University of the Basque Country (UPV/EHU)

Universidad del País Vasco
Euskal Herriko Unibertsitatea
The University of the Basque Country

Department of computation sciences and artificial intelligence
University of the Basque Country (UPV/EHU)

On the use of context information for an improved application of data-based algorithms in condition monitoring

Kerman López de Calle Etxabe

1. Supervisor Basilio Sierra
Department of computation sciences and artificial intelligence
University of the Basque Country (UPV/EHU)

2. Supervisor Susana Ferreiro
Intelligent Information Systems
Tekniker

July 27, 2020

Kerman López de Calle Etxabe
On the use of context information for an improved application of data-based algorithms in condition monitoring
July 27, 2020
Reviewers: Basilio Sierra and Susana Ferreiro
University of the Basque Country (UPV/EHU)

University of the Basque Country (UPV/EHU)
Department of computation sciences and artificial intelligence
Manuel Lardizabal Ibilbidea, 1
20018 Donostia (Gipuzkoa)

Abstract

The irruption of sensorized, connected and autonomous machinery has already marked a milestone in the manufacturing industry. This paradigm also known as Industry 4.0 consists in the digitalization of industrial processes and it is providing a new and previously underused source of knowledge: data.

The data streams that are massively being generated and stored are big sources of information. With the proper tools, the knowledge extracted from these data can enable monitoring and control, boost the decision making and customer satisfaction as well as the efficiency, the productivity, and the optimal use of facilities in manufacturing.

The process of knowledge extraction or data mining is carried out using algorithms that can handle large data volumes and are capable of extracting patterns that are hidden in the data. Once patterns are found, they can be captured in Machine Learning (ML) models to be later applied in a range of different fields with different purposes.

In the discipline of condition monitoring, the discipline that deals with the detection of the health status of the assets, the so-called data-based models have been around for a long time. Statistical methods as the quality control charts have been used to detect anomalies in operating machines since the last century. Later on, more sophisticated classification and regression algorithms such as artificial neural networks (ANNs) or support vector machines (SVMs) have been used to detect and predict failures with condition monitoring purposes. However, most of the works in the literature overlook one of the most important factors that are involved in the implementation of algorithms: the context. In short, we can say context is composed by the factors that have influence in the monitoring of a machine. Hence, a detailed understanding of the opportunities and limitations of the context of a particular application is needed to put algorithms in production. As they could limit the use of certain algorithms or could enable the use of some other algorithms that are not suited for the problem at a first sight.

Most of the algorithms are designed with a certain context in mind, nevertheless, they tend to be applied without considering that the final application context might change. In condition monitoring, this is the case of fault diagnosis algorithms that

are trained and validated in test rigs, without on-site data that will enable retraining the algorithm in the real-life application. Or other algorithms developed in steady conditions that do not consider the real operation is varying over time.

This work discusses the role of the context in condition monitoring algorithms in three different situations dealing with the constraints and the chances that are related to each case of study. The applications discussed are wind turbine gearboxes operating under varying operating conditions; rotating machinery operating under steady conditions with a lack of knowledge regarding its degradation; and, an electromechanical actuator that has been diagnosed with the help of a physical model. The contexts are studied and afterwards, solutions are proposed to provide ad hoc designed algorithms that compose condition monitoring systems.

Although this thesis project deals with three specific cases of study, the contexts that are addressed in this work can be found in many other condition monitoring problems, which makes the lessons obtained in this work be transferable to other real condition monitoring problems.

Laburpena

Sentsorizatutako, interkonektatutako eta autonomoa den makineriaren agerpena dagoeneko mugarri bilakatu da fabrikazioaren industrian. Prozesu industrialen digitalizazioan datzan eta Industry 4.0 izenaz ezagutzera eman den paradigma honek orain arte erabili gabea zen jakintza iturri batez hornitzen gaitu: datuak.

