role in the assets layout. This implies that faults in the transmission train may have ruinous consequences in terms of cost, environment and even safety, since the life of people can be put at risk. Additionally, in most of the cases, the replacement of a gearbox is generally quite expensive, and may involve the shutdown of the complete asset.

To overcome these problems, gearboxes are designed and manufactured to be robust and reliable, but still failures occur that can produce severe malfunctioning [1.12]. Among them, those caused by bearings and gear teeth account for about 41 % of the cases, constituting the dominant causes of fault in a gearbox, which is why it is very important to monitor these elements. Specifically, in the case of gears, both the manufacturer and the user must take a system approach to the specification, design, installation, operation and maintenance of a gearbox in order to avoid the main failure modes, which can be identified as:

- Pitting: Is a fatigue phenomenon that occurs also in bearings. It occurs as a result of repeated stress cycles, which lead to surface and sub-surface cracks. They eventually make particles from the gear to be detached, generating pits. It is also known as spalling or flaking.

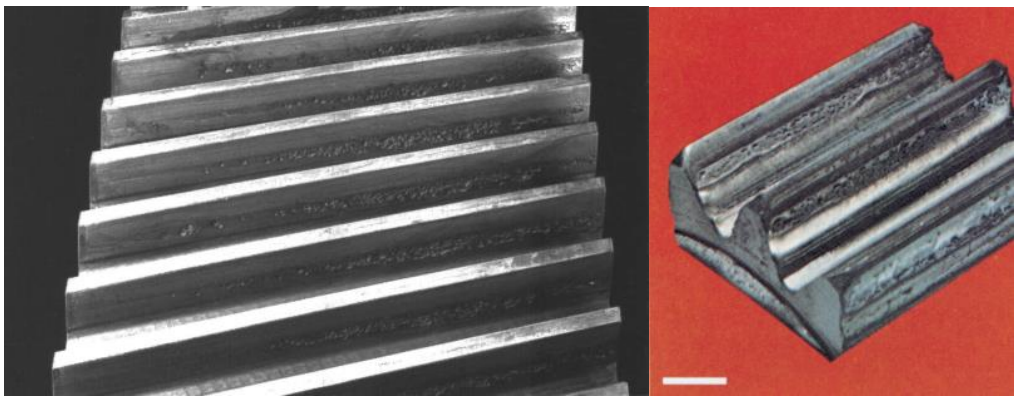


Figure 6. Macro-pitting on the surface of a gear [1.13].

- Breakage: The breakage of the gear teeth is caused by the root bending stress. Tooth breakage is more usually caused by bending fatigue rather than from overload exceeding the gear fracture strength.

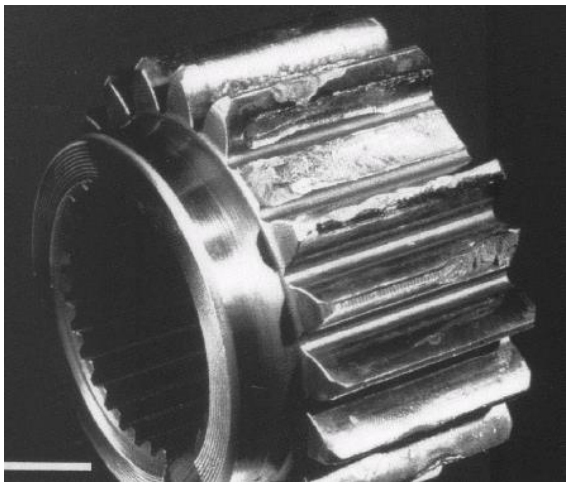


Figure 7. The breakage of a tooth in a gear [1.13].

- Scoring: Is a form of surface damage on the tooth flanks, which occurs when the lubricant film fails in its protective function, allowing metal to metal contact resulting in a local welding that is subsequently broken apart by the relative motion of the meshing gear teeth. Component surface material is transferred in macroscopic patches. It can also be known as galling or adhesive wear.

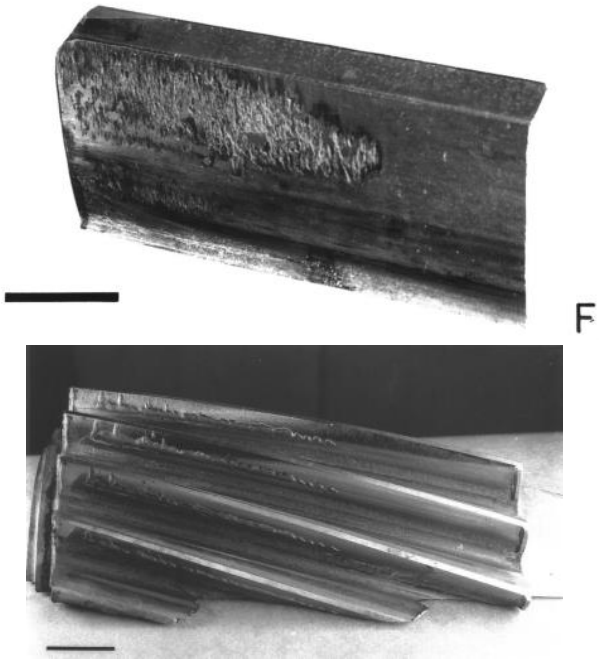


Figure 8. Scoring on helical gear teeth [1.13].

- Eccentricity: It is a failure mode usually related to an inadequate manufacturing processes leading to a misalignment between the centre of the shaft and the geometrical centre of the teeth corona.

Apart from the failures enumerated above, there are a number of additional problems that can also appear, regarding the complete condition

monitoring of a gearbox system. These can be related to misalignment or unbalance problems in the shafts, inadequate housing conditions, improper or corroded sealing systems that lead to lubricant fluid leakages, etc. As already pointed out, the present Thesis is focused on the monitoring of problems that are related either to the health state of the gears or to condition of the lubricant oil, which are introduced in section 3 below.

2.2 Maintenance and condition monitoring of gearboxes

If designed properly and used within the operational range, an adequate maintenance plan will prevent failures and keep the gearbox performing in proper conditions. In fact, the objective of the maintenance plan is to ensure a satisfactory performance at all times, including maintaining the transmission in a state of readiness during the periods that it is not in operation.

During the initial operation, there are several specific maintenance actions that must be performed. Typically, after some hours of operation (the actual period depends on the application), the coupling alignments of the shafts, the torque of the housing bolts, and the piping of the lubrication system must be checked. Besides, at this time the lubrication oil must be replaced, making sure that the oil reservoir is thoughtfully cleaned. During this process, every precaution must be taken to prevent foreign matter from entering the gearbox, since they can cause the degradation of the lubricant oil or the mechanical components.

In normal operation, a typical maintenance plan of gearboxes includes regular preventive maintenance actions. They are usually complemented by a manual monitoring of operation and inspection of the gearbox condition to determine its health state. This is a very important point since, typically, failures don´t appear suddenly, but they present an evolution that produces the appearance of certain symptoms in the assets. Among these, the most common ones are the raise of temperature and the onset of vibration.

According to Jantunen, the probability of a mechanical failure along the time follows a bathtub pattern (Figure 9) [1.14]. The evolution of the failures is as follows: there is an initial state where the failure rate is quite high but decays with time. In this stage, the dominant failure cause is the so-called early ‘infant mortality’ failure. After that, the dominant cause of failure becomes the constant (random) failures. Finally, in the last of the stages, an increase in the failure rate can be observed, with wear-out failures as the dominant ones.

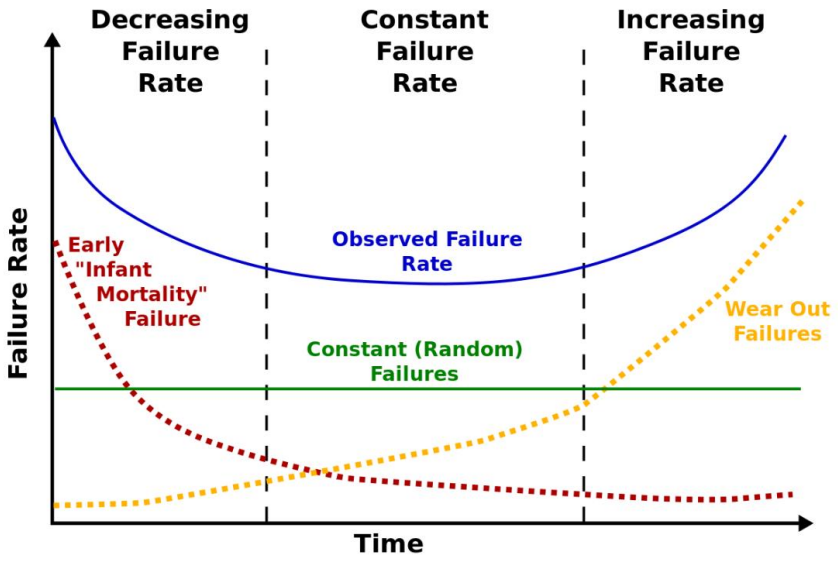


Figure 9. A bathtub curve illustrating typical pattern of failure probability (hazard) in gearboxes along the life time [1.14].

The mayor causes of a gear unit failure are improper lubrication and overload, which usually lead to the tooth wear problems indicated in previous section 2.1. Hence, the health state of the gearbox as a whole can be assessed through lubricant oil monitoring and by appraising the health state of the gears. This is the reason behind the traditional way of equipment monitoring, using the vibration signal and lubricant oil sampling [1.15].

The gear state in particular, is traditionally linked to the identification of certain vibration patterns as the development of gear vibration inside the gearbox (spur gears) is dominated by two phenomena:

1. Effects that are generated in the 'meshing tooth pair' – that is, the single tooth pair that is in full contact at each moment as the gears rotate. Typical effects are tooth deflection under load, uniformly distributed part of initial machining errors (also called 'running in'), or wear. They manifest in the tooth meshing frequency and its harmonics[4].
2. Variations between the teeth, which can be localized or distributed more uniformly around the gears. They manifest at other harmonics related to each gear rotational speeds. That is, vibration appears on the harmonics of the gear on which the faults are located. In this category we can find cracks, localized spalls, etc.

[4] Harmonic is any member of the harmonic series, the divergent infinite series. Every term of the series (i.e., the higher harmonics) after the first is the "harmonic mean" of the neighbouring terms [1.16].

Quantitatively, the basic vibration generating mechanism in gears is the so called “transmission error” [1.17]. It is the relative torsional vibration of the two gears, corrected by the gear ratio. In other words, it is the linear relative displacement along the line of action. In the initial phase of failures vibrations generate at high frequencies and low amplitudes. As the failure evolves, vibrations move to lower frequencies and higher amplitudes. In this transition they may become audible and, in the final stages, they can even be readily felt.

As the same geometric shapes always mesh in the same way, even with faults, the signals produced are deterministic. This means that the vibration signal can be used to determine the fault in the case of faulty gears.

The usual methods to monitor and measure vibrations involve the use of specific sensor devices (accelerometers) together with signal conditioning and data acquisition electronics that sample high frequency signals. The accelerometers, usually composed of piezo-resistive crystals, produce signals with specific frequencies and amplitudes depending on the existing failure, the stage of degradation or the operating conditions. The acquisition can be manual (through inspection activities where a portable device collects periodically samples) or on-line (where a specific acquisition infrastructure is continuously mounted to retrieve samples, usually within much shorter periods of time that in manual mode). Once the signal is acquired, two of the most typical analytic methods involve time-based analysis and frequency-based analysis.

Time based analysis relies in the use of statistical parameters obtained from the time domain signal. Typically, these parameters are root mean square, average, peak value, crest factor, skewness, kurtosis, median, minimum, maximum, deviation, variance, clearance factor, impulse factor, and shape factor [1.18].

In the case of the frequency-based analysis, the most common trend is to perform a fast Fourier transform (FFT). As the faults in a rotating machine often occur at a particular frequency, a frequency analysis is well suited. The problem with this type of analysis is to find the right compromise between the accuracy in frequency and the size of the frequency band studied.

This Thesis deals with novel methods to assess condition (lubricant viscosity and current) that will complement future application of on-line condition monitoring.

3 Assessment of lubricant condition

A lubricant is essentially a fluid composed by a base oil (from either a mineral or a synthetic origin) and a set of additives that enhance the base oil properties. Lubricant oil plays a major role in the working conditions of a gearbox. Its task is to prevent friction and wear among the different mechanical components that form the gearbox. At the same time, it provides refrigeration.

In normal operating conditions, lubricant oil forms a film between the mechanical components preventing direct contact [1.19]. This film must be the right thickness. If it is too thick, the mechanical components are subjected to extra stress, spoiling the work of the gearbox and, in the contrary, if it is too low, the mechanical components will contact, generating damage on them, excessive heat and general malfunction in the gearbox. Thus, viscosity directly determines the lubrication function.

Apart from preventing friction and wear, lubricants can also work as coolants, cleaning agents, electrical insulation and rust preventives.

3.1 Essential lubricant oil properties and degradation

Fluid degradation can be responsible for many kinds of equipment failure. A lubricant is subject to various operating conditions that may degrade it. Working conditions often include heat, moister, contamination (both internal and external), radiation, accidental mixing with other lubricant, trapped air and/or incompatible gases, etc.

The care of the lubricant oil and its selection is one of the principal objectives that must be achieved for the system to work in optimum conditions along its useful life.

Broadly, lubricant oil is adequate for further use if it has the specified viscosity, is dry (no water dissolved nor in emulsion), free of contaminants and contains sufficient concentration of key additives [1.20]. As a consequence, the main properties that require analysis are:

- Moist and water: Moist and water levels must be kept as low as possible, as they are harmful not only for the lubricant, but also for the mechanical components (gears, bearings…). Water ingress through sealings or condensations tend to create emulsions, which can be catastrophic to the asset. An amount of 0,1 % of water ingress can have disastrous effects.
- Acid number: Acid number measures the amount of alkaline needed to neutralize acid components in the lubricant oil. The increase of acid number can be an indicative of oxidation or contamination of lubricant

oil. For unused lubricant oil, the acid number is determined by the type and concentration of additives. As the additives are depleted, the acid number will varies. Comparing it to the total acid number gives an idea of the additives left. The oxidation of the base oil generally results in degradation products, with usually have an acid behaviour. Therefore, they may be spotted in the acid number. The organic acids that are created during the oxidation of the lubricant oil provoke corrosion, the formation of undesirable by-products such as varnish, sludge, etc.

- Additives: Lubricant oil additives are added to minimize oxidation of the oil, as well as to enhance the oil's anti-wear and extreme pressure performance. The chemistry of the additives is usually an industrial secret. As a result, the analysis will be limited to assess the oxidation inhibitors and anti-wear properties.
- Cleanliness: High level of particles indicates either an increase in wear, intrusion of external contaminants or improper function of oil filters in the circuit.
- Wear particles: Trend analysis of wear particles is a very effective tool to track changes in wear patterns. Analysis of wear particles should comprise analysis of individual elements and analysis of ferrous debris.
- Viscosity: This property is the most important property of any lubricant oil, as it determines oil film thickness, which affects directly the health state of the mechanical components, as it prevents components to have direct contact among them, causing mechanical degradation.

3.2 On-line lubricant condition monitoring

The control of the above properties during the lifetime of the lubricant is of crucial importance to assess the quality status of the lubricant during its use and to prevent failures and downtimes in mechanical systems, caused by lubricant degradation. Several other norms and recommendations can be applied, mainly depending on the type of application or the work environment of the asset.

The most common way of monitoring the lubricant oil is by means of manual sampling. However, during the last 15 years, as lubricant function and lubricated systems become more critical and miniaturization enables cost-effective on-line systems, on-line lubricant sensors have appeared. In this way, costly and slow analysis (that involves manual operation to extract and deliver samples to a laboratory) can be done in a matter of minutes/seconds. The monitoring can evolve from a monthly into a daily frequency. Several systems can be mentioned here:

Table 1. Some sensors for online monitoring the quality of lubricant oil.

Sensor	Properties measured	Measuring principle
Oilhealth [1.21]	Health index – Is an aggregated index taking into account oxidation (acid number), additives and water	Optical – Light passes through a RGB sensor array
Oilwear [1.21]	Wear particles - determining the origin and helping in the condition monitoring of the whole asset not just the lubricant oil	Optical – count and shape of particles
Fluidscan [1.22]	TAN, TBN, oxidation, water nitration, sulphation, glicol, additives AO, AW	Infrared analyser
ANALEXrs [1.23]	Water and acid numbers	Dielectric sensing
On-Line Metallic Wear Debris Sensor [1.24]	Simultaneous quantification of metallic composition, size category and particle count.	Not availeble
Eaton chip detection system [1.25]	Monitor cauntities of debris. Avility to burn off debris.	Dielectric

As introduced before and explained in detail in Chapter 2, viscosity reflects the dynamic (kinematic) and static properties of fluids.

It is especially interesting to monitor lubricant oil viscosity, as it is directly related with the oil film thickness between the mechanical components, which in turn is the root to many faults that lead to catastrophic failures.

However on-line viscosity measurement has not evolved as much as the monitoring of the rest of parameters. Two main types of devices exist for measuring viscosity: displacement-based and vibration- based [1.26]. Solid micro-displacement based sensors, which are based in methods used in laboratory conditions (capillary, rotary and dropping ball), are difficult to translate to on-line analysis. For on-line measurements, vibration based techniques, ranging from macro vibration mechanical methods to micro-vibration methods are more suitable. It is in this last family (micro-vibration method) where magnetoelastic viscosity sensors are inscribed.

In this Thesis, a magnetoelastic kinematic viscosity sensor for on-line or in-line detection is developed. The main advantage of the sensor proposed is the ability to measure in a wide range of viscosity values (32 to 320 cSt), which is not known to any other sensor commercially available.

4 Motor current signature analysis

The health monitoring by means of vibration analysis is not widespread because it has several drawbacks. First, sensors must be installed in the asset to be monitored, often in small spaces and difficult places to reach. Therefore, high knowledge of the mechanical design of the asset is needed for the installation of the sensor in order to find the right location that will maximize the desired signal and minimize other noises. Additionally, the cost of sensor purchase and installation can also be an issue. Although the price of the sensors has dropped dramatically in the past years, they are still perceived as an expensive piece of equipment, not worth their potential benefits.

In this Thesis, the use of internal signals is proposed for monitoring the health of the gearbox. In particular, it focuses on the current signal of the electric motor that drives the gearboxes. Current signal is already available in these systems where a control is used. Examples span from drivetrain connections between generator and gearbox in aerogenerators to actuation mechanisms in aircrafts (e.g. primary and secondary control systems in flaps, etc). The monitoring methods relying in current signals are especially focused on a first level of condition analysis related to 'anomaly detection' [1.2]: a 'screening' mechanism that identifies anomalous patterns and triggers further monitoring and diagnosis actions. In principle, internal signals give an economical approach regarding condition monitoring since they avoid the expenditure and implementation problems of adding sensors, although they may require complex signal processing.

Analysing the motor current signature is an indirect way of measuring the vibration generated by a mechanical fault. If a fault condition does exist, the effective load torque varies with the rotor position. Load fluctuations involve speed changes in electric motors. As a result of speed variations, adjustments in the per unit slip happen which consecutively cause shifts in the sidebands across the line frequency [1.27]. This is the principle in which the motor current analysis relies.

Motor current signature analysis has been used traditionally for the diagnosis of electric motor condition [1.28]. The condition of the winding, broken rotor bars and the internal bearings has been assessed. However, the analysis the health state of gears trough motor current has been scarcely used before in research works.

This Thesis assumes that the challenge in the identification of mechanical anomalies from current analysis may be related with the way that current signals are acquired and analysed. This is summarized in two aspects: The use transient operation intervals for data collection, and the analysis of such

data by traditional techniques (e.g. Time-domain, FFT) or by less traditional ones (i.e. wavelets)

Concerning data collection, the analysis of mechanical faults, as traditionally done through vibration signals, is normally based on a repetitive collection of signal 'samples' under continuous operating conditions (*e.g.* load, speed) in order to average them and reduce noise [1.29].

However, transients are commonplace in the life of assets, and especially in the case of gearboxes. In the normal basis operation, it is very difficult to find any moment were an asset is not working in transitory conditions, either in load or in speed. In the case of gearboxes, the most common strategy of condition monitoring is to avoid these variations and perform some specially designed tests, making the gearbox turn at constant speed and charge. This approach has two drawbacks: First, it implies an intrusive way of performing the data acquisition since the operator must set the machine to some specific working conditions where the continuous operation can be recorded. Second transient operations (especially when there is an increase in load or speed) can include important information about the machinery condition that can be unnoticed during continuous operation regimes. Thus, the idea of recording data in transient regimes seems very relevant to maximize information extracted from the current signal analysis.

Concerning the analysis of the acquired, the most common trend in current signal analysis is the application of time domain analysis (using characteristic values) and frequency domain (spectrum) analysis, as well as Cepstrum analysis (the result of taking the inverse Fourier transform (IFT) of the logarithm of the estimated spectrum of a signal [1.30]). These signal analysis techniques are similar to the techniques used in the case of detecting problems in the motor itself and to the analysis of vibration data coming from accelerometers.

However, the use of transients in the signal adds complexity to the analysis. Typical signal analysis techniques, such as FFT may not be adequate. Having this in mind, wavelet analysis has been suggested as more adequate method for the case of transient analysis [1.31]. The Fourier analysis consists in breaking up a signal into sine waves with different frequencies. Similarly, a wavelet analysis consists in the breaking-up of a signal into shifted and scaled versions of a function called the mother wavelet [1.32] which can be of different types as in [1.33]. Figure 10 displays two different mother wavelets of the types used in this work.

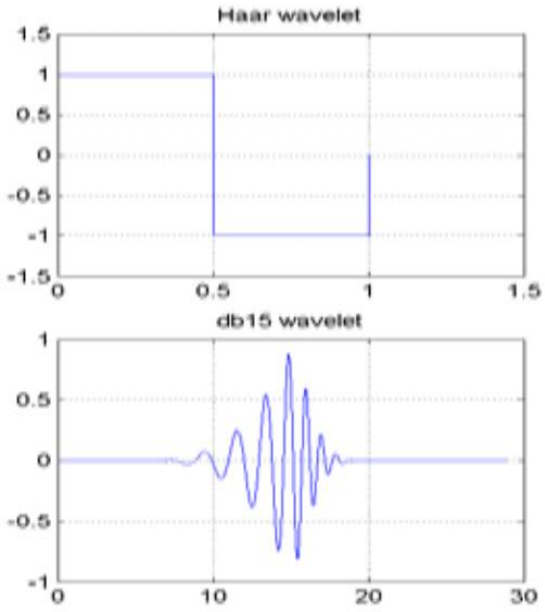


Figure 10. Haar and Daubechies 15 mother wavelets.

As in the case of the Fourier transform with FTT, a discrete wavelet transform exists, that is more efficient computationally.

The main advantage of the wavelet analysis over FFT is that transients can be analysed and that the resolution can be tuned depending on the frequency at high frequencies the resolution is higher in time and lower in frequency while, at low frequencies the inverse happens.

5 Objectives and structure of this thesis

This Thesis focuses on the analysis of novel technologies that can be utilized for on-line condition monitoring, and particularly for the monitoring of mechanical faults in gearboxes. The objective of the Thesis is to advance in the design of a system that monitors a gearbox in normal operation. The technologies presented measure two specific indicators of the health state: lubricant viscosity and motor current. The Thesis analyses their suitability, monitoring performance and potential applicability. The signal acquisition, processing and maintenance decision support have been analysed for each of the particular indicators measured.

Concerning lubricant viscosity, a magnetoelastic kinematic viscosity sensor for on-line or in-line measurements is developed. The main advantage of the sensor proposed is the ability to measure in a wide range of viscosity values (32 cSt up to 320 cSt), which is not known to any other sensor commercially available.

Concerning current analysis, being is a very novel tool for identifying gear anomalies, vibration analysis from accelerometers will be first used as a reference method to assess the performance of the novel current signals based analysis, as a tool for determining gear faults. Afterwards current signals are directly analysed using time-based and frequency based methods.

A gearbox test bench is used to reproduce different faults and acquire data at different operating conditions. The analysis of transients in speed, maintaining the load fixed, is addressed. Wavelet analysis is selected as a technique for the analysis of transients in speed. It is expected that this combination (the use of transient information from motor current, and the use of wavelets to analyses this data) will maximize the value of motor current signal for condition monitoring in a gearbox, enabling thus its widespread usage.

The structure of the chapters is as follows:

Chapter 1 Introduction presents the context of the Thesis inside the subject of the condition monitoring of gearboxes. First, an overview of different maintenance strategies is done, what gives the frame for the condition monitoring techniques that will be disclosed in detail in chapters 2 and 3.

The traditional approaches for determining the health state of a gearbox are presented., introducing the most common causes of failure and how they have been avoided by condition monitoring up to date.

The importance of lubrication is highlighted. Explaining the lubricant oil properties and how they change as the oil is degraded. Also, an overview of how lubricant oil is currently monitored is displayed.

Last, an overview of current signature analysis is done, explaining shallowly the working principles, and how they will be applied in the context of this Thesis.

Chapter 2 Magnetoelastic sensor for on-line monitoring of lubricant oil viscosity is centered in the development of a magnetoelastic sensor for the monitorization of lubricant viscosity. The principles of lubrication of a gearbox are introduced, with an explication of the different lubrications regimes that be found in gearboxes. The importance of viscosity for a correct lubrication is highlighted.

The difference between kinematic and absolute viscosity is revealed, together with the classification of the different viscosity of lubricants oils as determined by the most used indexes (SAE, ISO and military specification

index). The main laboratory methods for viscosity measured are also introduced in this chapter.

The importance of viscosity, and how an incorrect viscosity could affect the asset is accentuated, focusing on the causes for the inadequate viscosity.

A state of the art of on-line viscosity sensors is displayed. Special attention is given to other on-line magnetoelastic viscosity sensors. After this, the principles governing magnetoelastic materials and how it is possible to take profit of them for their use in the design of an on-line magnetoelastic viscosity sensor are explained.

The design of two prototypes is exposed. The first one is intended as a proof of concept of the technology, and another the second as a prototype of a practical device for the measurement of the viscosity of different oils. The signal processing to correlate the magnetoelastic response with the viscosity of the oils is described, including the use of a phenomenological model. Likewise, the relationship between the temperature and the measurements has been studied.

Chapter 3 Motor current signature analysis describes the procedure to monitor the health state of a gearbox analyzing the current feeding the motor that drives the gearbox.

The theoretical background is exposed, disclosing the mechanism that permits mechanical faults in the kinematic chain to be detected in the stator current.

The techniques used to analyze the signal, for the case of transient speed of the motor, are presented.

Then the experimental part of the work is explained describing the test bench, the modifications introduced, the design of experiments, the faults and how the data pool obtained is organized. The process for data reduction and analysis is explained. Finally, the results are examined and discussed.

As a complement, a similar analysis is performed.

Chapter 4 Conclusions resumes the main achievements of the work performed in this Thesis.

Bibliography

[1.1] JARDINE, Andrew KS; LIN, Daming; BANJEVIC, Dragan. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 2006, vol. 20, no 7, p. 1483-1510

[1.2] RAO, B. K. N. (ed.). *Handbook of condition monitoring*. Elsevier, 1996

[1.3] ROY, Rajkumar, *et al.* Continuous maintenance and the future–Foundations and technological challenges. *CIRP Annals-Manufacturing Technology*, 2016, vol. 65, no 2, p. 667-688

[1.4] JANTUNEN, Erkki, *et al.* Economical and technological prospects for e-maintenance. *International Journal of System Assurance Engineering and Management*, 2010, vol. 1, no 3, p. 201-209.

[1.5] MASOUMI, Shiba, BAGHERI, Seyedeh Maryam Babanejad, ARABI, Ali Mohammadpour, Role of maintenance and repairs to reduce production costs in the industries of Mazandaran from managerial view, *World of Sciences Journal*, 2013, vol. 2, p. 100–113.

[1.6] BEVILACQUA, Maurizio; BRAGLIA, Marcello. The analytic hierarchy process applied to maintenance strategy selection. *Reliability Engineering & System Safety*, 2000, vol. 70, no 1, p. 71-83

[1.7] FITCH, Ernest C. *Proactive maintenance for mechanical systems*. Elsevier, 2013

[1.8] DIN EN 13306: Maintenance – Maintenance terminology 13306:2008, DIN Deutsches Institut für Normung e.V., 2008

[1.9] MULLER, Alexandre; SUHNER, Marie-Christine; IUNG, Benoît. Formalisation of a new prognosis model for supporting proactive maintenance implementation on industrial system. *Reliability Engineering & System Safety*, 2008, vol. 93, no 2, p. 234-253

[1.10] RADZEVICH, Stephen P. *Dudley's handbook of practical gear design and manufacture*. CRC Press, 2016

[1.11] SAWYER, John W., HALLBERG, Kurt, *Sawyer's turbomachinery maintenance handbook*, Turbomachinery International Publications, 1980.

[1.12] HARALANOVA, Valentina; KHOSHABA, Samir. FAILURES IN GEARBOXES DUE TO MATERIAL PROBLEMS. *Materials, Methods & Technologies*, 2016, vol. 10, p. 112-136

[1.13] TALLIAN, Tibor E. The Failure Atlas For Hertz Contact Machine Elements. *Mechanical Engineering*, 1992, vol. 114, no 3, p. 66.

[1.14] JANTUNEN, Erkki, *et al.* Identification of wear statistics to determine the need for a new approach to maintenance. 2014.

[1.15] RANDALL, Robert B. State of the art in monitoring rotating machinery-part 1. *Sound and vibration*, 2004, vol. 38, no 3, p. 14-21

[1.16] Harmonic. En.wikipedia.org. Retrieved 19 January 2018, from https://en.wikipedia.org/wiki/Harmonic

[1.17] RANDALL, Robert B. State of the art in monitoring rotating machinery-part 2. *Sound and Vibration*, 2004, vol. 38, no 5, p. 10-17

[1.18] CHANDRAN, Praneeth, et al. Application of Laplace wavelet kurtosis and wavelet statistical parameters for gear fault diagnosis. *International Journal of Multidisciplinary Sciences and Engineering*, 2012, vol. 3, no 9, p. 1-7.

[1.19] IGARTUA, Amaya, *et al.*, 2012, *Tribology of gears*

[1.20] UKONSAARI, Jan and MOLLER, Hans, *Oil cleanliness in wind power gearboxes*. Elforsk rapport 12:52, 2012

[1.21] Description-Atten2. Atten2. Retrieved 19 January 2018, from http://www.atten2.com/

[1.22] FluidScan® 1000 Series - Handheld Infrared Oil Analyzer - Portable - Spectro Scientific. Spectrosci.com. Retrieved 19 January 2018, from https://www.spectrosci.com/product/q1000/

[1.23] ANALEXrs Oil Condition Sensor - Kittiwake - PDF Catalogues | Documentation | Boating Brochures. Pdf.nauticexpo.com. Retrieved 19 January 2018, from http://pdf.nauticexpo.com/pdf/kittiwake/analexrs-oil-condition-sensor/31223-93741.html

[1.24] On-Line Metallic Wear Debris Sensor | Parker Kittiwake. Kittiwake.com. Retrieved 19 January 2018, from http://www.kittiwake.com/metallic-wear-debris-sensor

[1.25] Eaton - Aerospace Sensor and controls - Smart Zapper System. Herberaircraft.com. Retrieved 19 January 2018, from https://www.herberaircraft.com/pdf/Glenolden/Smart%20Zapper%20System.pdf.

[1.26] WU, TongHai, *et al.* Progress and trend of sensor technology for on-line oil monitoring. *Science China Technological Sciences*, 2013, vol. 56, no 12, p. 2914-2926.

[1.27] SCHOEN, Randy R., *et al.* Motor bearing damage detection using stator current monitoring. *IEEE transactions on industry applications*, 1995, vol. 31, no 6, p. 1274-1279.

[1.28] BENBOUZID, M. El Hachemi. A review of induction motors signature analysis as a medium for faults detection. *IEEE transactions on industrial electronics*, 2000, vol. 47, no 5, p. 984-993.

[1.29] BEDIAGA, Inigo, *et al.* Ball bearing damage detection using traditional signal processing algorithms. *IEEE Instrumentation & Measurement Magazine*, 2013, vol. 16, no 2, p. 20-25

[1.30] Cepstrum. En.wikipedia.org. Retrieved 19 January 2018, from https://en.wikipedia.org/wiki/Cepstrum

[1.31] KAR, Chinmaya; MOHANTY, A. R. Monitoring gear vibrations through motor current signature analysis and wavelet transform. *Mechanical systems and signal processing*, 2006, vol. 20, no 1, p. 158-187.

[1.32] I ROURA, Jordi Cusidó; MARTÍNEZ, Jose Luis Romeral. Transient Analysis and Motor Fault Detection using the Wavelet Transform. En *Discrete Wavelet Transforms-Theory and Applications*. InTech, 2011.

[1.33] ESCAMILLA-AMBROSIO, P. J., *et al.* ANFIS-wavelet packket transform approach to structural healt monitoring. *Ratio*, 2010, vol. 10, p. 1.

2

Magnetoelastic sensor for on-line monitoring of lubricant oil viscosity

This chapter focuses on the development of a magnetoelastic sensor for the on-line determination of the viscosity of lubricant oils, as a proof of concept of the technology. The motivation and the objectives for measuring viscosity, the design of the magnetoelastic sensor, the development of the prototype, and the results achieved will be described in detail.

The term lubrication is applied to the process and/or technique employed to reduce friction between, and wear of one or both, surfaces in proximity and in relative movement to each other [2.1]. This is accomplished by interposing a substance, called a lubricant, between the moving surfaces. When the lubrication of a moving machine is performed by a fluid, the load applied to the system is, most of the times, transmitted by the pressure generated within the liquid, due to the frictional viscous resistance to motion of the lubricating fluid between the moving surfaces of the system. This is the case occurring in the lubrication of gearboxes where the lubricant is normally in liquid state and is commonly referred as *lubricant oi*l, in contraposition to lubricant 'greases', which have a denser appearance. Greases are usually employed in sliding mechanisms such as bearings, slides and guide ways.

Concerning the maintenance of gearboxes, including their rotating parts, such as rolling bearings and gears, it is important to state that almost any problem appearing during their function will be reflected in the state of the lubricant oil, as it collects the particles and other strange substances that can be formed due to the malfunction. In addition, we must take into account that the health state of the lubricated parts impacts not only on the correct performance of the gearbox, but on the whole asset. That is why it is of paramount importance to monitor the health state of the lubricant oil in order to assess the health condition of the gearbox.

The viscosity of the lubricant oil is one of the most important parameters to be monitored. Therefore, we have concentrated our efforts in establishing a method to determine the viscosity of the lubricant oil, which could be implemented as an on-line monitoring system for assessing the health state of gearboxes.

The first part of this chapter presents a general overview of lubricants, where the role of lubricants and the different types of lubrication are described. As this investigation is centered on gearboxes, we focus on petroleum based lubricant oils, since it is the most suited type of oil and practically the only one in use in such application. Some information on its manufacturing process and common classifications in use are also given. The importance of viscosity as an oil parameter and the need to monitor it on-line is highlighted.

Next, the existing approaches for on-line viscosity measurement are briefly reviewed, highlighting the benefits and suitability of the magnetoelasticity-based sensors for this task. The basic principles of magnetoelasticity and magnetoelastic resonance are examined, and the components and materials of a magnetoelastic viscosity sensor are presented.

The chapter ends with the most important part in which the design and development of a laboratory prototype of on-line magnetoelastic viscosity sensor is described. The experiments performed to test the demonstrator, and the results obtained in various situations are presented and analyzed. The applicability of this prototype in an on-line, real-time system to assess the lubricant oil health state is examined.

Content

1 Gearbox lubrication and viscosity monitoring

The lubricant oil plays a major role in the working conditions of a gearbox. Its task is to prevent friction and wear among the different mechanical components that constitute the gearbox. Additionally, lubricants can also work as coolants, cleaning agents, electrical insulation and rust preventives.

Historically, the first used lubricants were based on crude lumps of animal fat, in a time when the mechanical parts of early machinery were still made of wood [2.2]. Eventually, iron and brass we introduced in moving parts, needing new lubricants. Animal and vegetal oils and well as mixtures of both types of oils emerged for this second-generation of lubricants. Mainly tallow, olive oil, castor oil, peanut oil and rape oil were first used and after, about 16th century, whale oil and porpoise oil started to be widely used. In 19th century petroleum-based lubricants came into play, rapidly dominating the lubricant scene. Considering that early machines were relatively simple, there was no need for lubricants more sophisticated than simple crude oil. As larger, faster and more sophisticated machines were developed, the lubricants evolved concomitantly. The development of superior quality lubricants, the use of special additives and the introduction of automatic dispensations systems have accounted for the major advances in the science of lubrication.

1.1 Modern gearbox lubrication

Today, there are basically five different types of lubricants, depending on their composition, origin, or nature:

- Petroleum based oils.
- Synthetic oils.
- Vegetable based oils.
- Grease.
- Solid lubricants.
- Gas lubricants.

Petroleum-based oils and synthetic oils are commonly the preferred lubricants because they provide good resistance and superior lubricating capabilities, especially in comparison with vegetable and animal oils, and can be circulated to remove heat and contaminants, which is a big advantage over solid lubricants. They can also be easily cleaned by filtering and replaced, and are more adequate for small nearing clearances. Another benefit of liquid oils is that the quantity of lubricant can be monitored more precisely. Petroleum-based oils are refined from crude oil. They are predominantly composed of hydrogen and carbon (thus the name hydrocarbon), even though lesser amounts of other elements can be found, such as sulfur, nitrogen, etc. Synthetic oils consist of chemical compounds that are also obtained from the

synthesis of petroleum or other raw materials, but they are made from highly refined products, and as a consequence the final product is more homogeneous. They are more adequate for applications under extreme conditions, that is, very high or very low temperatures and/or extended service life.

Additionally, current lubricants also include diverse types of additives, chemicals that are added in quantities of a few weight percent to improve the lubricating capacity and durability of oils [2.3]. This practice gained general acceptance about 1940 and has since then been highly developed to provide, nowadays, an enormous range of different additives. The specific purposes of lubricant additives are:

- Improve the wear and friction characteristics.
- Enhance the resistance to oxidation, the control of corrosion, and the control of oil contamination by reaction products, wear particles and other debris.
- Upgrade the characteristics of the lubricant by reducing the pour point (the temperature at which a liquid becomes semi-solid, and loses its flow characteristics) and inhibiting the generation of foam.
- Prevent the excessive decrease of the viscosity of the lubricant at elevated temperatures.

Carefully chosen additives are extremely effective in improving the performance of the lubricant oil. The manufactures of lubricant additives usually maintain secrecy on the composition and other details of their products.

1.2 Lubrication regimes

Petroleum-based and synthetic oils are liquids at room temperature. In normal operating conditions, the lubricant oil forms a film between the moving mechanical components preventing direct contact. This film must be of the right thickness as, if it is too thick, the mechanical components are subjected to extra stress, spoiling the work of the gearbox, and if it is too low, the mechanical components will be in direct contact, generating excessive heat, possibly damaging the components and producing, in general, a malfunction of the gearbox.

There exist four generally accepted regimes (understood as ways for providing lubrication) that help obtaining the adequate film thickness [2.3]:

1. *Hydrodynamic lubrication* or thick film lubrication: in which the oil liquid layer is kept between the moving surfaces by viscous drag, that is, by the effect of the relative movement itself.

2. *Hydrostatic lubrication*, in which the oiled surfaces are fully separated by a lubricating film of liquid or gas that is forced between the moving surfaces by an external pressure, generated by an external pump. If a continuous supply of pressurized lubricant is maintained, a complete lubrication film is present between the surfaces even at null relative speed.
3. *Elasto-hydrodynamic lubrication*, which results from the combination of three effects: hydrodynamics, elastic deformation of the metal surfaces, and the increase in the viscosity of the lubricant oil under extreme pressures. These effects combine as follows: as pressure and charge on the mechanical components increase, the viscosity of the lubricant oil increases. As lubricant moves to the contact zone, both contact surfaces are elastically deformed due to the pressure created by the lubricant oil. In the contact zone, the hydrodynamic pressure developed in the lubricant oil causes an additional increase in viscosity, which is enough to separate both contact surfaces. Due to the high viscosity and to the small time between cyclic contact in the lubricated area, lubricant oil cannot escape, assuring the separation between the contact surfaces.
4. *Boundary and extreme pressure lubrication*, in which the lubrication mechanisms involved are classified in terms of relative load capacity and limiting frictional temperature. In general, this lubrication regime involves complex phenomena and several specialized modes of lubrication such as adsorption, surface localized viscosity enhancement, amorphous layers and sacrificial films.

In close relation with this, the way in which the lubricants are used in gearboxes also conditions the effectiveness of the lubrication. There are two basic modes of lubrication mainly employed in gearboxes: The most basic one is *splash lubrication*, one kind of hydrodynamic lubrication, which is the normal method for lubricating spur, helical, bevel and worm gears. In this mode, the gears simply dip into a bath of oil as they rotate. For higher speed units, hydrostatic lubrication is needed and normally *spray lubrica*tion is used. Engineered spray lubrication is generally provided using shaped nozzles with oil at a certain circulating pressure to ensure that the oil reaches the contacting surfaces, as centrifugal forces and escaping air flow will tend to deflect the oil jet from its objective.

In any case, for the regimes that usually are present in the lubrication of the gearboxes, the control of the viscosity is of extreme importance to avoid damaging the lubricated asset.

1.3 Definition of viscosity

The property of the lubricant oil that is directly related with the thickness of the oil film is the viscosity. The viscosity of a medium is defined as the resistance to the movement, or the change in shape of neighboring portions of such medium, relative to one another [2.4]. In plain words, viscosity denotes opposition to flow, which is caused because the part of a fluid that is forced to move carries along, to some extent, the adjacent parts. Viscosity can also be interpreted in terms of the internal friction between the molecules of the medium, which fight against the development of velocity differences in the medium.

Quantitatively, the viscosity of a medium can be represented by two related quantities: the kinematic and the dynamic or absolute viscosity. The *dynamic viscosity* represents the tangential force per unit area required to move one horizontal plane of the medium with respect to another plane, while maintaining a unit distance apart [2.5]. This definition is exemplified in Figure 1. The *kinematic viscosity* is the ratio of the dynamic viscosity to the density of the medium.

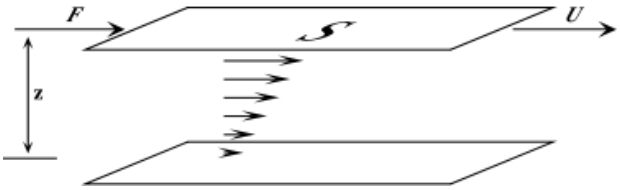


Figure 1. Definition of dynamic viscosity: it is the tangential force per unit area required to move one horizontal plane with respect to another plane, when maintaining a unit distance apart in the fluid [2.5].

The viscosity of the oil (both kinematic and dynamic) is determined mainly by the size of the molecules composing the lubricant: the larger the molecule structures are, the higher the viscosity is. In the case of the petroleum-based lubricants, the viscosity may vary depending on the method of refining, generating different viscosity rating lubricants (from 2000 cSt to 10 cSt measured at 40 °C). The viscosity and its dependence with the temperature can be largely modified by the incorporation of adequate additives.

1.4 Viscosity measurement

1.4.1 Kinematic and dynamic or absolute viscosity

Although several engineering units are used to express viscosity [2.6], the most common by far for lubricant oils are the centistoke (cSt) for kinematic viscosity and the centipoise (cP) for dynamic (absolute) viscosity. The Stoke (St) is the measuring unit of the centime-gram-second (CGS) system of units

for measuring the kinematic viscosity. Therefore, one cSt corresponds to 1 mm^2/s:

$$1\ St = 1\ cm^2/s = 0.0001\ m^2/s \ \Rightarrow\ 1\ cSt = 1\ St/100 = 1\ mm^2/s$$

The Poise (P) is also a unit based in the CGS system. One centipoise (cP) corresponds to one mPa s:

$$1\ Poise\ (P) = 1\ g/(s\ cm) = 1\ dina\cdot s/cm^2 = 0.1\ Pa\ s \ \Rightarrow 1\ cP = 1\ P/100 = 1\ mPa\ s$$

Numerically, the kinematic viscosity expressed in cSt is related to absolute viscosity expressed in cP by the fluid's specific gravity (SG), which is unit-less:

$$cSt = cP/SG.$$

1.4.2 Viscosity index

The viscosity of any fluid changes with temperature, increasing as temperature is lowered, and decreasing as temperature is increased. The viscosity can also change with the variation of the specific operating conditions (e.g. effort, cutting speed, etc.). Additionally, as stated above, the use of additives can significantly improve the viscosity temperature behavior. These additives are usually high molecular weight polymers that are dissolved in the oil and can change its shape from spheroidal to linear as the temperature is increased. This effect partly offsets the decline in base oil viscosity with temperature. The advances in the quality and properties of additives gave way to the development of a completely new family of lubricants known as *multigrade oils*.

The parameter *viscosity inde*x was introduced in 1929 with the intention of quantifying the variation of the viscosity of lubricant oils with temperature [2.7]. In those days, it was known that Pennsylvania crude oils were better than the Gulf Coast (Texan) crude oils [2.3]. Pennsylvania crude had the best viscosity temperature characteristics while the Gulf Coast crude had the worst, since its viscosity varied much more with temperature. Both oils have the same viscosity at 210 °F (98.9 °C) and they were initially selected as reference oils.

The viscosity index is an entirely empirical parameter, which compares the kinematic viscosity of a given oil with the viscosities of those reference oils. The Gulf Coast oil, whose viscosity varied a lot with temperature, was assigned a viscosity index equal to zero (VI = 0) at 100 °F (37.8 °C). The Pennsylvania oil was assigned a viscosity index equal to one hundred (VI = 100) at the same temperature. The viscosity index of the oil is given

relative to these two extreme values. An explanatory graphic can be seen in Figure 2.

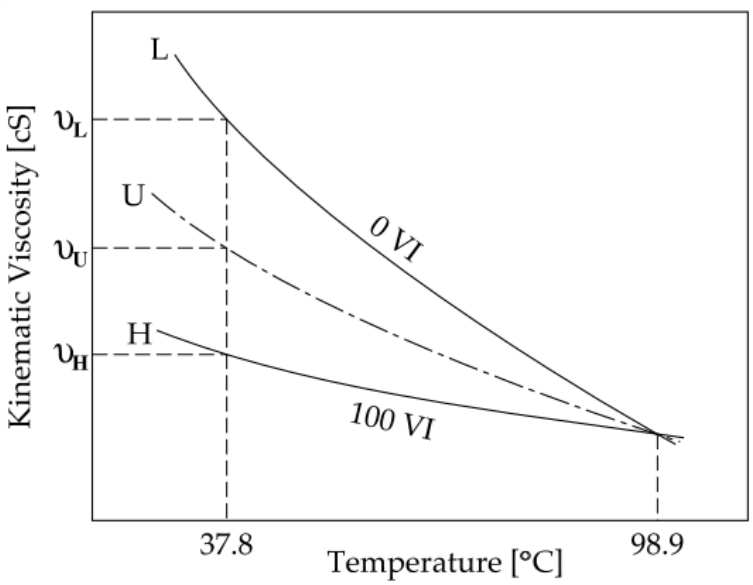


Figure 2. Evaluation of viscosity index [2.3]. U is the viscosity of the oil of interest, L the viscosity of the 0 VI index oil and H is the H is the viscosity of the 100 VI oils.

Today, the viscosity index of a petroleum-based oil is determined according to the ASTM D 2270 norm [2.8]. This norm generates a number used to characterize the variation of the cinematic viscosity of a petroleum product with temperature. The determination of the viscosity index is based on measurements of the cinematic viscosity to 40 °C and to 100 °C for oils of similar kinematic viscosity. The bigger the viscosity index is, the lower the effect of the temperature is.

1.4.3 Classification of lubricant oils

The most used indexes to classify the viscosity of lubricant oils are: the SAE (Society of Automotive Engineers); ISO (International Organization for Standardization); and military specification indexes [2.3]. These classifications are principally based on the viscosity of the oil, and are indirectly related to the quality of the oil, content of additives, performance or field of application. The most recent SAE viscosity grade classifications establish eleven engine-oil and seven transmission-oil grades [2.7].

Concerning engine oils, the so-called *monograde oils* only meet only one SAE viscosity grade. On the contrary, *multigrade oils* have low viscosities at low temperatures but higher viscosities at higher temperatures. This is achieved by adding viscosity improvers (polymeric additives) to the oil. The viscosity at low temperatures is noted with a *W* suffix, which stands for *winter*. For climates where the temperature regularly drops below zero Celsius, engine and transmission oils are formulated in such a manner that

they give low resistance at start, *i.e.* their viscosity is low at the starting temperature. For example, the nomenclature SAE 10W50 indicates the viscosities of the oil measured at -18 °C and at 100 °C. The oil behaves as a SAE 10 at low temperatures and as a SAE 50 at high temperatures.

The SAE J306 classification for transmission (gearbox) oils is very similar to that of engine oils. The only difference is that the winter grade is defined by the temperature at which the oil reaches the viscosity of 150.000 cP. This is the maximum oil viscosity that can be used without causing damage to the gears. The classification also permits multigrading.

It should also be noted that transmission oils have higher classification numbers than engine oils, simply to make it easier to differentiate between both types of oils. It does not mean that transmission oils are more viscous than the engine oils. The transmission oil viscosities for different SAE grades are shown in Table 1.

Table 1. SAE classification of transmission oils [2.3].

SAE viscosity grade	Max. temp. for viscosity of 150 000 cP [°C]	Kinematic viscosity [cS] at 100°C	
		min	max
70W	-55	4.1	-
75W	-40	4.1	-
80W	-26	7.0	-
85W	-12	11.0	-
90	-	13.5	< 24.0
140	-	24.0	< 41.0
250	-	41.0	-

The ISO (International Standards Organization) viscosity classification system [2.9]was developed in the USA by the American Society of Lubrication Engineers (ASLE) and in the United Kingdom by The British Standards Institution (BSI) for all industrial lubrication fluids. It is now commonly used throughout industry. Figure 3 displays an equivalence table of the different viscosity classifications, along with the corresponding values of the Kinematic viscosities and the Saybolt viscosities, called the *Saybolt Universal second* (*SUS*) and specified in the ASTM[1] D2161 procedure [2.10].

[1] ASTM stands for American Society for Testing Materials, now simply ASTM international.

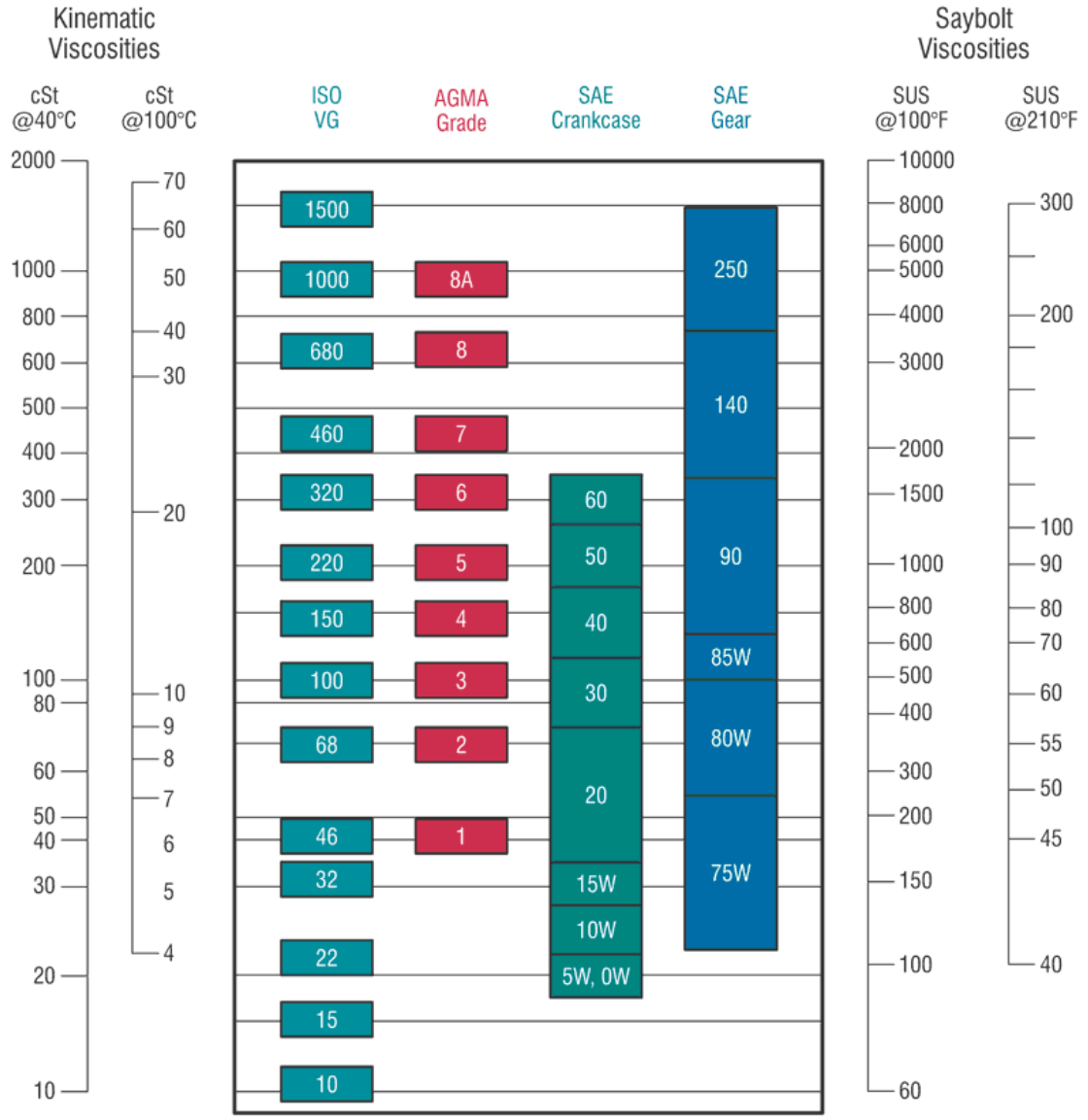


Figure 3. Equivalence table of the ISO VG, AGMA Grade, SAE crankcase and SAE Gear viscosity classifications with their corresponding values of the Kinematic and Saybolt viscosities [2.11].

In definitive, a lubricant oil with the first number as low as possible offers a better protection at low temperatures, whereas for a better protection at elevated temperatures a high number in the second position is required.

1.4.4 Measuring in laboratory

To determine experimentally the viscosity of oil, the kinematic or the absolute viscosity can be measured. Kinematic viscosity is traditionally quantified by measuring the time that it takes a sample of oil to travel through the orifice of a capillary under the force of gravity. The procedure is schematized in Figure 4. The orifice of the kinematic viscometer tube produces a fixed resistance to flow. The time taken for the fluid to flow through the capillary tube can be converted directly to a kinematic viscosity using a simple

calibration constant provided for each tube. For fluids with different viscosity, capillaries with different range of sizes exist in different norms. The dominant procedure for performing kinematic viscosity measurements is ASTM D445 [2.12]. The value of the kinematic viscosity, expressed in cSt, and measured at 40 °C, is the one established in the norm ISO 3448 (the international standard) for the kinematic viscosity grading system.

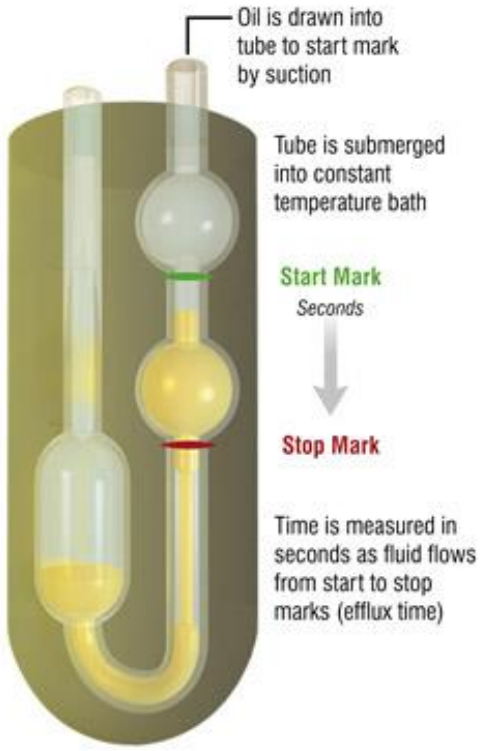


Figure 4. Gravity Flow U-shaped Glass Tube Capillary Viscometer ASTM D445-97 [2.13].

Absolute viscosity is measured in relation to the resistance to flow when a controlled external force (pump, pressurized air, etc.) forces oil through a capillary (ASTM D4624). Alternatively, it can be determined by forcing the movement of a body through the fluid, such as a spindle driven by a motor. In either case, the resistance to flow (or shear) as a function of the input force is measured, reflecting its dynamic viscosity.

There are several types and embodiments of absolute viscometers. The Brookfield rotary method depicted in Figure 5 is the most common. Absolute viscosity measurement has historically been used for research applications, quality control and grease analysis within the field of machinery lubrication. Procedures for testing absolute viscosity in the lab by the traditional Brookfield method are defined by ASTM D2983, D6080 and other norms.

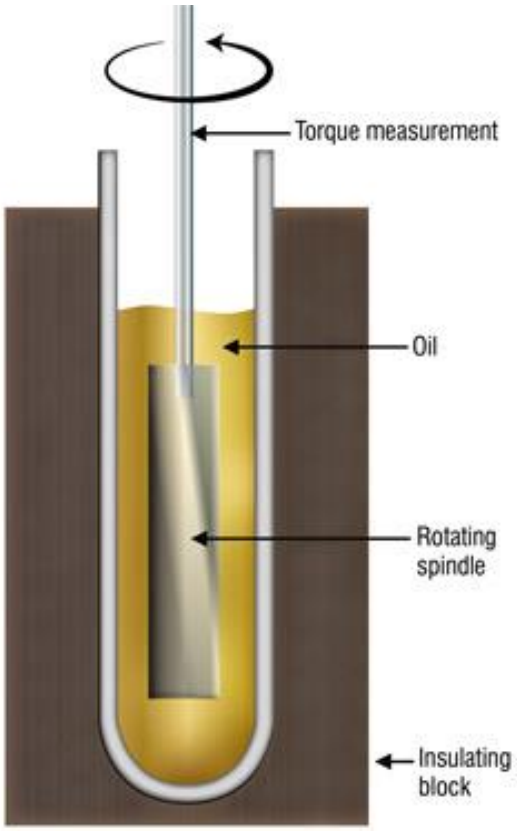


Figure 5. Rotary (Brookfield) Viscometer conforming to the procedure ASTM D2983 [2.13].

The lubricant oils utilized in this work were characterized using laboratory equipment. The kinematic viscosity was measured as described before using the procedure ASTM D445 [2.12], in a CANNON CAV 2100 fully automated viscometer, shown in Figure 6. It has a nominal accuracy of 0.1 cSt.

Figure 6. CANNON CAV 2100 automatic viscosimeter [2.14].

The density of oils was obtained using the norm ASTM D4052 [2.15]. A small volume (approximately 0.7 mL) of liquid sample is introduced into an oscillating sample tube and the change in oscillating frequency caused by the change in the mass of the tube is used in conjunction with calibration data to determine the density of the sample.

1.5 Variation of viscosity with degradation

Most mineral oils have molecules of an assortment of sizes. Oils with high viscosity have molecules with a predominantly large size while the contrary holds for low viscosity oils. The change in oil viscosity is caused by the modification of the average size of its molecules.

Although small changes in viscosity may be normal (for example seasonal changes in oils with added viscosity index improvers [2.16]), usually the change in oil's viscosity is the first indication of a more serious problem. Depending on the nature of the fault, the viscosity of the oil may increase or decrease. There are also malfunctioning situations in which the oil viscosity doesn't change.

1.5.1 Effects on the asset of inadequate viscosity of the lubricant oil

An oil viscosity different than that recommended by the manufacturer may cause a problem in the asset. Low viscosity may open the way for contact between the surfaces that are prevented of doing so with the right viscosity, potentially producing undesired effects as wear, high mechanical friction, loss of energy, heat generation, oxidation, leakage, increase of sensitivity to particle contamination, and worse performance at high temperatures, high charges and low speed. On the contrary, an increase of viscosity may prevent the oil from arriving to the interfaces to be lubricated, generating problems as excessive heating, oxidation, generation of varnish and sludge, cavitation, inadequate flux for bearings and gears, higher energy consumption, lower anti-foam and anti-emulsification efficacy and lower capacity to be pumped, specially at low temperatures.

1.5.2 Causes for inadequate viscosity

Possible causes for a low viscosity value are: the mixing of fuel with oil; electrostatic removal of insoluble oxides; shear of improvement additives; lubricant cracking; inadequate SAE/ISO grade of the lubricant oil; or contamination. On the other hand, the causes for high viscosity are: addition of a more viscous make-up oil; refrigeration fluid contamination; presence of water; contamination by other lubricant oils; inadequate SAE/ISO grade lubricant oil; light boiling hydrocarbon fractions; together with other chemical phenomena as oxidation or polymerization. As mentioned before, a degraded oil can show no viscosity change, due to a compensation of two

counterbalancing effects happening simultaneously. This is the case in a lubricant oil contaminated by fuel and by soot: the presence of fuel decreases the viscosity, but soot increases it.

In the case of an asset in good condition, and providing that the lubricant oil is not contaminated by water, the degradation process evolves as detailed in Figure 7. The degradation process is the oxidation of the molecules forming the lubricant oil in the case that it is not contaminated. It is measured with the Fourier transform infrared spectrometer (FTIR) oxidation number, depicted in blue in the image. There are three steps towards a *normal* degradation process of the oil (i.e. there is no external contamination, working conditions are normal, etc.). First, a depletion of the antioxidants is always occurring, but the viscosity of the oil is not affected, as the base lubricant oil is not being degraded. In a second step, after the antioxidants have been oxidized, other additives and the base oil start to react, causing an increase in viscosity. In the third stage, oil degradation produces varnishes and sludge, which can potentially cause damage to the asset and may have catastrophic effects. Therefore, the increase of viscosity is an excellent indication of the loss of the optimum properties of the lubricant oil.

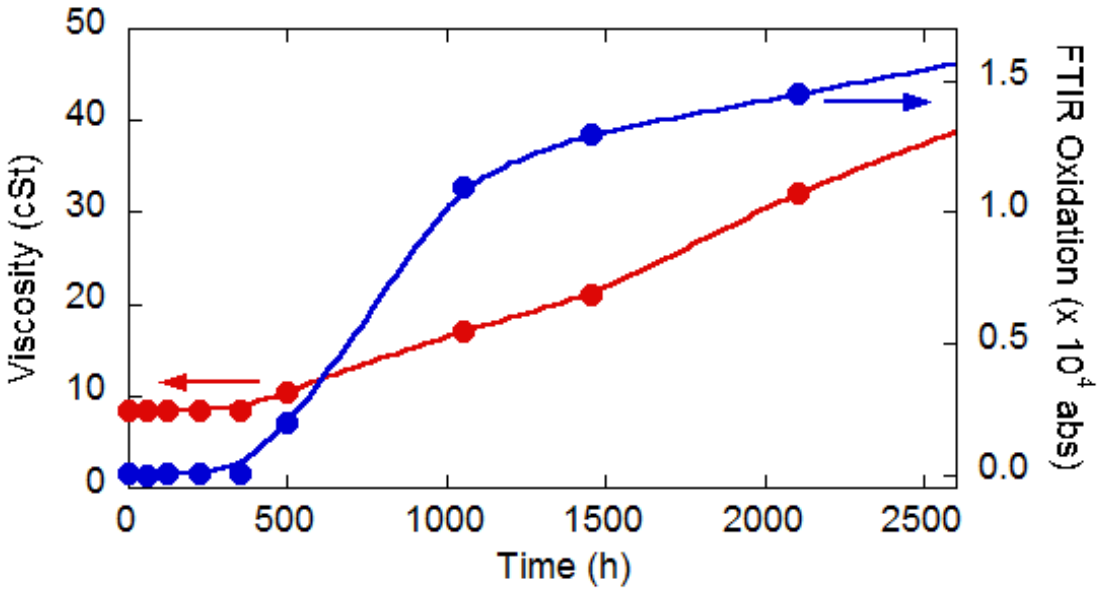


Figure 7. Viscosity variation with the oxidation of lubricant oil.

The usual strategy to monitor the health state of the oil is to take samples in a predefined time basis and send them to analyze in a laboratory. This is made manually, involving a considerable economical cost and other inconveniences since the asset may be in a difficult access area (as it happens, for example, with offshore wind-turbines). Besides, this practice may be also risky because a catastrophic failure can take place before the results are

received, if the oil is already degraded. Therefore, there is a clear need of developing sensors capable measuring the oil viscosity in a wide range of values, producing information in near real time, to implement an on-line condition monitoring strategy to reduce the possibility of a catastrophic failure. This is especially important in the case of gearboxes, which are critical elements whose malfunction or failure may have dramatic economical and security consequences.

1.6 Objectives

According to the ideas exposed previously, the main objective of this part of the Thesis is to develop a sensor to measure the viscosity of the lubricant oil on-line, in-situ and in real time, within a wide range of viscosities. For that purpose, a review of the available technologies must be performed, in order to select the basic operating principle of the sensor. Then the materials to be used and the appropriate design must be implemented in a working prototype, which could be used to validate its operation. The performance of the prototype must be assessed in conditions resembling those of the real environment in which it is to be used, in particular, being able to measure a wide range of viscosities. The work performed to fulfill those objectives and tasks is described in the next section.

2 Magnetoelastic on-line measurement sensors

After a brief description of the available technologies for on-line viscosity measurement, the fundamentals of the magnetoelasticity and magnetoelastic resonance are introduced, since this is the principle used in the prototype of viscosity sensor that has been developed in this work. The design and the selection of materials are described, together with the results obtained in different test conditions using oils in a wide range of viscosities. A new analytical method based in the fitting of the complete resonance curve is proposed to analyze the data. The results demonstrate the excellent correlation of different features of the resonance curve, such as the amplitude and the width, with the viscosity of the oils under tests. The prototype is therefore validated as a functional candidate for the development of on-line systems to assess the health state of the lubricant oil in gearboxes.

2.1 Types of viscosity sensors for on-line measurements

There are currently different approaches proposed for on-line viscosimeters [2.17]:

1. Methods relying in macro-displacement of a body in the fluid.
2. Methods relying in vibration.
3. Methods based in the propagation of mechanical waves (so-called acoustic viscometers).

In each approach, different methods can be used. They are summarized in Table 2 and explained in the following sections.

Table 2. Different approaches used in on-line monitoring of viscosity, and the main methods used within them [2.18].

On-line monitoring of viscosity				
Macrodisplacement	Vibration		Acoustic	
• Capillarity • Rotational • Falling piston	• Torsional oscillation • Oscillating console/cantilever	• Turning fork	• Acoustic Bulk	• Acoustic solid-state

2.1.1 Macro-displacement on-line viscosity monitoring methods

These methods are based on the laboratory measuring methods, but modified to be able to fit the requirements of an on-line monitoring device. These modifications make this family of sensors quite complex. Their size tends to be large in comparison with other sensors, and the existence of moving parts decreases the reliability of such sensors. Several devices using diverse working principles can be found under this category. The most common ones are: capillarity viscometer, rotational viscometer, and rolling-ball viscometer.

In capillarity-based sensors, the viscosity is calculated from the pressure drop resulting at a known flow rate in a straight section of tubing [2.19]. The inconvenience of these types of viscometers is that the flow velocity is usually variable depending on the working conditions of the asset. Besides, the tubing section of the sensor tends to get blocked, reducing its reliability.

Figure 8. In-line Capillary Viscometer PV100 from Brookfield [2.20].

In rotational viscometers, the viscosity is calculated form the torque necessary to rotate, at a known rate, a spindle of a known geometry immersed in the fluid.

Figure 9. Left, in-line rotational viscometer AST-INK [2.21]. Right, Rolling ball in-line viscometer TT-100 from Brookfield. [2.22].

Rolling ball viscometers, which are a modification of the falling ball laboratory viscometers, measure the travelling time of a piston (or a similar object) from one side to another of a tube. This time is directly correlated with the viscosity of the fluid were the piston is immersed [2.23].

2.1.2 Vibration viscometers

Vibration methods are based on the measurement of the amplitude or phase shift of the oscillations of a probe immersed in a fluid. The presence of the fluid damps the response of the probe. They operate at frequencies of about 102 to 106 Hz and function either by simple analysis of the displacement, or by analyzing the frequency difference in response to a given phase shift between excitation force and velocity near the resonant frequency. Generally speaking, vibration viscometers are robust, relatively small, simple to install (do not require additional flow loops or pumps), and provide continuous real-time data (updated every 1 to 10 seconds).

There are several types of viscometers that fall under this description. Viscometers based on torsional oscillation consist on a vibrator (tube, plate, etc.) situated at one end of a torsional tube with mechanisms mounted at the top to generate and measure vibration. These mechanisms usually rely in piezoelectric elements to generate and measure vibrations. Viscometers based on the use of a vibrating console cantilever use a cantilever beam with a permanent magnet fixed at its end. An electromagnet situated just under the beam creates short pulses to excite the movement of the beam. The response of the beam, affected by the viscosity of the fluid it is immersed in, is measured by the same electromagnet. Finally, tuning fork viscometers use a mechanical resonator made from piezoelectric elements operating at a frequency of about 75 kHz. Some of the elements are used to excite the fork and others to measure the answer. Immersed in a liquid, this answer depends on the viscosity of the surrounding medium.

2.1.3 Piezoelectric acoustic viscometers

The working principle of this family of sensors relies on the variation of the propagation mode of elastic compressional waves (acoustic waves in the liquid) depending on the viscosity of the medium. The waves are usually generated by piezoelectric excitation and, after propagating through the oil, are detected by another or the same piezoelectric transducer. The detected variations of the amplitude and velocity of the waves are related with the viscosity of the lubricant oil. Different configurations based on this general principle have been proposed for commercial use. Basically, there are two main types of acoustic on-line viscometers, which are denominated *bulk* and s*olid-state* in accordance with other types of piezoelectric acoustic sensors [2.24]. In the bulk ones, the waves propagate through the medium, whereas in the solid-state ones, the acoustic waves propagate on the surface of the substrate. They are a special type of SAW (surface acoustic wave) sensors. Both types are schematized in Figure 10.

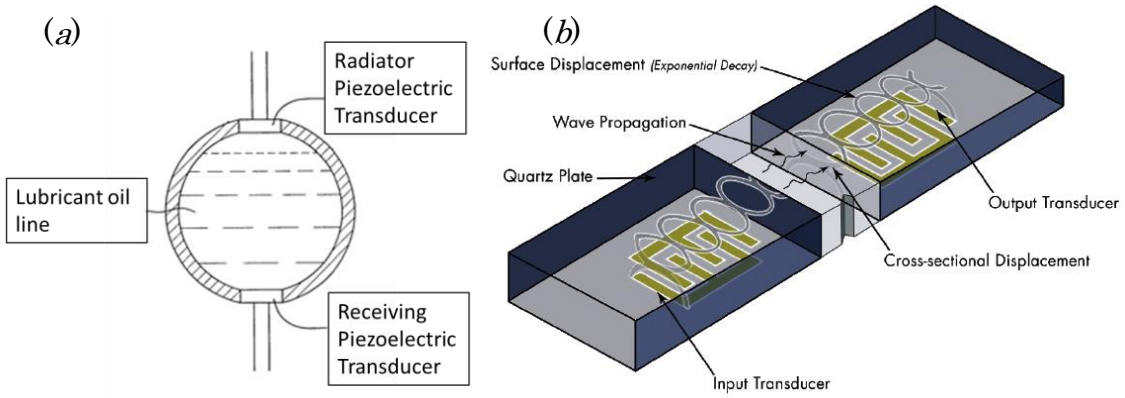


Figure 10. (*a*) Bulk acoustic viscometer [2.25], (*b*) Solid-state viscometer [2.26].

Viscometers based on acoustic waves present a number of problems and limitations. First, the readings may not be accurate as high-molecular fluids behave like gels, due to the operating frequencies used, that coincide or are lower than the oscillation frequency of the large molecules. Besides, in the solid-state sensors, the device is only sensitive to a thin layer of oil in the interface area. In general, this type of devices presents quite an elevated price due to the materials and technology used, the sophisticated electronics and, most of all, for their scarce industrial implantation.

2.1.4 Magnetoelastic viscometers

Magnetoelastic viscosity sensors are an alternative type of acoustic vibration sensors in which the oscillations are not created piezoelectrically, but by the coupling between the magnetic and elastic properties of a magnetic material [2.27]. The excitation produced by the magnetoelastic material is generated by an alternating magnetic field that, depending on the frequency and the

characteristics of the material, produces forced mechanical oscillations or mechanical resonances. The presence of the dissipative force caused by the lubricant oil produces an attenuation of the forced oscillations or a reduction of the magnitude and quality factor of the resonance, which can be correlated to the viscosity of the oil. Since this is the working principle of the prototype sensor developed in this Thesis, the fundamentals of the detection will be explained in much more detail in following sections.

The first reference of the use of the magnetoelastic effect for the in-laboratory and on-line determination of the viscosity dates back to 1958 [2.28]. The sensor consisted of a sheet of a magnetoelastic material, such as nickel, immersed partially in the fluid. At the other side of the sheet, the oscillations are produced by an excitation coil and their attenuation measured by detection one. A scheme of the working principle, as published in the original Patent document, can be seen in Figure 11.

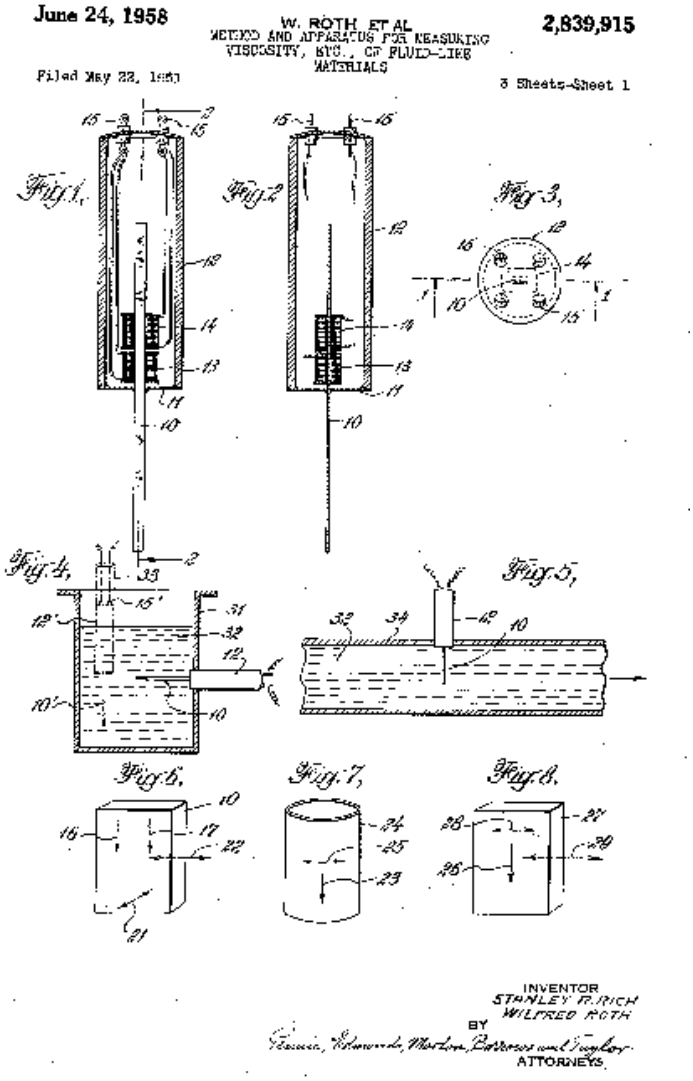


Figure 11. US Patent 2839915: Method and apparatus for measuring viscosity, etc., of fluid-like materials.

Two more patents that describe the use of the magnetoelastic effect to detect viscosity are dated in 1962 [2.29] and 1989 [2.30]. Both use a vibrating magnetoelastic element immersed in a fluid, and measure the damping of the vibration caused by the fluid being measured, to obtain an indication of viscosity. Neither one has been commercialized to the knowledge of the author.

An online viscosity sensor based in similar principles has been developed at research level by the Fraunhofer ICT-IMM: a magnetoelastic pin is immersed in the lubricant oil, where it is magnetoelastically excited. The decay of the signal or the change in the resonance is measured and related with the viscosity of the oil [2.31].

Figure 12. Image of the magnetoelastic on-line sensor of the Fraunhofer - ICT IMM.

A quite complete description of an on-line magnetoelastic sensor is provided by Markova and co-workers [2.18], in which a magnetoelastic amorphous ribbon is used as sensing element, immersed in lubricant oil. The ribbon is magnetically excited and queried to detect its resonant frequency. The resonant frequency of the ribbon in air is used as a reference while, when immersed in oil, the measured changes in the resonant frequency are correlated with the viscosity. An alternative detection schemes, based in the decay of the oscillation signal was additionally explored in this work. The system has been filed as a patent [2.32], although there is no commercial exploitation to the knowledge of the author.

A schematic representation of the device, taken from the patent application, can be seen in Figure 13. The probe head consists of the magnetoelastic material, a 37 mm x 6 mm x 0.03 mm strip of a commercial amorphous alloy (Metglass 2826MB) and a coil for excitation and measurement. A thermocouple is included to measure the temperature of the oil. The magnetic strip, electromagnetic coil and temperature sensor are installed in nonmagnetic housing with two holes that let the oil flow into the

chamber where the measurements are performed. Mechanically, the viscometer probe can be screwed either in an oil tank or in a pipe and connected to an electronic module.

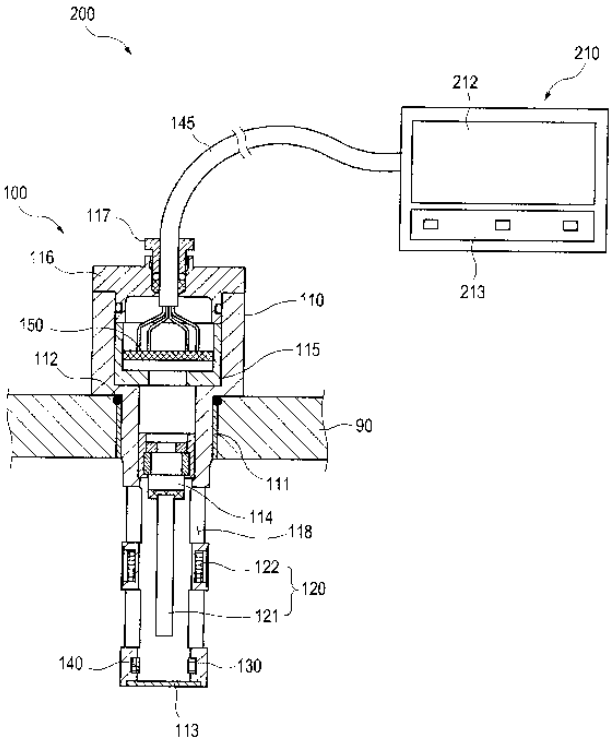


Figure 13. Scheme of the magnetoelastic probe, obtained from patent US 8521451 [2.32].

The published results, comparing the output of the sensor with laboratory measurements performed on synthetic and mineral lubricant oils with viscosities ranging from 17 cSt to 500 cSt showed a rather good correlation in the whole viscosity range [2.33].

2.2 Principles of magnetoelasticity and magnetoelastic resonance

The coupling between the magnetic and elastic properties of ferromagnetic materials gives rise to a number of physical effects that can be used for applications [2.34]. The effect that accounts for the change of dimensions produced in a magnetic material upon magnetization is denominated *magnetostriction*. This effect was first described by J.P. Joule in 1842. The reverse effect, that is, the change in the magnetic state of a material caused by a mechanical stress, is known as the *magnetoelastic effect*, *inverse magnetostrictive* or *Villari effect*, after the scientist that first described it in 1865. In this section, an introductory description of these effects, which are used in several sensor and actuator devices, is first presented. These effects are also the foundation of the existence of magnetoelastic waves and the magnetoelastic resonance phenomena that are described next. A brief review

of the more suited materials for magnetoelastic applications closes this introductory section.

2.2.1 Magnetostriction and magnetoelastic effect

Magnetostriction is defined as the phenomenon where a magnetic body shrinks or expands in the direction of magnetization as a function of the applied magnetic field. The fundamental physical origin relies in the spin-orbit coupling but, at a macroscopic level, it depends heavily in the magnetization mechanisms of the material. An exhaustive treatment of magnetostriction and related phenomena can be found in the book by Trémolet de Laichesserie [2.35]. Basically, the magnetostriction causes the deformation inside the material of each magnetic domain according to its own magnetization, whose direction is essentially determined by the magnetic anisotropy. When an external magnetic field is applied, the local distortions caused by magnetostriction sum up to produce a macroscopic deformation. As the external magnetic field is increased, the macroscopic deformation increases until the material reaches the magnetic saturation and the deformation takes its maximum value. In its simplest formulation, the magnetostriction is quantified as the strain suffered by the magnetic specimen, that is, the relative increment of length $\lambda = \Delta l/l$. The maximum deformation $\lambda_s = (\Delta l/l)_{\text{sat}}$, is denominated *saturation magnetostriction* or *magnetostriction coefficient* and is a characteristic of the material. Figure 14 illustrates the typical magnetostrictive behavior that occurs in an elongated sample. The magnetostrictive strain can be positive or negative, depending on the nature of the material. For instance, the magnetostriction is positive for iron but negative for nickel and cobalt. Its magnitude is generally small, of the order of some parts per million, except for certain special alloys as Terfenol, as described below.

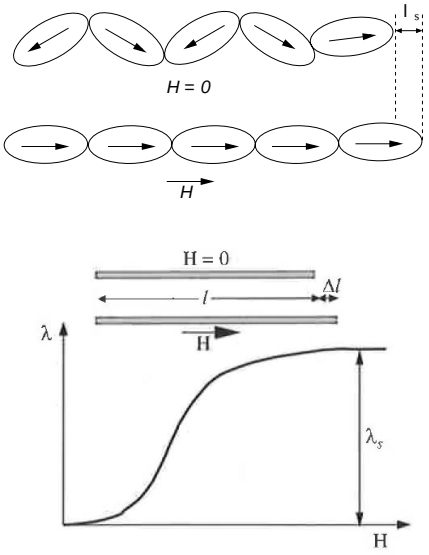


Figure 14. Magnetostrictive coefficient λ versus magnetic field H [2.34].

In crystalline samples, the magnetostriction is usually anisotropic, meaning that the deformation upon magnetization is different in each crystallographic direction. In the case of isotropic materials, the deformation in a certain direction can be expressed as [2.36]

$$\lambda = \frac{3}{2}\lambda_S \left(cos^2\theta - \frac{1}{3}\right) \tag{2.1}$$

where θ is the angle between the direction of magnetization and the direction in which the deformation is measured. This expression is also valid for polycrystalline and amorphous materials where the deformation is averaged over different regions.

Magnetostriction is used in different types of actuators [2.37], some of them in quite a similar way that piezoelectric ones. In fact, the magnetostrictive properties of nickel were used to develop the first SONAR (sound navigation and ranging) devices during World War I.

The inverse effect, magnetoelasticity, refers to the magnetization changes caused by the mechanical deformation of a ferromagnetic material. In particular, the magnetoelastic effect makes that the magnetic permeability of the material depend on the applied stress. Its fundamental advantage over other principles of sensing is its intrinsic non-contact nature, since the permeability of the sensing material is detected inductively.

The magnetoelastic effect can be easily quantified in a simple, but interesting case: an isotropic material with an uniaxial magnetic anisotropy K defining an easy magnetization axis in a direction perpendicular to the applied magnetic field H and stress σ [2.34]-[2.36]. The equilibrium position of the magnetization vector M_s is the one that minimizes the total energy of the system $E_T = E_m + E_a + E_{me}$, being E_m the magnetoelastic energy ($E_m = -\mu_0 HM_s \cos\varphi$), E_a the anisotropy energy ($E_a = K\cos^2\varphi$) and E_m the magnetoelastic energy ($E_{me} = -3/2\,\lambda_S\sigma\cos^2\varphi$), where φ is the angle between M_s and H (and σ). The equilibrium angle of the magnetization is then given by:

$$\cos\varphi = \frac{\mu_0 M_S H}{2K - 3\lambda_S\sigma}, \tag{2.2}$$

and the magnetization of the sample (measured in the direction of the applied field) is then given by $M = M_s \cos\varphi$ as:

$$M = \frac{\mu_0 M_S^2}{2K - 3\lambda_S\sigma} H = \chi^\sigma H, \tag{2.3}$$

where χ^σ is the magnetic susceptibility, related to the permeability by $\mu = 1 + \chi^\sigma$. The sensitivity of the permeability $\mu = M/H$ to the applied stress in this case is therefore,

$$\frac{d\mu}{d\sigma} = \frac{d\chi^{\sigma}}{d\sigma} = \frac{\mu_0 M_S^2}{(2K - 3\lambda_s\sigma)^2} 3\lambda_s, \tag{2.4}$$

which, for small values of the applied stress, reduces to

$$\left(\frac{d\mu}{d\sigma}\right)_{\sigma=0} = \frac{3\chi_0^2}{\mu_0 M_S^2}\lambda_s, \tag{2.5}$$

with $\chi_0 = \mu_0 M_s^2/2K$.

The expressions above demonstrate that the magnetoelastic effect in materials with high magnetostriction (λ_s) and low anisotropy (K) can be readily used to sense stress or any other related magnitude. For instance, the inductance changes caused by the magnetoelastic response to an applied strain can be directly implemented to construct a force sensor, as the one schematized in Figure 15, which has been commercialized by ABB under the name *Pressductor* for more than sixty years. Other examples of the application of the magnetoelastic effect into successful commercial devices are described elsewhere [2.34].

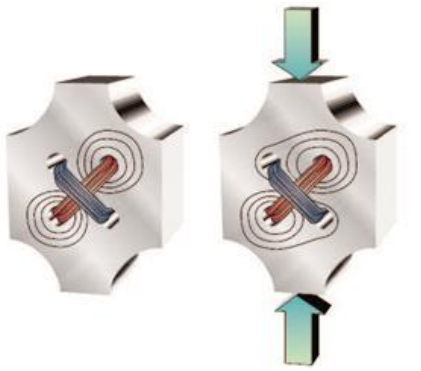


Figure 15. Pressductor operating principle. Two coils are wound at right angles through holes perforated in the sensible magnetostrictive element. One of the coils is used to excite the sensing element, and as it is stressed it becomes magnetically anisotropic, therefore the flux distribution becomes distorted and a measurable electrical output is generated in the secondary coil [2.38].

The Pressductor belongs to the category of *static magnetoelastic sensors*, since the magnitude to be measured produces directly a magnetic output through the magnetoelastic coupling.[2] The *dynamic magnetoelastic sensors*, in contrast, are based on the propagation of magnetoelastic waves through the magnetostrictive material. The sensors based on the magnetoelastic resonance are included in this last category.

[2] Despite their name, the working frequency of the magnetoelastic static sensors can reach up to some kHz.

2.2.2 Magnetoelastic waves and magnetoelastic resonance

The magnetoelastic coupling makes that a stress perturbation (mechanical or sound wave) travelling through a magnetoelastic material, is accompanied by a corresponding magnetic perturbation, building up a *magnetoelastic wave*. Due to its dual nature, magnetoelastic waves can be generated either mechanically or magnetically and detected likewise. This has a direct application in sensing since the magnetic excitation and detection allows contactless position and distance measurement, configuring the so-called *magnetostrictive delay line*. As an example, *Temposonics* position sensors developed by the German company MTS Sensor Technologie GmbH utilize the measurement of the time of travel of a magnetoelastic wave in a magnetostrictive material to determine the distance between the place where the wave is generated and the point where it is detected. Surface acoustic wave (SAW) sensors coated with a magnetostrictive thin film (also called MSAW for Magneto-SAW) have also been proposed to measure any magnitude (such as force or temperature) that modifies the amplitude or velocity of the travelling wave [2.39].

When the magnetoelastic waves are excited in a bounded material, for instance in a thin ribbon of a given length, stationary waves can be produced if the wavelength of the mechanical oscillation matches the length of the sample. This situation produces sharp resonances at which large strains and magnetization changes takes place. The effect is called *magnetoelastic resonance* (MER) and is the basis of well-established technologies (as acoustic anti-shoplifting labels) and the core of the detection technology proposed for the detection of the oil viscosity.

Mathematical description

A magnetoelastic wave can be described by combining the wave equation of motion

$$\rho \frac{\partial^2 u_z}{\partial t^2} = \frac{\partial \sigma}{\partial z} \tag{2.6}$$

(where ρ is the density of the material, u_z the displacement in the z direction, and t the time), with the coupled equations of state of the magnetoelastic effect. They mathematically express that the strain (ε) and the magnetization (m) in the material are produced by both the applied stress (σ) and magnetic field (h)

$$\epsilon = \frac{1}{E^H}\sigma + dh \tag{2.7}$$

$$m = \frac{d}{\mu_0}\sigma + \chi^{\sigma} h \tag{2.8}$$

where d is the magnetoelastic coefficient that characterized the mechanical and magnetic coupling, E^H the Young modulus at constant field, μ_0 the vacuum permeability, and χ^{σ} the magnetic susceptibility at constant stress. Note that this last equation implicitly contains several approximations and simplifications, and is only valid in the linear regime, implying that the amplitude of h and σ are small.

Let's consider the case of a magnetoelastic wave excited by a harmonic magnetic field of frequency $\omega = 2\pi f$, in a ribbon of length L oscillating with free ends. The boundary conditions impose that he stress must vanish at both ends of the ribbon:

$$\sigma(z = \pm L/2) = 0. \tag{2.9}$$

If the wavelength of the exciting magnetic field matches the length of the sample, standing waves build up, giving rise to resonances. This is mathematically evident in the solution for m/h obtained from equations (2.6, 2.7 and 2.8). These equations can be integrated along the length of the sample, to obtain the mean value of the magnetic susceptibility χ^{σ} [2.40]

$$\chi^{\sigma} = \frac{\chi^{\epsilon}}{1-k^2}\left|1 - \frac{8k^2}{\pi^2}\sum_n \frac{1}{n^2}\frac{1}{1-(\frac{\omega_n}{\omega})^2 + iQ^{-1}\frac{\omega_n}{\omega}}\right| \tag{2.10}$$

where χ^{ε} is the susceptibility at constant strain, that is to say, the pure magnetic susceptibility, without the magnetoelastic contribution. The ω_n represent the frequencies of the different modes of oscillations or resonance frequencies and are defined below. The parameter k is the magnetoelastic coupling coefficient, defined as the ratio of the magnetoelastic energy to the mechanical and magnetic energies of the system

$$k = \frac{U_{me}}{\sqrt{U_e U_m}} = d\sqrt{\frac{E^H}{\mu_0 \chi^{\sigma}}} \tag{2.11}$$

that expresses the capacity of the system to convert from one type of energy to the other.

The Q factor that enters in Equation 2.10 can be viewed as the quality factor of the resonance. It accounts for the energy losses, and is directly related with the width of the resonance peak. The losses can be mechanical as well as magnetic, mainly due to eddy currents in soft magnetic materials. If the sample is immersed in a viscous liquid the increased mechanical

damping makes that the amplitude of the resonance decreases while its width increases.

The solution given by equation 2.10 reveals the existence of consecutive magnetoelastic vibration modes, that occurs at the resonance frequencies

$$\omega_n = \frac{n\pi}{L}\sqrt{\frac{E^H}{\rho}} \tag{2.12}$$

producing consecutive modes of mechanical oscillations, which are schematized in Figure 16.