Masiboki sortzen eta pilatzen ari diren datuok erraminta egokiekin baliatuta, datuek duten baliozko informazioa monitorizazio eta kontrola ahalbidetzeko, erabaki hartzea eta bezeroaren gogobetetzea sustatu, zein fabrikazio instalazioen efizientzia, produktibitatea eta erabilera optimizazioan laguntzeko erabil daiteke. Informazio erauzketa prozesua edota datu meatzaritza datu bolumen handiak maneiatu eta datuetan izkutaturik dauden patroiak topatzeko gai diren algoritmo bitartez egiten da. Behin patroiak topatu ostean, Makina Ikasketa (Machine Learning) modeloen bitartez patroiok kapturatu eta geroago hainbat alorretan eta beste horrenbeste xederekin aplikatzen dira.

Osasun-egoeraren monitorizazioaren (Condition Monitoring) diziplinan, hau da, makinen osasunaren detekzioaz arduratzen den diziplinan, datuetan oinarritutako modeloak aski ezagunak dira. Kalitate kontrolerako grafikoak, adibidez, operatzen dauden makinetan anomaliak detektatzeko erabiliak izan dira azken mendeetatik. Beranduago, sofistikatuagoak diren klasifikazio zein erregresio algoritmoak, hala nola sare neuronalak edota bektore oinarri makinak, osasun-egoera monitorizazioa helburu zelarik erabili izan dira. Halaber, literaturan ageri diren lan gehientsuenek algoritmoen inplementazioan berebiziko garrantzia duen faktore bat aintzat hartzea ahazten dute: kontestua. Laburki esanda, kontestua makina baten monitorizazioan eragina duten faktoreek osatzen dutela esan genezake. Horregatik, aplikazio errealetan algoritmoak produkzioan jarri ahal izateko kontestuak dauzkan muga eta baliabideak ongi ulertzeak berebiziko garrantzia du. Izan ere, algoritmo jakin batzuen erabilera mugaturik egon baitaiteke kontestu zehatz batzuetan, aldi berean, lehen begirada batean bazterturiko algoritmo bat erabil liteke baliabideei ondo erreparatuz gero.

Algoritmo gehienak kontestu bat helburu delarik diseinatzen dira, halabaina, amaierako aplikazioaren kontestua ezberdina izan litekeela kontuan hartu gabe erabili ohi

dira. Osasunaren-egoeraren monitorizazioaren kasuan hau gertatzen da diagnosirako entrenaturiko algoritmoekin entsegu-bankuetan entrenatu eta testeatzen direlarik aplikazio errealean bertan datuak sortzeko dauden mugak kontuan hartu gabe, ondorioz, berriz entrenatzeko inongo aukerarik gabe. Edota egoera egonkorrean operatzen ari dela ontzat hartuta garatutako algoritmoak, amaierako aplikazioan operazioa aldakorra dela kontuan hartu gabe.

Lan honek kontestuak osasun-egoeraren monitorizazio algoritmoengan duen rola aztertzen du 3 egoera ezberdinetan bakoitzak dituen baliabide eta mugei erreparatzen zaielarik. Hurrengo aplikazio eremuak aztertu dira: operazio aldakorreko haize errota baten biderkagailua; degradazio bilakaera ezezaguna duten eta operazio egonkorrean diharduten makina birakariak; eta, modelo fisiko baten laguntzaz diagnostikatutako eragingailu elektromekanikoa. Kasuon kontestuak aztertu eta ondoren soluzioak proposatzen dira espresuki garatutako algoritmoen bidez osasun-egoeraren monitorizazioa hobetzeko bidean.

Lanak 3 kasu espezifiko aztertzen baditu ere, bertan aztertutako kontestuak beste hainbat monitorizazio problematan topa daitezke, ikerkuntza honetan ikasitako lezioak bizitza errealeko beste problema batzuetara transferitu daitezkeelarik.

Acknowledgement

Confronting the obstacles of the path and learning from them, txikirri-txikirri, this match has reached its 35th point, leaving behind another stage of this “txuin” life. In the meantime, in some way or another many of you have helped me carry out this research, and for that reasons, I would like to thank you.

Starting from those of you who have helped me technically and have directed my work, my tutors, thank you Susana for you have been so close to me during all this time and you have been ready to help when I have needed it, and thanks to you too, Basi, for ending all the e-mail with encouraging words. Thank you also to both of you, Aitor and Ana, even if your name is not official in the paperwork, this thesis project has the footprint of your guidance, as you have done it yours many times. Despite being for a short period of time, thank you very much, Cristobal, for welcoming this foreigner and helping me as much as you have helped me. At the same time, I would like to specially thank those of you that have mentored me, even by leaving aside your own duties to help me with my work, Ion, Ruben, Meritxell, and there are many others I could mention, as the unexpected help provided by Alexandre when only a “light-review” of this work was required. Thank you too.

This phase, however, has required of some relaxing to balance the other side of researching and learning, and that would not be possible without you, my travel companions. So thank you members of the Tekniker coffee brigade (Santi, Álvaro, Egoitz, Iñaki, Cristina, Edu and to the other short term members), the Lunchtime committee of Cranfield (Pavan, Iñigo, Alex, Sergio, Lawrence eta Dedy) and the group of Greek philosophers (Giorgios, Giorgios, Yeris and Alozie). Thank you for the simple disconnection chats. Thank you for showing me different countries and cultures. And thank you for helping me find the sense of live and the solution to the universal salvation.

At the same time, I want to specially thank those of you that have been on my side, not just in this phase, but in the whole journey, I would like to thank you too.

On the one hand, I want to thank friends and kuadrilla members, for making me smile on the weekend after the worst week. And I am sorry, but, as you are so many, you will have to be satisfied with a “cheers Julaspas” this time.

On the other, I want to thank you, the ones that have risen and educated me and have kept me on the right track during all this time. From Grandpa-grandma, uncle-aunts, and cousins for the unforgettable time we have had, in particular, mother and father, four your encouragement and your efforts to keep me focused, and, Ishmael and Patxi, for making me part of your little mischiefs. Finally, thanks to you, Janire, for sharing my tears and my joy and for being my support when I most needed it.

Thank you all!

Eskertzak

Bideko oztopoei aurre eginez eta haietatik ikasiz, txikirri-txikirri, partidu honi ere ailegatu zaio 35. tantua, bizitza “txuin” honen beste etapa bat atzean utziz. Tartean, era batera edo bestera ikerketa hau aurrera eramaten lagundu nauzuenak asko zarete, eta horregatik, eskerrak eman nahi dizkizuet.

Teknikoki lagundu eta lana bideratu duzuen tutoreengandik hasita, eskerrik asko Susana denbora guzti honetan horren gertukoa izan eta behar zenerako prest egotearren; eta zeuri ere Basi, idatzitako mezu guztietan animoak ematearren. Baita zuei ere, Aitor eta Ana, paperetan agertu ez arren, tesi honek baduelako zuen gidaritzaren aztarna, zeuena egin baituzue hainbat unetan. Epe laburrez izan bada ere, eskerrik asko, Cristobal, arrotz honi ongietorria eman eta lagundu nauzun beste laguntzeagatik. Era berean bereziki eskertu nahi nizueke mentore izan zaretenei, eskuartean zeneukaten lana utzi eta nire dudak argitzeko prestu agertu zaretenei, Ioni, Rubeni eta Meritxelli, eta beste asko aipatu litezke, ustekabean, tesiaren errebisio azkarra eskatuta prestutasun eta laguntza osoa eskaini didazuenoi, Alexandre kasu, eskerrak zuei ere.