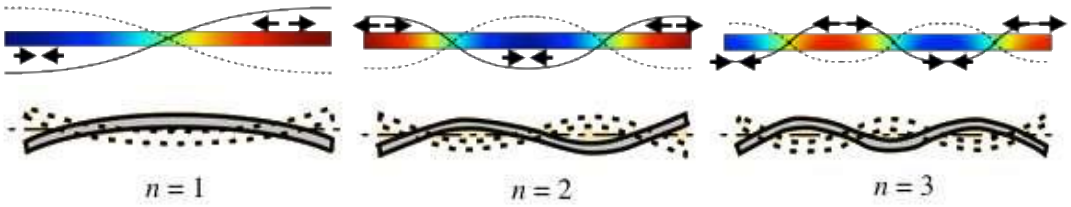


Figure 16. First oscillation modes of the magnetoelastic resonance of a free ribbon. On the top row, the longitudinal vibration modes are depicted. Blue zones experience compression and red ones expansion. The solid and dotted lines represent the extreme values that the longitudinal deformation can take. The modes are more easily seen in transverse vibration as depicted below, although the magnetoelastic waves are longitudinal.

The magnetic susceptibility expressed in equation 2.10 can be measured experimentally using, basically, a pick-up coil (the measuring system will be described in detail below, in section 2.2.5). Figure 17 shows a typical magnetoelastic signal measured in an amorphous ribbon 5 cm long, displaying four clear resonances (modes n = 1 to 4).

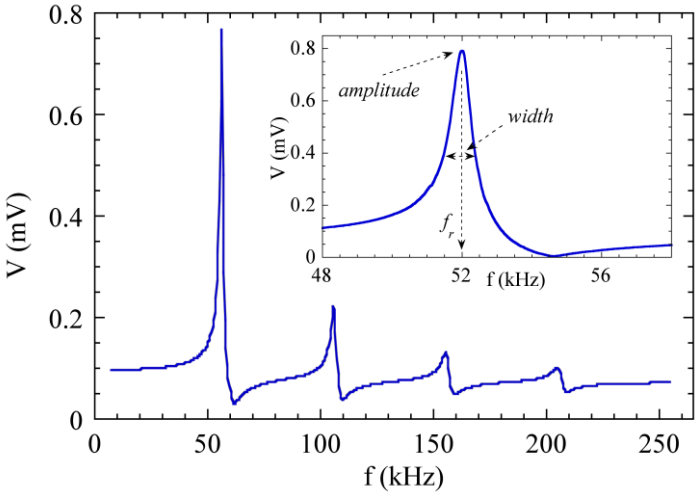


Figure 17. Magnetoelastic resonance modes of a free amorphous ribbon showing the fundamental resonance frequency at 52 kHz (n = 1) and the subsequent overtone resonances (n = 2 at 104 kHz, n = 3 at 156 kHz, and n = 4 at 208 kHz). The inset reveals the main features of the first mode.

The inset of Figure 17 shows an enlarged view of the first resonance mode, which is usually the one with larger amplitude and, consequently, the one used for detecting purposes. The resonance peak is characterized by the value of its resonance frequency (that is given by equation 2.12), its maximum amplitude, and its width, which is related to the quality factor Q. It is also noteworthy that there is a frequency at which the amplitude drops to zero, due to the out-of-phase coupling between the strain and the magnetization. The frequency at which this phenomenon takes place is usually called the anti-resonance frequency.

Any change in the external conditions such as temperature, magnetic field, stress, humidity, pressure, etc., determines a variation in the amplitude or frequency of a magnetoelastic sensor in which a resonant stationary magnetoelastic wave is established [2.41]. Therefore, the measurement of the amplitude and/or the frequency of the resonance can be used to sensor these physical effects.

2.2.3 ΔE effect

In the magnetoelastic resonance, it is very important to emphasize that the resonant frequency is determined by the value Young's modulus E^H (equation 2.12), that is, the Young modulus at given value of the applied magnetic field. Therefore, the resonant frequency can be varied at wish, within a certain range, by magnetically biasing the sample by a controllable magnetic field. The phenomenon behind this tuning capability is the so-called ΔE effect. It can be described as follows.

The Young's modulus E expresses the relation between the applied stress (σ) applied to a material and the concomitant elastic deformation (ϵ_{el}) that it generates. In magnetoelastic materials, however, the coupling between the strain and the magnetization makes that the applied stress generates an extra magnetoelastic deformation (ϵ_{me}). Besides, this extra contribution depends on the magnetic state of the material that can be greatly modified by an applied magnetic field H. Therefore, the Young Modulus in a magnetoelastic material depends on both H and σ: $E(H,\sigma)$. When the material is magnetically saturated, no extra magnetoelastic deformation is generated and, therefore, the relation between the stress and the deformation is given by $E_S = \sigma/\epsilon_{el}$, were the subscript S stands for saturation. In any other magnetic state, Young's modulus has a different (smaller) value. The ΔE effect expresses the difference between these values of the Young modulus and is quantified as

$$\frac{\Delta E}{E_S} = \frac{E_S - E(H,\sigma)}{E_S} \tag{2.13}$$

The ΔE effect is very small in normal ferromagnetic materials (about 6 % in nickel, and about 1 % in iron), but it can very large in special materials as metallic glasses, most of all after a proper thermal treatment is applied.

The functional dependence of the Young's modulus can be expressed mathematically in the simple case analyzed in section 2.2.1 for a sample with transverse anisotropy. Using the expression in equation 2.1 for isotropic materials, the total deformation of a magnetoelastic material can be expressed as:

$$\epsilon = \epsilon_{el} + \epsilon_{me} = \frac{\sigma}{E_S} + \frac{3}{2}\lambda_S\left(cos^2\varphi - \frac{1}{3}\right) \tag{2.14}$$

The equilibrium angle of the magnetization under the simultaneous action of a magnetic field H and a stress σ is given in equation 2.2, which is repeated here for clarity:

$$\cos\varphi = \frac{\mu_0 M_S H}{2K - 3\lambda_S\sigma}. \tag{2.15}$$

Thus, we can write

$$\frac{1}{E(H,\sigma)} = \frac{d\epsilon}{d\sigma} = \frac{1}{E_S} + \frac{3}{2}\lambda_S\frac{d(cos^2\varphi)}{d\sigma} \tag{2.16}$$

and, therefore, $E(H,\sigma)$ becomes:

$$E(H,\sigma) = \frac{1}{1 + \frac{9\lambda_S^2\mu_0^2M_S^2H^2}{(2K-3\lambda_S\sigma)^3}E_S} \tag{2.17}$$

Note that the magnetoelastic resonance frequency is directly coupled with Young's modulus, as evidenced in Equation 2.12. Therefore, the field dependence of E^H makes that the position of the resonance peak (that is, the value of the resonance frequency) also depends on the applied magnetic field. Figure 18, displays this dependence measured experimentally in an amorphous ribbon.

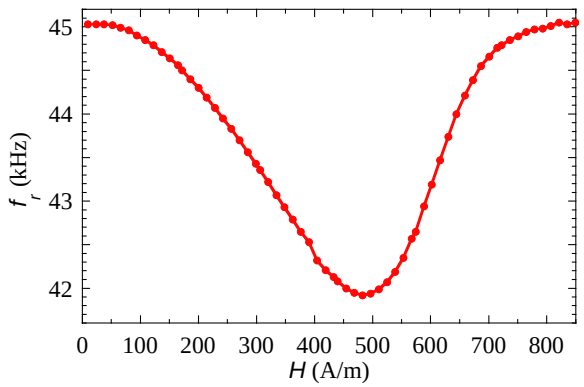


Figure 18. Dependence of the resonance frequency of the first mode of a free amorphous ribbon on the applied magnetic field, illustrating the great influence of the ΔE effect in this type of samples.

Since the magnetic state of the sample affects the magnetoelastic coupling, not only the resonance frequency depends on the applied field, but also the amplitude of the resonance and the magnetoelastic coupling coefficient k. The largest values of these two magnitudes correspond to the field at which Young's modulus is minimum (which is also the field at which the f_r-H curve reaches the minimum value). In the case of a sample with transverse anisotropy, that point corresponds to an applied field close to the anisotropy field $H^k = 2K/\mu_0 M_s$, at which the sample becomes magnetically saturated.

2.2.4 Magnetoelastic materials

Although all ferromagnetic materials exhibit a certain degree of magnetostriction, for sensing applications a combination of the best magnetic and elastic properties are required. In this sense, amorphous magnetic alloys, also called *metallic glasses*, are the best-suited materials for developing sensors based on the magnetoelastic resonance. Their characteristics and properties will be described in this section but, for completeness, a brief review of other magnetostrictive materials is presented before.

Pure transition metals and alloys

The first materials to be investigated were the three known magnetic materials at the time where the magnetostrictive effect was discovered, *i.e.* Fe, Co and Ni. Table 3 complies their magnetostriction constant values. As single crystals are anisotropic, their magnetostriction coefficients are different in different crystallographic directions.

Table 3. Magnetostriction constant at room temperature for magnetic transition metals.

	λ_{100} ($\times 10^{-6}$)	λ_{111} ($\times 10^{-6}$)	Polycristal, λ_S ($\times 10^{-6}$)
bcc-Fe	21	-21	-7
hcp-Co	-140	50	-62
fcc-Ni	-46	-24	-34

Piezomagnetic Ferrites

Piezomagnetic ferrites are a class of magnetic ferrites with high resistivity, high magnetostriction, and very low magnetocrystalline anisotropy and, in general, good magnetoelastic properties. Cobalt ferrites are among the most suited materials of them all. In Table 4 some values of the magnetostriction constants are presented. However, their mechanical properties are, in general, rather poor.

Table 4. Magnetostriction constant at room temperature for piezomagnetic ferrites.

	$\lambda_S (\times 10^{-6})$
$Ni_{50}Fe_{50}$	28
$Ni_{95}Co_5$	-35
$Co_{49}Fe_{49}V_2$	-65
$Fe_{80}B_{20}$	32
$Fe_{40}Ni_{40}B_{20}$	14
$Co_{80}B_{20}$	-4

Al-Co-Fe alloys

Alcofer alloys are alloys containing 2 to 14 wt. % aluminum, 2 to 3 wt. % cobalt and 83 to 86 wt. % iron. They show high saturation values ($\mu_0 M_S$ up to 1.4 T), high Curie temperature (about 500° C) and high magnetostriction (λ_S above 40×10^{-6}). The magnetoelastic coupling coefficient ranges between 0.3 to 0.45.

Giant magnetostriction materials

The highest magnetostriction values correspond to compounds containing Rare Earth elements such as Tb, Dy and Sm at cryogenic temperatures and high applied magnetic fields. Alloyed with Fe, large strains (about 1500 ppm) at room temperature are achieved with fields of some kOe. The most popular of such alloys is Terfenol-D, with nominal formula $Tb_{1-x}Dy_xFe_2$. In particular, Terfenol with $0.7 \leq x \leq 0.73$ presents an anisotropy compensation about room temperature, which allows to minimize the required magnetic field to achieve large deformations. Unfortunately, Terfenol displays rather poor mechanical properties, preventing its application in magnetoelastic sensors.

Amorphous alloys

Amorphous alloys combine excellent magnetoelastic and mechanical properties. They offer such a good combination of properties that many of the sensors relying in magnetoelastic materials use this category of materials. The lack of long-range order in the atomic arrangement is the reason for the good mechanical properties (high ductility and Young's modulus) and magnetic properties (high permeability, low coercivity and null magnetocrystalline anisotropy). They are also called metallic glasses, since they are obtained by rapid quenching from the melt, achieving cooling rates of 10^6 degrees per second, that freezes the disordered state of the liquid phase when solidifying. However, to obtain the amorphous state, it is necessary to include non-magnetic atoms in the structure. Commercial amorphous alloys are typically composed of about 70 to 80 % transition metals (Fe or Co) and

20 to 30 % metalloids (Si, B or P), with occasional additions of other elements. Table 5 compiles some commercially available metallic glasses and their properties.

Table 5. Magnetic and magnetoelastic properties of some commercial glasses.

Name		Composition	$\mu_0 M_s$ (T)	T_c (K)	P (g/cm³)	E (GPa)	λ_S (× 10^{-6})
Vitrovac[a]	7505	$Fe_{81}B_{13}Si_4C_2$	1.5	693	7.1	150	30
	4040	$Fe_{39}Ni_{39}Mo_4Si_6B_{12}$	0.8	533	7.4	150	8
	6025	$Co_{66}Fe_4(MoSiB)_{30}$	0.55	523	7.7	150	0.5
Metglass[b]	2826	$Fe_{40}Ni_{40}P_{14}B_6$	0.87	520	7.5	147	11
	2605	$Fe_{80}B_{20}$	1.6	647	7.4	172	31
	2826 B	$Fe_{29}Ni_{49}P_{14}Si_2B_6$	0.49	408	7.5	135	4

[a]Fabricated by Vaccumschmeze GMBH, Hanau, Germany
[b]Fabricated by Metglas Inc.

Small changes in composition affect the value of the saturation magnetostriction, which can even change its sign depending of the concentration of different components. Although the magnetoelastic properties can be excellent in the as-prepared state, thermal treatments under the influence of applied magnetic fields or stress allow enhancing and tailoring them.

Nanocrystalline materials are obtained by adding small quantities of Cu and Nb to Fe-rich metallic glasses and processed by thermal treatments. The growth of Fe or Fe-Si precipitates is hindered by the presence of Cu and Nb atoms, and thus nanometric crystals of about 20 nm in size are embedded in the amorphous matrix. The magnetoelastic coupling is high, but they are extremely brittle and so they have not been used in magnetoelastic sensing to the knowledge of the author.

The materials used in this work are the amorphous ribbons, as they provide a good balance between magnetoelastic properties and resiliency, making them ideal for the working conditions of a lubricant oil viscosity sensor.

2.2.5 Measuring system

The basic set-up for measuring the magnetoelastic resonance must have the capability to produce the magnetic excitation in the sample and to detect its magnetic response. This functionality is usually implemented using two sets

of coils: the primary coil which is feed with an alternating current and a secondary or pick-up coil in which the voltage related to magnetization changes is induced. This basic scheme is depicted in Figure 19. It can be advantageous to connect a compensation coil in series (but wounded in opposite direction) with the pick-up coil in order to suppress the voltage induced directly by the excitation, as well as any other background signal. In this way, the signal to noise ratio is considerably improved.

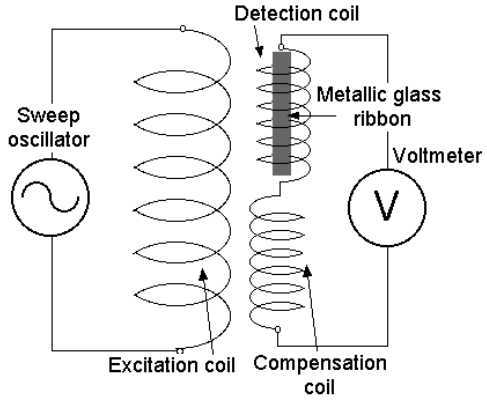


Figure 19. Scheme of the exciting and measuring set-up.

The system of coils is designed in such a way that the sample is placed inside them, making as large as possible the effect of the excitation and the level of the induced signal, therefore maximizing the performance of the measuring set-up. However, other arrangements are also possible, depending on the application. The emitter and the receiver can also be planar coils acting as antennas. This is the se-up implemented in anti-shoplifting systems, which are the most extended applications of the magnetoelastic resonance (Figure 20). In this configuration, the coils ate situated in the posts at the interrogation zone, activating and detecting the response of the magnetoelastic label when it crosses the space between the posts.

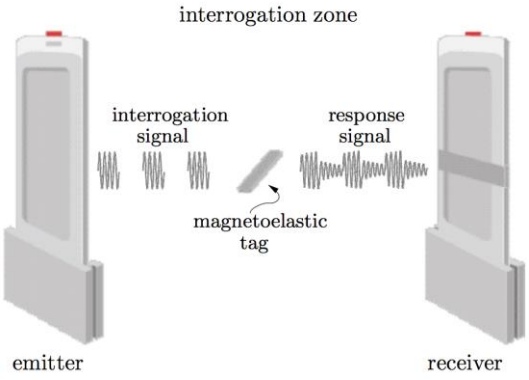


Figure 20. Anti-shoplifting system based in magnetoelastic materials.

There are commercial systems, using the magnetoelastic resonance as the sensing effect as in the integrated compact *MERMAID* system developed by ACREO (Figure 21) [2.42].

Figure 21. MERMAID system developed by ACREO.

Alternatively, the magnetoelastic resonance can be detected using a single coil arrangement, which provides simultaneously the excitation of the sample and the detection signal. The resonance is measured in this configuration through the variation of the self-inductance L of the coil, which is directly dependent to the susceptibility of the sample $L=L(\mu_0\chi)$ (χ is given in equation 2.10)[3]. The changes in the self-inductance can be easily measured as a function of the frequency using an impedance analyzer, obtaining the whole resonance curve in a single sweep. Alternatively, the curve can be recorded by feeding the coil with a signal generator and measuring the current in the circuit (through the voltage drop in a series resistance, for instance). The experimental set-up in this single-coil configuration is simpler and can be used if the amplitude of the resonance is large. If the resonance is highly damped, a two-coil set up is usually needed to improve the signal to noise ratio.

In any measuring system using the magnetoelastic resonance, the set up must take into account that the ΔE effect modifies the resonance conditions, as explained in section 2.2.3. In this sense, not only the resonance frequency depends in the applied bias field, but also the amplitude of the resonance. For that reason, to maximize the detection capability, a bias magnetic field must be applied to magnetize the sample in the optimum operating point on the magnetization curve. Additionally, this operation point can also be selected as the one in which the temperature dependence of the magnetoelastic resonance is cancelled, as it will be explained in section 2.4.3. The suitable bias field can be provided either by additional coils or by a permanent magnet situated close to the sample (as in the case of the anti-shoplifting application).

[3] Note that in the conventional system using two coils described in Figure 19, the electrical quantity involved is the mutual inductance *M*.

The magnitude of the optimum bias field depends on the sample, but for the amorphous ribbons used in this work it is in the range of 1-10 mT.

Finally, concerning the way in which the magnetoelastic resonance is used in sensor devices, there are two different approaches to determine the dynamic magnetic response of the resonators. In the first one, the response of the material is continuously measured while the excitation field is on. By sweeping the frequency of the excitation and measuring the magnetic response at the same frequency, the complete resonance curve is determined, from which the frequency, amplitude and width of the resonance can be used in the analysis of the signal. In the second approach, the response of the resonators can be measured during the ring-down period after a field pulse has been applied to the sample. During that period, the magnetoelastic sample keeps oscillating at its resonance frequency. From the time variation of the ring-down oscillations, the resonance frequency, the amplitude and the damping can be obtained.

2.3 Prototype of magnetoelastic viscosity sensor

We have designed, constructed and test two prototypes to evaluate the function and performance of the magnetoelastic approach in developing a sensor for the on-line monitoring of lubricant oil. The first one was conceived to be as simple as possible to function as a proof-of-concept prototype. In the second, a more elaborate design was implemented to fulfill some essential requirements, such as

- Enhanced sensitivity. The design must permit a certain degree of flexibility in the configuration of the coil systems to optimize the magnetoelastic response.
- Guarantee the repetitiveness of the measurement. The magnetoelastic resonators must operate in the same conditions even in test with different oils.
- Easiness to change the oil probes. The design must allow for a easy access to the vessel containing the oil under test.

Additionally, we sought a final prototype design that could be scalable to an actual sensor to be implemented in an on-line test facility. This implies to try to maintain a compact and economical design. Being a prototype, however, it contained parts that don`t satisfy this requirement. For instance, bulky Helmholtz coils were used to easily select the optimum bias field acting on the magnetoelastic sample. This is not a constrain for real systems since the bias field can be applied by a small magnet once the optimum value of the bias field has been determined for a given configuration.

In the following section, the different elements of both prototypes are described, starting from the magnetoelastic materials selected for them. The data acquisition and analysis strategies are also described.

2.3.1 Materials

Two different commercial amorphous ribbons were used in the experiments:

- Vitrovac 4040, with a nominal composition $Fe_{39}Ni_{39}Mo_4Si_6B_{12}$ (material datasheet available at producer [2.43]).
- Vitrovac 7600, with a nominal $Fe_{64.5}Co_{18}Si_1B_{12}C_{0.5}$ (material datasheet available at producer [2.44]).

In our prototype, the oil under test is contained inside a 4 ml glass vial with a plastic cap (the oils used in the experiments are described in the results section below). The magnetoelastic ribbons are made to oscillate submerged in the oil, where their response is modulated according to the viscosity of the medium. Therefore, the commercial magnetoelastic materials, supplied in 6 mm wide and 23 μm thick ribbons, were cut to 30 mm long pieces, as shown in Figure 22.

Figure 22. Magnetoelastic ribbons cut in 30 mm long and 6 mm wide pieces, together with the vial used to contain the oil under test.

The magnetoelastic ribbons were precisely cut using a laser facility [2.45] that guaranties a precision in the micrometer range. Besides, the laser cut process prevents the crystallization of the amorphous magnetoelastic ribbon. Figure 23 shows scanning electron microscope images of the edges of the sample after the cutting process.

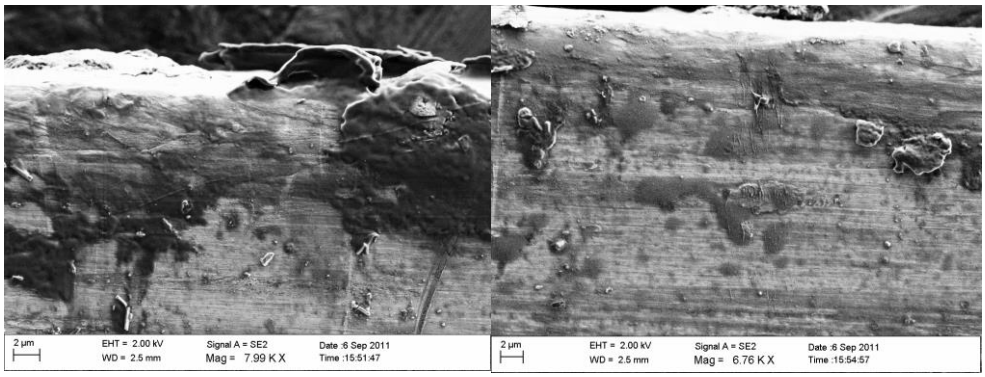


Figure 23. SEM images of the border of the Vitrovac 4040 magnetoelastic ribbon.

In the first set of experiments, using the proof-of-concept prototype, the magnetoelastic ribbons were simply inserted in the vial as shown in Figure 24.

Figure 24. Oil samples in the prove were they were measured in the experiments.

In the second approach, the magnetoelastic ribbons were clamped to the cap of the vial, to assure the repeatability of the positioning of the ribbon in relation to the system of coils. The ribbons were glued to a glass rod. In one end of the rod, a portion was cut using a diamond saw to obtain a planar region in which one end of the sample (about two millimeters) was glued after a careful alignment with the rod. The ribbon is free to oscillate in the non-clamped end. The cap of the vial was perforated to accommodate the rod with the sample. Figure 25 shows the final arrangement of the sample holder.

Figure 25. Magnetoelastic ribbon clamped on one end to a glass tube, enclosed in a plastic tap of the lubricant oil recipient.

When the cap is secured in the vial, the sensing material is immersed in the oil as illustrated in Figure 26.

Figure 26. The sensing element, attached to the cap of the vial, is completely immersed in the oil under test.

According to equation 2.12, a clamped ribbon resonates at frequencies that are half of the ones corresponding to a free ribbon of the same length. This is graphically justified in Figure 27 for the first two modes: the oscillation of a clamped ribbon in each mode is the same of the one of a free ribbon of double length.

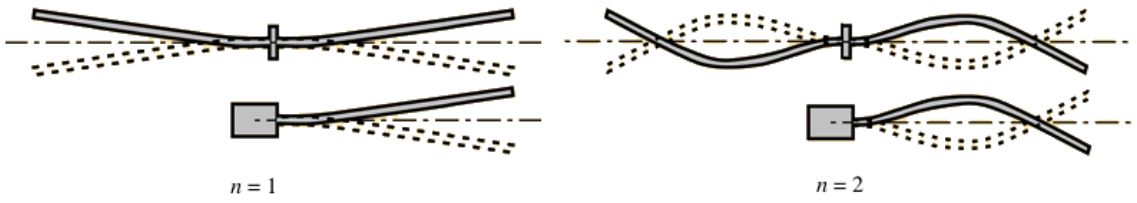


Figure 27. Sketch comparing the oscillation modes in a clamped ribbon and in a free one of double length for the first two modes (n = 1 and n = 2). According to equation 2.1, the resonance frequencies of a clamped ribbon are half of the free one of the same length. For clarity, the figure depicts transverse oscillations although the magnetoelastic oscillations are longitudinal.

2.3.2 Prototype design and experimental set-up

As previously explained, the prototype for the on-line measurement of the oil viscosity was designed in two steps. First, a simple arrangement was developed, conceived as a proof-of-concept system. Then a more elaborated laboratory prototype was built, incorporating the information gathered during the previous stage.

Proof-of-concept design

The objective of this preliminary design was to verify the possibility of making an online sensor taking advantage of the properties of magnetoelastic materials. It uses a single coil for excitation and measuring, together with a pair of Helmholtz coils to set the bias field. The excitation and measuring coil is made of 170 turns of AWG28 copper wire (0.32 mm in diameter) wound in two layers over a plastic cylindrical support 6 cm long and 20 mm in diameter. The electrical characterization of the coil using a LCR meter yielded a self-induction of 655 μH at 1 KHz, and a resistance of 5.7 Ω measured at 120 Hz. The oil-filled vial with the sample freely lying in the bottom was inserted horizontally inside the coil and positioned in its central part. For testing different oils, the magnetoelastic strips are changed from one vial containing one oil sample to another. In between, the strip was thoroughly cleaned using a jet of ether.

The Helmholtz coils provide a uniform magnetic field in the central part of the excitation coil, where the sample is positioned. The Helmholtz coils shown in Figure 28 produce a magnetic field of 19.05 Oe/A (1520 Am^{-1}/A).

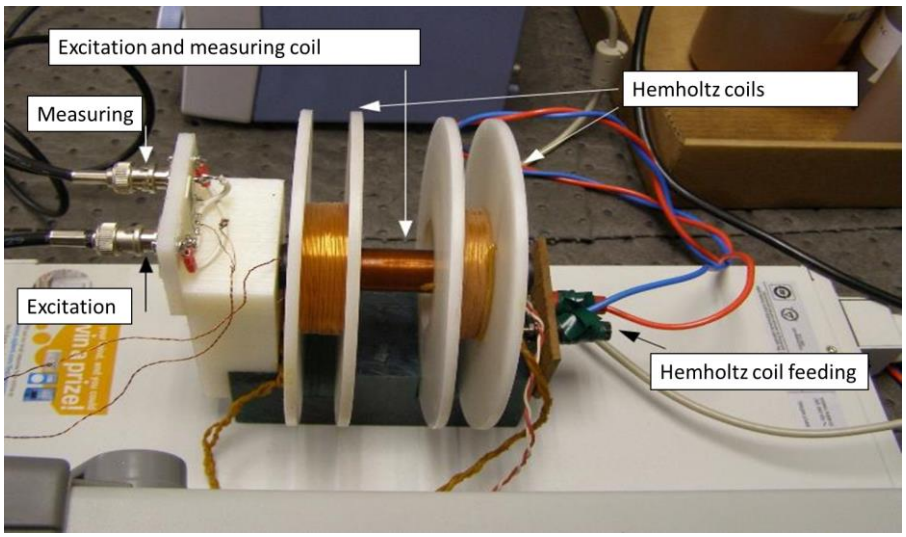


Figure 28. Proof-of-concept design with indication of the different components.

An Agilent N5769A power supply was used to feed the Helmholtz coils. The excitation was provided by a Keithley 3390 function generator. The

detection was performed using a Yokogawa DLM2022 digital oscilloscope. Figure 29 presents a general overview of the complete experimental set-up.

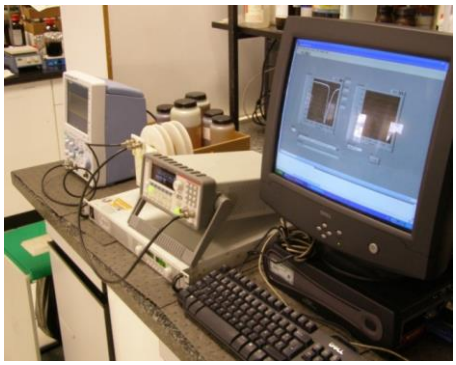


Figure 29. Laboratory set up of the proof-of-concept prototype, displaying the instrumentation and the computer running the control program.

A dedicated LabView program was developed to perform the experiments and to analyze the measured signals. It is explained in section 2.3.3.

Laboratory prototype

This second prototype was designed after the possibility of measuring the viscosity of lubricant oil using magnetoelastic materials was successfully established using the proof-of concept design. This second prototype was built to be close to a functional on-line measuring system, demonstrating that the sensor could work in a wide viscosity range and with different lubricant oils. In addition, this prototype was also used to determine the variation of viscosity with temperature.

The set-up was configured as a three-coil system, as schematized in Figure 19. The excitation was provided by a primary coil, whereas the detection was performed by two secondary coils connected in series but in opposition, one acting as pick-up and the other as compensation (see section 2.2.5). Figure 30 shows a close view of the system of detection coils. The vial containing the oil is positioned vertically inside the detection coil, with the sample clamped to the cap of the vial as explained before. In this way, the changing of the vial was straightforward. Following the procedure developed in the proof-of-concept set-up, when changing from one oil to other, the magnetoelastic ribbons were cleaned using ether, to eliminate any trace of the previous lubricant oil.

The system was designed to allow for a great degree of flexibility to displace the detection and compensation coils in order to optimize the signal. The measurements were performed by a Spectrum-Network Analyzer (HP

8714ET), which provided the excitation performing a frequency sweep and measured the amplitude of the induced voltage in the detection coils.

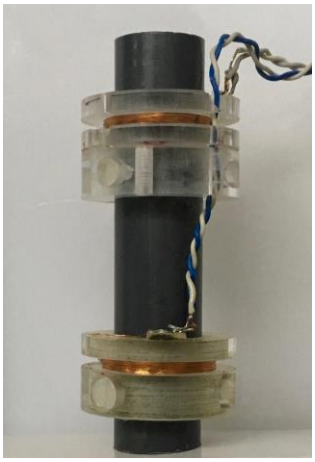


Figure 30. Pick-up (top) and compensation coil (bottom) in the sensor prototype.

In the prototype, the bias field was provided by a pair of large Helmholtz coils, fed by a power supply (Agilent N5769A). They produced a field of 19.41 Oe/A (1548 Am^{-1}/A).

Figure 31 shows a general view of the prototype, together with an exploded scheme indicating its main parts.

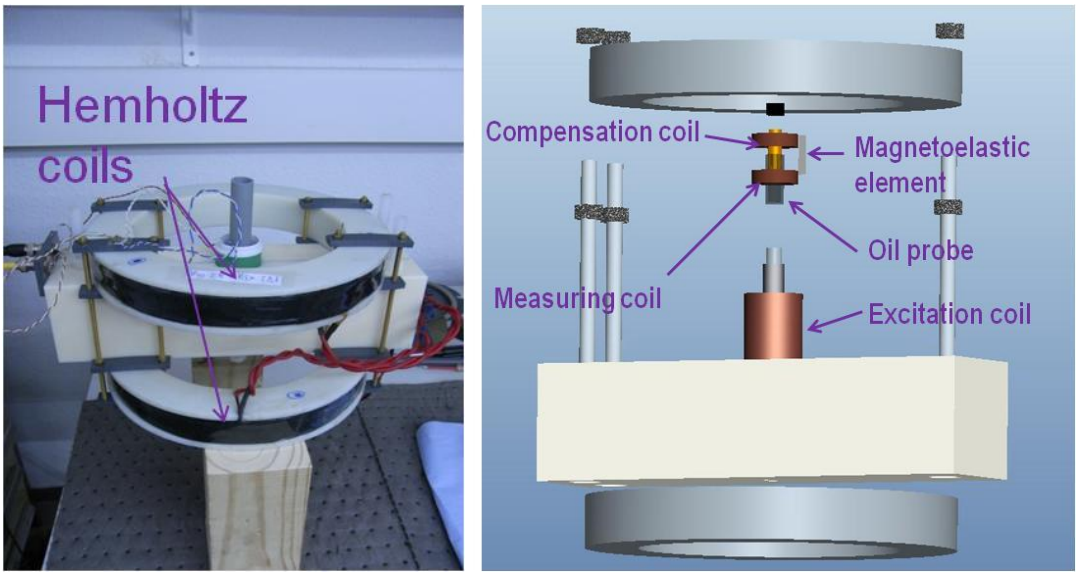


Figure 31. General view of experimental set-up (left) and explicative drawing (right).

The viscosity sensor was also used to determine the variation of the viscosity with temperature. For that purpose, the complete measuring system (without the instrumentation) was introduced into a climate chamber (Angelantoni Industry Challenge 250).

A quite elaborated code was developed in LabView (National Instruments) to control the experiment and acquire the results. In particular, it was necessary to create a special communication protocol with the proprietary software controlling the climate chamber, as explained in next section.

2.3.3 Data acquisition

An important part of the development of the prototypes was dedicated to generate and improve the computer programs for the control and data acquisition. These programs were developed in LabView from National Instruments. The following sections explain the most important features of the two versions implemented, one for each prototype.

Proof of concept prototype

In this set-up, the signal generator provides the excitation. The measurement program commands it to perform a frequency sweep between 10 and 100 kHz. A DC power supply feeding the Helmhotlz coils provides the bias field. The DC current is varied from 0 to 0.5 A, in steps of 0.01 A (0 to 9.52 Oe in steps of 0.19 Oe; 760 Am^{-1} in steps of 15 Am^{-1}). In each step, a full frequency sweep is performed. The voltage drop in the measuring coil is measured in the oscilloscope, which registers and displays its amplitude for each value of the frequency.

First, the instruments are initialized and the conditions of the experiments are established. Then the program sequentially sets the corresponding current in the Helmholtz coils. After a fraction of a second, allowed to stabilize the current, the program starts the frequency sweep and, for each frequency, records the data registered by the oscilloscope. To improve the signal to noise ratio, the program averages the values taken in five consecutive measurements performed from the same value of the bias field. The recorded signal in then processed by the program to determine the amplitude of the response and the values of the resonance and anti-resonance frequencies. These results are stored in an output file together with the corresponding value of the bias field.

For the series of experiments performed with this prototype, the first step was to determine the optimum bias point for which the measured amplitude of the resonance was the greatest. This was performed for the sample in air. In the case of the Vitrovac 4040, the greatest amplitude was obtained for a current in the Helmholtz coils of 22 A, corresponding to an applied magnetic field of 4.2 Oe (288 Am^{-1}). This bias magnetic field was kept constant in the subsequent measurements with the sample immersed in different lubricant oils.

Laboratory prototype

In the laboratory prototype the data is acquired using a Spectrum-Network Analyzer (HP 8714ET). The quality of the measured data is considerably improved over the previous proof-of-concept version. The Spectrum-Network analyzer is equipped with a built-in signal generator that is used to produce the excitation of the sample, performing a frequency sweep. It is also capable of measuring the amplitude of the induced voltage in the detection coils for each frequency and displays, in a single sweep, the complete resonance curve.

For the measurements, a similar strategy is followed. According to the length of the samples, the frequency was swept from 20 to 40 kHz, range in which the first harmonic of the resonance is located. Note that, as the samples are clamped by one of their ends, the resonance frequencies are half than in the previous prototype. The bias field is swept in steps of 0.01 A (0.194 Oe; 15.5 Am^{-1}) and in each step the complete resonance curve is registered. For each sample, the optimum bias field is found, at which the amplitude of the resonance with the sample is air is the greatest. This bias field is kept constant when performing the measurements with different oils.

The control program, also written in LabView, first initializes the instruments and sets the conditions of the experiment. Then, it sequentially changes the bias field and, for each value of the field, perform a frequency sweep. The Spectrum-Network Analyzer can be set to use averaging or different integration times to perform the measurements. These parameters are selected to obtain a good signal to noise ratio, making unnecessary to make several measurements programmatically. The complete resonance curve is saved in a data file, since the data analysis, explained in a section below, is performed after the acquisition, using a Matlab routine.

This prototype was also used to determine the variation of the viscosity with temperature (section 2.4.3). A climatic chamber (Angelantoni Industry Challenge 250) was used for this purpose. The temperature was varied between 0 and 60° C in steps of 10° C. To control the climatic chamber, its own proprietary software (Winkratos) was used. It was programed to set the desired temperature and wait until the magnetoelastic measurements at that temperature were finished. It was necessary to synchronize the data acquisition program in LabView with the Winkratos control program. The only way to do it that we could find was to set different alarm status in the climate chamber when the target temperature was reached and stabilized. These alarms appeared as digital signals in a communications interface of the chamber, which the LabView program could read through a data acquisition board. Therefore, the main LabView program waited until the target

temperature was set, and then launched the magnetoelastic measuring program. There was no possibility of communicating back to the chamber when the acquisition at each temperature finished. Therefore, ample time was provided between temperature changes. In any case, the temperature of the chamber was continuously being recorded by the LabView program using the acquisition board, so it was possible to determine *a posteriori* if the acquisition was completely done at the prescribed temperature and finished before the chamber begun to change to the next set-point. Using this automated procedure it was possible to perform the experiments necessary to get the data for a plot such as that of Figure 51, that could last about 24 hours for a given oil.

2.3.4 Data processing

The data processing is designed to extract useful parameters from the acquired resonance curves. One or several of those parameters can be used to establish a calibration of the sensor as discussed in next section.

Proof-of-Concept Prototype

In the case of the proof-of concept prototype, the data processing is embedded in the data acquisition program. It extracts the amplitude and the frequency of the resonance from each curve measured. Before explaining the procedure to do so, we must make a remark about the shape of the resonance curve measured using this set-up.

Figure 32 shows a typical resonance curve obtained using a Vitrovac 4040 sample in air (without oil). Due to the special configuration of the detection circuit used in this prototype, the resonance curve doesn't display the usual shape, which is the one represented in Figure 17. The measured curve presents a deep minimum at the resonance. It can be easily demonstrated that this response curve is the one to be expected with the detection circuit used.

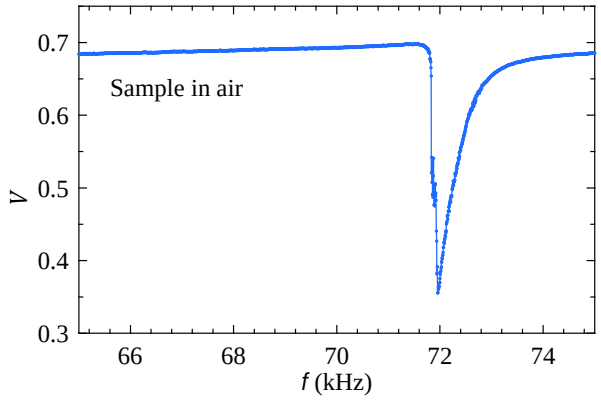


Figure 32. Typical resonance curve measured using the proof-of-concept prototype. It corresponds to a sample of Vitrovac 4040 in air.

The electrical model of the circuit used in this prototype is schematized in Figure 33. V_g and R_g (50 Ω) correspond to the excitation source that feds the measuring coil. The response V_L is taken as the voltage drop in the coil. L_μ and R_L represent its self-inductance and resistance respectively. R is the resistance of the rest of the circuit. Let's consider that the intrinsic response of the magnetoelastic material is represented by its relative permeability μ_r as depicted in Figure 34 (*a*). It displays the typical resonant behavior of a magnetoelastic sample illustrated in Figure 17. We can impose that $L_\mu = L_0\mu_r$, being L_0 is the self-inductance of the coil when $\mu_r = 1$, that is, the self-inductance of the coil without sample. It was measured to be $L_0 = 774.6$ µH. The resistance of the coil was measured to be $R_L = 5.4$ Ω. The circuit of Figure 33 can be modeled in Matlab and the frequency dependence of the measured voltage V_L calculated. The result, displayed in Figure 34 (*b*) shows a clear resemblance with the actual curves measured (Figure 32).

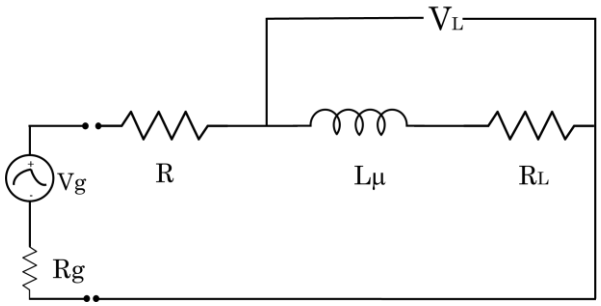


Figure 33. Model of the measuring circuit in the proof-of-concept set-up.

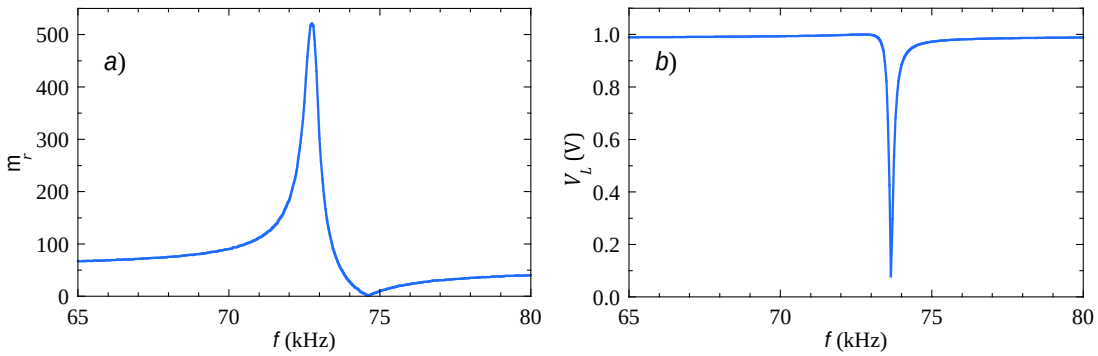


Figure 34. (*a*) relative permeability of a magnetoelastic simple, displaying the typical resonance behavior. (*b*) Output voltage calculated in the circuit of Figure 33 using the permeability in (*a*).

In the proof-of-concept prototype the data processing consists exclusively in subtracting the background from the measured data. The background should correspond to the signal measured by the system without sample. This signal is displayed in Figure 35.*a*. However, for simplicity, the background is calculated as a straight line joining by the first and the last point. In fact, as the data is quite noisy, ten points are selected in each extreme of the data array and their value averaged to determine the first and last points (Figure 35.*b*). After the subtraction of the background, the amplitude and the frequency of the resonance are determined and saved in the output data file.

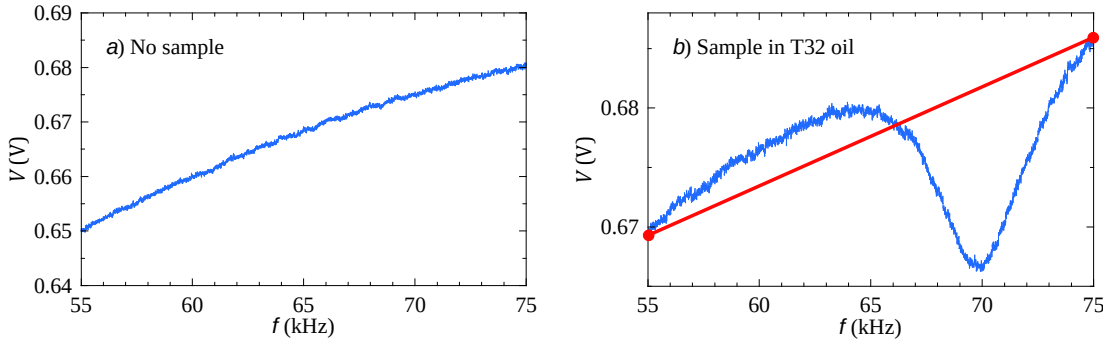


Figure 35. Background removal during the data analysis in the proof-of-concept prototype. *a*) Signal measured by to the system without sample. *b*) The background is calculated using straight line between the first and the last measured points.

Laboratory prototype

In this set-up, the registered signal is the voltage induced in the pick-up coil, which is directly related to the permeability of the sample. Therefore, the measured resonance curve has the usual shape as displayed in Figure 36 for both types of magnetoelastic ribbons measured in air.

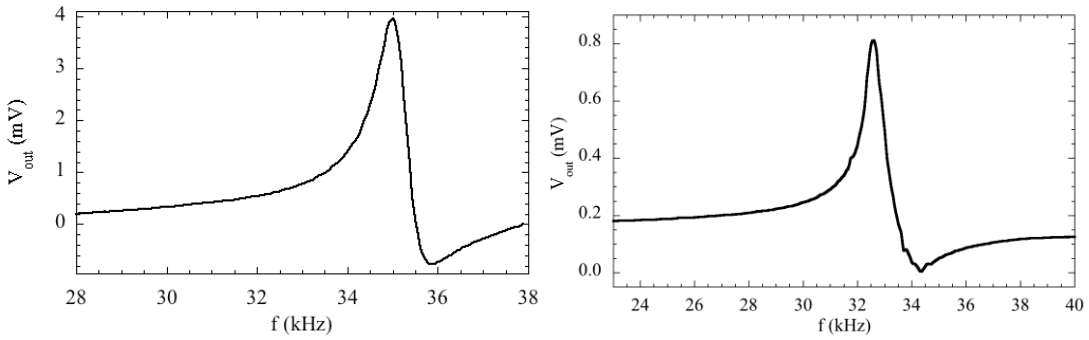


Figure 36. Resonance curves measured with the Laboratory prototype corresponding to the types of magnetoelastic ribbons used in this work: *left*: Vitrovac 4040; *right*: Vitrovac 7600.