Fase honek ordea, ikertze eta ikasteaz gain, erlaxatzearen oreka ere behar izan du eta hori ez litzateke posible izango bidelagun izan zaretenengatik izan ez balitz. Eskerrak beraz Teknikerreko kafe brigadakoei (Santi, Álvaro, Egoitz, Iñaki, Cristina, Edu eta denbora laburrez brigada-kide izan zareten besteoi), Cranfild-eko bazkalorduko komitekoei (Pavan, Iñigo, Alex, Sergio, Lawrence eta Dedy) eta Greziar filosofoen taldeari (Giorgios, Giorgios, Yeris eta Alozie). Mila esker eguneroko deskonexio hizketaldi sinpleengatik. Mila esker herrialde eta kultura ezberdinak erakusteagatik. Eta mila esker bizitzaren zentzua eta munduaren salbamena topatzen laguntzeagatik.

Era berean, etapa honetaz gain, bide osoan zehar hor egon zaretenoi, batez-ere, zuoi eman behar dizkizuet eskerrak.

Alde batetik, koadrilako kide eta bestelako lagunei eskertu nahi nizueke, asterik okerrenaren osteko asteburuan negar egin arteko barreak eragitearren. Eta, sentitzen dut baina, horrenbeste zaretenez gero, “Zuregatik doa Julaspas” batekin konformatu beharko zarete oraingoan.

Bestetik, eskerrik mila etxekooi, hazi, hezi eta denbora guzti honetan bide zuzenean mantendu nauzuenoi eskertu nahi dizuet. Aiton-amona, osaba-izeba, lehengusu-lehengusina, zuekin egindako otordu paregabeetatik hasi eta, bereziki, aita eta amari, zuen animo eta zentratzeko esfortzuarengatik baita ere, Patxi eta Ishmaeli, zuen bihurrikerietan parte izateagatik. Azken-azkenik, zuri, Janire, bideko malkoak eta pozak elkarrekin bizitzeagatik eta indarrik gabe nengoenean nire sostengu izateagatik.

Eskerrik asko zuoi guztioi!

Contents

Part I

Part I: The rationale

1 Background

> *Because the thing about repairing, maintaining, and cleaning is not an adventure.*
>
> — **Dr. Wong, Therapist**
> (Rick and Morty)

In the engineering world, one might think that the challenges in the development of the machines of the future reside only in the design, the manufacturing or in the commissioning of the machines. Nothing further from the truth, the recent advances in Cyber-Physical Systems (CPS), big data, cloud computing, and industrial wireless networks are turning product life-cycle management into a reality. Therefore, in contrast to what Dr. Wong believed, repairing and maintaining are definitely an adventure. One that has taken for almost 3 years and has provided the author with broad and varied experience.

1.1 Industry: From the rock to the binary code

Since the beginning of the times, humans had some hidden advantages in comparison to the rest of living beings of the world that eased their subsistence. Our ancestors had no weapons integrated in their bodies such as long fangs, sharp claws or poisonous bites that helped them survive. Instead, they had the ability to ***make*** and ***use*** sharp stones flakes and to hold and carry sticks and food, while they could work together to get things done. This ability to handcraft tools for hunting and/or farming greatly increased the chances of early humans to survive. Hence, the following generations developed more and more sophisticated tools.

In that way, some thousands of years later, approximately around the year 1760, the advantages provided by alternative sources of power such as steam engines changed the paradigm of manual work to the creation of machines for mass production. Consequently, during that first industrial revolution period, textiles, chemicals, metallurgy and concrete industries emerged, and the standards of living of the general population consistently increased.

Around 1850, the second industrial revolution or the technology revolution began, thanks to the use of railroads; large-scale iron and steel production; and the use of electrically powered production, that started in an American slaughterhouse in 1870 and soon spread to other industries. Again, living standards increased and, additionally, the prices of goods fell dramatically due to the increase in productivity.

At the end of the 20th century, about 1970, the digital revolution or third industrial revolution began. This time it was triggered by the invention of the transistor, which allowed the further developments of programmable logic controllers (PLC) and the consequent automation of industrial processes.

Finally, in our era, the beginning of the 21st century, the so called Industry 4.0 or the forth industrial revolution has started to take shape thanks to the initial boost of the German government. This incoming revolution wants to turn the industry smart by making the entire supply chain accessible and controllable through the internet. In that direction, there are 5 components that are adopting key roles in the industry of the future:

- Cyber Physical Systems (CPS): CPSs are systems that allow the interaction between the physical machines and computer system models while there is a data exchange between them.
- Internet of Things (IoT): This concept refers to the trend of connecting every single device to the internet, which gives accessibility advantages by providing new data exchange possibility to devices.
- Internet of Service (IoS): Through the internet of services, systems are able to receive data from online resources that are related to their domain.
- Smart factory: The goal behind the Smart Factory is to make production flexible by having total and constant control over the status of each product. Consequently, resources will be better planned and the quality of the final products will be continuously assured.
- Cyber security: Due to the full connectivity of the systems, new risks will appear, as networks would be directly in control of the integrity of the assets. Therefore, new efforts would be needed to avoid vulnerabilities in the connections and the protection of systems/assets.

This progression has pushed the humanity from using stone or wooden tools to coding in the smart and connected industry for the survival of the species.

1.2 Maintenance: When our tools fail

In parallel to the development of the first tools by ancestors, the effect of wear and strain appeared and made the tools fail. Hence, our ancestor began the new and "boring" art of repairing and maintaining the tools as soon as those first tools broke down. Over time, machines and tools became more complex, but the policy of waiting till they failed to fix them with spare parts lasted for centuries.

However, following the trend of developing more complex and useful tools, and adding on the top of that our growing dependence on the correct functioning of these tools, new needs emerged. As the shutdowns of the machines could cause fatal failures or huge economical losses, industries started to focus on reducing downtime after World War II. For that purpose, they started to approximate the life of the assets and to replace components following either schedules or wear related parameters such as cycles or kilometres to avoid breakdowns. Nevertheless, the fixed interval replacements lead to unneeded replacements which produced additional expenses. Therefore, since 1980 the efforts are being made to detect the health status of the assets, which is done by means of Condition Monitoring, and to perform maintenance actions only when required.

And this is how maintenance is going from being corrective (repairing when failures are given) to condition based (detecting health status and planning maintenance accordingly) going through preventive (performing fixed time repairs).

1.3 Condition Monitoring: A brain for the senses

As explained in the previous section, Condition Monitoring aims to detect the current health status of our assets. The main advantages of condition monitoring are that, as the condition of the asset is known, replacements are only carried out when required; catastrophic failures can be detected; and, it is possible to plan replacements considering the present and the predicted wear of the assets (also known as Predictive Maintenance PdM).

Essentially, condition monitoring operates by taking physical measurements of the assets that are indicators of deterioration. Later, the evolution of these measurements over time is analysed, and patterns or trends are used to reflect the actual health status of the assets and to infer the evolution of the health status. For the measurement of these indicators, various sensor are typically used, such as: accelerometers, acoustic emission sensors, thermocouples, tachometers, oil debris sensors, infrared thermography, current sensors, etc.

Once measurements are available, the procedure in Figure 1.1 is typically followed:

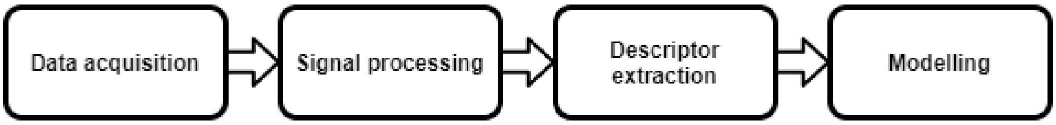

Fig. 1.1.: Elements of condition monitoring.