Due to the use of a compensation coil, the background signal is nearly zero and the values of the amplitude and the resonant frequencies can be easily obtained from the resonance curves. However, much more information can be obtained by analyzing the complete resonance curve. For that purpose, we have developed a phenomenological approach to describe the frequency response of a magnetoelastic material [2.46].

Using the formalism of linear systems, we can derive an analytical expression for the transference function of a system displaying a resonance at a frequency ω_r and an anti-resonance at a frequency ω_a. The resonance imposes a couple of complex conjugated poles in the denominator, while the anti-resonance is described with a couple of complex conjugated zeros in the numerator. The transference function is then expressed as

$$G(s) = \frac{w_r^2}{w_a^2} \times \frac{s^2 + 2d_a w_a s + w_a^2}{s^2 + 2d_r w_r s + w_r^2} \tag{2.18}$$

δ_a and δ_r being damping parameters. In this formalism $s = j\omega$, where $\omega = 2\pi f$ and $j = \sqrt{-1}$ is the imaginary unit.

Therefore, the experimentally measured resonance curve can be fitted to the following expression

$$V(\omega) = A \left| \frac{\omega^2 - 2j\delta_a\omega_a\omega - \omega_a^2}{\omega^2 - 2j\delta_r\omega_r\omega - \omega_r^2} \right| + a\omega + b \tag{2.19}$$

where A accounts for the amplitude of the curve, and a and b provide a linear contribution to deal with a possible remaining background. In this expression, the parameters to be fitted are the frequencies of resonance $\omega_r = 2\pi f_r$ and anti-resonance $\omega_a = 2\pi f_a$, the damping parameters δ_r and δ_a, the amplitude A, and the coefficients a and b.

This expression reproduces very accurately the experimentally measured resonance curves and allows extracting the resonance parameters by means of a least squares fitting. This is demonstrated in Figure 37 for the case of 37 mm long and 6 mm wide magnetoelastic ribbon resonating freely (both ends unclamped) in air. The best-fit parameters, which produces an excellent agreement with the measured curve, are $f_r = 56.73$ kHz; $\delta_r = 0.0022$; $f_a = 60.82$ kHz; $\delta_a = 0.0121$; $A = 38.07$ mV; $a = 8.57 \times 10^{-6}$ mV/Hz; $b = -55.97$ mV.

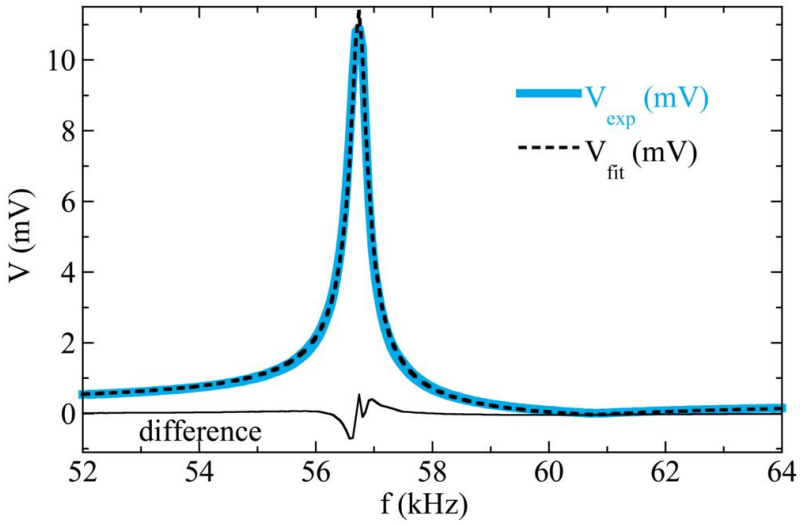


Figure 37. Resonance curve of a magnetoelastic ribbon and best fit to Equation 2.19.

This data analysis approach represents a consistent method to rigorously obtain the resonance parameters. Most importantly, it produces an objective value for the damping parameter δ_r, which is related to the width of the resonance and intimately connected with the viscosity of the medium in which the sample is resonating.

2.4 Results

In this section we will present the results obtained with both prototypes. In the first case, they are meant nearly exclusively to demonstrate that the resonance curve of the magnetoelastic material reflects important changes depending on the viscosity of the oils in which it is immersed. In the second prototype, using the analysis procedure describe in the previous section, a calibration curve will be determined. Also, a study of the viscosity as a function of the temperature will be presented.

2.4.1 Origen of the oils

Oils from different origins are used in this work. The goal is to test the prototypes using oils of different nature and, most of all, displaying a wide range of viscosities. These oils, together with the nomenclature used to identify them are:

- A base oil without additives, denoted as H150NS.
- A family of hydraulic oils specially created for modern transmission systems of power or control, where high work pressure, low tolerances and relatively high temperatures take place. They are denoted with the letter T (T32, T46, T68 and T100).
- A family of lubricant oils for gear-boxes under extreme pressure, specially designed for gears under extreme loads and speeds (they can be used even for gears that withstand shock loads). They are represented with the letter O (O220 and O320).
- Finally, to obtain viscosities that could not be obtained commercially, mixtures of O220 and T100 oils were performed. The resulting oils in this category are denoted as follows: 50/50 for a mixture of a 50 % (in volume) of both oils; 25/75 for a mixture containing a 25 % of O220; and 75/25 for a mixture with a 75 % of O220.

Some of these oils can be seen in the test vials in Figure 38.

Figure 38. Some of the oil samples used to test the prototypes.

The oils were thoroughly characterized using standard laboratory procedures: the viscosity was measured using the ASTM D445 standard test method with an estimated accuracy of 0.1 cSt. The density was determined under the ASTM D4052 standard test method. The measured values are compiled in Table 6.

Table 6. Laboratory results for the different oils tested.

Oil code	T32	T46	T68	T100	O220	O320	H150N S	25/75	50/50	75/25
Density (Kg/m^3)	875	879	886	891	898	902	874.2	894.5	896.8	895.9
Kinematic viscosity (cSt)	32.41	46.3	67.12	108.6	218.2	325.9	31.65	139.7	136.2	174.6
Absolute viscosity (Kg/m s)	0.0284	0.0407	0.0595	0.0968	0.1959	0.2940	0.0277	0.1250	0.1221	0.1564

2.4.2 Results

Proof-of-concept prototype

Only the magnetoelastic sample Vitrovac 4040 was used with this prototype. The 6 mm wide ribbon, cut in a piece 30 mm long, was inserted in the vial containing the oil. The optimum bias field, determined with the sample in air, was obtained with 0.22 A, corresponding to a magnetic field of 4.19 Oe (334 Am^{-1}).

The oils coded T32, T46, T68, T100, O220 and O320 displaying viscosities in the range from 32 to 326 cSt, were successively tested. The raw resonance curves, measured with this prototype are displayed in Figure 39. The resonance in air is also shown for comparison. Just from the observation of the curves, it is evident that the different viscosities of the oils cause important changes in the response. The amplitude and the frequency-position of the resonance vary significantly.

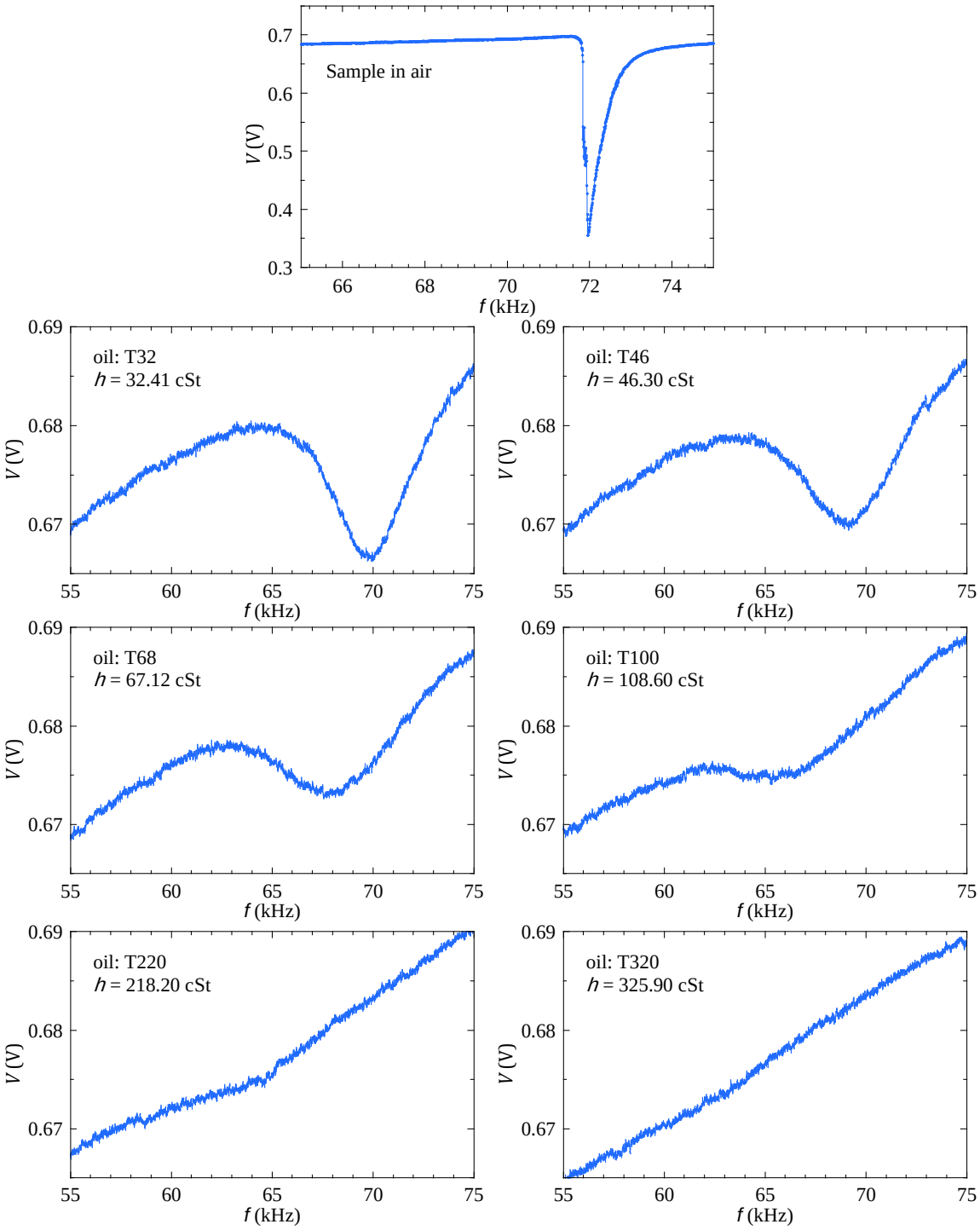


Figure 39. Resonance curves of Vitrovac 4040, in air and in oils of increasing viscosities.

The processing of the measured curves produces quantitative values for these parameters, which are compiled in Table 7. It is to be noted that, for large viscosities, the amplitude of the resonance is quite small, which makes it difficult for the processing algorithm to determine accurately the values of the amplitude and the resonance frequency.

Table 7. Parameters of the resonance after processing the measured curves.

Oil code	Kinematic viscosity (cSt)	Resonance Amplitude (mV)	Resonance Frequency (kHz)
Air		355	71.98
T32	32.4	21	69.52
T46	46.3	16	69.19
T68	67.1	13	68.45
T100	108.6	9	67.23
O220	218.2	5	65.41
O320	325.9	3	64.59

The obtained data can be used to determine a calibration curve in terms either of the amplitude or the resonance frequency. Figure 40 displays the plot of the data and their corresponding least-squares fittings. The amplitude of the resonance as a function of the viscosity of the oil is given by the expression

$$V\,(\text{mV}) = 385.5 \times \eta^{0.82} \tag{2.20}$$

where the viscosity η is expressed in cSt.

The obtained values of the resonance frequency are quite well represented by a second order polynomial in the measured viscosity range:

$$f_r\,(\text{kHz}) = 70.67 - 0.04\ \eta + 5.51 \times 10^{-5}\ \eta^2 \tag{2.21}$$

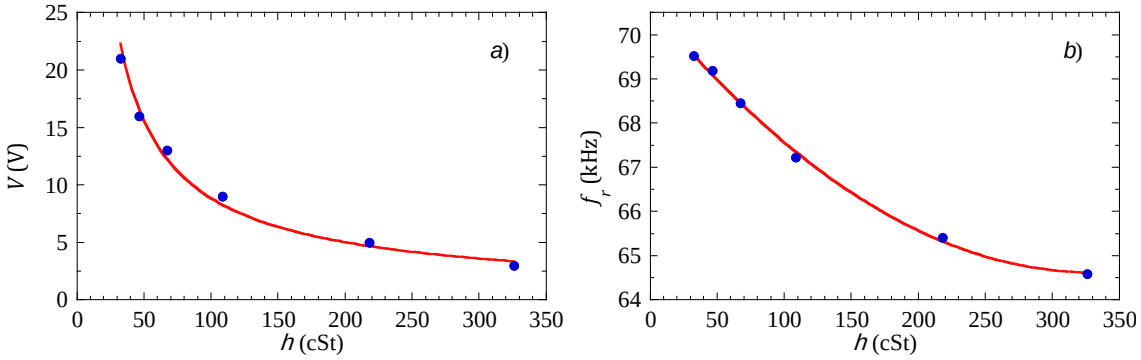


Figure 40. Resonance parameters as a function of the viscosity obtained in the proof-of-concept prototype using VitroVac4040 ribbon. *a)* amplitude of the resonance. *b)* resonance frequency.

Laboratory set-up

With the final prototype, both the Vitrovac 4040 and 7600 amorphous ribbons were used as sensing elements.

Using the samples in air, without introducing them in the oils, the bias field that produced the larger resonance amplitude was determined. For both samples, it resulted H_{bias} = 535 Am^{-1}.

The quality of the resonance curves obtained with this prototype is much greater that those obtained with the previous one. Figure 41 shown the resonance curves obtained for both sensing materials when inserted in oils with different viscosities. The plotted data represent the raw voltage measured by the system. Only a linear background has been subtracted to remove slight amplitude deviations among the different curves (that are inevitability produced by the process of extracting a reposition the test vial when changing the oil).

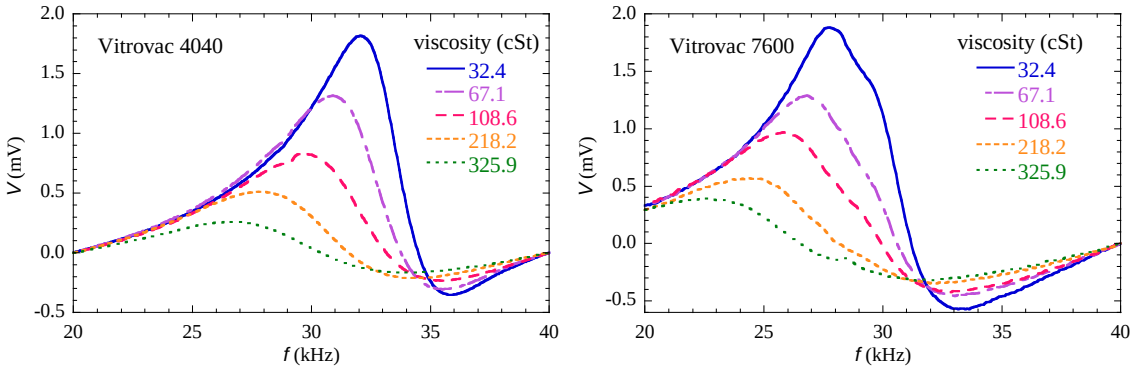


Figure 41. Resonance curves obtanied in oils with diffferent viscosities using Vitrovac 4040 (left) and Vitrovac 7600 (right) ribbons.

The curves display a clear variation of both the value of the resonance frequency and the amplitude of the resonance, when the viscosity of the oil in which they are inserted varies. The small shoulder that can be observed in the curved measured with the Vitrovac7600 sample is caused by an imperfect clamping of the sample to the supporting glass rod (see Figure 25), which allows exciting two different modes of resonance with close resonance frequencies.

A simple data analysis that determine the value and the position of the maximum of the curves allows obtaining the calibration curves represented in Figure 42 and Figure 43.

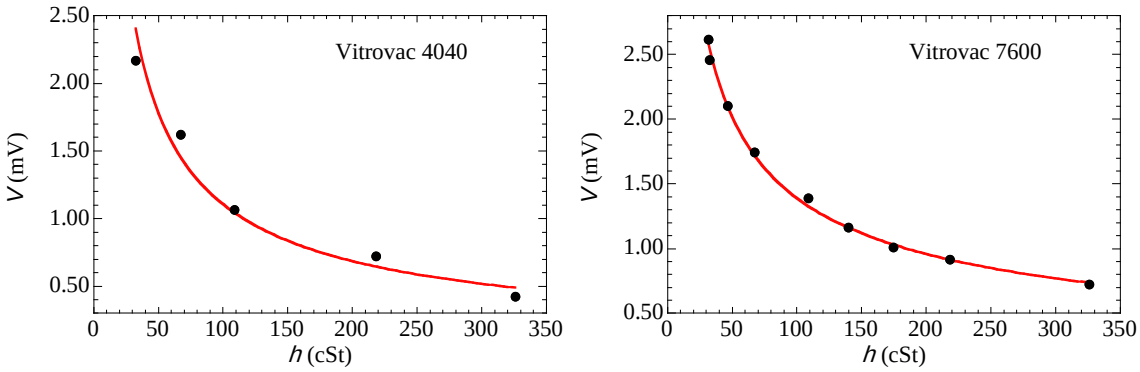


Figure 42. Amplitude of the resonance curves shown in Figure 41 as a function of the oil viscosity for Vitrovac 4040 (left) and Vitrovac 7600 (right).

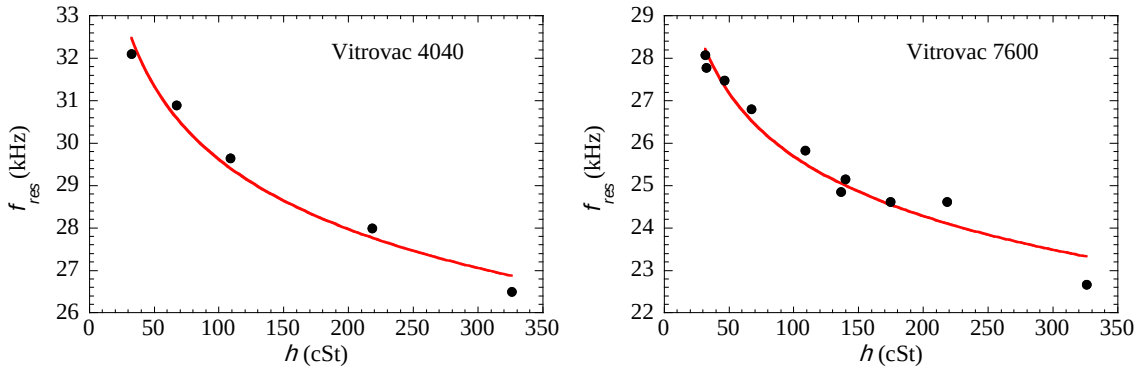


Figure 43. Value of the resonance frequency of the curves shown in Figure 43 as a function of the oil viscosity for Vitrovac 4040 (*left*) and Vitrovac 7600 (*right*).

Stoyanov and Grimes [2.27] proposed that the frequency shift follows a linear relation with the square root of the product of the viscosity and the density. Figure 44 demonstrated that the proposed dependence holds for our results obtained with Vitrovac 4040 and 7600 ribbons.

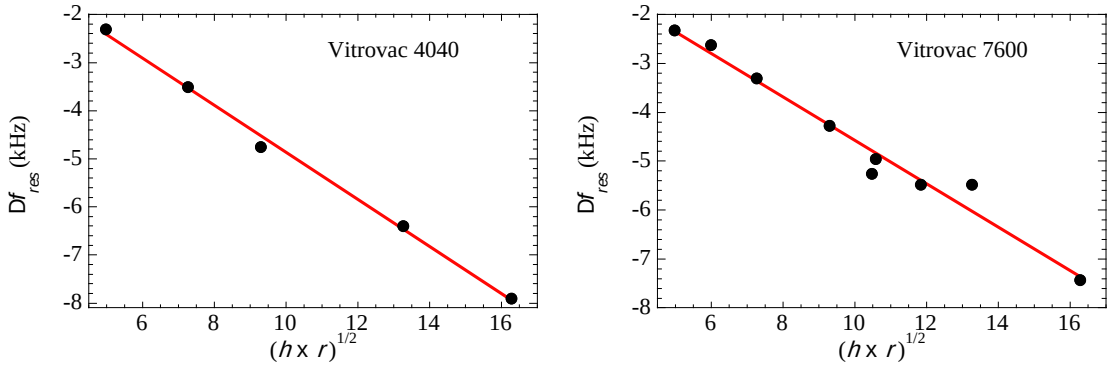


Figure 44. The change in the resonance frequency of the curves shown in Figure 41 follows a linear dependence with respect to the square root of the product of the viscosity and the density of the oil, for both for Vitrovac 4040 (*left*) and Vitrovac 7600 (*right*).

We can apply the data analysis described in section 2.3.4.2 to the data measured for the sample Vitrovac 4040. The analysis is based on the fit to Equation 2.19 and allows determining not only the amplitude and the frequency of the resonance but also a damping parameter, which associated with the width of the resonance curve. This is a parameter that is directly influenced by the viscosity of the oil.

Figure 45 shows the resonance curves obtained with the sample Vitrovac 4040, immersed in different oils, fitted to equation 2.19. The fitting has been performed in Matlab using the *lsqcurvefit* function from the optimization toolbox. The equation 2.19 is implemented in the fitting routine through the *ft* function of the Matlab control toolbox.

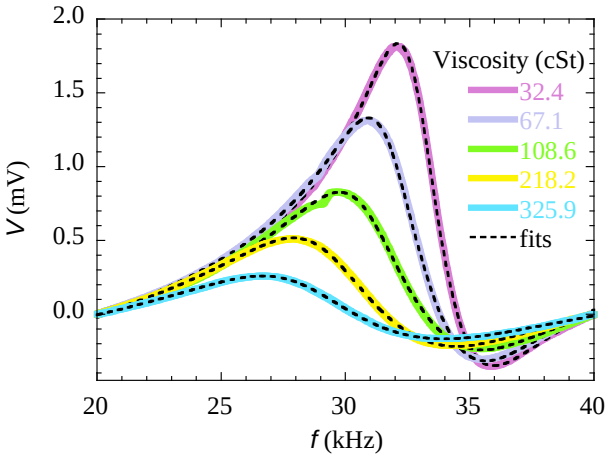


Figure 45. Resonance curves of the sample Vitrovac 4040 in oils with different viscosity and best fits to equation 2.19.

The best-fit values of the amplitude of the resonance, A, and the resonance frequency, f_r, are plotted as a function of the viscosity of the oils in Figure 46. The values obtained from the fit are very close to those obtained directly form the maxima of the resonance curves.

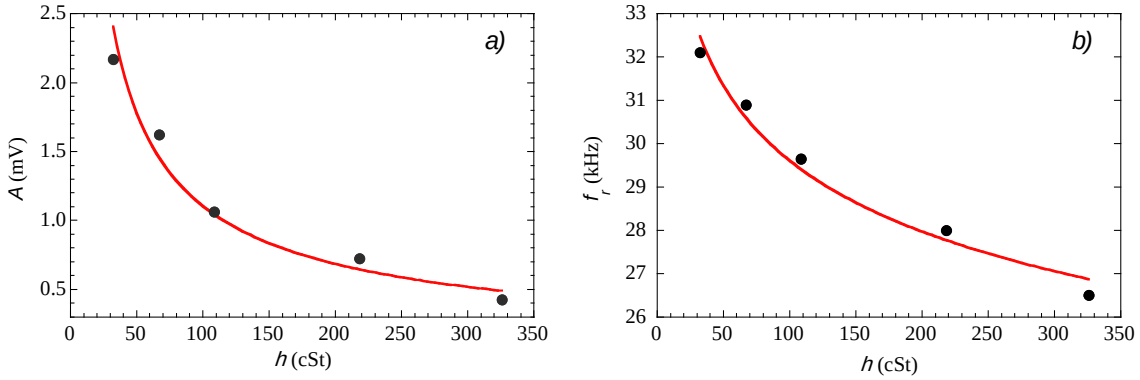


Figure 46. Amplitude (*a*) and resonance frequency (*b*) obtained from the fit of the resonance curves the sample Vitrovac 4040 in oils with different viscosity.

The obtanied values of the amplitude and the resonance frequency as a function of the viscosity of the oil can be used as calibration curves. Figure 46 displays the best-fit curves to these values. The amplitude of the resonance as a function of the viscosity of the oil is given by the expression

$$A\,(\mathrm{mV}) = 26.49 \times \eta^{0.69} \tag{2.22}$$

where the viscosity η is expressed in cSt. The resonance frequency is described by the equation:

$$fr\,(\mathrm{kHz}) = 43.23 \times \eta^{-0.08} \tag{2.23}$$

As stated before, the fit to equation 2.19 allows to determine the value of the damping parameter δ of the magnetoelastic resonance. Figure 47 shows the values obtained for the different oils. It is noteworthy to observe that the hydraulic oils (from the family T, with viscosities up to 108.6 cSt) display a different slope that the oils from gearboxes lubrication (family O, with larger viscosities). This indicates that the damping parameter is very sensitive not only to the viscosity but also to the nature of the oil.

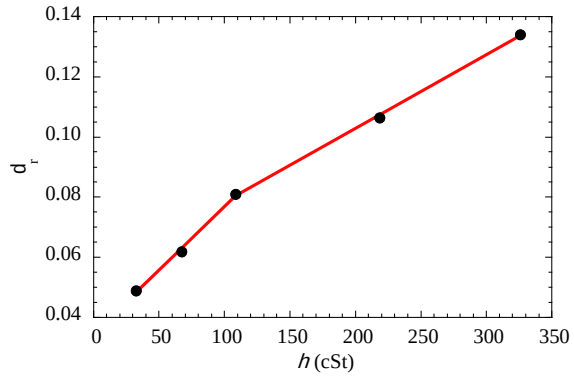


Figure 47. Damping parameter obtained from the fit of the resonance curves.

The linear trends are described by

$$\delta = 0.035 + 4.2 \times 10^{-4}\ \eta \tag{2.24}$$

for oils of the family T ($\eta \leq 108.6$ cSt), and

$$\delta = 0.054 + 2.4 \times 10^{-4}\ \eta \tag{2.25}$$

for oils of the family O ($\eta > 108.6$ cSt).

As seen in Figure 41, the resonance curves obtained with the ribbon Vitrovac 7600 present a small shoulder, indicating the existence of two different resonance modes. The excitation of two different modes is probably caused by an imperfect clamping of the sample to the glass rod. If we intend to apply the analysis procedure that uses the fit to equation 2.19, we must consider two different contributions, one for each of the resonance modes.

Figure 48 resume the results obtained by the fitting procedure for increasing values of the viscosity. In each plot, the sum of the resonance curves of the modes 1 and 2 reproduces perfectly the measured curve. It is very interesting to observe that the relative amplitude of the mode 2 decreases rapidly when the viscosity increases (see Figure 49.*a* where the amplitudes of each contribution are plotted against the viscosity). It seems that higher viscosities impose the existence of only one mode of oscillation (mode 1), damping the other one (mode 2).

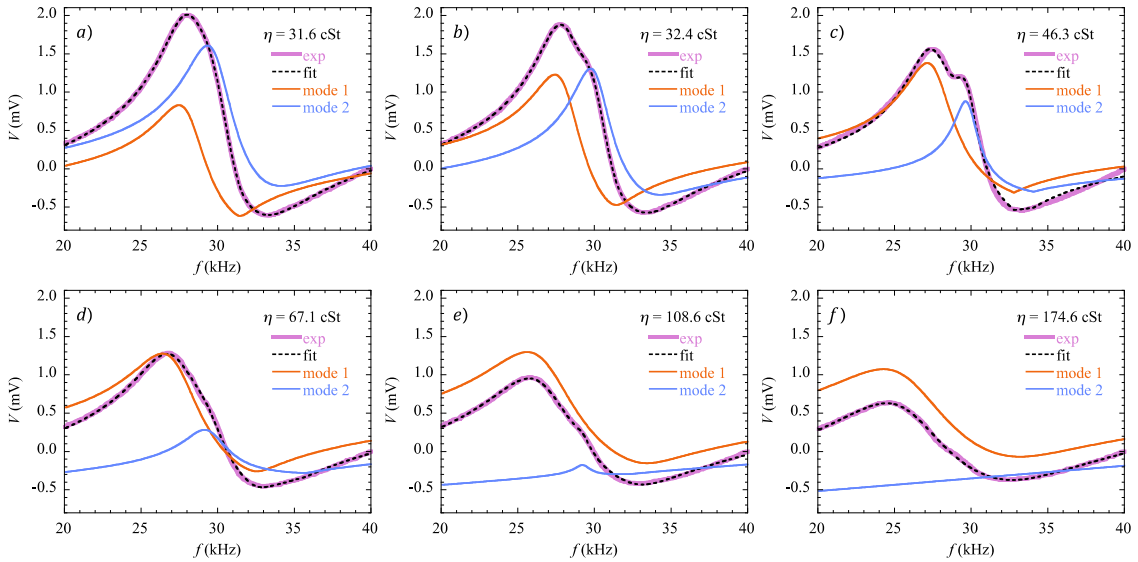


Figure 48. Fit of the measured resonance curve for Vitrovac 7600 to equation 2.19 using two different modes of oscillation. For each viscosity, the calculated resonance curves of mode 1 and mode 2 are displayed. The fit to the experimentally measured curve is the sum of both contributions.

The fitting produces also the value of the damping parameters. They are represented in Figure 49.*b*. The damping parameter of the first mode is the one that reflects clearly the effect of the viscosity of the oil. It is interesting to observe that the damping presents different slope for T and O types of oils, as it happened with the Vitrovac 4040 sample.

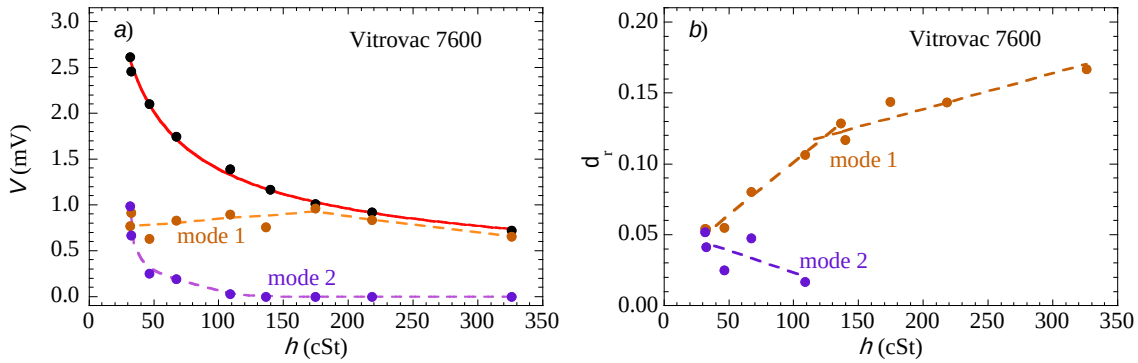


Figure 49. *a*) Amplitude of the two modes involved in the resonance of the Vitrovac 7600 sample as a function of the viscosity, as obtained in the fittings shown in Figure 48. The raw amplitude of the resonance curve as displayed in Figure 42.*b* is also plotted. *b*) Damping parameters obtained from the fits for both modes.

2.4.3 Results as a function of the temperature.

The viscosity of oils is normally dependent on temperature. It is desirable that the on-line monitoring viscosity sensor can be used to determine the viscosity at different temperatures. However, the magnetoelastic effect is also temperature dependent, and the magnetoelastic resonance can be severely affected by temperature changes. This is evident in equation 2.17 (section 2.2.3) that gives, for a simple model, the analytical expression of the young modulus of a magnetoelastic material:

$$E(H,\sigma) = \frac{1}{1+\frac{9\lambda_S^2\mu_0^2M_S^2H^2}{(2K-3\lambda_S\sigma)^3}E_S} \tag{2.17}$$

The materials parameters λ_s (saturation magnetostriction), M_s (saturation magnetization) and K (anisotropy constant) are all temperature dependent. Therefore, the frequency at which the magnetoelastic resonance takes place f_r changes appreciably with temperature, since f_r and E are intimately coupled (see equation 2.12). In Ref [2.47] this dependence was established experimentally. Figure 50.*a* exemplifies the variation of the ΔE curve with temperature in a $Fe_{66}Co_{18}Si_1B_{15}$ amorphous ribbon in air. As illustrated in Figure 50.*b*, the relative variation of the resonance frequency with respect to its value at 20° C can be as large as 13 %, depending on the bias field. The amplitude of the resonance curve is affected similarly. Therefore, any sensor relaying in the magnetoelastic resonance must take into account the temperature response of the magnetoelastic sensing material. Fortunately, Figure 50.*b* also reveals that there is a value of the bias field for which the variation of the resonance frequency is null. In the case of Figure 50, this compensation point takes place with an applied bias field of 827 A/m. The strategy is, therefore, to determine the compensation bias field in the samples used in our laboratory prototype. Then we can test the possibility of measuring the viscosity of the oils at different temperatures.

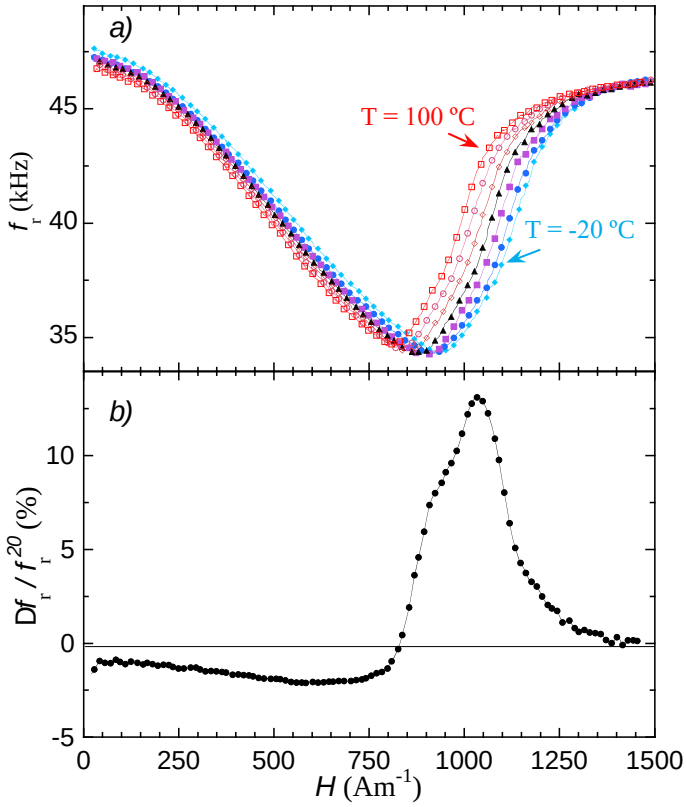


Figure 50. *a)* Dependence of the ΔE curve with temperature in a $Fe_{66}Co_{18}Si_1B_{15}$ amorphous ribbon. *b)* Variation of the sensitivity of the resonance frequency to the field in the range of temperatures from -20° to 100° C, relative to the value at 20° C. Adapted from [2.47].

The complete measuring system was situated inside a climatic chamber and the ΔE curves of the sample Vitrovac 4040 in air (without oil in the vials) measured at different temperatures, from 0° to 60° C. Figure 51 compiles the results obtained.

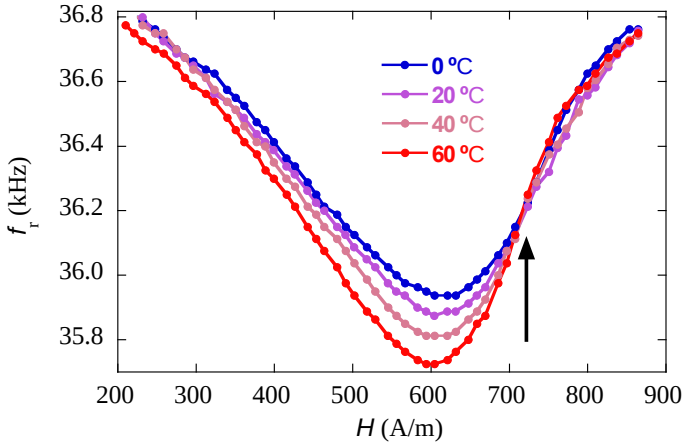


Figure 51. ΔE curves of the sample Vitrovac 4040 measured in air at different temperatures, from 0° to 60° C. The compensation point occurs at a bias field of H_{bias} = 722 A/m (indicated by an arrow).

The ΔE curves describing the variation of the resonance frequency with the applied field show a clear dependency on temperature. The compensation point, that is, the field at which the curves collapse and the frequency of resonance becomes insensible to temperature, takes place when the sample is biased at about 722 A/m.

The complete resonance curves of the magnetoelastic sample Vitrovac 4040, inserted in oils of the T family with different viscosities were measured at different temperatures, with the sample biased at the compensation field of 722 A/m. The curve corresponding to the oil T46 (46.3 cSt at 20° C) is shown in Figure 52 as an example.

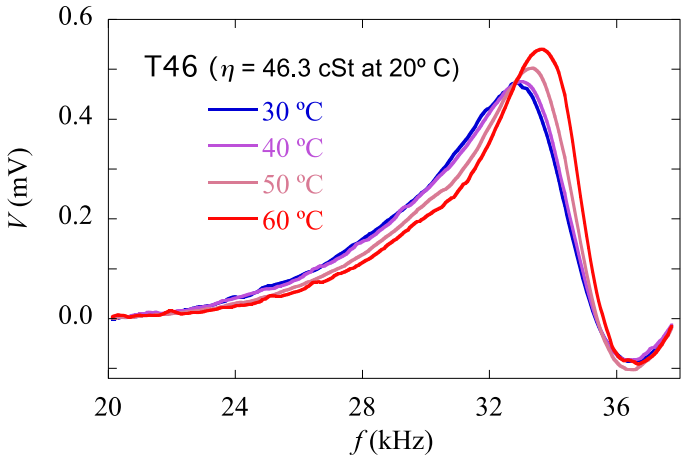


Figure 52. Resonance curves measured with the sample Vitrovac 4040 inserted in the T46 oil, measured at different temperatures while biased at 722 A/m.

The values of the resonance frequencies as a function of the temperature obtained from curves similar to that of Figure 52 for the oils of the T family, are represented in Figure 53. Since the resonance has been measured at the compensation point in which the magnetoelastic behavior is independent of the temperature, the variation in the resonance frequency must be caused by the changes in the viscosity of the oils due to temperature variations.

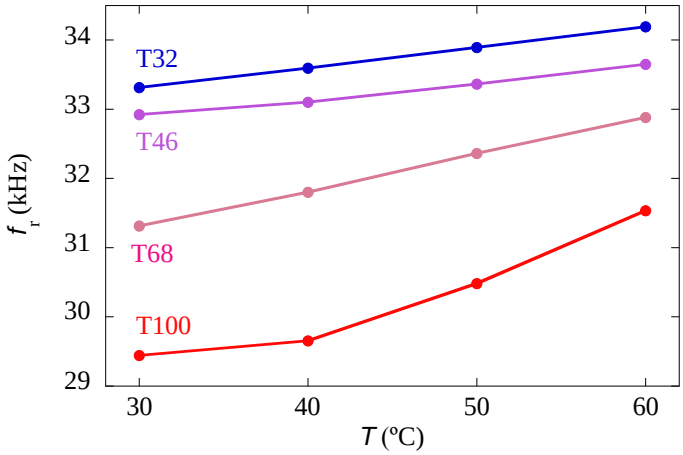


Figure 53. Resonance frequencies measured at different temperatures for the oils of the T family.

From the measured resonance frequencies, we can determine the viscosity of the oils at each temperature using the calibration between the resonance frequency and the viscosity. This relation was established for the Vitrovac 4040 sample at room temperature in Figure 46.*b* and its corresponding Equation 2.23. Note that this calibration data was obtained with the sample biased at 533 A/m. Fortunately, it is not necessary to obtain a new calibration curve at room temperature biased at 722 A/m, since the resonance frequency at both bias fields (722 and 533 A/m) is approximately the same (see Figure 51).

Figure 54 displays the values of the viscosity of the oils at different temperatures calculated through the calibration curve. Note that the oils with lower viscosity, T32 and T46, are close to the lower limit of the calibrated curve. When the temperature increases, the viscosity is reduced, and its value gets out of the calibrated range. Therefore, for these two oils, the values displayed in Figure 54 are calculated by extrapolating the calibration curve.

The estimated values of the viscosity as a function of the temperature displayed in Figure 54 are affected by an undetermined uncertainty for all the reasons stated above. However, the presented results must serve as an example of the procedure that can be followed to use the viscosity sensor in a wide range of operating temperatures.

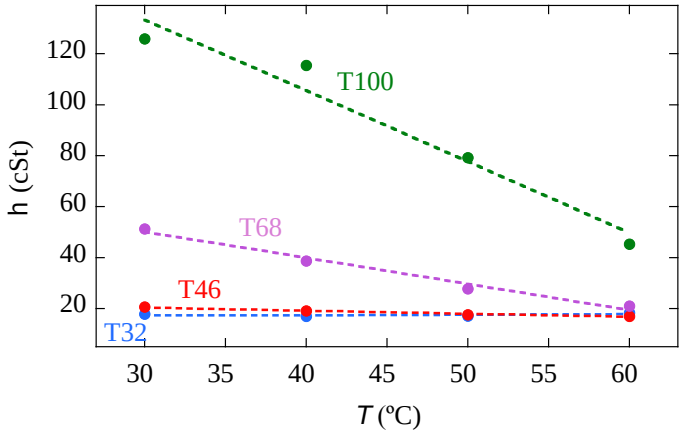


Figure 54. Viscosity as a function of the temperature as deduced from the measured resonance frequencies and the calibration performed at room temperature.

2.5 Conclusions

In this chapter we have shown that the magnetoelastic resonance is a useful technique to determine the viscosity of lubricant oils in a wide range. Both the frequency and the amplitude of the resonance can be used to determine the viscosity of the oils. A working prototype of a sensor based on this principle, with on-line and real-time operation, have been designed, constructed and tested. A new analytical method for extracting the relevant parameters from the resonance curves has been developed. Additionally, we have proved the possibility of determining the viscosity at different temperatures using the same principle.

Bibliography

[2.1] Lubrication. en.wikipedia.org. Retrieved 18 January 2018, from https://en.wikipedia.org/wiki/Lubrication

[2.2] ANDERSON, Kevin J., 1991, A History of Lubricants. *MRS Bulletin*. 1991. Vol. 16, no. 10p. 69–69.

[2.3] STACHOWIAK, Gwidon; BATCHELOR, Andrew W. *Engineering tribology*. Butterworth-Heinemann, 2013.

[2.4] viscosity | Definition, Facts, & Examples. Encyclopedia Britannica. Retrieved 18 January 2018, from https://global.britannica.com/science/viscosity.

[2.5] EESC V2100 - The Life System. Eesc.columbia.edu. Retrieved 18 January 2018, from http://eesc.columbia.edu/courses/ees/life/lectures/lect03.html

[2.6] Understanding Absolute and Kinematic Viscosity. Machinerylubrication.com. Retrieved 18 January 2018, from http://www.machinerylubrication.com/Read/294/absolute-kinematic-viscosity.

[2.7] THIBAULT, R.L., 2018, Certified lubrication specialist exam course preparation handbook - Lubrication Training and Consulting. Lecture.

[2.8] ASTM D2270-10, Standard Practice for Calculating Viscosity Index from Kinematic Viscosity at 40 °C and 100 °C, ASTM International, West Conshohocken, PA, 2016

[2.9] ISO 448:1992(en) Industrial liquid lubricants — ISO viscosity classification, ISO International

[2.10] ASTM D2161-17, Standard Practice for Conversion of Kinematic Viscosity to Saybolt Universal Viscosity or to Saybolt Furol Viscosity, ASTM International, West Conshohocken, PA, 2017

[2.11] ISO Viscosity Grades. Machinerylubrication.com. Retrieved 18 January 2018, from http://www.machinerylubrication.com/Read/213/iso-viscosity-grades

[2.12] ASTM D445-17a, Standard Test Method for Kinematic Viscosity of Transparent and Opaque Liquids (and Calculation of Dynamic Viscosity), ASTM International, West Conshohocken, PA, 2017

[2.13] DOKTER, M. Anatomy of a Viscometer. Machinedokter.blogspot.com.es. Retrieved 18 January 2018, from http://machinedokter.blogspot.com.es/2014/12/anatomy-of-viscometer.html

[2.14] *CAV®-2100 Single-Bath Kinematic Viscometer*. *Cannoninstrument.com*. Retrieved 18 January 2018, from https://www.cannoninstrument.com/Image/GetDocument/586?language=en

[2.15] ASTM D4052-16, Standard Test Method for Density, Relative Density, and API Gravity of Liquids by Digital Density Meter, ASTM International, West Conshohocken, PA, 2016

[2.16] FITCH, Jim, Trouble-Shooting Viscosity Excursions, *Practicing Oil Analysis Magazine*, 2001, vol. 5.

[2.17] MARKOVA, L. V., et al. On-line monitoring of the viscosity of lubricating oils. *Journal of Friction and Wear*, 2010, vol. 31, no 6, p. 433-442.

[2.18] MARKOVA, L. V., et al. On-line acoustic viscometry in oil condition monitoring. *Tribology International*, 2011, vol. 44, no 9, p. 963-970.

[2.19] KRESS-ROGERS, Erika; BRIMELOW, Christopher JB (ed.). *Instrumentation and sensors for the food industry*. Woodhead Publishing, 2001.

[2.20] PV-100 Viscosity Control System. Brookfieldengineering.com. Retrieved 18 January 2018, from

http://www.brookfieldengineering.com/products/viscometers/in-line-process-viscometers/pv-100-viscosity-control-system

[2.21] AST-INK Viscometer. Brookfieldengineering.com. Retrieved 18 January 2018, from http://www.brookfieldengineering.com/products/viscometers/in-line-process-viscometers/ast-ink-viscometer

[2.22] TT-100 Viscometer. Brookfieldengineering.com. Retrieved 18 January 2018, from http://www.brookfieldengineering.com/products/viscometers/in-line-process-viscometers/tt-100-viscometer

[2.23] Vinci Technologies | Rolling ball Viscometer (RBV 1000) | Laboratory and field instruments for Petroleum Industry. Vinci-technologies.com. Retrieved 18 January 2018, from http://www.vinci-technologies.com/products-explo.aspx?IDM=536755&IDR=82291&IDR2=82560

[2.24] Acoustic Wave Technology Sensors | Sensors Magazine. Sensorsmag.com. Retrieved 18 January 2018, from http://www.sensorsmag.com/sensors/acoustic-ultrasound/acoustic-wave-technology-sensors-936

[2.25] FARONE, William A.; SACHER, Robert F.; FLECK, Charles. *Acoustic viscometer and method of determining kinematic viscosity and intrinsic viscosity by propagation of shear waves*. U.S. Patent No 6,439,034, 27 Ago. 2002.

[2.26] BiODE, M. Surface Acoustic Wave Sensor for Viscosity Measurement - Tech Briefs. Techbriefs.com. Retrieved 18 January 2018, from http://www.techbriefs.com/component/content/article/ntb/tech-briefs/physical-sciences/1285

[2.27] STOYANOV, Plamen G.; GRIMES, Craig A. A remote query magnetostrictive viscosity sensor. *Sensors and Actuators A: Physical*, 2000, vol. 80, no 1, p. 8-14.

[2.28] ROTH, Wilfred, et al. *Method and apparatus for measuring viscosity, etc., of fluid-like materials*. U.S. Patent No 2,839,915, 24 Jun. 1958.

[2.29] BRADFIELD, Geoffrey; MCKENNELL, Raymond. *Viscometers*. U.S. Patent No 3,062,040, 6 Nov. 1962.

[2.30] THOENE, Ernst, Viscometer, DE Patent No 3,725,034,2 February 1989

[2.31] Oil Sensor Technology - Fraunhofer IMM. Fraunhofer Institute for Microengineering and Microsystems IMM. Retrieved 18 January 2018, from https://www.imm.fraunhofer.de/en/innovation-fields/oil-sensor-technology.html

[2.32] KONG, Ho Sung, et al. *Method and apparatus for measuring oil viscosity*. U.S. Patent No 8,521,451, 27 Ago. 2013.

[2.33] MARKOVA, L.V., et al., On-line magnetoelastic oil viscosity detector. In : *3rd European Conference on Tribology*. 2011. p. 503-504.

[2.34] GRIMES, Craig A., et al. (ed.). *Encyclopedia of sensors*. American Scientific Publishers, 2006.

[2.35] DU TRÉMOLET DE LACHEISSERIE, Etienne. *Magnetostriction: theory and applications of magnetoelasticity*. CRC press, 1993.

[2.36] CULLITY, Bernard Dennis; GRAHAM, Chad D. *Introduction to magnetic materials*. John Wiley & Sons, 2011.

[2.37] OLABI, Abdul-Ghani; GRUNWALD, Artur. Design and application of magnetostrictive materials. *Materials & Design*, 2008, vol. 29, no 2, p. 469-483.

[2.38] Pressductor® Technology. ABB. Retrieved 18 January 2018, from http://www.abb.com/product/ap/seitp330/7c2dbedf705e56d1c1257899003180d0.aspx

[2.39] CHIRIAC, H.; PLETEA, M.; HRISTOFOROU, E. Magneto-surface-acoustic-waves microdevice using thin film technology: design and fabrication process. *Sensors and Actuators A: Physical*, 2001, vol. 91, no 1, p. 107-111.

[2.40] VANDERBURGT, C. M. Dynamical physical parameters of the magnetostrictive excitation of extensional and torsional vibrations in ferrites. *Philips Research Reports*, 1953, vol. 8, no 2, p. 91-132.

[2.41] AUSANIO, Giovanni; IANNOTTI, Vincenzo; LANOTTE, Luciano. Magnetoelastic Stress and Strain Sensors. 2006.

[2.42] Magneto-Elastic Resonance (MER) Sensors | RISE Acreo. Acreo.se. Retrieved 18 January 2018, from https://www.acreo.se/our-offer-0/magneto-elastic-resonance-mer-sensors

[2.43] VITROVAC 4040 P40. Vacuumschmelze.com. Retrieved 18 January 2018, from http://www.vacuumschmelze.com/fileadmin/Medienbiliothek_2010/Unternehmen/Qualitaet/SD_Sicherheitsdatenblaetter/HT/IB37E.pdf

[2.44] *VITROVAC 7600 T70. Vacuumschmelze.com*. Retrieved 18 January 2018, from http://www.vacuumschmelze.com/fileadmin/Medienbiliothek_2010/Unternehmen/Qualitaet/SD_Sicherheitsdatenblaetter/HT/IB42E.pdf

[2.45] QUINTANA, I., et al. Laser micromachining of metallic glasses: investigation of the material response to machining with micro-second and pico-second lasers. En *Laser Applications in Microelectronic and Optoelectronic Manufacturing XV*. International Society for Optics and Photonics, 2010. p. 75840Y.

[2.46] GARCÍA-ARRIBAS, Alfredo, *et al*. Sensor applications of soft magnetic materials based on magneto-impedance, magneto-elastic

resonance and magneto-electricity. *Sensors*, 2014, vol. 14, no 5, p. 7602-7624.

[2.47] GARCÍA-ARRIBAS, A., *et al*. Selectable temperature sensitivity of the magnetoelastic resonance. *Sensors and Actuators A: Physical*, 2003, vol. 106, no 1, p. 111-116.

3

Motor current signature analysis

This chapter focuses on proving the suitability of motor current signature analysis (MCSA) as a method for condition monitoring of gearboxes.

The expression condition monitoring denotes the collection and analysis of the information that is associated to the health state of a system. However, the condition monitoring strategy is not free of imperfections. For instance, the analysis of the obtained information can take time, and a failure in unexpectedly degraded equipment may occur before the results are available. Additionally, the data collection and the analysis must often be done manually, by specifically trained personnel, with the associated increase of costs. Finally, the placement of the sensors needed to retrieve the required information on the system may not only be expensive but also difficult to implement in many cases.

The key steps of condition monitoring are [3.1]:

- Data acquisition: collecting information relevant to the health state of the system.
- Signal processing: for handling and analysing the information collected in step one, for understanding and interpreting it.
- Maintenance and decision-making: for recommending the most suitable maintenance action.

In the present work, the focus is set in data acquisition and signal processing, with the objective of automating and enabling the application of these steps in MCSA, and ultimately validating the suitability of the method. The work includes two different testing scenarios: First, the testing is focused on distinguishing gears in good condition from gears with severe damage. The second testing is focused on the diagnostics of exact faults in the gearbox.

The chapter is therefore divided in four mayor parts.

In the first part of the work, the theoretical background of MCSA and of the signal processing techniques used is explained. Signal analysis mainly takes place from three points of view: Time domain analysis from the raw signal; pre-processing with dual level time synchronous averaging; and pre-processing with wavelets analysis. After that, a feature selection is done, followed by a classification of the features using different techniques. Each of the techniques used are explained in detail.

In the second part, the experimental set up is described. The test bench and its modifications are explained. Each of the components with certain importance for the work presented are gone through in detail (sensors, way of organizing the data pool, etc.). This includes the experimentation itself,

clearing up the design of experiments, and the faults tested. Finally, the procedure of data analysis, reduction and classification is shown.

In the third section, the results are explained. Results from the concept proof as well as from the gear condition diagnostics are discussed, including the analysis of the different techniques. In particular, wavelet analysis and dual level time synchronous averaging are compared by mapping the results and studying their statistical significance. Different percentages of success and deviations are obtained from the classification of both approaches. This particular part of the work is a result of a collaboration performed during a research stage at the University of Cincinnati's Intelligent Maintenance Systems Center, in Cincinnati, OH, in the United States of America.

A fourth section is dedicated to analysing additional results of the application of MCSA, this time to bearing fault detection.

Finally, the last section describes de concept of fingerprint, and some possible applications in different fields: machine tool, electro-mechanical actuators and electric vehicles.

Content

1 Theoretical background

Usually, the condition of a gearbox is assessed through the use of accelerometers [3.2]. This approach however, has inconveniences: accelerometers are quite expensive and their installation at the proper places within the gearbox may be cumbersome, requiring also a deep knowledge of the machinery to deploy the sensors and to interpret the data.

The inconveniences of accelerometers can be avoided by using internal signals for monitoring. In the particular case of electric motors with gearboxes, the analysis of the current feeding the motor that moves the gearbox may be used. These signals are correlated with the health state of the gearbox, as it is evidenced in this work, and readily available without the use of additional sensors. In fact, motor current signature analysis is already used for the condition monitoring of electrical motors [3.3] although, to the best knowledge of the author, there is still no commercial product for the condition monitoring of the mechanical components situated downstream of the electric drive.

1.1 Motor current signature analysis for mechanical faults

The approach of examining the electrical current feeding the motor, as a mean for establishing the health condition of mechanical elements in rotating machinery, became known as Motor Current Signature Analysis (MCSA). Developed in the 80's in the Oak Ridge National Labs [3.4], MCSA has been studied, applied end extended by many other authors both from a theoretical and an experimental point of view [3.5]-[3.6]. It relies on the fact that an electric motor driving a mechanical load operates as an efficient and permanently available transducer, capable of sensing mechanical load variations and converting them into induced currents generated in the motor windings. For example, as described in Ref. [3.7], the vibration caused by mechanical faults in downstream equipment induces changes in the magnetic field and, as a consequence, in the inductance of the motor.

Mechanical (torsional) vibrations transmitted to the shaft of electric motors can be detected and analysed through the specific frequencies that appear in the feeding current as a consequence of modulation products between the power system frequency and the vibrating frequencies. Therefore, useful information about the torsional vibrations can be extracted from the analysis of the stator current using the theory of electrical machines. This can be shown mathematically, following the treatment given in [3.8].

To analyse the influence on the feeding current of vibrations at a single frequency, it is necessary to derive the air-gap torque T_e in the electrical machine, since it links the electromagnetic and mechanical behaviour of the system. In general, when a vibration of frequency f_0 is present, the air-gap torque T_e can be written as

$$T_e = T_{e0} + A_T \cos(2\pi f_0 t + \varphi) \tag{3.1}$$

constituted by two contributions: the average torque T_{e0} and the alternating torque ($A_T \cos(2\pi f_0 t + \varphi)$) produced by the vibration.

The torque is produced by the interaction between the magnetic field and the currents in the machine. In a balanced three phase machine, the flux linkage of the stator λ_s has a constant magnitude and rotates at $2\pi f$ rad/s, where f is the frequency of the electrical supply. The torque can be expressed as [3.9]

$$T_e = \frac{3P_p}{2} \lambda_s i_{sT} \tag{3.2}$$

where i_{sT} is the component of the stator current that produces the torque and P_p is the number of pole pairs. The total stator current i_s is composed by the magnetizing current i_{sM}, which is in phase with λ_s, and the torque producing component i_{sT}, which is 90º ahead (see Figure 1). Therefore, according to equations (3.1) and (3.2), both components of the stator current can be expected to be composed of two terms:

$$i_{sM} = i_{sM0} + A_{sM} \sin(2\pi f_0 t + \varphi_M) \tag{3.3}$$

$$i_{sT} = i_{sT0} + A_{sT} \sin(2\pi f_0 t + \varphi_T) \tag{3.4}$$

The stator current i_s can be measured when projected over the phase currents of the machine, with frequency f. For example, for the first phase (phase a), the phase current becomes (see Figure 1):

$$\begin{aligned} i_s^a = {} & i_{s0} \sin(2\pi f t + \varphi_{s0}) \\ & + \frac{1}{2}\{A_{sM} \cos[2\pi(f - f_0)t - \varphi_M] + A_{sT} \cos[2\pi(f - f_0)t - \varphi_T]\} \\ & - \frac{1}{2}\{A_{sM} \cos[2\pi(f + f_0)t + \varphi_M] - A_{sT} \cos[2\pi(f + f_0)t + \varphi_T]\} \end{aligned} \tag{3.5}$$

where i_{s0} is the magnitude of the stator current given by

$$i_{s0} = \sqrt{i_{sM0}^2 + i_{sT0}^2}, \tag{3.6}$$

and φ_0 is given by

$$\varphi_{s0} = tg^{-1} \frac{i_{sT0}}{i_{sM0}}. \tag{3.7}$$

Equation (3.5) clearly establishes that the effect of a vibration with a frequency f_0 can be readily observed in the feeding current as sideband components of frequency $f + f_0$ and $f - f_0$.

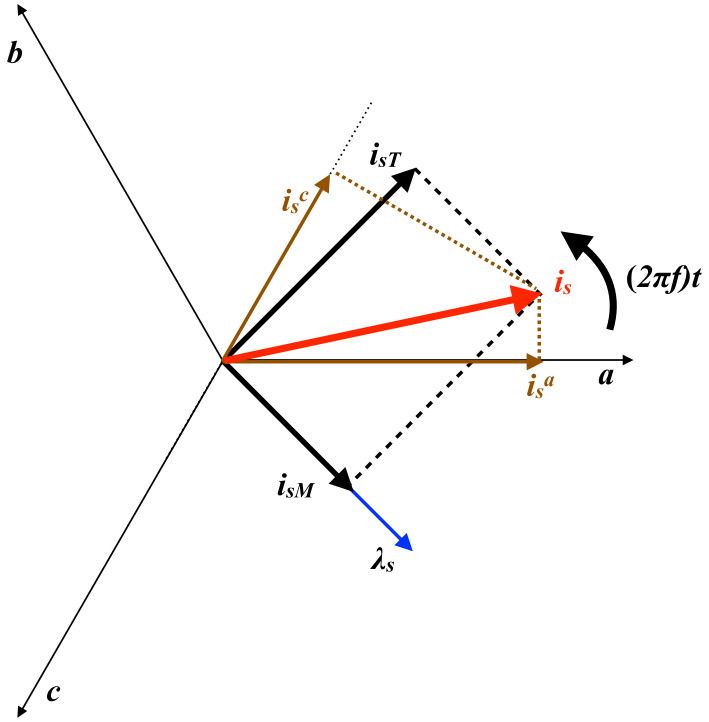


Figure 1. Vector diagram of the current components in the induction motor. Adapted from [3.8].

The data generated in the test bench used in this work, that will be explained in section 2.1, support the above theoretical concept. As an example, the analysis of the measured signal from a test performed at a 1000 rpm proves the existence of mechanical frequencies (Figure 2). A discrete Fourier transform was applied to the data, and a specific mechanical frequency was searched. In this case, the mechanical component searched for was the rotational speed of the intermediate shaft (see Figure 10 below for a detailed description of the gearbox). The rotational speed of the 2-pole electric motor is 33.32 Hz (2×16.66 Hz corresponding to 1000 rpm). The relationship between the first and the second shafts is 2.5, so the rotation frequency is 13.33 Hz. The sidebands are therefore expected at 46.65 Hz and 19.99 Hz respectively. In Figure 2, the feeding current is 33.6 Hz, with the top sideband at 48.8 Hz and the bottom sideband at 18.33 Hz. The small discrepancies may come from small variations in the rotating speed and from the resolution of the DFT.

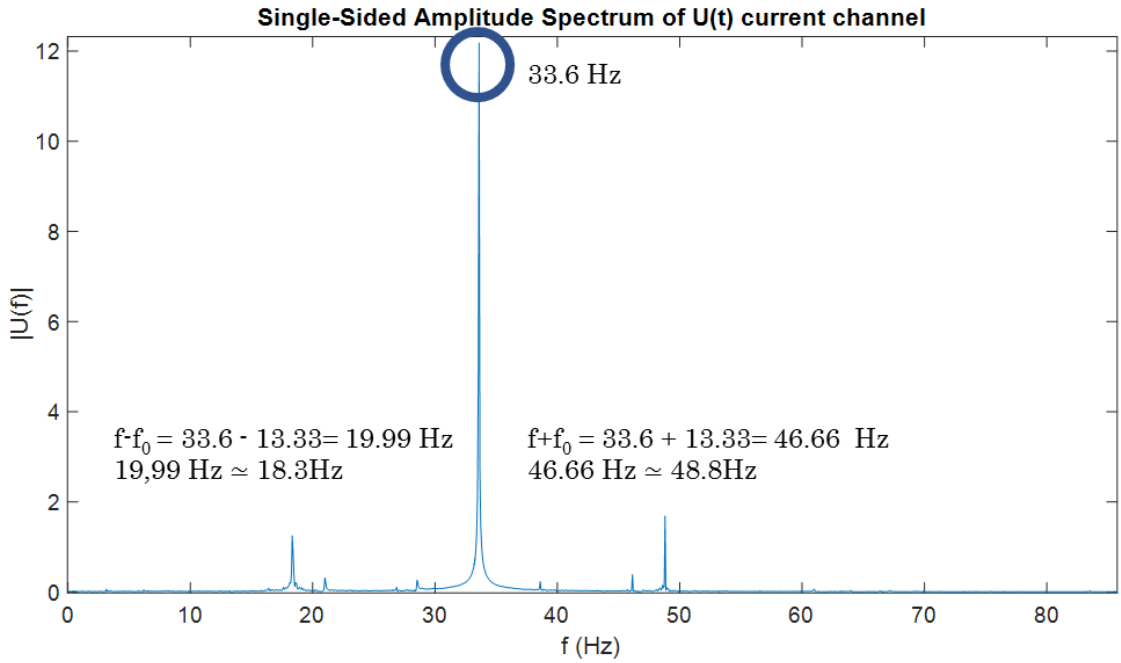


Figure 2. Single sided amplitude spectrum of *U* current channel of a test performed with the motor turning at 1000 rpm constant speed. (The data correspond to a test using a healthy gear, repetition number 12).

1.2 Signal processing

Signal processing serves to analyse the measurement signals obtained in the data acquisition step, using adequate techniques and methods [3.10]. The task of signal processing is to extract useful information that can reveal the health condition of the asset in which the signals were acquired. Several techniques have been proposed in the literature that can be roughly classified in three categories: time domain, frequency domain and time-frequency domain. Figure 3 schematizes the process of signal analysis.

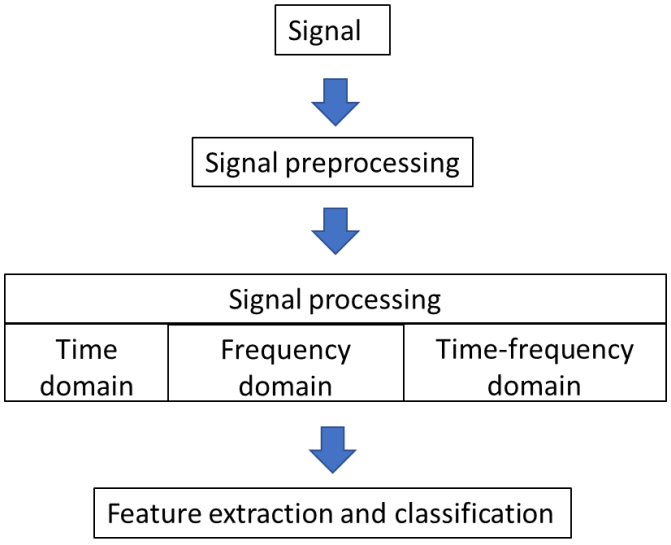


Figure 3. Scheme showing the process of data analysis.

Signals are commonly divided between stationary and non-stationary. Stationary denotes that the statistical properties of the signal are invariant with time. Consequently, non-stationary refers to anything that doesn´t satisfy these conditions, and can be separated into two categories: continuous and transient. The signal that will be the focus of the analysis performed in the present work will be transient, non-stationary. The signal will be the result of the transients generated in speed changes, as will be explained in section 2.2.1. Therefore, the techniques selected for signal analysis in this Thesis will be valid for such conditions.

Generally, fault characteristics are embedded in heavy background noise. This is especially true for early faults. That is why it is recommended an initial step of signal pre-processing. The most important pre-processing techniques are mean removal, trend removal and signal filtering. Mean and trend removal are based on eliminating the mean value carrying the signal, which doesn´t generally have any useful information. Signal filtering means removing unwanted frequency bands and eliminating the interference of noise. One method widely used nowadays is the time synchronous averaging (TSA), designed to extract periodic waveforms from noisy signals.

The time domain signal processing category, involves dealing directly with the pre-processed time wave form. This category includes auto-correlation (that captures the periodic characteristics of a signal) and cross-correlation analysis (which measures the relevance degree between two different signals). Another important processing technique under this category is statistical feature extraction. These features reflect the statistical properties from the time-domain waveform. The most common ones include: mean value, root mean square, peak value, kurtosis and crest factor, among others.

Time domain analysis only reflects the variations of the waveform of the signal, but is not able to disclose the distribution and the frequency components of it. To compensate the shortcomings of time domain analysis, frequency domain analysis was proposed, mainly through the frequency spectrum. Its fundaments rely on the Fourier transform that decomposes a periodic signal into the superposition of its sinusoidal components, which are projected in the frequency axis, creating the frequency spectrum. However, the Fourier transform can only process continuous signals, not discrete ones. The discrete Fourier transform was developed to process discrete signals with a computer. The Fast Fourier Transform (FFT) was developed to determine the frequency spectrum in a reduced computing time.

Considering that the Fourier transform is restricted to stationary signals and cannot reflect the changes in each frequency with time, alternative time-frequency domain processing techniques were developed. These techniques can reflect the change of the signal from time and frequency domain. Some of the techniques included in this category are short-time Fourier transform, Wigner-Vile distribution, Hilbert-Huan transform, continuous wavelet transform and the discrete wavelet transform.

In this thesis, the time domain statistical feature extraction and discrete wavelet transform were used, as they provide an optimum manner to analyse transitory signals, with a reduced computational cost. After the signals are extracted from the discrete wavelet transform, time-domain descriptors are obtained. So, we have effectively used discrete-wavelet transform as a pre-processing technique. The process will be explained in depth in sections 1.2.1, 1.2.2, 1.2.3 and 2.3.

Also, the dual level time synchronous averaging pre-processing technique was used to compare results with other well-known techniques. It will be explained more in depth in section 1.2.2.

As it has been seen, there are several ways of obtaining descriptors, and nowadays computational power makes easier their acquisition. It is of common knowledge that, for a good diagnostic, the use of more than one descriptor is needed [3.11], [3.12]. The shortcomings encountered in some of the descriptors are overcome by others. As a result, the problem appearing after analysing the signal is the high number of descriptors generated, and the selection of the most interesting ones is a challenging task.

For such, in the present work, four feature selection methods were used: a one-way analysis method; the correlation feature selection method; information gain method; and the relief method. They will be described in section 1.3.

In an effort to quantify the impact of both techniques and the feature selection methods, four different classification methods were used; Bayesian network; sequential minimal optimization; lazy instance based learning; and trees J48. The *t*-test was used to compare the results. They will be described in section 1.4.

1.2.1 Time domain analysis

Time domain analysis is the most straightforward technique as statistical values are calculated directly from the time-domain signal [3.10]. It is based on the assumption that, when a fault occurs in the mechanical component,

the stiffness of the mechanical structure around it changes. This variation can produce impulse or shock that will be reflected in the vibration of the component. The amplitudes and distribution of the measured signals may be affected. Typically, time domain methods have been used for the analysis of signals coming from accelerometers, but as mechanical faults generate vibrations in the stator, motor current signature analysis is an indirect way of measuring them [3.13].

Common statistical features in the time domain are compiled in Table 1. They include the mean value, reflecting the average of the signal, and the root mean square and peak value, which are related with the vibration amplitude and energy in the time-domain. These are generally good fault indicators as the mechanical fault excites mechanical vibrations, uplifting the values of these descriptors. They specially work better the more severe the fault becomes, being able even to distinguish the severity of the fault. On the other hand, kurtosis value, crest factor, clearance factor and impulse factor are more adequate descriptors for incipient faults. They are related with the impulse in the signal.

Table 1. Table with the main descriptors used in the present work, where $x(n)$ is a signal series for $n = 1,2\ldots, N$ data points.

Mean	Standard deviation	Peak value
$X_m = \frac{\sum_{n=1}^{N} x(n)}{N}$	$X_{sd} = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - X_m)^2}{N-1}}$	$X_{pv} = \max\lvert x(n)\rvert$
Kurtosis	**Skewness**	**Impulse factor**
$X_k = \frac{\sum_{n=1}^{N}(x(n) - X_m)^4}{(N-1)X_{sd}^4}$	$X_{SK} = \frac{\sum_{n=1}^{N}(x(n) - X_m)^3}{(N-1)X_{sd}^3}$	$X_{Im} = \frac{X_{pv}}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$
Clearance factor	**Shape factor**	**Root mean square**
$X_{Clf} = \frac{X_{pv}}{\left(\frac{\sum_{n=1}^{N}\sqrt{\lvert x(n)\rvert}}{N}\right)^2}$	$X_{Sf} = \frac{X_{rms}}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$	$X_{rms} = \sqrt{\frac{\sum_{n=1}^{N}(x(n))^2}{N}}$
	Crest factor	
	$X_{Crf} = \frac{X_{pv}}{X_{rms}}$	

1.2.2 Dual level time synchronous averaging

Time Synchronous Averaging (TSA) is a signal pre-processing technique especially designed to extract periodic waveforms from noisy signals. It is especially adequate for rotational machinery, as the signals generated are periodical. In essence, and applied to shaft vibration analysis, the technique consists on averaging the vibration signal measured in the shaft for one rotating period according to the phase information, by calculating the mean

of waveform samples that have been measured for a number of rotations. The phase information is extracted from the tachometer readings using a zero-crossing technique, by determining the time instants where one rotation starts and another ends. To have an equal number of data points in each revolution, maintaining the same angular separation between each point, angular re-sampling techniques are used after acquisition.

In the case of current signature analysis, the TSA technique faces several difficulties. First, the vibration information is embedded in the line current signal, which is far more intense than the vibration signal itself. Besides, line and shaft frequencies are very close together. This, in addition to the motor slip, makes it difficult to separate the shaft signal from the line current. Additionally, the motor slip is not constant but depends on speed and/or load changes, which affects the ergodicity and stochasticity of the noise components.

The Dual-level Time Synchronous Averaging (DLTSA) method, proposed in [3.14], is conceived to overcome these problems, using a data analysis whose steps are schematized in Figure 4. It is explained more deeply in the Thesis of H.D. Ardakani [3.15]. Initially, the phasor for the first level TSA is generated using an adaptive band-pass filter centred at the line frequency. Then, a usual TSA is performed on the raw signal with respect to the line frequency phasor. The useful information appears in the residual signal extracted during the TSA process. In our case, the residual is obtained by removing the synchronous component of the original signal, which is a different approach from the usual residual definition [3.16],[3.17]. It is to be noted that the residual signal contains the gearbox related harmonics and stochastic signals. At the same time, the original time stamp is not discarded, as it is used to interpolate the residual signal back to its original time domain. The second level TSA is applied to the residual signal with respect to the tachometer signal phasor, obtaining the final signal containing the shaft related harmonics. The signature analysis of this $X(\theta)$ signal is then performed in both angle and order domains.

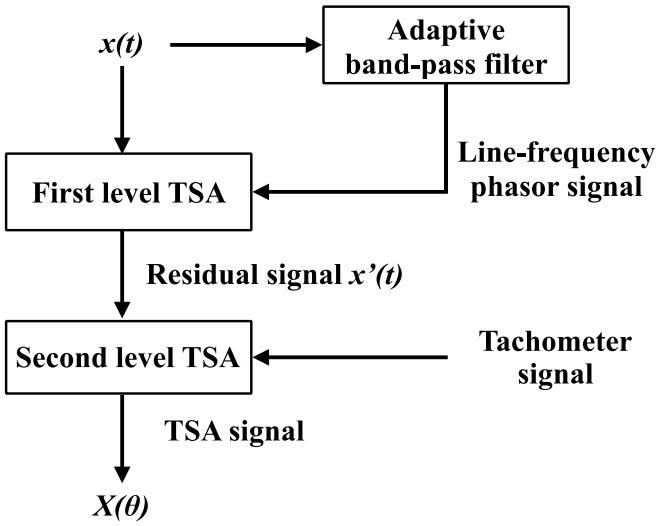


Figure 4. Scheme of the signal analysis using the DLTSA method [3.14].

Figure 5 graphically describes again the process, showing the result of the different steps of DLTSA on the current signal. Using the zero-crossings of the filtered current signal, the current signals are averaged. After replicating the averaged signals and constructing signals with the original length, tachometer signal is used to perform the final step of the averaging.

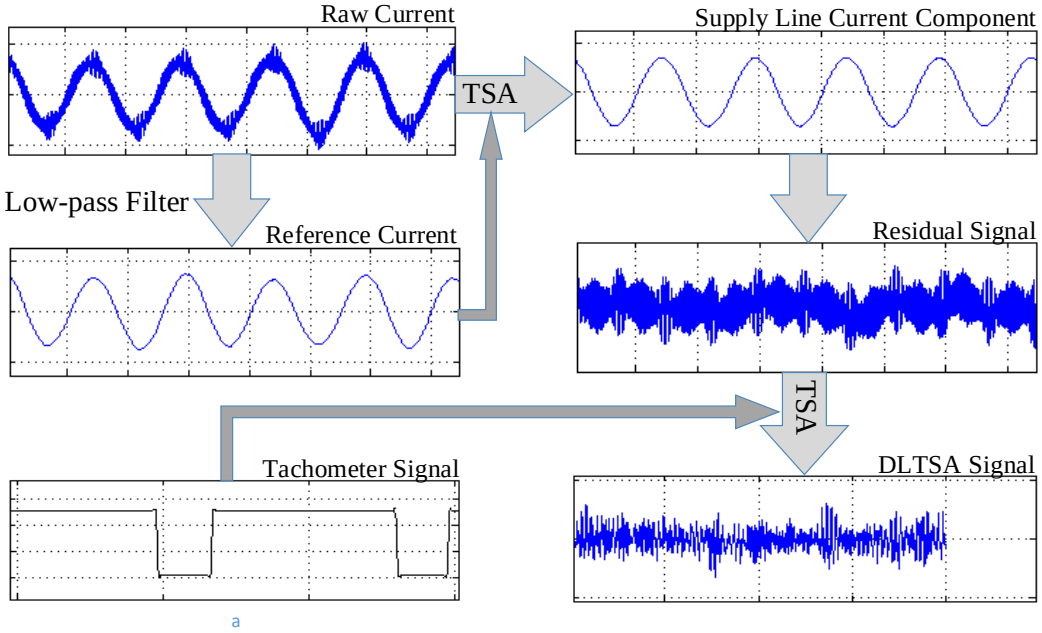


Figure 5. Motor current signal through the different steps of DLTSA.

After performing DLTSA, features in both angle and order domains are extracted from the signal data. The extracted features include standard deviation, kurtosis, peak-to-peak and crest factor in the angle domain, and magnitude of the signal at different orders in order domain. The same features were also extracted from classical residual and difference signals. Classical residual signal refers to the averaged signal having the

fundamental shaft frequency components removed from it. The classical difference signal refers to the averaged signal having the fundamental shaft frequency and the first order sidebands removed from it. From the whole set of extracted features, seven of them with the most discriminative power were selected for further steps of the process, using an analysis of variance (Fisher criteria), as described in section 2.3.2 afterwards.

1.2.3 Wavelets decomposition

As a pioneer attempt to overcome the limitations of windowed Fourier transform to analyse signals, *i.e.* fixed resolution in the time and frequency domains [3.18], Grossmann and Morlet proposed an analysis procedure based on the signal decomposition into a family of functions [3.19]. The evolution of this approach gave place to wavelet analysis, which consists basically on expressing the signal as a linear combination of a particular set of functions, obtained by shifting and dilating one single function called a mother wavelet [3.20].

The most prominent and convenient feature of the wavelet analysis is the variation of the time-frequency aspect ratio, giving good frequency localization at low frequencies, and good time location at high frequencies. It demands fewer processing resources than windowed Fourier transform for the analysis. Besides, the segmentation produced by this variation of the time-frequency aspect ratio makes the wavelet analysis particularly suited for transient signals.

In the continuous wavelet transform, the result is the sum over time of the signal multiplied by scaled and shifted versions of the wavelet. This process produces wavelet coefficients that are a function of scale and position [3.21]. However, an alternative approach is used commonly, denominated discrete wavelet transform, in which a discrete set of orthogonal wavelets are employed, associated to orthonormal bases of $L^2(\mathbb{R})$ [3.22]. In this base, the wavelet transform is performed only on a discrete grid of the parameters of dilation and translation. Within this framework, from a given mother wavelet $\psi(t)$, the base of orthonormal functions is constructed by dilation and translation as

$$\psi_n^m(t) = 2^{m/2}\psi(2^m t - n), \tag{3.8}$$

where *m* and *n* are the dilation and translation indices, respectively.

An arbitrary signal *x(t)* of finite energy can be written in this wavelet basis as

$$x(t) = \sum_m \sum_n d_n^m \, \psi_n^m(t), \tag{3.9}$$

where the coefficients of expansion are given by:

$$d_n^m = \int_{-\infty}^{\infty} x(t)\psi_n^m(t)dt. \tag{3.10}$$

Each dilation index m designates a particular wavelet level. The contribution of the original signal $x(t)$ to each level is obtaining by summing over the translation indices as

$$d_m(t) = \sum_n d_n^m \psi_n^m(t). \tag{3.11}$$

The usefulness of the wavelet analysis is that different wavelet decomposition levels provide information on the time behaviour of the original signal within different scale bands.

In order to implement a computationally efficient method to perform the wavelet decomposition, the so-called multi-resolution analysis was developed, which is the basis of the Fast Wavelet Transform algorithm used habitually [3.23]. Basically, the method constructs iteratively the wavelet functions using scaling functions and scaling coefficients a_n^m defined in such a way that the coefficients of expansion at a given decomposition level m, can be calculated from the ones at the previous level m-1:

$$a_n^m = \sum_l h[l - 2n] a_l^{m-1} \tag{3.12}$$

$$d_n^m = \sum_l g[l - 2n] a_l^{m-1} \tag{3.13}$$

where h and g are called low-pass and high-pass filters from the associated filter bank. These coefficients allow determining the different wavelet decomposition levels (see Figure 6)

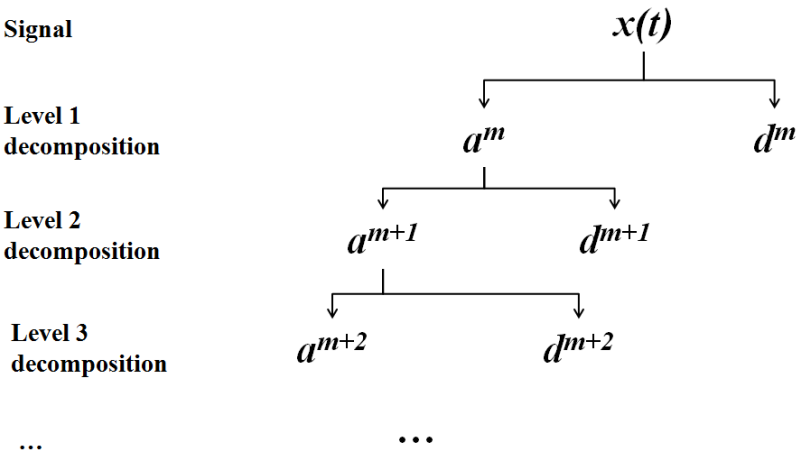


Figure 6. Discrete wavelet decomposition.

In Figure 7, the process of analysing the current signal by means of discrete wavelet decomposition is exemplified. Starting from the raw current signal, it is submitted to a high pass filter and to a low pass filter. The resulting signal from the high pass filtering is then obtained and time domain descriptors are generated from it. They are stored in a vector. At the same time, the resulting signal from the low pass filtering is submitted to low pass filtering and to high pass filtering. In an analogue way to the analysis made in the first stage of the wavelet decomposition, the descriptors obtained from this second level are saved in a vector and attached to the previously generated one, creating a result matrix. The process is repeated until the last of the levels is reached where the signal resulting from the low pass filtering is also obtained. At each of the levels that is gone down, the number of points of the generated signal is reduced to half of the one in the previous level. At the end of the process, a matrix with the descriptors linked with each of the decomposition levels is obtained.

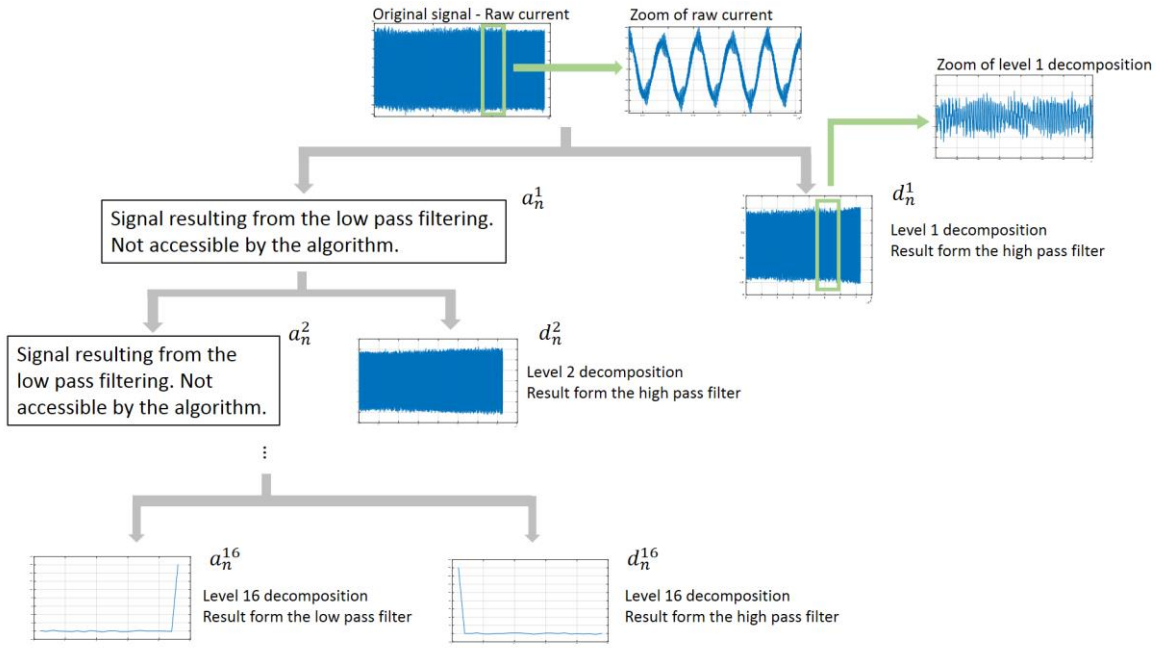


Figure 7. Discrete wavelet decomposition through different steps.

1.3 Feature selection

Not all the indicators will produce useful information, although each of them will be a better fit for different faults [3.24]. So, it is necessary to make a selection between the most suitable indicators and the rest. Feature selection or variable selection consists of reducing the available features to a set that is optimal or sub-optimal and capable of producing results which are equal or better to that of the original set [3.25]. Reducing the feature set scales down the dimensionality of the data which, in turn, reduces the training time of the induction algorithm selected and the computational cost, improves the accuracy of the final result, and makes the data mining results easier to understand and more applicable.

In this Thesis four feature selection methods were used: a *one-way analysis method*; the *correlation feature selection* method; *information gain* method; and the *relief* method.

1.3.1 ANalysis Of VAriance – ANOVA

The analysis of variance [3.26] is a collection of statistical models, and their associated procedures, that are used to analyse the differences among group means [3.27]. They were first developed by statistician Ronald Fisher [3.28]. Particularly one-way analysis of variance is a technique that can be used to compare means of two or more samples.

The *F*-test or *Fisher criteria*, in one-way analysis of variance, is the tool to assess how much the expected values of a quantitative variable within

several pre-defined groups differ from each other. The result is a *F number* associated with each of the descriptors.

The equation to calculate the one-way ANOVA F-test statistic is:

$$F = \frac{between - group\ variance}{within - group\ variability} = \frac{\sum_{i=1}^{K} n_i(\bar{Y}_{i\cdot} - \bar{Y})^2/(K-1)}{\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2/(N-K)} \tag{3.14}$$

where $\bar{Y}_{i\cdot}$denotes the sample mean in the i^{th} group, n_i is the number of observations in the i^{th} group, $\bar{Y}$ represents the overall mean of the data, $K$is the number of groups, Y_{ij} is the j^{th} observation in the i^{th} out of K groups and $N$is the overall sample size.

The F number will be large if the between-group variability is large relative to the within-group variability, i.e. the dispersion between the same descriptor relative to the dispersion between the descriptors of the rest of the health states. As a result, the bigger the F number for the studied descriptor is, the better it distinguishes between health states of the gear.

1.3.2 Correlation Feature Selection (CFS)

The correlation feature selection measure evaluates subsets of features under the hypothesis that good feature subsets contain features highly correlated with the classification, but uncorrelated to each other.

If the correlation between each of the components in a test and the outside variable is known, and the inter-correlation between each pair of components is given, then the correlation between a composite test consisting of the summed components and the outside variable can be predicted from [3.29]:

$$r_{zc} = \frac{k\overline{r_{zi}}}{\sqrt{k + k(k-1)\overline{r_{ii}}}} \tag{3.15}$$

where r_{zc} is the correlation between the summed components and the outside variable, k is the number of components, $\overline{r_{zi}}$ is the average of the correlations between the components and the outside variable, and $\overline{r_{ii}}$ is the average inter-correlation between components.

The above obtained correlation is known as the *merit*. The merit is used in the so-called *first search* algorithm. It starts calculating the merit of all the variables individually, selecting the biggest one. After this, it adds the variables one by one, selecting the variable that provides the biggest merit, and in case that the merit lowers that variable is discarded. In the case that it doesn´t find a bigger merit in five attempts, the algorithm starts again but

it selects another set of variables. The solution is the set of variables with the biggest merit.

1.3.3 Information Gain

Information gain algorithm compares the reduction in the entropy that is archived by adding a feature to the feature set ***F*** *[3.30]*. It is expressed by:

$$IG(\boldsymbol{F}) = H(S) - \sum_i \frac{S_i}{S} H(S_i) \tag{3.16}$$

where S is the data group, $H(S)$ is the entropy of the given dataset and $H(S_i)$ is the entropy of the i^{th} subset generated by partitioning S, based on all features in the joint set $\boldsymbol{F}$.

Usually a feature with high information gain should be ranked higher than other features because it has a stronger power in classifying the data.

It is necessary to explain the concept of entropy to fully understand this technique. In information theory, entropy measures the amount of information that is missing before reception. The entropy we will use to identify feature interactions is *Shannon Entropy*. For a data group S with n class labels, the Shannon Entropy is a measure of its unpredictability or impurity:

$$H(S) = -\sum_{i=0}^{n} p(i)\ log\ p(i) \tag{3.17}$$

where $p(i)$ is the probability of class i in the data group S.

1.3.4 Relief

The classical relief algorithm estimates the quality of attributes according to how well their values distinguish between instances that are near to each other [3.31], [3.32] and [3.33]. An instance is the set of the values of all the features that conform a label. In this work, a modification of the classical algorithm is implemented, where more than one hit/miss pairs are compared [3.34].

Given a randomly selected instance Ri, relief searches for its k nearest neighbors, from the same class, called nearest hits H, and the others from the different class, called nearest miss M. Here, k was set to 10. It updates the quality estimation $W[A]$ for all attributes A depending on their values for Ri, M, and H. If the instances Ri and H have different values of the attribute A, then the attribute A separates two instances within the same class, which is not desirable, so the quality estimation $W[A]$ is decreased. On the other hand, if instances Ri and M have different values of the attribute

A, then it separates two instances with different class values, which is desirable, so the quality estimation *W[A]* is increased. The whole process is repeated *m* times, where *m* is a user defined parameter. In our case the parameter was set to fulfill all of the labels, *i.e.* $m = 45$. In each repetition the starting instance in *Ri* is different.

1.4 Feature Classification

After the selection of the features, the obtained feature sets must be tested and classified according to their performance to discriminate the different types of faults. This is done using different algorithms denominated *classifiers*. The way of testing the performance of the feature sets is the following: from all the experimentally obtained data, a small part is used to train a classifier, that is, to generate a model, and then the trained classifier examines the rest of the data according to that model, trying to identify the fault to which the data corresponds. The test determines for each classifier and feature set the percentage of hit rate with its standard deviation.

Four different classifiers are used in this Thesis: bayesian network; sequential minimal optimization; lazy IBk instance based learning; and trees J48. The process of training the classifier and testing the model generated is performed for all of them by a technique denominated *cross-validation*.

It is important to highlight that the hit-rate results obtained by the classifiers cannot be directly used to decide which of the analysis methods, Wavelet or DLTSA, is better. This is because the feature sets selected from the Wavelets analysis and DLTSA are different. The number of elements used and their expected distribution must be considered to determine if the differences produced by the classifiers are statistically relevant of not. This study was performed with the help of the *Student's t-test*.

All the methods enumerated above are described briefly in the following paragraphs.

1.4.1 Cross-validation

Cross validation is a technique for assessing how the results of a statistical analysis will generalize to an independent data set [3.35]. It is used to estimate how accurately a predictive model will perform in practice.

In *k*-fold cross-validation process [3.36], the original sample is randomly partitioned into *k* equal size subsamples. A single subsample is retained as the validation data for testing the model, and the remaining *k*-1 subsamples are used as training data. The cross-validation process is repeated *k* times (each time is called a *fold*), using each of the different *k* subsamples as the

validation data. The *k* results from the folds can then be averaged (or otherwise combined) to produce a single estimation. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once.