Data acquisition systems record the signals. Then, this signals are first processed (such as denoised and transformed to frequency domain) and descriptors (statistical moments such as mean and standard deviation) are extracted. Additionally, if the relation between these descriptors and the condition of the assets is known, it is possible to develop a model that reflects the condition of the asset.

Depending of the understanding of the faults of the machine and the historical data of the machine itself or similar ones, these models can be categorized into three stages, regarding their complexity and level of detail of CM, as represented in Figure 1.2.

Fig. 1.2.: Condition monitoring stages.

Each category represents a higher degree of knowledge provided by the model and the categories can be explained in short as:

- Detection: The model distinguishes normal/abnormal conditions.
- Diagnosis: The model knows which fault is causing the abnormal condition.

- Prognosis or Remaining Useful Life (RUL): The model approximates how long the machine will operate before fault causes catastrophic damages.

These models, that are the core of the condition monitoring systems, can be developed in three different manners. Either the model is based on physical principles (physics based), or the model is inferred from experimental data (data-based) or it combines both (hybrid).

Particularly, due to the recent advances in computing, cloud data storage and sensing, added to the complexity of developing physics based models of certain assets, data-driven modelling is gaining particular attention. For that reason, considerable amount of works using latest state of the art data-driven models are flourishing in the condition monitoring literature. From simple k-nearest neighbour, going to the more complex Support Vector Machines, going through random forests and reaching to the latest Deep Learning models.

Nevertheless, most of the works use data obtained in test rigs, with faults that have already been seeded; run to failure tests, where experiment conditions are steady; or data from physical models, which tend to be quite cleaner than signals from real applications. Therefore, there is a poor representation of data-based models in production environments, despite the fact that the presence of Computer Maintenance Management Software (CMMS) is in increasing trend.

1.4 Research environment: The fertiliser that grows the plants

The development of this research work is specially bonded to two particular entities: The University of the Basque Country (EHU/UPV) and the research center Tekniker. Both entities are situated in the Basque Country, a region with a solid industrial base where industry produces 23.5 % of the GDP whit a strong governmental bid for advanced manufacturing [@Gru17]. In addition, the region has a particularly strong machine tool sector, accounting for 78% of the whole machine tool production of Spain [@Pre08].

At the same time, the figures provided by the Spanish Maintenance Association (Asociación Española de Mantenimiento - AEM) reflect that, even though the awareness of adopting improved maintenance techniques exists in Spain (with CMMS use going from 63% in 1995 to 98% in 2005), its adoption is yet in infant stages.

The latest poll (2015) reveals that the expenditure in corrective maintenance (%44) is in decreasing trend giving place to increasing preventive maintenance (%46), that has surpassed corrective for first time. Additionally, predictive maintenance appears for first time, accounting for the 10% of the remaining expenditure. Furthermore, only 22% of the total of the industries answering the poll used the CMMS for actual monitoring of the machines [Man15]. These figure collide with the ones gathered by Advanced Technology Services (ATS), as, according to the 2018 Plant Engineering survey (surveying worldwide but mostly American industries), the use of predictive maintenance with analytics has increased from 47% in 2017 to %51 in 2018 [@Vav18], which suggests Spanish market is way behind its more competitive counterparts.

In this scenario, applied research that reduces the gap from the theory of maintenance to its final application and makes our industry competitive gains special interest.

1.4.1 TEKNIKER: Growth makers

Tekniker is a research center located in Eibar. The research areas of Tekniker include Advanced manufacturing, surface engineering, ICTs and product engineering. Its closeness to Eibar, cradle of the Basque machine tool industry, has centred big part of its research in machine tools and manufacturing. Nevertheless, other industrial sectors such as aeronautics, agrofood, energy, infrastructure and health are also covered by their research. During the development of the thesis, the author has grown both technically and personally thanks to the support of the surrounding professional research team.