For classification problems, one typically uses *stratified k*-fold cross-validation, in which the folds are selected so that each fold contains roughly the same proportions of class labels (in our case, the class labels are the different faults of the gears).

In *repeated* cross-validation, the cross-validation procedure is repeated *n* times, yielding *n* random partitions of the original sample. The *n* results are again averaged (or otherwise combined) to produce a single estimation.

In this Thesis we use ten fold cross-validation repeated ten times, because it has become a *de-facto* standard and is well-suited to the case studied.

1.4.2 Bayesian Network

A Bayesian network is a combination of two different mathematical areas: graph theory and probability theory. It is a representation of a joint probability distribution defined on a finite set of random variables that can be discrete or continuous.

The representation is a directed acyclic graph [3.37], [3.38], that is, one in which all of the edges in the graph are *directed* (*i.e.* they point in a particular direction) and there are no *cycles* (*i.e.* there is no way to start from any node and travel along a set of directed edges in the correct direction and arrive back at the starting node).

The graph is formed by nodes, usually discrete, and arcs. Nodes correspond to random variables, and arcs represent probabilistic dependencies between the variables. A Conditional Probability Table (CPT) is associated with each node and describes the dependency between the node and its parents. This CPT is converted into a conditional probability distribution (CPD) in case of continuous nodes. To derive the probability of a node, the probabilities of its parent nodes and conditional probability distribution functions on their connecting edges are computed [3.39].

Figure 8 shows an example of Bayesian network structure for a classification task, where x1; x2; x3; x4; x5 are the variables (nodes) of prediction and C the class to be predicted [3.40].

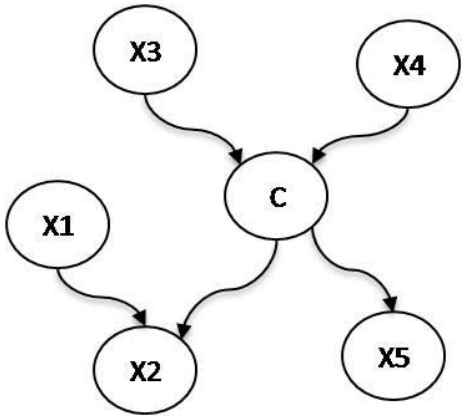


Figure 8. Example of Bayesian network structure for a classification task.

Even though the combination of graph and probabilistic theories is one of the key benefits in Bayesian networks, a wide application of this technology would have not been possible without the development of efficient inference methods to calculate the posterior probabilities, also called belief updating. Exact inference methods (such as enumeration and variable elimination algorithms) are feasible tools for low dimension unconstrained networks. However, computational complexity of probability inference becomes intractable in large and multiply connected networks [3.41]. Therefore, it has been of special relevance the development of approximate inference methods, where there is normally a trade-off between time and accuracy (normally linked to the *space* of solutions that can be sought). Different approaches are used today, such as direct sampling, loopy belief propagation and changeable methods, though stochastic simulation (*e.g.* probabilistic logic sampling, likelihood weighting, Gibbs sampling...) may be the preferred choice in most cases. These algorithms generate samples of data sets from random configurations of the existing Bayesian network, and estimate the posterior probabilities from the sampled configurations. Here, the issue is to construct a database with enough case samples as to have a valid distribution of probability over the variables, which follows the probability distribution specified in the CPTs (or CPDs). Jensen and Nielsen [3.42] provide a good introduction to these methods.

In the case considered in this Thesis, a simple Bayesian network classifier learning algorithm, implemented by Weka, has been used and applied in data set under study in order to classify the state of the component from the set of features extracted from signals. Some of the relevant characteristics of this algorithm are: K2 hill climbing algorithm has been selected as a search algorithm to search and learn the structure of the network; BAYES score metric has been utilized to measure and judge the quality of the structure; and finally, Simple Estimator method from data

has been used once the structure has been learning to estimate the conditional probability tables of the network.

1.4.3 Sequential minimal optimization

Sequential minimal optimization is a modification of the support vector machine technique [3.43], which in turn is a supervised learning algorithm. It uses supervised learning to generate a map. A map is a low-dimensional discretized representation of the input space, useful for visualizing low-dimensional views of high-dimensional data. Multi-class problems, as the one considered in this Thesis, are solved using pairwise classification (*i.e.* one versus another) [3.44].

The algorithm gets a set of training examples, labeled with the category to with they belong. With this information, it generates a model that, when new examples are added, assigns them to one of the categories. The categories are separated by a gap, which will be as wide as possible.

However, support vector machine techniques have inconveniences: they are slow and training algorithms are complex, subtle and sometimes difficult to implement [3.45]. The training of a support vector machine requires the solution of a very large quadratic programming optimization problem. In the sequential minimal optimization technique, the problem is broken into a series of the smallest possible problems, enabling to solve them analytically, saving time and computational resources.

1.4.4 Lazy IBk instance based learning

Lazy learning is a method in which the learning, beyond that done with the training data, is done after the system is questioned. The opposite would be *eager learning* where the system tries to generate a classification model before receiving new data. [3.46] , [3.47].

Instance based learning algorithms are derived from the nearest neighbor pattern classifier [3.48], [3.49]. In the nearest neighbor classifier, the object is assigned to the class most common among its nearest neighbors. It is a type of lazy learning where the function is only approximated locally, and all computation is deferred until classification. It works by assigning a weight to the contributions of the neighbors, being bigger the contributions of the nearer neighbors and lower as they get distant.

Unfortunately, it has also drawbacks: if the search for the nearest neighbors is done exhaustively, it is time consuming, and the storage of all instances can take a big space of memory. They also present problems to work with noisy datasets.

To overcome these drawbacks instance based learning algorithm was developed. It works with a partial dataset, which is used as training, and fits the new data to that dataset. In this way, a smaller space is taken inside the memory. The training dataset is selected so that there is no noise in it.

1.4.5 Trees J48

It is a type of decision tree. Decision trees are a type of decision support tool that uses a graph similar to a tree [3.50]. They can be represented as a compilation of *If* and *Then* rules.

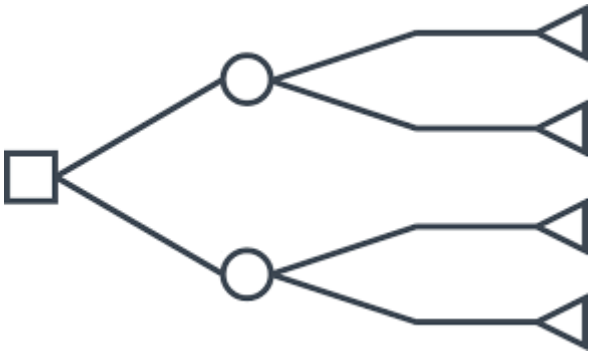


Figure 9. Example of a decision tree [3.51].

As can be seen in Figure 9, the tree starts from a single entity from which branches grow. The branches are generated in function of the values taken by one or various variables of the tree. Each of the nodes that is not a leave (the end of a branch, where no other branches grow from), represents a variable (in the case of the image they would be the circles). Each of the leaves is a value of classification (in our case the health state of a gear), represented by a triangle in the figure. Each of the branches that leave a node with an attribute represent the possible values of that attribute (in our case a variable).

Decision tree J48 is an open source Java implementation of the C4.5 algorithm in the Weka data mining tool. At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other [3.52]. The splitting criterion is the normalized information gain. Information gain measures the amount of information contained in a set of data [3.53]. It gives the idea of the importance of an attribute in a dataset. Thus, the attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurs on the smaller sub-lists.

1.4.6 Student's t-test

As explained before, the values from the classification of the Wavelets analysis and DLTSA cannot be compared as such. The number of elements in the study and their expected distribution must be considered to be able to

quantify if the difference in the hit-rate obtained by the classifiers is statistically significant or not. A technique that considers these parameters for the comparison is the Student's *t*-test. [3.54] The *t*-test is used to determine if two sets of data are significantly different. The comparison was made using R language [3.55].

In the case studied in this Thesis, the *unpaired t* method is the most appropriate. It is used to test if the population means related to two independent, random samples from an approximately normal distribution are equal.

The confidence interval and variances are calculated assuming unequal variances, to be later compared with the quantile function. The quantile function is the value at which the probability of a variable is less than or equal to the given probability.

Assuming unequal variances, the test statistic is calculated as:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} \tag{3.18}$$

where $\overline{x_1}$ and $\overline{x_2}$ are the sample means, s^2 is the sample variance, n_1 and n_2 are the sample sizes, t is a Student t quantile with $n_1 + n_2 - 2$ degrees of freedom. (Note that in this work n_1 and n_2 are 10, as the cross-validation was done with a ten fold).

2 Experimental

2.1 Test bench

The test stand used for this work is the gearbox prognostic simulator (GPS) from the Spectra Quest Company [3.56], which is displayed in Figure 10. It is designed to mimic the real working conditions of actual machinery while providing great versatility for implementing specific experiments with easy access to the mechanical and electrical parts. The original software and hardware were modified to fit better experimental needs. In particular, software and monitoring set-up were modified to include new sensors and the gears were changed for others made from harder steel.

2.1.1 General description

The test rig is composed by two confronted electrical motors, one acting as drive and the other as load, and two gearboxes, complemented with sensors and couplings. The drive and load motors have the same characteristics. Both are 10 Hp (7.35 kW), three-phase, induction motors having two pair of

poles. The monitored electrical signals are the ones feeding the driving motor. The two gearboxes are also quasi-identical, having three shafts and four gears as described in Figure 11. In our configuration, the first gearbox after the drive motor is the one under test, in which gears in different health states are inserted. The second gearbox works as a reducer and load motor protector in this set-up. The load motor protector has a bigger safety coefficient that prevents it to be damaged.

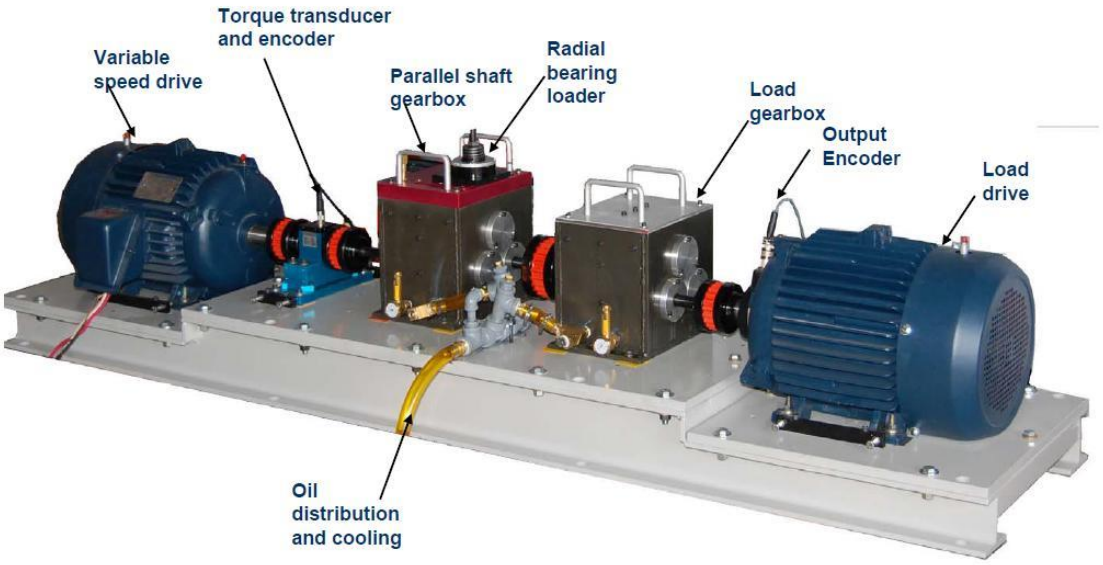


Figure 10. Detail of the Gear Prognostics Simulator test rig. (Figure from [3.57]).

The monitored gearbox is composed by four spur gears (Figure 11). They are built by *Martin company* [3.58]. The first gear, as it comes from the motor that drives the test bench, has 32 teeth. It is the one substituted by gears in different health state, leaving the rest un-changed. It is followed by a gear with 80 teeth. In the same axle, a gear with 48 teeth is found, connected to a gear with 64 teeth, resulting in a global transmission relationship of 3.33. Several tests were performed, and in experiments performed in a second stage, gears manufactured by *Juaristi Engranajes* [3.59] were used. These were built with F-155 steal, and underwent a cementation treatment.

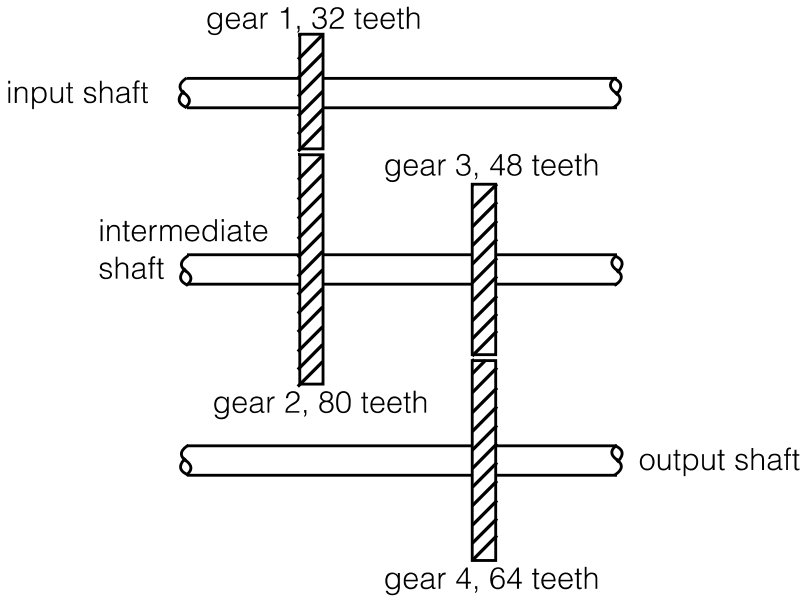


Figure 11. Scheme of the gear disposition inside the monitored gearbox.

2.1.2 Variator descriptor

The basic scheme of the electrical configuration of the test rig is displayed in Figure 12. The drive motor is fed by a power converter unit, connected to the grid and constituted by a rectifier and an inverter. The power converter delivers the current to the motor with the amplitude and frequency determined by the control unit that drives the motor at the desired speed [3.60],[3.61]. The GPS test rig use a scalar "voltage-frequency" open-loop control in which the magnetic flux in the motor is keep constant by using a value of the ratio amplitude to frequency which is fixed by design, and remains unchanged regardless of the load of the motor.

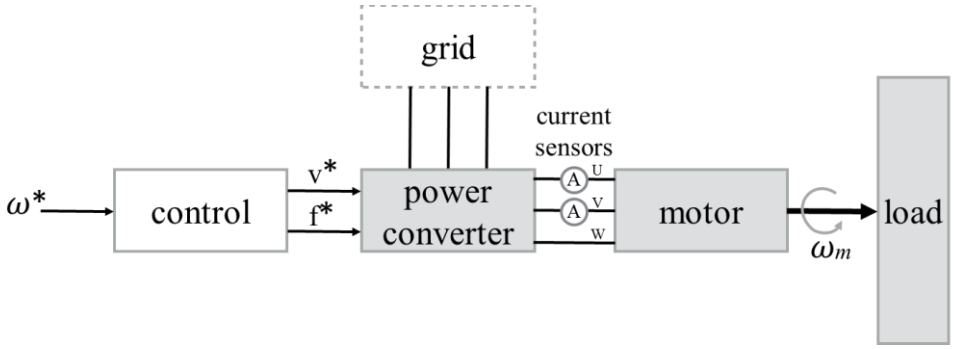


Figure 12. Elements of the power configuration of the experimental test-stand. ω^* represents the speed command for the motor from which the control unit produces the corresponding voltage V^* and frequency f^* for the power converter. ω_m is the real speed of the motor.

It is important to note that the inverter is constituted by thyristors that work at a fixed frequency, typically in the 3-6 kHz range. Thus, the effects of the load perturbations may only be seen in the current at frequencies up to the fixed frequency at which the inverter works.

2.1.3 Sensors

The test bench used in the present work has a computer to govern the data acquisition equipment and the control of the two motors. Several sensors are used in the test bench. Two *uni-axial piezoelectric accelerometers* (PCB piezotronics, model 608A11 [3.62]) are accommodated perpendicularly in the intermediate shaft of the monitored gearbox to measure transverse vibrations. An additional accelerometer is situated on the top of the gearbox. A *torque sensor* (Kistler, model 4502A100RA [3.63]) and a *speed encoder* (Scancon, model SCH68B [3.64]) in the input shaft, outside the gearbox, determine the mechanical input to the gearbox. A supplementary encoder is situated after the reducer gearbox in front of the load motor. Figure 13 shows the actual position of the sensors in the test rigs.

Table 2. Sensors installed in the GPS test bench, and their characteristics.

Sensor	Manufacturer / Model	Sensitivity / Resolution /Accuracy	Measurement range
Accelerometer	PCB 908A11	100 mV/g (10.2 mV/(m/s^2))	±50 g (±490 m/s^2) 0.5 to 10000 Hz
Torque sensor	KISTLER 4502A100RA	0.2 class*	100 Nm
Encoder	Scancon SCH68B	± 0.8 arc-min	1 to 10000 ppr
Current sensor	Lem HTA 100S	± 1 %	100 A rms

*0.2 class as specified by IEC designates that the highest permissible percentage current error shall not exceed 0.2% for current values from 100% to 120%.

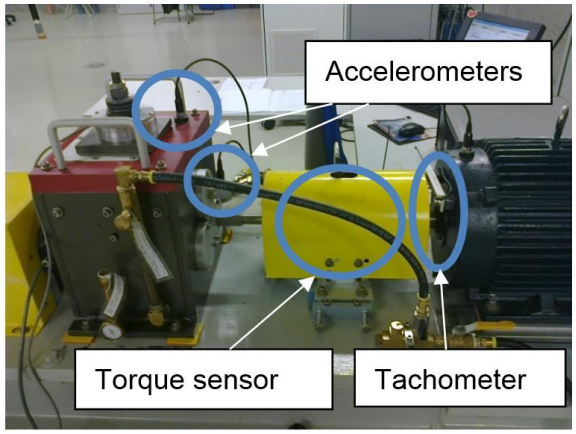


Figure 13. Position of the sensors on the test rig.

The *current sensors* LEM HTA 100 S [3.65] are used for the measurement of the feeding current. They were selected because of their optimum characteristics: good linearity, good accuracy, low temperature drift, wide frequency bandwidth, immunity to external interference and

wide dynamic range. Additionally, they are robust enough to be installed in the electric cabinet of the test rig, as can be seen in Figure 14. There are two installed, for the measurement of U and V electrical lines, but only the measurements from the U line are analysed in the present work. The studied motor is an equilibrated motor, so U and V are equal signals, with a difference in phase of 120º. Figure 15 shows a typical example of the raw current signal obtained from the LEM HTA transducer during operation.

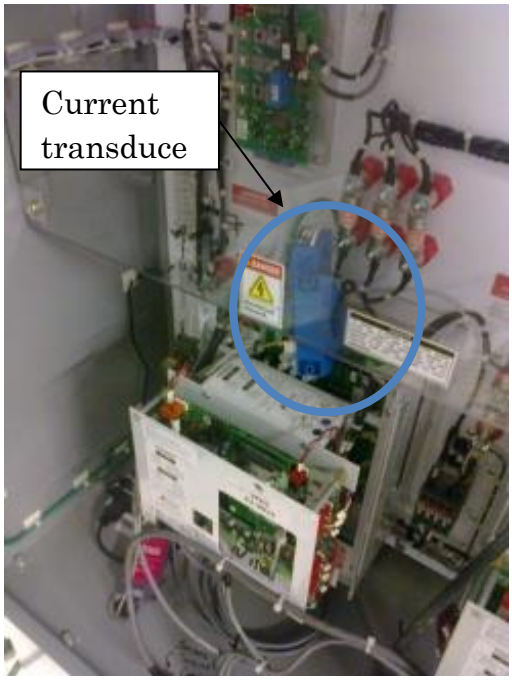


Figure 14. LEM HTA 100 S sensor installed in the test rig's electrical cabinet.

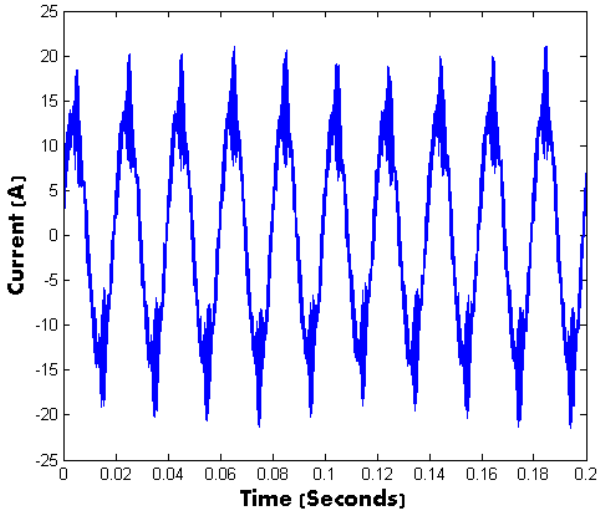


Figure 15. Motor current signal from the U channel, for the case of 1.500 rpm, and 0 % axle load.

2.1.4 Data acquisition

The data was recorded using a computer with a National Instruments acquisition card (NI 4472 series). It offers 24-bit resolution, with a maximum sampling rate of 102.4 kilosamples per second and up to 8 acquisition channels.

A data acquisition and control program has been developed using Labview from National Instruments [3.66]. It provides the necessary control over the data acquisition process requested by this work. In particular, the objective of the program was to synchronize the control of the test rig with the acquisition providing, at the same time, an automatic way of generating a comprehensive data pool. It is also important to underline that it was designed to make each of the repetitions statistically independent, by making the machine to stop after each test has been completed. Additionally, it allows some seconds for the signal to stabilize before starting to measure.

Essentially, the program works in the following way. First, an input test file with different lines is generated. In those lines, each of the tests is programmed. Different parameters describing the tests are recorded: the range of speeds, the running time, the time it takes to achieve the designed speed, the stabilization time allowed before the measure, the axle load, the acquisition frequency, and the number of repetitions. After one test line has been performed, the program saves the results, and passes to the next one, until the input test file is finished. The program flow is described in Figure 16. The data is recorded at periodic intervals, with a selected length. The organization of the measured data is explained in detail in section 2.2.3, organization of the data pool.

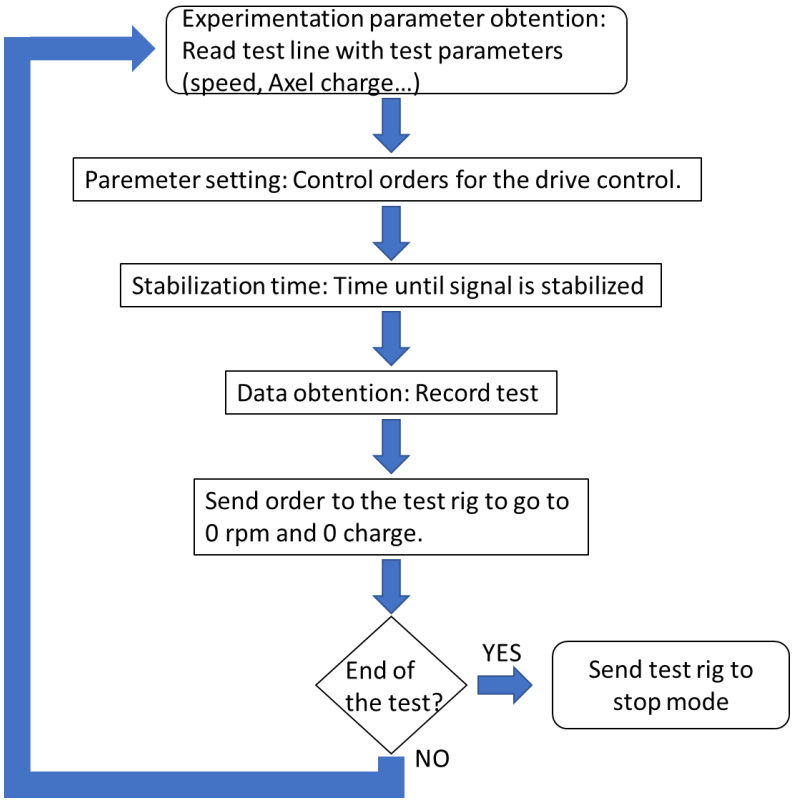


Figure 16. Diagram of the way that the data acquisition takes place.

2.2 Experimentation

One of the advantages of the test rig is the possibility of mimicking the working conditions or real machinery, such as spindles, electro-mechanical actuators and so on, which is done by selecting the most appropriate working conditions. A design of experiments is performed, fitting it to the requirements of the machine to be modelled to analyse the footprint of the different faults in the signals generated. In this Thesis, the focus has been on the spur gears, so faulty gears have been inserted in the monitored gearbox of the GPS. The faults could have been generated by degradation in the proper test rig, but it is hard to generate an isolated and controlled fault in the gear under study, and it would have been very costly in time, as safety coefficients and superior design in mechanical components make it difficult to produce degradation. In section 2.2.1 the design of experiments performed in the present work is described in depth.

On the other hand, a big data pool is generated during the tests, so a good file naming strategy and organization is needed to be able to process the data in an efficient way. The time invested in this will save later a lot of effort, as it will permit to analyse the data in an organized way.

2.2.1 Design of experiments

Condition monitoring is generally performed by making determined tests where the assets are put to work in known conditions, for example at constant speed. So, the performance of the test requires the asset to stop the production process and pass to a testing mode, sometimes referred as *fingerprint* [3.67]. The challenge of eliminating these tests is high since usually there are variables that are not under control and is hard to replicate the exact working conditions during the activity of the asset. For instance, in machine-tool it is hard to replicate loads conditions, as machined materials can be different each time.

The rationale behind this Thesis and the design of experiments is to permit the condition monitoring of gearboxes to avoid dedicated tests or, at least, to make the tests in the most repetitive conditions. For that reason, the focus is set in analysing speed transitions and especially in the no-load condition. The speed gain (acceleration) of the asset before operation is an action that is replicated in similar conditions, allowing the comparison between the signals generated each time. For example, looking into the machine-tool field, we can easily observe these speeding up phases, during which the tool is still not cutting material. Another application may be in the electro-mechanical actuators in aerospace. In this case, the tests can be performed in ground, without aerodynamic charge. Besides, the movements of the actuators are generally short enough so that transitory signal analysis techniques are more convenient. Both cases, machine tool and electro mechanical actuators, are just two examples of many others where the research performed in this Thesis can be applied.

The theory and application of MCSA have been mainly developed for the case of constant speed, while transients have only been considered in terms of load variations [3.68], [3.69]. In fact, the implementation of transient analysis in MCSA is still a big challenge. In this work, we purposely employ transients of speed to assess the health status of the gearbox using MCSA. The speed ramps are applied to the motor driving the gearbox in the no-load condition (as a particular case of constant load).

The maximum speed that the GPS test bench can reach is 1500 rpm. Different speeds were selected for the test to implement the transitory, within the limitations of the test bench. The minimum selected speed is 250 rpm. Two more intermediary speeds were selected, 500 and 1000 rpm. The variations were implemented among these speeds:

- 0-250 rpm
- 250 rpm - 500 rpm
- 500 rpm - 1000 rpm
- 1000 rpm - 1500 rpm

The time length of the transitory test was set differently in each test round. In the first round, each test was done with a length of 15 seconds in order to allow the slowest gear in the gearbox to be able to perform at least 10 revolutions. In the second round the time was set to 30 seconds.

Each test condition is repeated 15 times to enable statistical robustness. Each repetition is independent to the rest as, between two repetitions, the speed is set to zero before the following test is launched. All the tests were performed in the same speed and load conditions, thus eliminating the influence of these two parameters. This was true for both rounds of tests.

Note that, as the motor in the GPS test rig has two pair of poles, to convert to frequency, the speeds (in frequency) for feeding the motor must be multiplied by two, apart from being translated into hertz. The signal that the current sensor is going to measure displays then the following carrying frequencies:

- 250 rpm → 8.33 Hz
- 500 rpm → 16.66 Hz
- 1000 rpm → 33.33 Hz
- 1500 rpm → 50 Hz

Together with the carrying frequency, the characteristics of the faults are within the signal. But different gears with dissimilar faults cannot be distinguished at naked eye, as was made obvious in Figure 17.

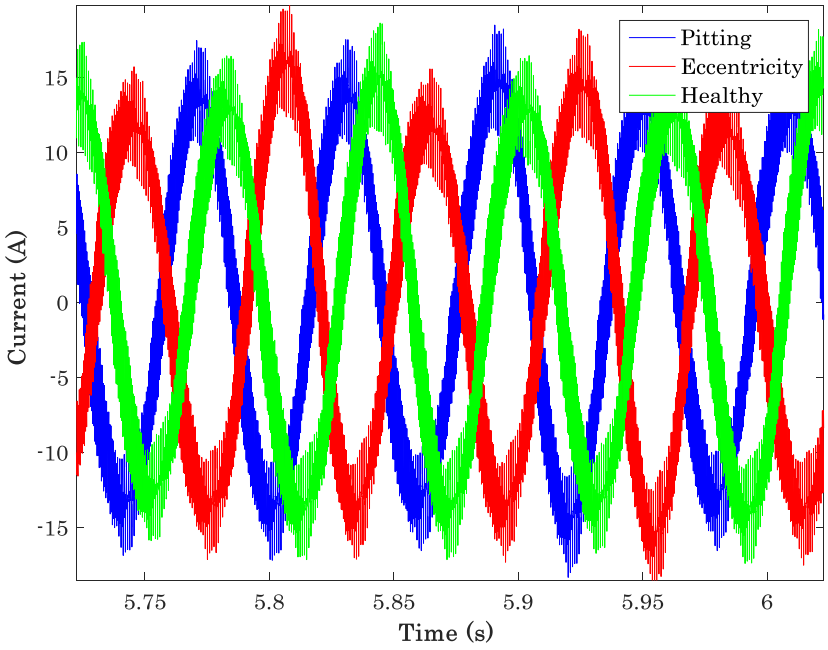


Figure 17. *U* channel raw signal, form gears with eccentricity in the bore, with pitting and in healthy state. The image corresponds to the 10th repetition of the 500 rpm constant speed signal.

2.2.2 Faults

The GPS stand is used to test gears in different fault conditions. During the experiments, only one gear is the subject of the test, the one inserted in the position of the gear 1 of the first gearbox (the 32 teeth gear in Figure 11). Gears from two manufacturers have been used in the tests: from *Martin Company* [3.58] and from *Juaristi Engranajes* [3.59]. The gears from Martin were made of steel (lacking more specification) without any heat treatment, while the gears from Juaristi were made from F-155 steel and suffered a cementation treatment. Martin gears were notably softer.

Two different rounds of tests were performed, using a separate set of faulty gears in each one. The first round was conceived as a proof of concept, to see how motor current signature analysis compares to vibration signal analysis, whereas in the second round the objective was to directly perform the diagnostic of the health state of the gears through MCSA.

For the first-round testing the faults were generated initially by machining, but the gears were afterwards intensively used in the GPS and, as a result, they were further degraded. At this state, a health assessment using the signal form the accelerometers was performed and three distinct fault categories were identified using the accelerometer data mapping: severe damage, moderate damage and little damage. There are two exceptions: the first one is the gear numbered 0001G, whose surface has been degraded (machined) on purpose, and the gear numbered 0006G which has an eccentricity, but no degraded surface. Each of the tested gears was assigned a unique code, which is revealed in Table 3. The code was engraved on the side of the gears by mechanical means. The turning sense was engraved as well. The fault categories were validated by the tribological laboratory in IK4-Tekniker [3.70].

Table 3. Code and health assessment of the gears used.

Gear number	Health assessment	Manufacturer
0001G	Degraded surface	Martin
0003G	Severe damage	Martin
0005G	Severe damage	Martin
0006G	Eccentricity	Martin
0007G	Severe damage	Martin
0010G	Severe damage	Martin
0011G	Little damage	Juaristi
0012G	Moderate damage	Martin
0013G	Moderate damage and little damage	Martin
0014G	Little damage	Martin

In a second-round testing new faulty gears were used. In particular, three gears were used in this part of the study: one in a healthy state manufactured by Juaristi, which is used as reference; one with pitting in the teeth manufactured by Martin; and one with eccentricity in the shaft hole from Juaristi. Figure 18 displays photographs of these gears highlighting their representative health features. The pitting of the second gear was generated from intensive use in the gearbox. The eccentricity was intentionally produced by machining the gear.

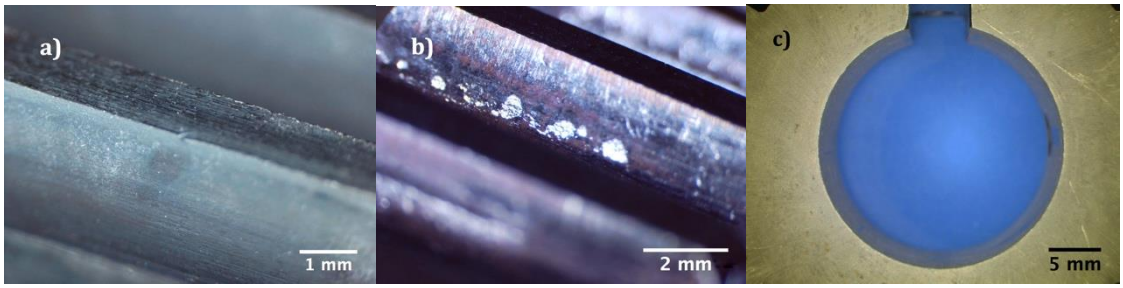


Figure 18. Images of the representative characteristics of the gears tested in the second round of the tests. (a) healthy gear. (b) gear with pitting. (c) gear with eccentricity.

2.2.3 Organization of the data pool

In order to store the tests in a clear way, avoiding undesired data mixing and helping to make the signal analysis more competent, a test codification was defined. The tests are saved in a directory tree, so they are classified and saved in a known, ordered way. The objective is to facilitate further analysis.

The directory tree and the filename codification provide the way to automate the signal analysis since an algorithm was developed to search the files in the directories and automatically perform the analysis to extract the selected features of the signals sequentially from each test.

The directory tree is organized in this way: All data is saved in a directory named *Test data*. Under this directory, the data corresponding to each gear under test is saved in a folder with the coded name given to each of the gears.

The name of each data set was uniquely selected so that it provides information on the test conditions. If the working conditions are the same for different test, they use the same filename, together with a counter growing with each of the tests performed. The codification of the data sets was done in the next way:

X_YR_AZPCT_REKN_Irpm_AHz,

where:

- X is the codification of the gear (for redundancy).
- Y represents the number of the repetition of the test condition and gear.
- Z is the axle charge in percentage.
- E is the radial load in kN.
- I is the speed in rpm.
- A refers to the acquisition frequency, in Hz.

For example, in the case of testing the G0005 gear, 1st repetition, with an axle load of 50 % at a speed transitory between 500 rpm and 1000 rpm, a radial load of 0 kN and an acquisition frequency of 50000 Hz, the filename would be

0005G_1R_A50PCT_R0KN_500-1000RPM_50000HZ.

2.3 Data reduction and analysis

For the analysis, the data set is divided by the variables present in the test. These variables are:

- Speed
- Axle charge
- Radial charge
- Gear

Each combination of these variables is denominated an *experimental condition*. That is, the experiment condition is represented by one unique set of variables: one speed, one axle charge, one radial charge and one gear. The results of the analysis are stored in a results file for each experiment condition. The different repetitions of the test under the same experimental conditions are saved in the same result file.

For the signal analysis, an algorithm was coded on Matlab. Two toolboxes were used, the Signal Processing Toolbox and the Wavelet Toolbox. The code worked in steps. According to the organization and structure of the data pool, the algorithm goes through the data tree opening the folder of each of the tested gears. There, following the name pattern explained in 2.2.3, the algorithm searches each of the data sets generated, opens them, perform the data analysis and save the result in an output file. The details are given in the next section.

2.3.1 Signal analysis implementation

To exemplify the process of analysing the data, only time domain analysis and the pre-processing using wavelet analysis are described. The DLTSA analysis follows a similar procedure. The code in this last case was implemented at University of Cincinnati's Intelligent Maintenance Systems Center.

Signal pre-processing and processing

The analysis is centred in obtaining a set of descriptors from the time domain analysis without any pre-processing and using the wavelet decomposition as a pre-processing technique and after obtaining statistical descriptors from each of the decomposition levels. Table 4 complies the descriptors obtained in the time domain analysis, and on the processing performed after decomposing the signal using wavelet, together with the Matlab function used to obtain each of them (the corresponding formulas were collected in Table 1).

Table 4. List of descriptors calculated in the analysis and Matlab functions used for obtaining them. *y* represents the analysed signal.

Descriptor	Matlab Function
Root mean square	Rms = sqrt (mean (y.^2))
Average	Avg = mean(y)
Peak value	Pv= (1/2) * (max(y)-min(y))
Crest factor	Crf= Pv/Rms
Skewness	Sk = skewness (y)
Kurtosis	ku = kurtosis (y)
Median	Median = median (y)
Minimum	Mini = min(y)
Maximum	Maxi = max(y)
Deviation	Devi = std(y)
Variance	Vari = var(y)
Clearance factor	Clf = Pv/(mean(sqrt(abs(y)))) ^2
Impulse factor	Imf = Pv/(mean(abs(y)))
Shape factor	Shf = Rms/(mean(abs(y)))

Beside these descriptors, in the case of the wavelet pre-processing the ratio of the absolute mean values of adjacent sub-bands, which represents the difference between one level and the next [3.71], is also calculated using the Matlab code

$$\text{ratio} = \text{abs}(\text{mean}(y1))/\text{abs}(\text{mean}(y2)),$$

where y1 is the wavelet decomposition of the level being analysed, and y2 is the wavelet signal from a superior level.

The flow of the pre-processing and processing code for the signals is described in Figure 19.

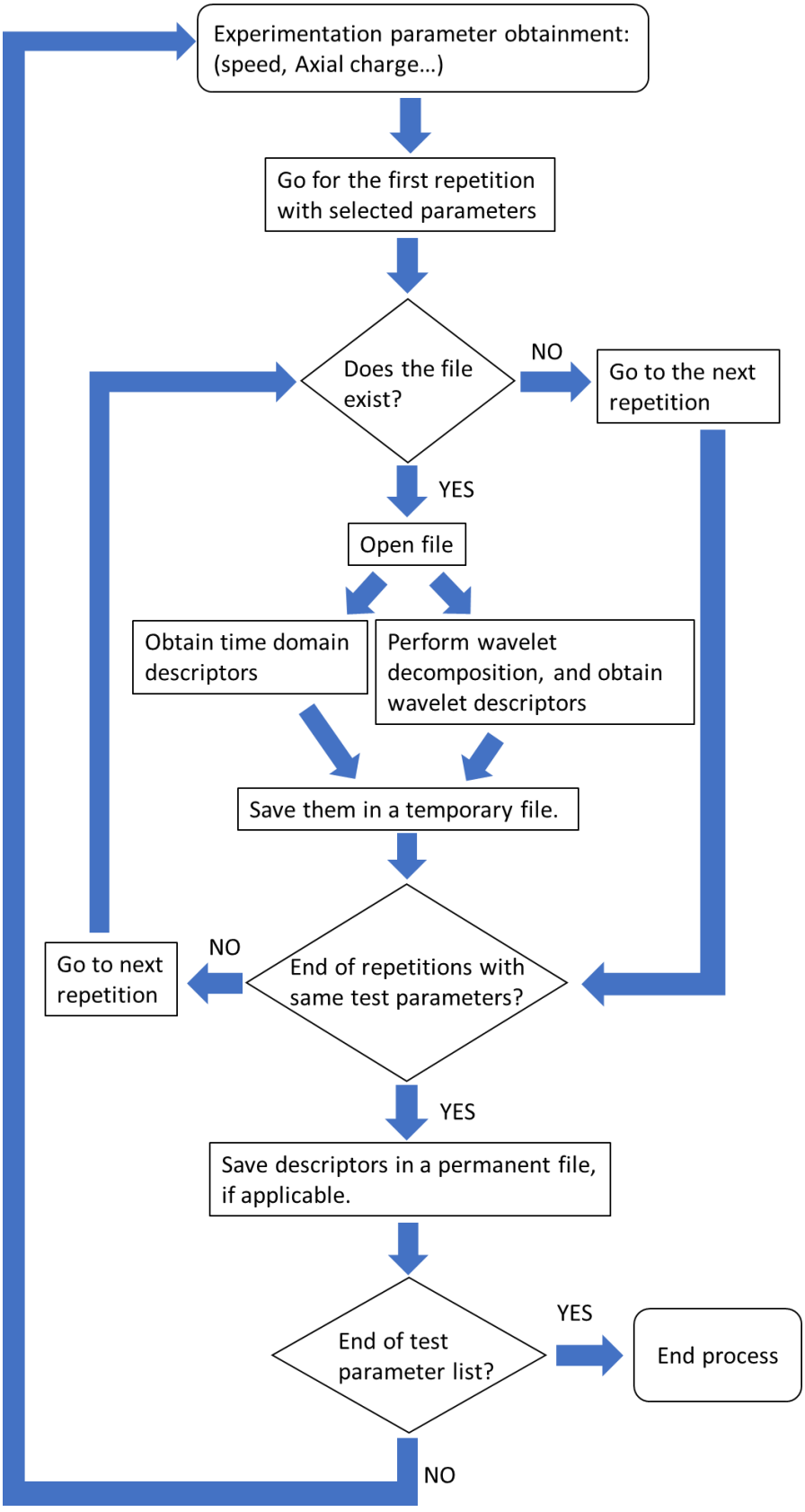


Figure 19. Signal wavelet analysis process flow.

It starts with the selection of the experiment condition (speed, charge, etc.) that are going to be analysed. The files containing all the repetitions corresponding to that experiment condition are sequentially scanned. The algorithm accounts for the possibility that the repetition numbers are not consecutive. The descriptors are evaluated first in the time domain, and then the wavelet decomposition is performed, obtaining the descriptors right after. The descriptors obtained in each decomposition level are saved in the temporary variable, which is sequentially appended with the results from the subsequent levels. When all the repetitions of the experimental condition have been analysed, the results are saved in a file. Each result file then corresponds to a unique experimental condition.

It is to be noted that in the first-round experimentation, a unique wavelet decomposition was performed, using always the same mother wavelet. In the second round, several decompositions were performed using different mother wavelets. In this second round, we made the decision of separating the result files corresponding to diverse types of mother wavelet, since this simplified the comparison of the results. Therefore, for each experimental condition in the second run, there are several results files, one for each wavelet decomposition.

2.3.2 Feature selection

After the analysis, the useful information among the huge data set generated must be selected. For this purpose, a feature selection analysis is carried out, as described in section 1.3. Four feature selection methods were used: a one-way analysis method, the correlation feature selection method, information gain method and the relief method. The analysis was performed using *Weka* data mining software [3.72]. Weka is a collection of machine learning algorithms for data mining tasks. It contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. One of the feature selection methods (one-way analysis) was made on Matlab software, as it is not provided as so in Weka.

The correlation feature selection method searchs for the variables that provide additional information. The result is a set of independent variables, but it doesn´t discriminate which one of them is the most or less useful. On the other hand, the result of the one-way analysis, information gain and relief is a score chart with a parameter calculated and related with each feature. The highest the number, the best that the feature works to distinguishing the health stage of the gears.

In the case of the one-way analysis, it was noticed that some of the features reflect bigger differences for one of the gears in comparison to the

other two gears. It was masking the results, as the features that differentiated the other two gears would not appear among the best ones. To solve this, the one-way analysis was relaunched, but in this case the gear that was most obviously differentiated was kept out. In this way, the features obtained from the first analysis will provide valuable information for the difference of one the gears, and the remaining two will be differentiated with the features obtained in the second stage analysis. The final features selected from the one-way analysis were a mixture from both analysis.

2.3.3 Feature classification

After the selection was done, the results were four different sets of features. To be able to quantify the adequateness of such sets, some feature classification algorithms were used, using Weka data mining [3.72]. The classifiers used in this Thesis are Bayesian network, sequential minimal optimization, lazy instance based learning and trees J48, as explained in section 1.4.

The classification models selected in this Thesis work all in a similar way: A subset of features is selected to generate the model, the so-called *training set*, and the rest of the data pool is classified, normally called the *validation dataset*. The classification is then compared with the labels on each of the instances and a hit rate is obtained.

The models that are used, at some point, have a heuristic step, *i.e.* one or more of the parameters are selected randomly. Thus, the result may differ from the same data set and model. Another factor to be taken in mind is that the data pool is limited (it has 45 instances). If the data pool would be validated conventionally (for example using 70% of the data set for training and the remaining 30% for validation), a significant validation or training capability would be lost. Therefore *cross-validation* is used in the process of training. It was explained in detail in section 1.4.1.

The results obtained from the cross-validation study were the mean of the classifications linked with the standard deviation. The values from the classification of the Wavelets analysis and DLTSA cannot be compared as so. The number of elements in the study and their expected distribution must be considered, to be able to quantify if the difference is significant or not. This was done with the help of the *Student's t-test*. It was done using R [3.55], as explained in detail in section 1.4.6.

3 Results and discussion

As explained above, the tests were performed in two different rounds. In the first one, the faulty gears were classified according to the results of the vibration signal analysis. This set of tests was conceived as a proof of concept, to assess the performance of motor current signature analysis compared to vibration signal analysis. In the second round of test, the objective was to obtain the diagnostic of the health state of three gears with different faults directly from the analysis of the motor current signature.

For conciseness, the results that will be presented here correspond to only one set of working conditions. They, for both round of tests, are no-load condition, and speed transitory from 1000 rpm to 1500 rpm.

3.1 First round testing

As the objective is to see how the motor current signature analysis compares with the analysis from the vibration signal, the reading from both sensors, hall (current) sensors and accelerometers were recorded from the same tests. The vibration analysis produced the fault classification that is resumed in Table 3.

The current sensor data processed comes from the *U* line of the drive motor. Time domain analysis and wavelet analysis was performed as described in section 2.3.1. The mother wavelet in this round of test was always a *daubechies 44* [3.73]. The 14 features compiled in Table 4, together with the ratio between successive levels were obtained in the original signal and in each level of the wavelet decomposition. As explained before, a one-way analysis of the variance was performed to reduce the number of levels and the number of features for further analysis.

The highest value of the *F* number obtained is plotted in Figure 20. It is observed that the analysis of the first decomposition (level 1) produces meaningful results, being levels 4 and 14-15 the next ones with high *F* numbers.

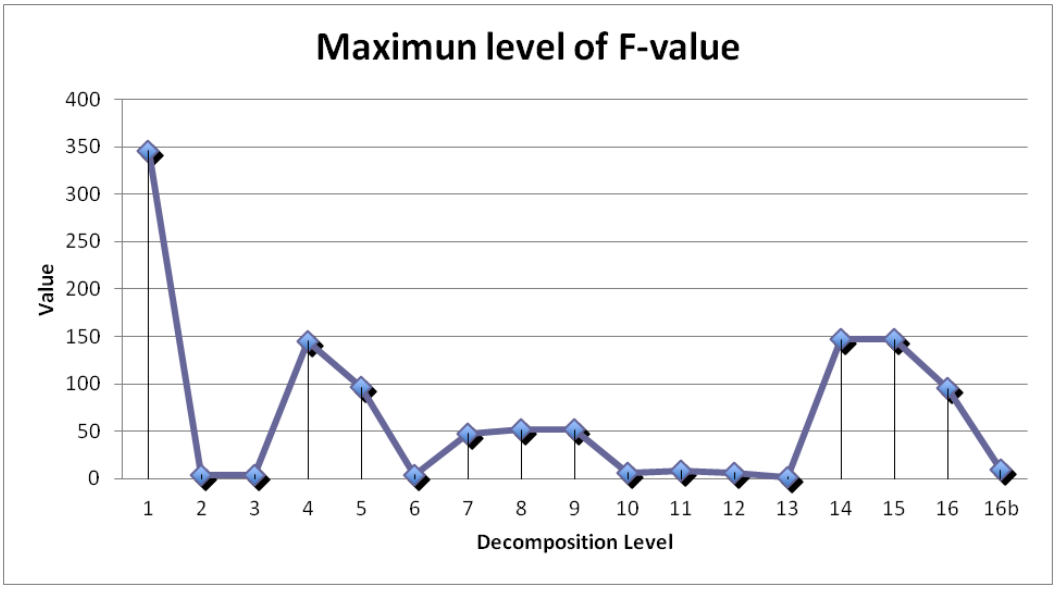


Figure 20. The maximum level of the F value in each of the decomposition levels.

The most interesting values of the variables in level 1 are the shape factor, the variance and the crest factor. For the case of level 4 decomposition, the skewness, the average and the ratio are the most interesting. In levels 14 and 15, the ratio, the clearance factor and the median provide the information.

From these pre-selected features, a further analysis allows to select the ones that provide the most accurate and useful information. For instance, in Figure 21, we can see that the values corresponding to the variance obtained in level 15 of the wavelet decomposition are grouped in two ranges for the different gears. The highly damaged gears appear in the upper range. The gears with lower values of variance include 0011G and 0014G which are the ones in good state. However, gear 0013G and 0006G are in the same area as the low damage gears. As compiled in Table 4, they are somehow peculiar. The gear 0013G has a number or repetitions classified as moderate damage and other as low damage. The gear 0006G has a machined fault but doesn´t show a big level of superficial pitting. It is somewhere between the low-level damage and the high-level damage. In Figure 22 compiling the results of the shape factor in decomposition level 1, the gear 0006G appears grouped with the high damage gears. However, in Figure 23, corresponding to the skewness in decomposition level 4 it is again classified among the low damaged gears.

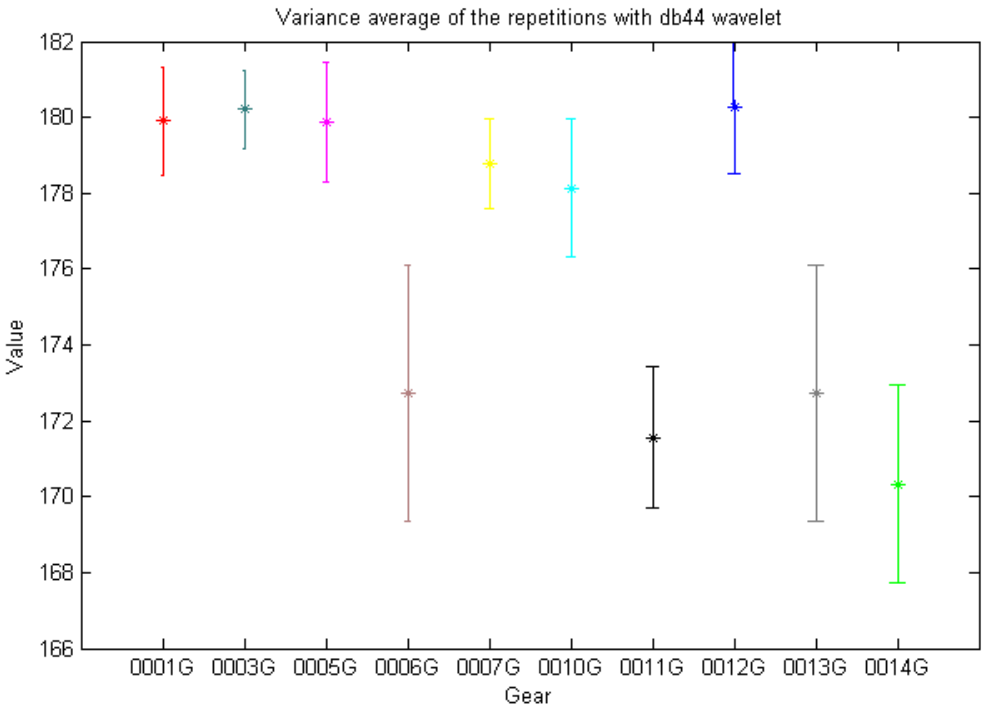


Figure 21. Variance of the signal obtained in the level 15 wavelet decomposition.

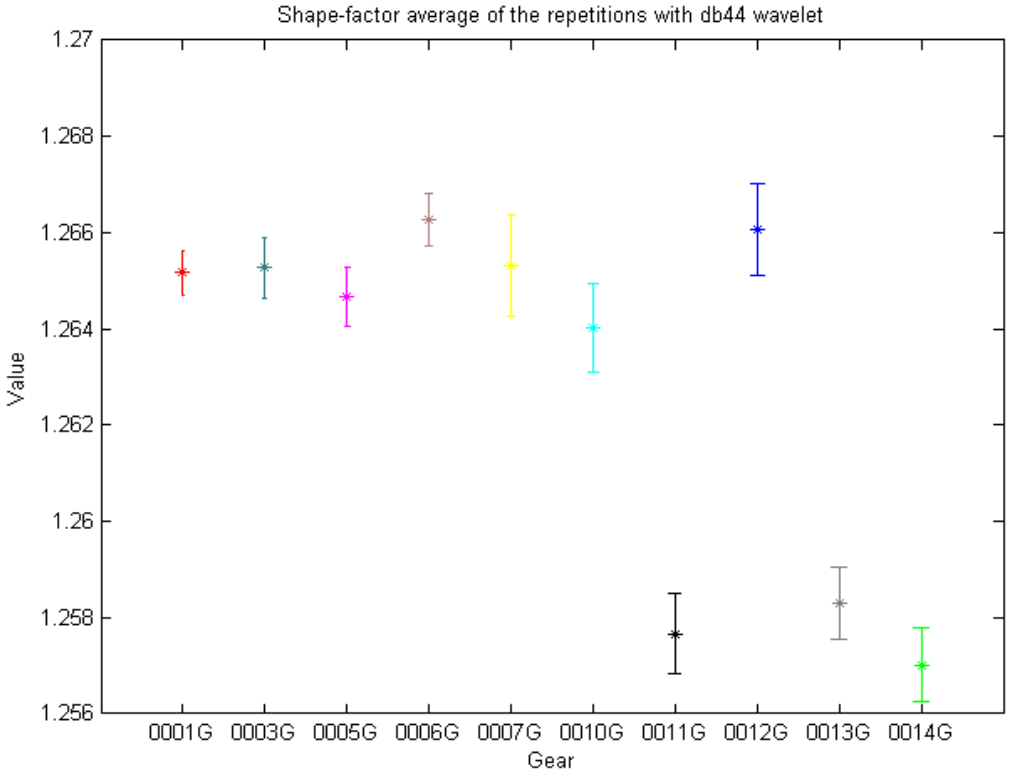


Figure 22. Shape factor of the signal obtained in the level 1 wavelet decomposition.

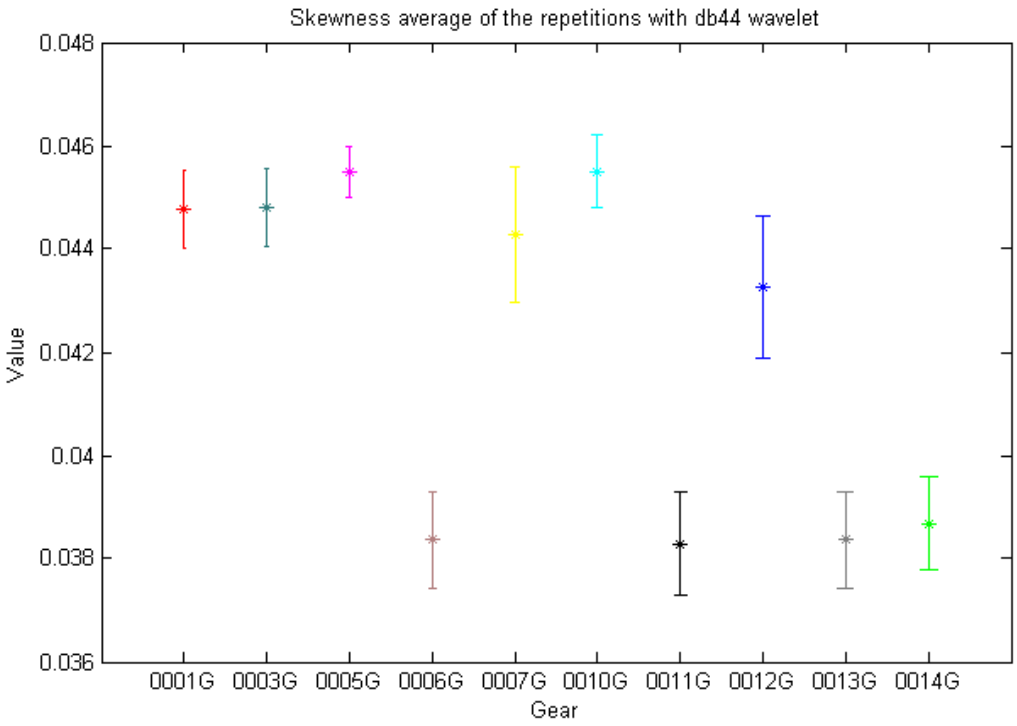


Figure 23. Skewness of the signal obtained in the level 4 wavelet decomposition.

3.2 Second round testing

In the second-round two faults were tested. The objective was to be able to detect the two different health states from the healthy one. Considering the results from the first-round testing, a new set of tests was performed, focusing on the different health states. The analysis was performed in two different ways. On the one hand, a wavelet analysis was performed. On the other hand, a DLTSA analysis was used. The results were compared using several classifications methods.

3.2.1 Wavelet analysis

The selection of the mother wavelet conditions greatly the results obtained. Habitually, it is selected according to a criterion of similarity between the shapes of the mother function and the fault signal, because in this way, the result of the convolution of both functions is reinforced. However, this strategy is not always a guarantee of success [3.73]. Our approach has been to test a wide range of mother wavelets from Haar, Daubechies, Symlets, Coiflets, BiorSplines, ReverseBior and DMeyer families summing a total of 106 different mother wavelets. In each case, 16 decomposition levels were obtained, and for each decomposition level, as explained in section 2.3.1, 14 parameters were calculated: rms, average, peak value, crest factor, skewness, kurtosis, median, minimum, maximum, deviation, variance, clearance factor, impulse factor, and shape factor. In total, 23744 descriptors were calculated for each one of the data sets.

To deal with such a big number of descriptors, a preliminary analysis was performed to disclose the ones that reveal greater differences between gears. Four different techniques were used for the selection of the most relevant features: analysis of variance, correlation feature selection, information gain and relief. They have been explained in section 1.3.

Analysis of Variance

An analysis of the variance through the *F*-number [3.26] was performed in two stages. First, using the descriptors for the three gears, a selection was made that is collected in Table 5. This analysis, however, resulted somehow biased, since the descriptors of the data from the gear with eccentricity appeared much more clearly distinguished than the ones from the other two gears. To better discriminate between healthy and pitting gears, a second analysis of the variance was performed, this time using only the descriptors of those two gears, with the results shown in Table 6. Top ranking descriptors from both tables were then used to distinguish the faulty gears.

Table 5. Data descriptors ranked by variance analysis, including the three gears.

Mother wavelet	Descriptor	Decomposition level	Band range (kHz)	Variance F value
DMeyer	Average	1	12.8-25.6	72.4
Daubechies 41	Average	1	12.8-25.6	63.7
Daubechies 42	Average	1	12.8-25.6	61.4
Daubechies 36	Average	1	12.8-25.6	50.8
Daubechies 43	Average	1	12.8-25.6	49.1
Symlets 24	Average	1	12.8-25.6	47.5
Symlets 22	Average	1	12.8-25.6	44.8
Daubechies 33	Average	1	12.8-25.6	43.8
Symlets 25	Average	1	12.8-25.6	43.8

Table 6. Data descriptors ranked by variance analysis for healthy and pitting gears.

Mother wavelet	Descriptor	Decomposition level	Band range (kHz)	Variance F value
Symlets 10	Impulse factor	15	0.78-1.56 $x10^{3}$	17.9
Symlets 10	Variance	16	0.39-0.78 $x10^{-3}$	17.8
Symlets 10	Median	4	1.6-3.2	18.7
Symlets 8	Impulse factor	15	0.78-1.56 $x10^{-3}$	17.3
Symlets 8	Variance	16	0.39-0.78 $x10^{-3}$	17.3
Symlets 8	Median	4	1.6-3.2	17.0
Daubechies 8	Median	7	0.2-0.4	8.0
Symlets 19	Kurtosis	8	0.1-0.2	7.0

Figure 24 displays a representative collection of the results obtained. The values of the selected descriptors for the 15 data sets obtained for each gear are compared. As expected, the data descriptors selected from Table 5, allow to distinguish the gear pitting, whereas, the descriptors from Table 6, separated the faulty gears from the healthy one. It can be observed that the value of each descriptor presents a large fluctuation along the data sets corresponding to the same gear. The differences between gears must be established within these fluctuations, and this can be done by visual inspection in Figure 24. In the two plots at the top, which correspond to data descriptors selected from Table 5, it is possible to distinguish clearly the gear with eccentricity, but the other two gears produce descriptor values that cannot be separated within the fluctuation of the values. (This was expected according to the way that these descriptors were selected). In the two plots at the bottom of Figure 24, the three types of gears can be visually distinguished, even with the great fluctuation of values, because of the different mean level of the descriptor for each gear.

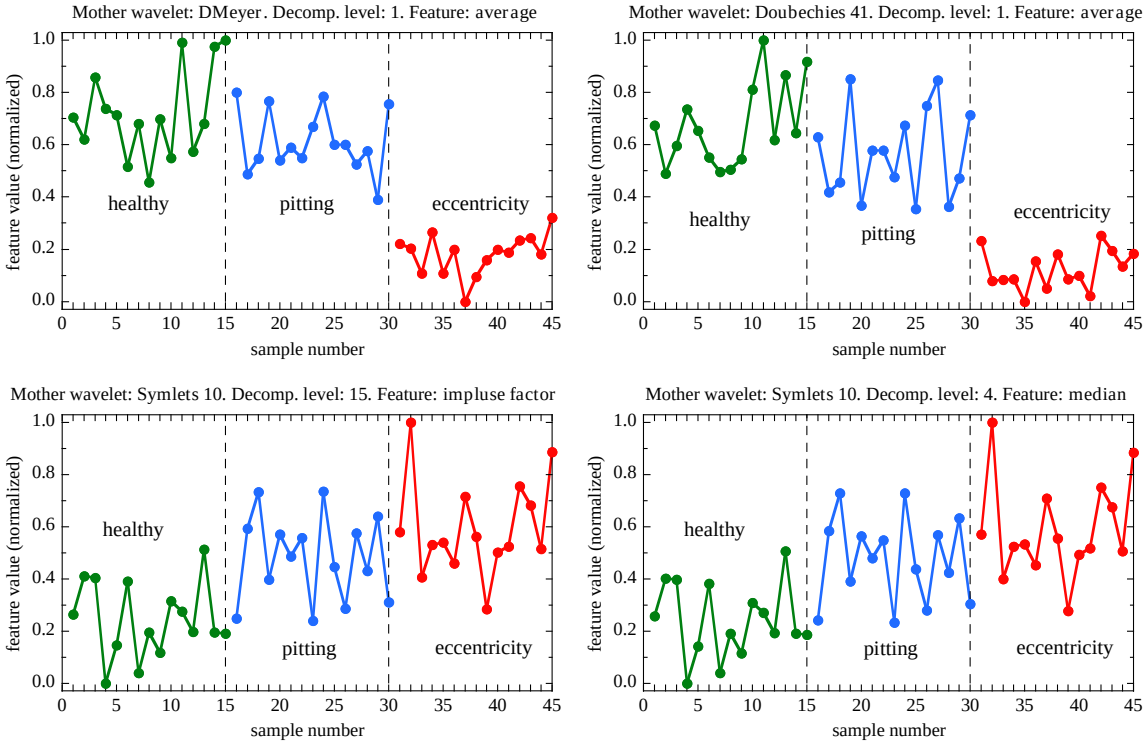


Figure 24. Selected results from wavelet analysis. For the sake of easier comparison between descriptors, their values have been normalized.

Correlation feature selection (CFS)

In this case the algorithm searches for the features that add information on the difference between the labels. It also discards features that are related, *i.e.* don´t add information. The method was described in section 1.3.2.

The results of the algorithm are a selection of features. It selected 62 features, although this method doesn´t rank them. The first 8 are represented in Table 7.

Table 7. Some of the features obtained from the correlation feature selection method (Wavelet analysis).

Mother wavelet	Descriptor	Decomposition level	Band range (kHz)
Raw signal	Impulse factor	-	-
ReverseBior 2	RMS	16	0.39-0.78 x10^{-3}
Daubechies 5	Median	4	1.6-3.2
DMeyer	Average	1	12.8-25.6
Daubechies 41	Average	1	12.8-25.6
Daubechies 42	Average	1	12.8-25.6
Daubechies 27	Median	7	0.2-0.4
Symlets 7	Kurtosis	8	0.1-0.2

Information Gain

Information gain algorithm compares the reduction in the entropy (related with information theory) that is achieved by adding a new feature to the feature set. The features that contribute to the reduction of entropy achieve a higher information gain number. The result is a ranked set of features. It is described in detail in section 1.3.3.

In this work, 62 features were used for the classification. The first 8 with the highest information gain number are presented in Table 8.

Table 8. Data descriptors ranked by information gain method, including the three gears (Wavelet analysis).

Mother wavelet	Descriptor	Decomposition level	Band range (kHz)	Inf. Gain value
DMeyer	Average	1	12.8-25.6	1
Symlets 18	Average	1	12.8-25.6	1
Daubechies 43	Average	1	12.8-25.6	1
Daubechies 42	Average	1	12.8-25.6	1
Symlets 24	Average	1	12.8-25.6	1
Daubechies 38	Average	1	12.8-25.6	0.875
Daubechies 45	Average	1	12.8-25.6	0.875
Daubechies 36	Average	1	12.8-25.6	0.875

Relief

The classical relief algorithm estimates the quality of attributes according to how well their values distinguish between instances that are near to each other. The features are ranked according to that quality estimation.

In this work, 62 features were used for the classification. The first 8 features ranked are presented in Table 9.

Table 9. Data descriptors ranked by relief method, including the three gears (Wavelet analysis).

Mother wavelet	Descriptor	Decomposition level	Band range (KHz)	Quality estimation
DMeyer	Average	1	12.8-25.6	0.243
Daubechies 41	Average	1	12.8-25.6	0.236
Raw signal	Crest factor	-	-	0.228
Raw signal	Impulse factor	-	-	0.225
Raw signal	Clearance factor	-	-	0.221
Raw signal	Kurtosis	-	-	0.216
Daubechies 42	Average	1	12.8-25.6	0.203
Daubechies 44	Average	1	12.8-25.6	0.203

3.2.2 DLTSA analysis

The DLTSA approach described in section 1.2.2 was applied to the same data sets analysed in section 3.2.1 by the wavelet method. First, to extract the supply line current component, a second-order Butterworth low-pass filter with cut-off frequency of 500 Hz was applied to the current signals. Note that this a simplified approach from the general scheme displayed in Figure 4, where a band-pass filter is used. In our case, the line frequency is limited to 50 Hz, whereas the gear fault related component frequency (such as amplitude modulation effect to Gear Mesh Frequency) is at much higher frequency range, above 800 Hz. Therefore, it is reasonable to use low-pass filter to the original signal to extract the supply line current component.

After performing DLTSA, features in both angle and order domains were extracted from signal data, as explained in section 1.2.2. The extracted features included standard deviation, kurtosis, peak-to-peak and crest factor in the angle domain, and magnitude of the signal at different orders in order domain. The same features were also extracted from classical residual and difference signals.

As explained in section 1.3, the selection of the most relevant features was performed by four different techniques: analysis of variance, correlation feature selection, information gain and relief.

Analysis of variance

Figure 25 shows the first four selected features with the Fisher criteria for different classes of the faults. As it can be observed in Figure 25, features extracted from signals during the presence of eccentricity fault are significantly different from healthy and pitting conditions. Although compared to eccentricity, healthy and pitting gear faults are not significantly different, their values are easily distinguishable. This trend is observable among all of the features selected from the different methods.

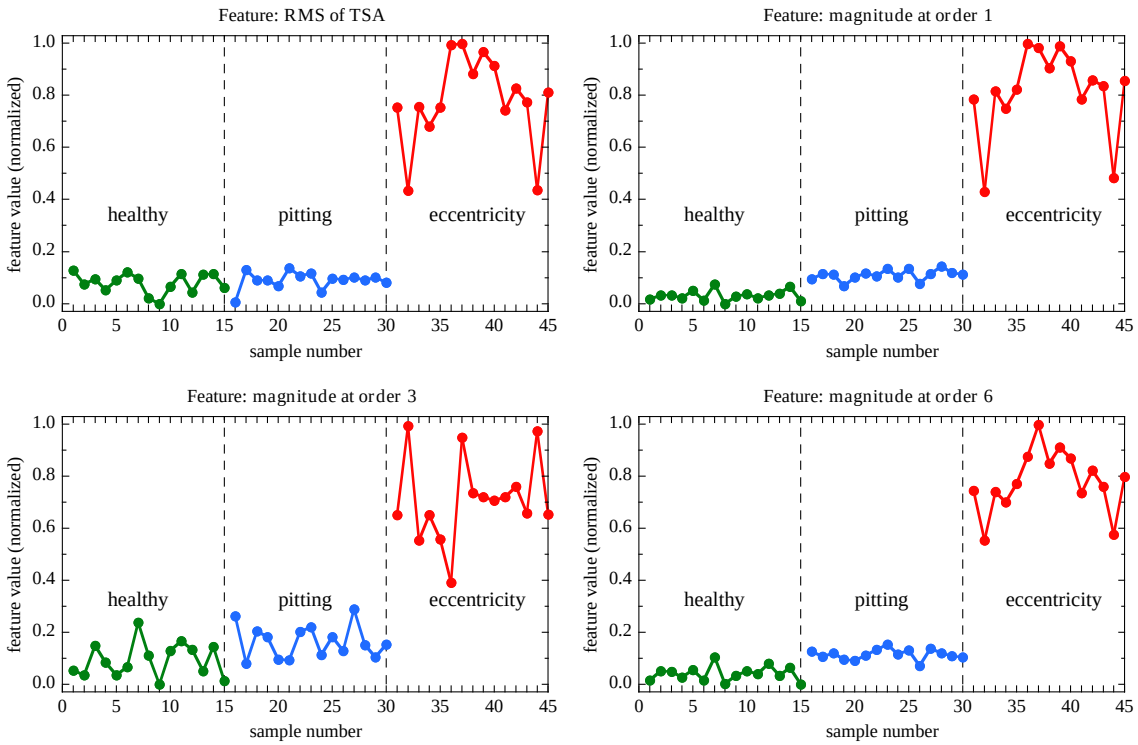


Figure 25. Selected results of the DLTSA analysis. For the sake of easier comparison between descriptors, their values have been normalized.

The features selected by all the selection techniques are compiled in Table 10, together with the index that rank them (note that CFS, don't provide a ranking index).

Table 10. Data descriptors obtained from the different selection methods from DLTSA analysis.

Fisher criteria	Ranking	CFS	Information gain	Ranking	Relief	Ranking
Sum of signal below Order 5	506,3	Kurtosis of residual	RMS of TSA	1	Max. of signal below Order 5	0.447
Order 1	298	Peak to peak of residual	Peak to peak of residual	1	Order 1	0.446
Max. of signal below Order 5	295.3	Crest factor of difference signal	Kurtosis of difference	1	Sum of signal below Order 5	0.406
RMS of residual	231.43	Order 1	RMS of residual	1	RMS of residual	0.38
RMS of TSA	229.1	Sum of signal below Order 5	RMS of difference	1	RMS of TSA	0.356
Order 3	146.1	Max. of signal below Order 5	Kurtosis of TSA	1	RMS of difference	0.322
Kurtosis of residual	123.7					

3.2.3 Wavelet and DLTSA analysis

Although we wish to determine which of the two analysis procedures (wavelet or DLTSA) yield better results, it is also interesting to examine the results that can be obtained by combining all the features obtained from both analysis methods. Therefore, both feature sets, obtained from the wavelet and DLTSA analysis, were added in a common data set, in which the feature selection and classification methods were applied.

The same four selection methods used with the separated feature sets were applied to the combined set. The results are compiled on Table 11.

Table 11. Data descriptors obtained from the different selection methods in the combined feature set from wavelet and DLTSA analysis.

Fisher criteria	Ranking	CFS	Information gain	Ranking	Relief	Ranking
Sum of signal below Order 5	506,3	RMS.of.TSA	RMS.of. TSA	1	Max. of signal below Order 5	0.457
Order 1	298	Kurtosis.of .TSA	Dmey Level 1 Average	1	Order 1	0.457
Max. of signal below Order 5	295.3	Raw signal Crestfactor	Kurosis of difference	1	Sum of signal below Order 5	0.433
RMS of residual	231.43	Raw signal Kurtosis	Db43 Level 1 Average	1	RMS of residual	0.418
RMS of TSA	229.1	Haar Level 11 Median	Sym24 Level 1 Average	1	RMS.of. TSA	0.395
Order 3	146.1	Haar Level 5 Ratio	Sym18 Level 1 Average	1	RMS of differenc e	0.368
Kurtosis of residual	123.7	db3 Level 4 Median	RMS of difference	1	Peak to peak of TSA	0.356
Crest factor of raw signal	72.7	db4 Level 4 Ratio	Db42 Level 1 Average	1	Order 3	0.306

3.2.4 Mapping of the results

To classify and allow an easy evaluation and interpretation of the parameters extracted from both types of analysis (wavelet and DLTSA), different classification algorithms (*classifiers*) were used; Bayesian network, J48 tree, intense based learning and sequential minimal optimization, as explained in section 1.4. Weka data mining software [3.72] was used for this purpose. The best features obtained from wavelet and DLTSA analysis have been previously selected using the analysis of variance, correlation feature selection, information gain and the relief algorithms, as explained in section 1.3. The features selected by each one of these methods constitute

the *feature vectors* that will be given as input to the classifiers. For the wavelet method, these vectors contain: 23 features from the analysis of variance, 62 features from the correlation feature selection, 62 features from information gain and 62 from relief. For the DLTSA method, the feature vectors consisted of 7 features from the analysis of variance, 23 from the correlation feature selection, 23 from the information gain method, and 23 from the relief methods.

Each feature vector represents one stream of experimental signals collected from the test-stand. Remind that there are 15 repetitions, or labels, from each class of fault, i.e. 15 labels from the healthy state, 15 labels from the eccentricity fault and 15 labels from the pitting fault. The classifiers were tested using the cross-validation method, which was explained in section 1.4.1. It divides the data set in 10 parts and uses one of them for testing and the rest for training. Each of the parts contains an approximately equal number of labels from each class. It does it 10 times, and the result is the mean and deviation of the testing.

Corresponding to the analysis based on wavelets, Table 12 compiles the results obtained. The classification accuracy is shown, along with the deviation linked with the feature selection and classification method used. The results are highly dependent on the feature selection algorithm and on the classifier used. But the overall results are quite high, having the lowest classification rate at 87.80 % and the highest at 100.00%

Table 12. Results from the cross validation of the features selected for the Wavelet analysis.

	Wavelet analysis							
	Classifier							
	Bayesian Network		**Sequential minimal optimization**		**Intense based learning**		**J48 tree**	
	μ (%)	**σ**	**μ(%)**	**σ**	**μ (%)**	**σ**	**μ (%)**	**σ**
Anova	85.05	14.49	88.65	15.60	90.85	13.01	72.95	19.68
CFS	100.00	0.00	97.25	8.05	86.90	13.78	90.30	12.98
Gain	88.00	16.13	90.20	14.30	90.05	12.72	90.80	13.35
Relief	87.80	16.10	92.35	13.38	95.50	10.58	89.30	14.32

The classification results for the case of the DLTSA are shown in Table 13. The results again show that there is a high dependence on the methods used for feature selection and classification. The lowest mean in this case corresponds to 59.60 %, and the highest to 94.75 %.

Table 13. Results from the cross validation of the features selected for the DLTSA analysis.

DLTSA analysis								
	Classifier							
	Bayesian Network		Sequential minimal optimization		Intense based learning		J48 tree	
	μ (%)	σ	μ(%)	σ	μ (%)	σ	μ (%)	σ
Anova	92.05	10.76	74.55	16.56	94.75	9.88	91.05	12.54
CFS	94.70	9.56	63.75	17.87	84.90	15.36	91.60	11.26
Gain	59.60	9.58	61.55	14.83	58.35	18.18	64.20	14.66
Relief	94.45	9.74	69.20	19.88	94.00	10.00	92.50	10.98

As explained before, we cannot say which analysis method (wavelet or DLTSA) is better based on the results complied in Tables 12 and 13. To determine if the differences in the results from each method are statistically relevant, a Student's *t*-test was performed, as explained in section 1.4.6. The values obtained from the tables are used to calculate the t value using equation 3.18. The critical value for t is searched in R, using the *qt* command, corresponding to the 95 % of certainty (the degrees of freedom $n_1 + n_2 - 2$ is, in this case, 10 + 10 - 2 = 22). This critical value is 2.100922. The differences in the results form tables 12 and 13 are statistically sound when the calculated t value (complied in Table 14) are larger (in absolute value, since a two-tailed distribution is used) than the critical value 2.100922.

As can be seen in Table 14, the results from the analysis of variance show significant differences only for the case of J48 classifier. In the case of the CFS selection method, sequential minimal optimization classifier is the only one to show significant differences. For the cases of the Gain feature selection method, all the classifiers show significant results. With the relief method the significant differences come from sequential minimal optimization classifier and the intense based learning classifier. The rest of the results don't show significant differences and can be denoted as being statistically equal.

Table 14. Results from the Student's *t*-test analysis.

Student's *t*-test analysis				
	Classifier			
	Bayesian Network	Sequential minimal optimization	Intense based learning	J48 tree
	t value	*t* value	*t* value	*t* value
Anova	1.226490	-1.959858	0.754938	2.452777
CFS	1.753145	5.405058	0.306491	-0.239240
Gain	5.107274	4.607240	2.651491	2.918252
Relief	-0.530402	2.544566	-2.367907	-1.078985

Finally, the cross-validation method was applied to the classifiers using the combined set of features obtained from wavelet and DLTSA analysis. The same feature selection methods and classifiers as the above cases were used. The results obtained showed that the lowest mean is 87.75 % and the highest is 100 % for this case, they can be seen in Table 15. These results are better than the ones obtained from each of the data sets separately (shown in Tables 12 and 13).

Table 15. Results from the cross validation of the features selected for the sum of Wavelet and DLTSA analysis.

Wavelet +DLTSA analysis								
	Classifier							
	Bayesian Network		Sequential minimal optimization		Intense based learning		J48 tree	
	μ (%)	σ	μ(%)	σ	μ (%)	σ	μ (%)	σ
Anova	96.05	8.54	92.50	12.34	85.25	14.27	81.80	15.71
CFS	100.00	0.00	98.25	6.64	95.70	8.99	87.15	13.62
Gain	95.35	9.14	92.30	12.58	90.25	12.72	87.15	12.84
Relief	87.80	16.10	93.70	11.21	98.40	5.90	88.00	11.98

In summary, we can conclude that, in general, motor current signature analysis is readily capable of experimentally of distinguishing between gears in different health states with a good percentage of accuracy. Some of the results appear to be better when using wavelet decomposition and others when using DLTSA, but the results from both types of analysis are strongly dependent on the feature selection and classification methods selected. The combination from both pre-processing techniques, wavelet and DLTSA, provides an improvement of the results.

4 Application of motor current signature analysis to bearing defect detection

Along with gears, bearings are one of the principal elements present in gearboxes. And, similarly to gears, it is important to monitor their health state. As the faults in bearings generate characteristic frequencies, the classical way of detecting them has been vibration analysis. Motor current signature analysis technique has also been used by some authors for the monitoring of such elements outside of the electric motor [3.74]-[3.75]. The transmission of the characteristic frequency of the fault to the stator current is identical to that explained in section 1.1: the fault frequencies affect the torque, which is related to the stator current.

In the present section, the experimental work performed in two different test benches to detect faults injected in bearings is explained. The motor current signal is analysed, and related with the health state of the bearing tested. One of the experiments was developed in collaboration with IK4-Ikerlan Technological Center. The second one was performed in collaboration with the German Aerospace Center (DLR).

4.1 Experimental

The experiments were carried out in two different test benches, one of them being the GPS test bench already explained in section 2.1. The second one is the Multi-Specimen Test Machine from Falex [3.76]. Unlike the work done with the gears, this time the signal analysed comes from stationary speeds, and the tests were performed under load conditions, that is, charges were applied to the gearbox during the experiments. Besides, the tests done in each of the machines differ in speed, charge and size of the tested specimen.

4.1.1 Experimental set 1

Test bench

The first series of experiments used the GPS test bench explained in section 2.1. In this case, the gears in good condition were used and remained unaltered during the tests. The only change that was performed in the gearbox was the bearing being tested.

The bearing under test is an ER16k bearing form the Rexnord manufacturer [3.77]. It is a radial ball bearing with a shaft of 1 inch in diameter. It is classified as *standard duty*. It has a seal and its own lubrication by grease. The dimensions of the monitored bearing are shown in Table 16, and an image is displayed in Figure 26.

Table 16. Specifications of the REXNORD ER16K bearing.

Parameter	Value
Number of balls, Z	9
Ball diameter, D_w	7.94 mm
Inner race diameter, d_i	31.38 mm
Outer race diameter, d_o	47.26 mm
Pitch diameter, d_m	39.32 mm
Race groove radius, r	4.1 mm
Material density, ρ	7750 kg m^{-3}
Elastic modulus, E	210 MPa
Poisson's ratio, v	0.25

Figure 26. Image of the bearing installed in the test bench.

The bearing was inserted in the intermediary shaft, in the position marked with a blue circle in Figure 27. Note that two accelerometers are installed in this shaft, in radial position, near the bearing although they were not used in the present analysis.

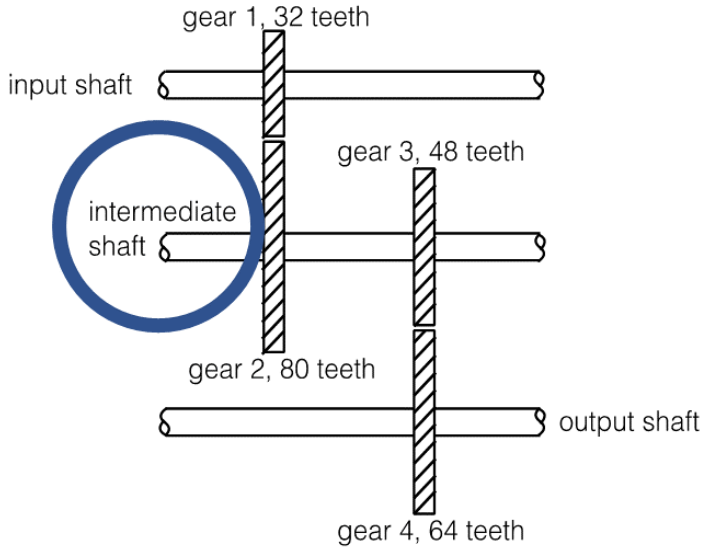


Figure 27. Position of the bearing under test in the gearbox.

Design of experiments

Four different levels of the speed of the drive motor were used: 250 rpm, 500 rpm, 1000 rpm, and 1500 rpm. Three different values of the axle load were applied by the second motor in the test bench: no load, 30 % of the load applied by the drive motor, and 60 %.

For statistical robustness, the duration of the tests was enough to allow at least sixty revolutions of the intermediate shaft in the gearbox. As a result of this condition and the speed used in the tests, the time span of the acquired data is 36, 18 and 10 seconds long, respectively. Each test condition is repeated 24 times. To make each repetition independent, the speed was brought to zero before launching the following test.

Faults

Four bearings in different conditions were tested in this study: a healthy bearing and three damaged ones. The damages have been seeded to the outer ring of the bearing by a drilling process from its outer surface to its raceway, taking care of not damaging any other component of the bearing. The seeded damages have different diameters: 0.6 mm, 1 mm and 2 mm, and have been denoted as F1, F3 and F5 respectively, whereas the healthy state is denoted as H. Figure 28 shows the tested bearing with the F3 damage.

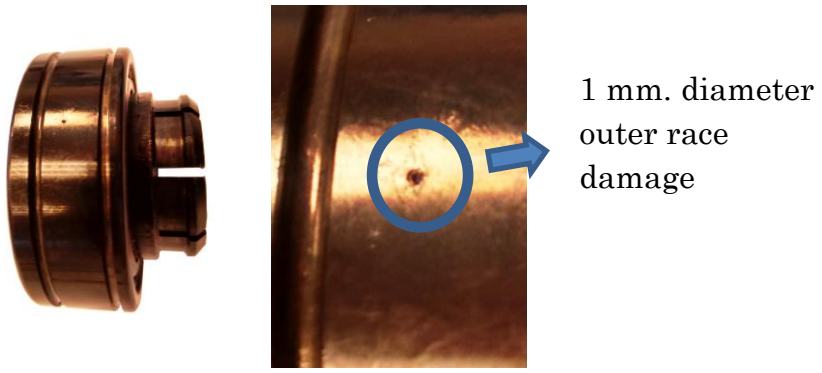


Figure 28. Bearing with F3 damage.

Results

The same approach used and explained in section 2.3 is employed for the case of this dataset. Only the wavelet analysis was performed, using the descriptors and mother wavelets described in section 1.2, and the feature selection and classification techniques used in sections 1.3 and 1.4. For this case the data was obtained at 50 kHz, and the wavelet decomposition was performed until the 16th level.

Figure 29 displays, superimposed, two different signals from a faulty and a healthy bearing. The small differences are barely appreciable at naked eye, but the signal analysis is performed to make these subtle differences comparable one with each other.

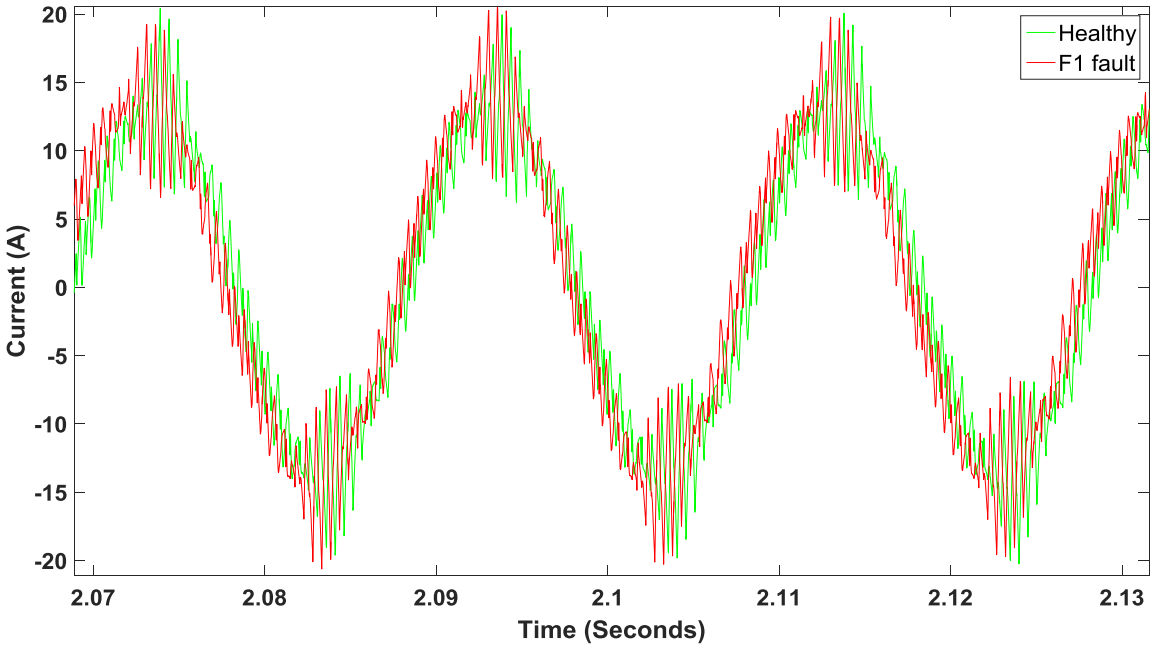


Figure 29. Healthy and faulty signal comparison.

In a first attempt, the analysis was performed using the same procedure used in the analysis of the gears. The results, displayed in Table 17, show that the obtained hit ratios range from 62 % to 96 % depending on the feature selection and classification method used. In general, the hit rates are smaller than the ones obtained in the case of the gear analysis.

Table 17. Results from the cross validation of the features selected from the experimental set 1 and the Wavelet analysis.

Experimental set 1 - Initial								
	Classifier							
	Bayesian Network		Sequential minimal optimization		Intense based learning		J48 tree	
	μ (%)	σ	μ(%)	σ	μ (%)	σ	μ (%)	σ
Anova	72.10	13.33	75.94	13.59	75.91	12.54	68.26	13.76
CFS	95.79	12.93	85.88	11.44	81.51	10.59	71.36	16.03
Gain	68.20	12.93	73.63	12.61	79.91	12.50	65.89	14.38
Relief	67.88	13.81	78.86	12.95	62.54	13.62	67.49	14.44

The low hit rate was understood by a detailed analysis of the feature sets that have been used, which made it clear that they were not appropriate. On the one hand, the features couldn't distinguish between the healthy bearing and the one with the smallest fault (F1). On the other, they couldn't differentiate between bearings with larger faults (F3 and F5), although clearly separated them from the healthy one and the bearing with F1 fault. In consequence, a new analysis was launched in which the original data set was separated in two, one for the healthy bearing and the one with the F1 fault, and the second for the bearings with F3 and F5 faults. In the second data set, new features were selected again using the analysis of variance, correlation feature selection, information gain, and relief methods, compiled in Table 18, Table 19, Table 20 and Table 21 respectively.

Table 18. Data descriptors ranked by the analysis of variance.

Mother wavelet	Descriptor	Decomposition level	Band range (Hz)	Variance F value
ReverseBior 2.2	Shape factor	9	48.83 - 97.66	225.3
Daubechies 2	Shape factor	9	48.83 - 97.66	134.7
Symlet 2	Shape factor	9	48.83 - 97.66	134.7
Coiflet 1	Shape factor	9	48.83 - 97.66	118.1
ReverseBior 2.4	Impulse factor	9	48.83 - 97.66	115.0
BiorSplines 2.4	Impulse factor	9	48.83 - 97.66	93.0
ReverseBior 4.4	Impulse factor	9	48.83 - 97.66	89.0
BiorSplines 2.2	Shape factor	9	48.83 - 97.66	87.8

Table 19. Some of the features obtained from the correlation feature selection method.

Mother wavelet	Descriptor	Decomposition level	Band range (Hz)
ReverseBior 2.2	Shape factor	9	48.83 - 97.66
Daubechies 2	Shape factor	9	48.83 - 97.66
Symlet 2	Shape factor	9	48.83 - 97.66
ReverseBior 2.4	Impulse factor	9	48.83 - 97.66
BiorSplines 2.4	Impulse factor	9	48.83 - 97.66
ReverseBior 4.4	Impulse factor	9	48.83 - 97.66
BiorSplines 2.2	Shape factor	9	48.83 - 97.66
Daubechies 8	Peak value	10	24.41 - 48.83

Table 20. Data descriptors ranked by information gain method.

Mother wavelet	Descriptor	Decomposition level	Band range (Hz)	Inf. Gain value
ReverseBior 2.2	Shape factor	9	48.83-97.66	0.75
Symlet 2	Shape factor	9	48.83-97.66	0.622
Daubechies 2	Shape factor	9	48.83-97.66	0.622
Coiflet 1	Shape factor	9	48.83-97.66	0.507
ReverseBior 2.4	Impulse factor	9	48.83-97.66	0.445
Daubechies 8	Peak value	10	24.41-48.83	0.425
BiorSplines 2.4	Shape factor	9	48.83-97.66	0.413
Daubechies 15	Peak value	1	12.5×10^3–25×10^3	0.411

Table 21. Data descriptors ranked by relief method

Mother wavelet	Descriptor	Decomposition level	Band range (Hz)	Quality estimation
ReverseBior 2.2	Shape factor	9	48.83 - 97.66	0.244
Symlet 2	Shape factor	9	48.83 - 97.66	0.198
Daubechies 2	Shape factor	9	48.83 - 97.66	0.198
Coiflet 1	Shape factor	9	48.83 - 97.66	0.155
Daubechies 8	Minimum	8	97.66- 195.31	0.14
Daubechies 25	Minimum	8	97.66- 195.31	0.137
Daubechies 28	Minimum	8	97.66 - 95.31	0.136
BiorSplines 2.4	Impulse factor	9	48.83- 97.66	0.135

Using the new selected features much-improved hit rates were obtained from the classifiers as evidenced in Table 22. The hit rates were comprised between 81.82 % as the lowest, and 99.79 % as the highest, improving significantly the ones obtained before.

Table 22. Results from the cross validation of the grouped features selected from the Wavelet analysis.

Experimental set 1 - Grouped								
	Classifier							
	Bayesian Network		Sequential minimal optimization		Intense based learning		J48 tree	
	μ (%)	σ	μ(%)	σ	μ (%)	σ	μ (%)	σ
Anova	88.98	9.53	98.04	4.49	94.01	7.66	93.32	7.47
CFS	99.79	1.49	96.10	5.65	92.47	8.68	90.02	9.11
Gain	87.40	10.21	95.42	6.27	87.48	10.37	91.58	7.10
Relief	81.82	14.03	95.34	6.01	88.92	9.79	91.97	8.43

4.1.2 Experimental set 2

Test bench

The multi-specimen test machine used in the second series of experiments was built by the Falex Company [3.76] (Figure 30). It was designed for several types of analysis, including bearing tests, and offers the possibility of producing different axle forces (up until 10.000 N) and rotation speeds (up to 10.000 rpm). It is capable of isolating the effects of the defect in the bearing from other mechanical defects.

Figure 30. General view of FALEX Multi Specimen Test Machine.

The test rig, illustrated in Figure 31, is composed by a motor that provides de movement, a transmission belt to transmit the movement from the motor to the test probe, and the test probe itself that include the bearing in the middle.

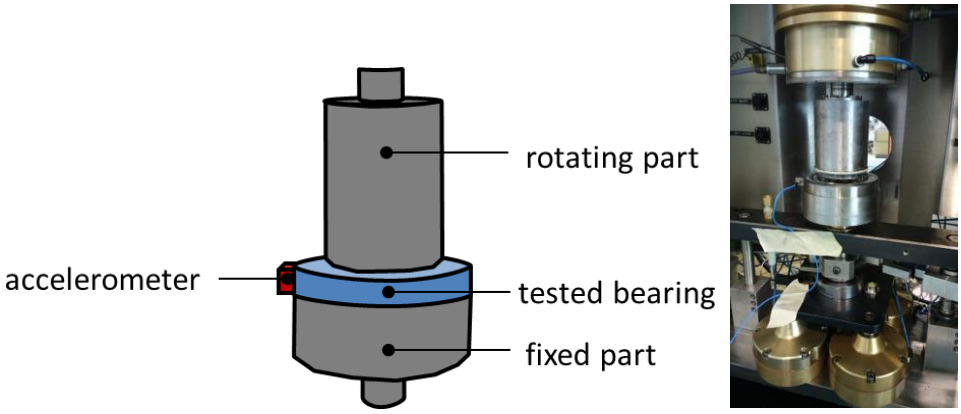


Figure 31. Setting to test the bearing in the FALEX test rig [3.78].

For the monitorization, a computer was used with an acquisition CDAQ-9189 chassis. For the control and the data acquisition, a program with Labview from National Instruments was developed, providing the synchronization between the control and the acquisition. The sampling rate used for the measurements is 25600 Hz.

Several sensors were installed. Vibration was measured by a 356A32 PCB PIEZOTRONICS accelerometer. The current consumed by the motors was monitored using two LEM HTA 100 sensors (as the system is equilibrated, a third one is not needed). A thermocouple was used for measuring the temperature in the upper surface of the bearing. A Dynapar encoder (model H220300022509) was used for measuring the speed. The axial load was measured with a load cell, which can measure up to 10.000 N.

In this case, the bearing under test was a FAG QJ212TVP (Figure 32). It is an angular contact ball bearing (35° of contact angle), featuring a single row, open, with a polyamide/nylon cage and normal clearance. The size of the bearing is 60 mm of internal diameter, 110 mm of outer diameter and 22 mm wide. The nominal maximum rotational speed is 6300 rpm. The maximum loads that can be applied are 94300 N in static mode and 96080 N in dynamic force.

Figure 32. The bearing from FAG opened.

Design of experiments

Two different speed levels were tested, 60 rpm and 600 rpm, together with two different axle charge levels, 5000 N and 8835 N. Each test was repeated three times.

Faults

In this experiment, the faults consisted in pits created in the inner and outer races of the bearing. Overall, four different health states were considered: A healthy bearing, a bearing with an outer race fault, a bearing with inner race fault and a bearing with both inner and outer faults. Both inner and outer race pits were ellipses machined in the surface using electrical discharge, a technique providing good machining precision (see Figure 33). The depth of the faults was 0.40 mm and the large diameter of the ellipse was for both 6.80 mm. The small diameter of the inner fault was 4.00 mm, while the one of the outer fault was 4.40 mm.

Figure 33. Detail of the inner race fault.

Results and discussion

Just three repetitions were made of each health state. But experience showed that they were not enough for a good analysis of the signals generated. Thus, each repetition was sliced in three, generating three times more signals, thirty in this case. At least 10 turns of the bearing were contained in each of the repetition slices. Although this partially compromises the independence of each of the repetitions, the benefits of having a bigger data pool, outweigh the drawbacks.

The descriptors (features) and mother wavelets used for the analysis are the same as the ones explained in section 1.2. The rest of the analysis is the one explained in section 2.3. The wavelets decomposition in this case is performed until the 8th level, as the acquisition frequency is 25600 Hz. The descriptors generated, were selected and classified as explained in section 1.3. The sets of features selected by the distinct methods are collected in Table 23 to Table 26. All of them correspond to the test case of 60 rpm and 5000 N charge.

Table 23. Data descriptors ranked by variance analysis.

Mother wavelet	Descriptor	Decomposition level	Band range (Hz)	Variance F value
BiorSpline 5.5	Impulse factor	4	800 -1600	53731.8
Daubechies 3	Impulse factor	4	800 -1600	36375.0
Symlet 3	Impulse factor	4	800 -1600	36375.0
BiorSplines 3.1	Peak value	3	1600 -3200	26153.1
Daubechies 21	Impulse factor	4	800 -1600	24846.2
Daubechies 12	Impulse factor	4	800 -1600	19835.5
Daubechies 22	Impulse factor	4	800 -1600	18933.3
Symlet 22	Impulse factor	4	800 -1600	16502.8

Table 24. Data descriptors ranked by the correlation feature selection method.

Mother wavelet	Descriptor	Decomposition level	Band range (Hz)
Daubechies 39	Crest factor	8	50 -100
Haar	Kurtosis	3	1600 -3200
Symlet 1	Impulse factor	3	1600 -3200
Haar	Kurtosis	8	50 -100
Raw signal	RMS	-	-
Daubechies 1	Kurtosis	8	50 -100
Daubechies 8	Variance	3	1600 -3200
BiorSplines 3.3	Ratio	4	800 -1600

Table 25. Data descriptors ranked by information gain method.

Mother wavelet	Descriptor	Decomposition level	Band range (Hz)	Inf. Gain value
Coiflet 5	Clearance factor	4	800 -1600	1
Daubechies 41	Variance	1	6400 – 12800	1
BiorSplines 6.8	Skewness	2	3200 –6400	1
Daubechies 39	Kurtosis	8	50 -100	1
Daubechies 39	Skewness	8	50 -100	1
Daubechies 39	Crest factor	8	50 -100	1
Daubechies 39	Shape factor	6	200 -400	1
Daubechies 41	Deviation	1	6400 – 12800	1

Table 26. Data descriptors ranked by relief method.

Mother wavelet	Descriptor	Decomposition level	Band range (Hz)	Quality estimation
Daubechies 1	Ratio	8	50 -100	0.521
ReverseBior 1.1	Ratio	8	50 -100	0.521
Symlet 1	Ratio	8	50 -100	0.521
BiorSplines 1.1	Ratio	8	50 -100	0.521
Haar	Ratio	8	50 -100	0.521
Haar	Kurtosis	3	1600 -3200	0.518
Daubechies 1	Kurtosis	3	1600 -3200	0.518
Symlet 1	Kurtosis	3	1600 -3200	0.518

The classifiers, using the correlation variation method, produced the hit rate values and deviations compiled in Table 27. The worst result is 70.47 % and the best is a 100.00 %. The features selected by the correlation feature algorithm offer particularly good results.

Table 27. Results from the cross validation of the features selected for the Wavelet analysis of the second experimental set.

	Experimental set 2							
	Classifier							
	Bayesian Network		Sequential minimal optimization		Intense based learning		J48 tree	
	μ (%)	σ	μ(%)	σ	μ (%)	σ	μ (%)	σ
Anova	70.47	15.15	66.67	7.34	83.07	16.86	77.17	16.84
CFS	100.00	0.00	100.00	0.00	100.00	0.00	94.10	10.66
Gain	99.80	2.00	99.80	2.00	100.00	0.00	94.70	8.71
Relief	94.03	9.36	98.20	5.45	98.20	5.45	94.17	8.92

In summary, we can state that the results obtained from the motor current signature analysis for the analysis of bearings in gearboxes are very promising, probing that the technique is a valid approach for the condition monitoring of such elements.

5 Example of practical implementation of *fingerprint* based on motor current signature analysis

As was mentioned before in chapter one, condition monitoring has the potential to reduce maintenance costs and upgrade the uptime of the assets. The most common strategy to implement it so far, has been periodic inspection by an expert, but this is slow and may not be cost effective. In other cases, where the benefits of adding sensors were clearer, they have been added, but it is not widespread mainly for two reasons, the cost and the complexity of the installation. Therefore, the motor current signature analysis is proposed as a cost effective continuous monitoring method.

In this context, the fingerprint concept is taken into consideration [3.67]. The asset in a good health state is monitored as a set of pre-defined operating conditions. The monitorization can be triggered during the whole asset's life-time in the same pre-defined operating conditions. As a result, a reference value is taken which accounts for normality. This value will be compared with measurements made throughout the life. The goal is to be able to detect and determine the appearance of abnormalities.

The process for the implementation of the condition monitoring tool consists in four steps and it is depicted in Figure 34.

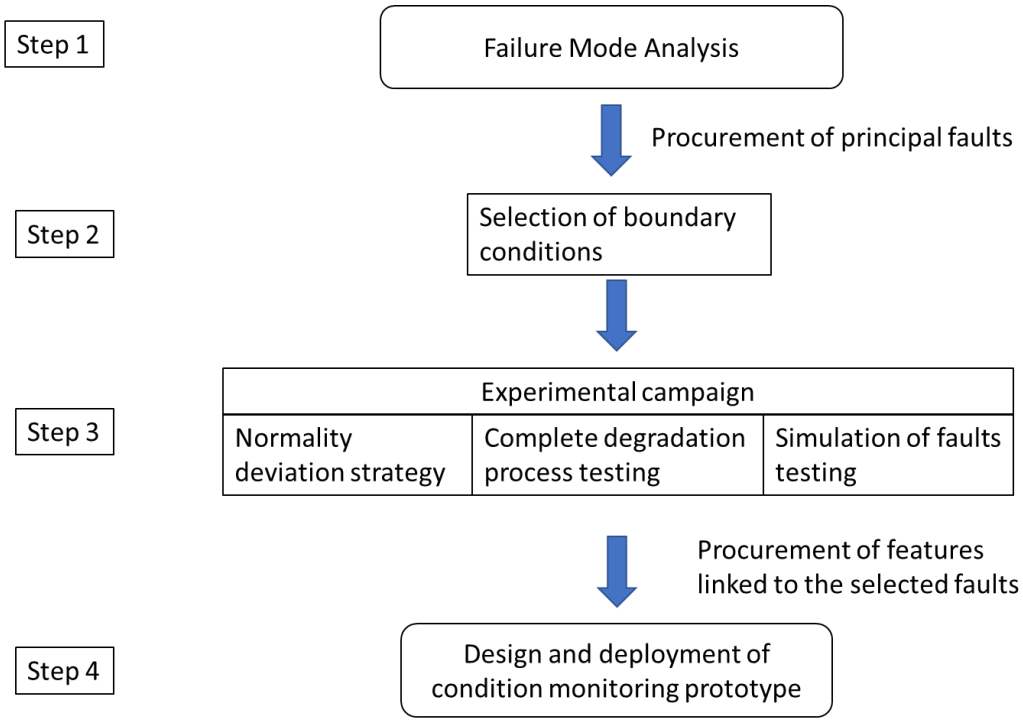


Figure 34. Scheme of the application of the fingerprint concept.

First, the principal potential faults of the asset must be identified. For that, the recommended tool is an analysis of *failure mode and effects analysis* (FMEA). The analysis provides a ranked set of faults that become the focus of the condition monitoring due to their higher potential to generate a failure.

In step two, the boundary conditions (*i.e.* speed, charge…) are analysed, with the goal of them being comparable. The most convenient strategy is to set the asset in a *monitorization* mode, where main operating conditions are kept fixed.

In step three an experimental campaign is launched. The asset in good state is monitored, to establish the baseline of the health indicators. With this information deviations on the original signal can be monitored and abnormalities can be detected. Additional experience can provide further insight on specific faulty conditions.

The ideal situation for the experimental campaign is to be able to test the asset through the process of generating the fault, but this is an extremely demanding situation, as the quantity of data needed is very high and the possibility of obtaining the desired isolated fault is quite low. Thus, faults can be simulated, either by setting a seed from where the fault will arise or by machining the fault directly. The parameters measured with the fault will set the limit until it is safe to use the asset, additionally providing information about the fault indicators.

In step four, the condition monitoring prototype is deployed. The parameters measured before are monitored periodically to control their evolution. If there is an anomaly they should inform before it becomes a catastrophic failure.

Motor current signature analysis (MCSA) can play a key role in the implementation of the *fingerprint* concept. Condition monitoring is becoming more and more popular, and MCSA can be the element to overpass barriers that stop it, such as the cost, the difficulty of installation and the decrease in reliability for the addition of elements.

Wavelet signature analysis will help in the implementation of MCSA, as the most common situations that are found in real-life applications are transients. The monitorization must be adapted to fit the asset monitored, and reduce as much as possible the downtime. Ideally it should enter monitorization mode seamlessly, without interrupting the *normal* operation of the asset.

In this section, some examples of implementation will be discussed.

5.1 Machine tool monitorization

The components that cause most of the trouble in machine tool industry are the gears in the spindle. Incorrect maintenance of the gears can have an impact in quality, cost and, more generally, in productivity.

The use of condition monitoring could contribute to improve the situation, although there are several obstacles that prevent its adoption. The main one is that they may not be cost effective. Although machine tool manufactures are steadily adding sensors to their equipment so that it can be monitored, they normally charge a premium for such capability, which prevents it from wide spreading.

The fingerprint concept combined with motor current signature analysis is especially suited for spindle monitoring. Following the steps above mentioned, a FMEA will point out the faults to monitor, and a fingerprint of the asset in the healthy state will be measured.

The comparison of the signal in the *fingerprint* and in the different health states will serve to select the most important parameters to be monitored. It is here where the importance of motor current signature analysis comes to life, as it is a particularly suited technique for spindle monitoring. The signal is readily available in the control scheme, and can be extracted and analysed without the introduction of additional sensors.

The implementation of the fingerprint requires comparable boundary conditions for the measurements. During the work of the spindle it is very hard to do this due to high number of different operations that can be performed with a single spindle. Even the material and the crystallographic plane where it is working have an effect in the vibration signal and the current consumed by the spindle.

So, the most convenient strategy would be to set the spindle in a *monitorization* mode, where the speed and charges will be kept equal. It seems that the most convenient strategy is to monitor the transient moments were the machine tool is gaining speed, but it is still not machining (load is zero). This way we may enter seamlessly into a *monitorization mode*, upgrading the up-time of the machine.

5.2 Electro-mechanical actuator for aerospace monitorization

Nowadays, large aeronautical multinationals, aircraft Original Equipment Manufacturers (OEMs) and their suppliers are immersed in the concept called *More Electric Aircraft*, where traditional hydraulic and pneumatic systems are being replaced by electrically driven systems [3.79].

In this context, the development and the implementation of advanced actuation systems has increased, as many factors are driving the migration from hydraulic actuators to electro-mechanical actuators (EMAs). EMAs provide significant advantages in complex applications because they increase the controllability of the system (shortening changeover times), provide re-configurability, and maintain functionality during faults (improving the accuracy and reliability).

EMA technologies are already being used in aeronautics, but for safety reasons they are limited to Secondary Flight Controls or military aircraft [3.80]. Their application to Primary Flight Controls will allow reductions in the weight of drives, gas consumption, and polluting emissions. The major step in moving from EHAs to jam-free EMAs is the prevention of potential jamming cases by appropriate technology and monitoring, thus giving the system aircraft availability for dispatch and failure sizing cases.

The drawback of EMA's is that the reliability is not high enough to be airworthy. One possible solution is to implement condition monitoring. The problem is that the addition of more elements reduces the overall reliability, what makes the inclusion of accelerometers (traditionally used element for condition monitoring of screwballs) undesirable. But as actuators are electrically driven, and the current signal is readily available in the command module, motor current signature analysis can avoid the addition of accelerometers, and in combination with the fingerprint concept it is ideal for this application.

EMA have a particularity regarding their function. The amount of work done at stationary speeds is limited or doesn´t exist. And again, the transient signal analysis is the only mode to monitor and the wavelet technique is particularly suited. The test should be designed so that repetitive conditions are achieved, so the boldest strategy would be to perform the tests, grounded in the pre-flight check phase. Additionally, simpler in-flight monitorization could be performed, for example triggering an alarm as certain parameters overpass a set limit.

The fingerprint process should be replicated, performing the FMEA analysis, enabling the realization of tests in good health condition and in faulty condition. The ideal indicators will be pointed out by this analysis, opening the door to the development of a prototype and further implementation.

5.3 Electric car mechanical parts monitorization

As this Thesis is written we are facing the dawn of the electric car. It is a new paradigm that will change the habitudes of millions of people. Furthermore, the trend is towards the servitization of the car. This is true especially for the electric car. Particularly, both of these facts will have a big impact in the maintenance of the cars, as there are less mechanical components, and the servitization requires having a live-time monitorization of the car [3.81]. The approach taken in this Thesis is especially appropriate for the monitorization of the drive train, as the current feeding the driving motor is readily available in the control module. It will avoid adding additional sensors, which, in a mass-produced asset like the car, provides high cost savings. The wavelet analysis may be a technique of choice, as it will be hard to obtain a signal of continuous speed and charge.

The process of condition monitoring of the drive train starts with the generation of the FMEA that will provide the elements to be monitored. The implementation of this technique is particularly tricky as it can be hard to obtain a situation where the car has comparable boundary conditions. The high range of speeds and charges, added to the irregularities of the pavement to which the car is submitted, represent a challenge. But making use of speed and other sensors that are of standard use, such as accelerometers in the traction control, equivalent conditions can be found. The information recorded in those conditions can be used to identify the *fingerprint* and later assess the health conditions pointed out in the FMEA analysis.

Another advantage of monitoring a mass market product as a car, is the availability of a bigdata base that will help provide a refined condition monitorization product, with a low number of false positives, and false negatives.

Bibliography

[3.1] JARDINE, Andrew KS; LIN, Daming; BANJEVIC, Dragan. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 2006, vol. 20, no 7, p. 1483-1510.

[3.2] ISO 8579-2:1993, Acceptance code for gears - Part 2: Determination of mechanical vibrations of gear units during acceptance testing, Last Review: 2010. ISO International.

[3.3] *UltraCheck Valve & Motor Diagnostic Products & Services, Areva, Empath 200 system for electric motor diagnosis*. *Us.areva-np.com*. Retrieved 19 January 2018, from http://www.us.areva-np.com/ultracheck/pdf/empath2000.pdf

[3.4] KRYTER, R. C.; HAYNES, H. D. *Condition monitoring of machinery using motor current signature analysis*. Oak Ridge National Lab., TN (USA), 1989.

[3.5] ARELLANO-PADILLA, Jesus, et al. A novel approach to gearbox condition monitoring by using drive rectifier input currents. In *Power Electronics and Applications, 2009. EPE'09. 13th European Conference on*. IEEE, 2009. p. 1-10.

[3.6] KIA, Shahin Hedayati; HENAO, Humberto; CAPOLINO, Gerard-Andre. A modeling approach for gearbox monitoring using stator current signature in induction machines. En *Industry Applications Society Annual Meeting, 2008. IAS'08. IEEE*. IEEE, 2008. p. 1-6.

[3.7] KAR, Chinmaya; MOHANTY, A. R. Monitoring gear vibrations through motor current signature analysis and wavelet transform. *Mechanical systems and signal processing*, 2006, vol. 20, no 1, p. 158-187.

[3.8] YACAMINI, R.; SMITH, K. S.; RAN, L. Monitoring torsional vibrations of electro-mechanical systems using stator currents. *journal of vibration and acoustics*, 1998, vol. 120, no 1, p. 72-79.

[3.9] XU, Xingyi; DE DONCKER, Rik; NOVOTNY, Donald W. A stator flux oriented induction machine drive. En *Power Electronics Specialists Conference, 1988. PESC'88 Record., 19th Annual IEEE*. IEEE, 1988. p. 870-876.

[3.10] LEI, Yaguo. *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*. Butterworth-Heinemann, 2016.

[3.11] BRAVO-IMAZ, Iñaki, *et al.* Mechanical fault detection in gearboxes through the analysis of the motor feeding current signature. In Comadem 2014 - Implications of life cycle analysis in asset and maintenance management*, 2014.*

[3.12] BRAVO-IMAZ, Inaki, *et al.* Motor current signature analysis for gearbox condition monitoring under transient speeds using wavelet

analysis and dual-level time synchronous averaging. *Mechanical Systems and Signal Processing*, 2017, vol. 94, p. 73-84.

[3.13] KAR, Chinmaya; MOHANTY, A. R. Multistage gearbox condition monitoring using motor current signature analysis and Kolmogorov–Smirnov test. *Journal of Sound and Vibration*, 2006, vol. 290, no 1, p. 337-368.

[3.14] ARDAKANI, Hossein Davari, *et al.* Motor current signature analysis for gearbox fault diagnosis in transient speed regimes. En *Prognostics and Health Management (PHM), 2015 IEEE Conference on*. IEEE, 2015. p. 1-6.

[3.15] DAVARI ARDAKANI, Hossein. *Prognostics and Health Management of Engineering Systems Using Minimal Sensing Techniques*. 2016. Doctoral Thesis. University of Cincinnati.

[3.16] ZAKRAJSEK, James J.; TOWNSEND, Dennis P.; DECKER, Harry J. *An analysis of gear fault detection methods as applied to pitting fatigue failure data*. National Aeronautics and Space Administration, Cleveland, OH, Lewis research center, 1993.

[3.17] BRAUN, S. The synchronous (time domain) average revisited. *Mechanical Systems and Signal Processing*, 2011, vol. 25, no 4, p. 1087-1102.

[3.18] WANG, W. J.; MCFADDEN, P. D. Application of wavelets to gearbox vibration signals for fault detection. *Journal of sound and vibration*, 1996, vol. 192, no 5, p. 927-939.

[3.19] GROSSMANN, Alexander; MORLET, Jean. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis*, 1984, vol. 15, no 4, p. 723-736.

[3.20] SUBASI, Abdulhamit. EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, 2007, vol. 32, no 4, p. 1084-1093.

[3.21] CUSIDO ROURA, Jordi; ROMERAL MARTÍNEZ, José Luis. Transient analysis and motor fault detection using the wavelet transform. In *Discrete Wavelet Transforms-Theory and Applications*. InTech, 2011.

[3.22] MURGUÍA, José S.; ROSU, Haret C. Discrete wavelet analyses for time series. In *Discrete Wavelet Transforms-Theory and Applications*. InTech, 2011.

[3.23] MALLAT, Stéphane. *A wavelet tour of signal processing*. Academic press, 1999.

[3.24] SIEGEL, David; LEE, Jay; DEMPSEY, Paula. Investigation and Evaluation of Condition Indicators, Variable Selection, and Health Indication Methods and Algorithms For Rotorcraft Gear Components. En *MFPT 2014 Conference*. 2014.

[3.25] VILLACAMPA, Osiris. *Feature selection and classification methods for decision making: a comparative analysis*. 2015. DoctoralThesis. Nova Southeastern University.

[3.26] LARSON, Martin G. Analysis of variance. *Circulation*, 2008, vol. 117, no 1, p. 115-121.

[3.27] HAHS-VAUGHN, Debbie L.; LOMAX, Richard G. *An introduction to statistical concepts*. Routledge, 2013

[3.28] FISHER, Ronald A. XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 1919, vol. 52, no 2, p. 399-433.

[3.29] HALL, Mark Andrew. Correlation-based feature selection for machine learning. 1999. Doctoral Thesis. The University of Waikato.

[3.30] SUI, Bangsheng. *Information gain feature selection based on feature interactions*. 2013. Doctoral Thesis. Unversity of Houston.

[3.31] KIRA, Kenji; RENDELL, Larry A. A practical approach to feature selection. *Proceedings of the ninth International Workshop on Machine learning*. 1992. p. 249-256.

[3.32] ReliefFAttributeEval. Weka.sourceforge.net. Retrieved 22 January 2018, from http://weka.sourceforge.net/doc.dev/weka/attributeSelection/ReliefFAttributeEval.html

[3.33] ROBNIK-ŠIKONJA, Marko; KONONENKO, Igor. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine learning*, 2003, vol. 53, no 1-2, p. 23-69.

[3.34] KONONENKO, Igor. Estimating attributes: analysis and extensions of RELIEF. *European conference on machine learning*. Springer, 1994. p. 171-182.

[3.35] Cross-validation (statistics) En.wikipedia.org. Retrieved 19 January 2018, from https://en.wikipedia.org/wiki/Cross-validation_(statistics)

[3.36] Vanschoren, J. OpenML - 10 fold crossvalidation. OpenML: exploring machine learning better, together. Retrieved 19 January 2018, from https://www.openml.org/a/estimation-procedures/1/1

[3.37] STEPHENSON, Todd Andrew. *An introduction to Bayesian network theory and usage*. IDIAP, 2000.

[3.38] JENSEN, Finn V. Bayesian networks basics. *AISB quarterly*, 1996, p. 9-22.

[3.39] Bayesian network. En.wikipedia.org. Retrieved 19 January 2018, from https://en.wikipedia.org/wiki/Bayesian_network

[3.40] DEL RÍO, Susana Ferreiro. *Aportaciones para el diagnóstico y pronóstico en problemas industriales mediante técnicas de clasificación supervisada*. 2011. Doctoral Thesis. Universidad del País Vasco / Euskal Herriko Unibertsitatea.

[3.41] COOPER, Gregory F. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial intelligence*, 1990, vol. 42, no 2-3, p. 393-405.

[3.42] NIELSEN, Thomas Dyhre; JENSEN, Finn Verner. *Bayesian networks and decision graphs*. Springer Science & Business Media, 2009.

[3.43] CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. *Machine learning*, 1995, vol. 20, no 3, p. 273-297.

[3.44] SMO. Weka.sourceforge.net. Retrieved 22 January 2018, from http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/SMO.html

[3.45] PLATT, John. Sequential minimal optimization: A fast algorithm for training support vector machines. Technical report MSR-TR-98-14. Microsoft. 1998.

[3.46] IRIDIA Projects: Lazy Learning for Modeling and Control. Web.archive.org. Retrieved 19 January 2018, from https://web.archive.org/web/20120216183916/http://iridia0.ulb.ac.be/~lazy/

[3.47] AHA, David W.; KIBLER, Dennis; ALBERT, Marc K. Instance-based learning algorithms. *Machine learning*, 1991, vol. 6, no 1, p. 37-66.

[3.48] Data driven modelling. Data-machine.com. Retrieved 19 January 2018, from http://www.data-machine.com/nmtutorial/introduction1.htm

[3.49] Instance based learning course. Cs.uccs.edu. Retrieved 19 January 2018, from http://www.cs.uccs.edu/~jkalita/work/cs586/2013/InstanceBasedLearning.pdf

[3.50] Arboles de decision - Razonamiento automatico. Cs.us.es. Retrieved 19 January 2018, from https://www.cs.us.es/~jalonso/cursos/ra-00/temas/tema-12.pdf

[3.51] What is a Decision Tree Diagram. Lucidchart. Retrieved 19 January 2018, from https://www.lucidchart.com/pages/decision-tree

[3.52] C4.5 algorithm. En.wikipedia.org. Retrieved 19 January 2018, from https://en.wikipedia.org/wiki/C4.5_algorithm

[3.53] J48 decision tree - Mining at UOC. Data-mining.business-intelligence.uoc.edu. Retrieved 19 January 2018, from http://data-mining.business-intelligence.uoc.edu/home/j48-decision-tree

[3.54] Unpaired (Two Sample) t Test. Statsdirect. Retrieved 19 January 2018, from https://www.statsdirect.com/help/parametric_methods/unpaired_t.htm

[3.55] R: The R Project for Statistical Computing. R-project.org. Retrieved 19 January 2018, from https://www.r-project.org/

[3.56] SpectraQuest Inc.: Gearbox Prognostics Simulator. Spectraquest.com. Retrieved 19 January 2018, from http://spectraquest.com/prognostics/details/gps/

[3.57] SpectraQuest Inc. Spectraquest.com. Retrieved 19 January 2018, from http://spectraquest.com/

[3.58] Power Transmission, Material Handling & Hand Tool Manufacturer. Martinsprocket.com. Retrieved 19 January 2018, from http://www.martinsprocket.com/

[3.59] Engranajes de Calidad y Elementos de transmisión - Engranajes Juaristi. Engranajesjuaristi.com. Retrieved 19 January 2018, from http://www.engranajesjuaristi.com/

[3.60] BOSE, Bimal K.; BOSE, Bimal K. (ed.). *Power electronics and variable frequency drives: technology and applications*. IEEE press, 1997.

[3.61] WU, Bin; NARIMANI, Mehdi. *High-power converters and AC drives*. John Wiley & Sons, 2017.

[3.62] PCB Model 608A11. Pcb.com. Retrieved 22 January 2018, from https://www.pcb.com/products.aspx?m=608A11

[3.63] Torque Sensor Mini-Smart Torque Sensor Type 4502A. Kistler.com. Retrieved 22 January 2018, from https://www.kistler.com/?type=669&fid=66908&callee=frontend

[3.64] Industrial Encoders Type SCH68B. Scancon.dk. Retrieved 22 January 2018, from http://www.scancon.dk/media/4389/sch68b-specifications-18.pdf

[3.65] Current Transducer HTA 100 ... 1000-S. Lem.com. Retrieved 22 January 2018, from http://www.lem.com/docs/products/hta_100-1000-s.pdf

[3.66] LabVIEW - National Instruments. Ni.com. Retrieved 22 January 2018, from http://www.ni.com/en-us/shop/labview.html

[3.67] FERREIRO, Susana, *et al.* INDUSTRY 4.0: Predictive Intelligent Maintenance for Production Equipment. In *European Conference of the Prognostics and Health Management Society, no*. 2016. p. 1-8.

[3.68] KAR, Chinmaya; MOHANTY, A. R. Multistage gearbox condition monitoring using motor current signature analysis and Kolmogorov–Smirnov test. *Journal of Sound and Vibration*, 2006, vol. 290, no 1, p. 337-368.

[3.69] KAR, Chinmaya; MOHANTY, A. R. Vibration and current transient monitoring for gearbox fault detection using multiresolution Fourier

transform. *Journal of Sound and Vibration*, 2008, vol. 311, no 1, p. 109-132.

[3.70] Tribology - IK4-TEKNIKER. Tekniker.es. Retrieved 22 January 2018, from http://www.tekniker.es/en/tribology

[3.71] SUBASI, Abdulhamit. EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, 2007, vol. 32, no 4, p. 1084-1093.

[3.72] Weka 3 - Data Mining with Open Source Machine Learning Software in Java. Cs.waikato.ac.nz. Retrieved 22 January 2018, from https://www.cs.waikato.ac.nz/ml/weka/

[3.73] RAFIEE, J.; RAFIEE, M. A.; TSE, P. W. Application of mother wavelet functions for automatic gear and bearing fault diagnosis. *Expert Systems with Applications*, 2010, vol. 37, no 6, p. 4568-4579.

[3.74] LESSMEIER, Christian, *et al.* Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification. En *Proceedings of the European Conference of the Prognostics and Health Management Society*. 2016. p. 05-08.07.

[3.75] SINGH, Sukhjeet; KUMAR, Amit; KUMAR, Navin. Motor current signature analysis for bearing fault detection in mechanical systems. *Procedia Materials Science*, 2014, vol. 6, p. 171-177.

[3.76] DEMEYERE, BLADE.BE, 2018, Falex Tribology Testing and Technical Support - Multispecimen tester. *Falexint.com* [online]. 2018. [Accessed 4 January 2018]. Available from: http://www.falexint.com/en/materials/multispecimen-tester

[3.77] ER16K Link-Belt 200 Series Standard Duty Ball Bearing - Rexnord, 2018. *Rexnord.com* [online]

[3.78] PHMA, Thu Hien et al., A strategy on selection of condition monitoring methods, In: *First World Congress Condition Monitoring*, 2017, London (UK).

[3.79] DERRIEN, Jean-Claude; SÉCURITÉ, Sagem Défense. Electromechanical actuator (EMA) advanced technologies for flight controls. En *International Congress of the Aeronautical Sciences*. 2012. p. 1-10.

[3.80] JENSEN, Stephen C.; JENNEY, Gavin D.; DAWSON, David. Flight test experience with an electromechanical actuator on the F-18 systems research aircraft. In *Digital Avionics Systems Conference, 2000. Proceedings. DASC. The 19th*. IEEE, 2000. p. 2E3/1-2E310 vol. 1.

[3.81] Arnaiz, Aitor. Advanced maintenance solutions and the impact in circular economy. In: *T-rex workshop Industry 4.0 Optimization of machinery life-cycle, component re-use and servitization*, 2016, Oñati, Lecture.

4

Conclusions

As stated in the first chapter, section 1.5, the primary objective of the thesis was to analyse the potential of two novel technologies to monitor the health state of a gearbox, the first one by means of screening the quality of the lubricant oil (viscosity), and the second by monitoring vibrations of the gearbox by examining the current signal feeding the electric motor. We consider that, to a considerable extent, these objectives were satisfactorily fulfilled.

The conclusions obtained regarding the potential of the techniques used, are summarized below, mostly organized along the two main subjects of research that constitute the work.

The most convenient strategy to prevent failures in a gearbox is the combined monitoring of the lubricant oil and the vibrations generated in the system. Replacing the lubricant before degradation works as a protection for failures, whereas vibration monitoring screens existing faults, and prevent them from becoming catastrophic.

A) Concerning the oil viscosity monitoring, using the magnetoelastic sensor concept developed in this work, the main results and conclusions are:

A.1 The parameters of the lubricant oil were studied, and viscosity was selected as the most important one for the correct function of the gearbox.

The relationship between the variation in viscosity and the problems in the lubrication oil was displayed, and linked with the faults generated in the gears when working with such broken-down oil.

A.2 The magnetoelastic effect in suitable materials offers a good base to develop an on-line viscosity sensor capable of a wide measuring range.

The on-line viscosity measurement concept has been focused on magnetoelastic sensors. The mechanical oscillations in the sensing probe are created by means of the coupling between the magnetic and elastic properties of a magnetoelastic material. The excitation is produced by an alternating magnetic field that, depending on the frequency and the characteristic of the material, produces forced mechanical oscillations and eventually resonances. The dissipative force caused by the viscosity of the lubricant oil, produces an attenuation of the oscillations, and changes in the frequency, the magnitude and the quality factor of the resonance. All these parameters can be correlated to the viscosity of the oil.

A.3 These concepts have been experimentally verified in a proof-of-concept prototype, in which we were able to observe measurable differences between

the magnetoelastic response of the sensing material in the presence of lubricant oils with different viscosities.

Especially, the relation between the amplitude and the frequency of the resonance for different oil viscosities was determined.

The suitability of amorphous ribbons for the magnetoelastic measurement was confirmed. They were used in the form of strips, free to oscillate in the bottom of a vial containing the oil under test. Although oils with high viscosities present quite small amplitudes, which makes it difficult to be analyzed by the processing algorithm, calibration curves were established for the change in amplitude and resonance frequency as a function of the viscosity.

A.4 A second, more elaborated prototype, was designed to overcome limitations of the first one, considerably improving the results and the measuring range.

The new prototype permits much cleaner and repetitive results, easing the change of sample and adjusting the coil system for maximum sensitivity. It was tested with two different magnetoelastic materials which are now used in a cantilever configuration (clamped by one end). The amplitude and the value of the resonant frequency were confirmed to be linked with the viscosity of the lubricant oil tested. Both Vitrovac 4040 and 7600 amorphous ribbons showed a clear dependence between the amplitude and the frequency of the resonance with the viscosity of the lubricant oils, in a wide range from 32 to 326 cSt. It was confirmed that the frequency shift with respect of the resonance in air follows a linear relation with the square root of the product of the viscosity and the density of the oils.

A.5 A new phenomenological model for the magnetoelastic resonance has been proposed, which permits the detailed analysis of the experimental results by a least-squares fitting procedure.

The curve fitting of the data obtained with the magnetoelastic sample Vitrovac 4040 was used to determine new calibration curves. The amplitude of the resonance as a function of the viscosity of the oil is given by the expression A (mV) = 26.49 × $\eta^{-0.69}$, where the viscosity η is expressed in cSt. The resonance frequency is described by f_r (kHz) = 43.23 × $\eta^{-0.08}$.

The fitting procedure allows for the determination of the damping parameter δ of the magnetoelastic resonance. It was found that the damping parameter is very sensitive not only to the viscosity but also to the nature of the oil. Hydraulic oils (from the family T, with viscosities up to 108.6 cSt) display a linear trend with different slope that the oils from gearboxes lubrication

(family O, with larger viscosities). These dependencies are described by $\delta = 0.035+4.2 \times 10^{-4}\ \eta$ for oils of the family T ($\eta \leq 108.6$ cSt), and $\delta = 0.054+2.4 \times 10^{-4}\ \eta$ for oils of the family O ($\eta > 108.6$ cSt).

The curve fitting in the case of the data obtained using the sample Vitrovac 7600 was complicated by the presence of two overlapped resonant modes (probably caused by an imperfect clamp of the sample). Both resonant modes are fitted simultaneously. It is observed that the amplitude of the second mode decreases rapidly when the viscosity increases, implying that higher viscosities favours the existence of only one mode of oscillation. The damping parameter extracted from the fitting for the first mode reflects clearly the effect of the viscosity of the oil. It presents different slope for T and O types of oils, as it happened with the Vitrovac 4040 sample.

A.6 The evolution of the oil viscosity with the temperature in the range from 0° to 60° C was determined using the magnetoelastic sample Vitrovac 4040 biased in its temperature compensation point.

The on-line monitoring viscosity sensor must be able to determine the viscosity at different temperatures. In principle, the magnetoelastic resonance itself can be severely affected by temperature changes, but when the sample is biased at a specific compensation point the frequency of resonance becomes insensible to temperature. For the sample Vitrovac 4040 oscillating in air this has been found to happen when it is biased at about 722 A/m.

The viscosity of the hydraulic lubricant oils (T family) was measured at different temperatures, from 0° to 60° C using a climatic chamber. Biased at the compensation point, the observed changes in the resonance frequencies are originated by changes in the viscosity of the oils. They were translated to viscosity values using the calibration curve determined at room temperature. The results indicate that the most viscous oils display larger variations of viscosity with temperature.

B) Regarding the motor current signature analysis (MCSA) for the monitorization of the health state of the gears, the main conclusions are:

B.1 MCSA is not as accurate as accelerometers for vibration data analysis. But it provides other important advantages, it is more economical to implement, there is no need to install additional sensors in hard to reach areas and the reliability of the monitored system is not compromised by the addition of new equipment.

Vibration analysis by means of accelerometers is nowadays the standard practice in industry. A new technology as motor current signature analysis

must be benchmarked against it. In general terms MCSA is not as accurate as accelerometer vibration analysis, especially when the fault monitored is in its initial stages and it is not large enough. For example, we have shown that, for the case of bearings, MCSA hardly distinguishes faults consisting of pits below 1 mm of diameter (comparable results were reported in the literature). For larger faults, MCSA successes in detecting them and distinguish among different types of faults. Additionally, MCSA may need more advanced signal processing and larger computational resources than vibration analysis based on accelerometer data. However, MCSA provides important benefits. As no additional hardware is needed for the monitoring (the current signal is available from the control governing the movement), apart from the clear economic advantage, the overall reliability of the system is not compromised by additional elements. Besides, MCSA is also most convenient when the gearbox to be monitored is not easily reachable.

B.2 Motor current signature analysis, combined with wavelet analysis, is especially convenient for transitory working conditions and for the so-called *fingerprint concept* in the monitorization and screening of faulty states of a great variety of assets (machine tools, electro-mechanical actuators, electric cars among others).

Transient speeds are present in the working parameters of many of the assets present in industry, transport or energy sectors. For example, in certain actuators the constant speed regime may too short for monitoring under steady conditions. Also, in machine tools, the transient until reaching the working speed may be considered a much more repeatable condition than the constant speed regime used in machining the material. In these and similar cases, it is enormously interesting to adequately monitor, the transient regime. However, the data analysis of transitory speeds presents some challenges over the one with constant speed. The signal processing techniques must be more carefully selected, and the data acquisition sampling rate must be adapted to the transient generated. In the case of gear-boxes this implies that, for each physical revolution of the mechanical components, a sufficient number of points must be obtained to have enough signal information, which depends on the number of teeth of the gears and their interaction. This may be changeling if the transient is very fast.

B.3 Discrete wavelet analysis technique provides good results, and is confirmed as an adequate technique for its use with motor current signature analysis when transitory speed changes are to be analysed.

Different signal analysis procedures were critically evaluated for the desired working conditions (speed transitions). According to the literature, it was

decided to use wavelets as a pre-processing technique, and statistical parameters as a processing technique. The use of computer resources was minimized using the discrete wavelet transform. The adequateness of the approach was backed up by the results obtained.

B.4 The concept of motor current signature analysis implemented through the discrete wavelet analysis was validated experimentally. It has been proved to be useful for the condition monitoring of the health state of gears in a gearbox under transitory speeds.

As a proof of concept, a set of gears with different health states were used. They were grouped in three categories attending to a classification performed using accelerometer data. The objective was to see how the motor current signature analysis compares with the analysis from the vibration signal. Signals from both current sensors and accelerometers were obtained from the same tests.

The analysis performed shows the viability of capturing and extracting variations in motor current. The main conclusion of the wavelet analysis and feature selection procedure is the production of observations correlating with the health state of gears. The results from the feature selection indicate that the health condition of the gears cannot be determined unambiguously using only one descriptor, but when diverse of them, wisely selected, are combined, the diferent faulty conditions can be clearly indetifyied. It is to be understood that here the term descriptor is used in an ample sense, meaning the result of the analysis of the data obtained using a given mother wavelet and decomposition level, and a given statistical measure (average, variance, mean, etc).

B.5 Motor current signature analysis combined with discrete wavelet analysis is proved experimentally to distinguish between gears in three different health states with a good percentage of accuracy.

A healthy gear and gears with two faults (pitting and eccentricity) were tested. Two different strategies, discrete wavelet analysis and dual level time synchronous averaging (DLTSA), were compared as pre-processing techniques to analyse the experimental data acquired. From each analysis, the relevant features of the signals are extracted and catalogued using a self-organizing map. The resulting classification allows for an easy discrimination of the diverse health states of the gears, which is successful in both approaches.

Comparing the results from both techniques, using the results of the Student's t-test, some of the results appear to be better for the Wavelet

technique and others for the DLTSA, but both are strongly dependent on the feature selection and classification methods selected. Therefore, with the present data set, no clear differences exist in the performance of both techniques (although the wavelet analysis provides a 100% accuracy in one of the instances). The combination from both pre-processing techniques, wavelet and DLTSA, provides an improvement of the results.

B.6 Faulty conditions on bearings outside the electric motor can also be detected using motor current signature analysis combined with wavelet decomposition.

Additional experimentation was performed to assess the health state of bearings, in this case under steady speed conditions. The experimental work was performed in two different test benches to detect faults injected in the bearings. In the case of the bearings tested in the first bench, the analysis was performed in a similar way that in the case of the gears using the discrete wavelet transform as pre-processing technique. The first results only differentiate between bearings with big faults and bearings with small or no faults (it could not separate the bearing with the smallest faults from the healthy state, nor differentiate between bearings with the bigger faults). In a new analysis the original data set was separated in two, one for the bearings with larger faults and the other for the bearing with small faults, obtaining new features that provided much-improved discrimination.

In the second experimental test, using bearings with different faults, both a discrete wavelet decomposition and a traditional time domain statistical analysis were employed. The selection of the features and the classification of the schemes used in the case of the testing of the gears provided results that allowed to differentiate each of the health conditions of the bearings.

The results obtained were very promising, probing that motor current signature analysis is a valid technique for the analysis of bearings in gearboxes moved by electrical motors.