Intelligent Information Systems - SII

Particularly, thanks to the Intelligent Information Systems unit that has a vast experience on the field of condition monitoring as the various publications and their involvement in Regional, Statal and European projects prove. The following are some projects, in which I have participated, that have taken place during the development of this thesis project and are related to condition monitoring:

- Mainwind+: The goal of Mainwind+ project was to develop a non-destructive oil-debris sensing technique that would allow remote monitoring of Wind Turbine gearboxes. Tekniker took part in the development of the optical debris sensor, the data acquisition and storage platform, and, finally, in the analysis of the signals and their relation with the operation of the Wind Turbine. Mainwin+ was a HAZITEK project (industrial RD support project funded by the Basque Government) that lasted 2 and a half years from July 2016 to December 2018. It was leaded by Ingeteam Power Technology and 9 industrial partners and 7 basque research agents took part, being Tekniker one of them.

- Virtual: This ELKARTEK project (basic collaborative research funded by the Basque Government) dealt with the development of hybrid models. Tekniker developed models used for the commissioning as well as models that were later used in combination with operational data for monitoring purposes. It began on March 2018 and closed on December 2019 being Ikerlan the leading research partner and Tekniker one among the remaining 7 research partners.

- Mantis: This Horizon 2020 European project with call ECSEL-2014-1 consisted in the development of a proactive maintenance service platform and its associated architecture. Tekniker developed a smart monitoring device that monitored clutch-brakes and uploaded the data to a cloud based data repository, later analysing the content of the data with machine learning techniques. The project began in April 2015 and lasted for 36 months. It was leaded by Mondragon Goi Eskola Politeknikoa S. Coop. and a total of 54 European partners took part on it.

1.4.2 EHU/UPV: Give and spread

As stated by the motto "Give and spread", the University of the Basque Country has provided the author with the knowledge foundations for the development of this thesis. Firstly, thanks to the Renewable Energy Engineering bachelor, understanding of basic engineering concepts was acquired, and, later, the Masters and the PhD programs related to Computational Engineering and Intelligent Systems have provided the skills to deal with the understanding and modelling of complex data.

RSAIT

By means of the director of the thesis, the thesis has been especially attached to the RSAIT research group. This department has consistent knowledge of Machine learning and statistics techniques, as well as in their application of the previous to the field of condition monitoring.

1.4.3 Cranfield University: After clouds light

In addition to the EHU/UPV and Tekniker, part to this thesis project has taken place in collaboration with Cranfield University through a brief yet fruitful stay that lasted from September to December 2019. The support of the Through-life Engineering Services Institute has been invaluable, particularly their knowledge related to the monitoring of linear actuators which definitely brought the light to the long awaited success of the hybrid modelling.

2 Motivation

> *I can not train you the way I trained the Five. I now see that the way to get through you is with this. (Shows a bowl with bean buns)*
>
> — **Master Shifu**
> (Kung fu panda)

Shifu was the well-known trainer of the Furious Five, the guardians of the Valley of Peace. However, a new incoming threat forces Shifu to find a new disciple, Po, who, according to Oogway (the creator of Kung Fu), will be the next Dragon Warrior. Po is far from being anything that Shifu could have expected. He is a lazy panda that drowns his sorrows by eating. And Shifu's attempts to train Po as he did with the Five obtain disastrous results. However, Shifu's training outcome changes when he understands that the context he is facing now is different. That the driving force that makes Po an over-weighted lazy panda, his endless famine, can be used to turn him into the next Dragon Warrior.

2.1 Where to push the boundaries: Context

Algorithms are powerful tools that have been proven to be great solutions for some intricate problems. They are equivalent to a tireless group of workers, that can execute orders in a minimal amount of time with minimum resources with surgical precision. However, one of the big problems they have is the lack of intuition, the lack of capability of processing and adapting to the context they are surrounded by. In other words, if an algorithm trained to classify between dogs and cats is shown a cow, it will have no doubt and will classify it as canine or feline and sleep well that night. Due to this lack of intuition, engineers are needed to code the context in the best possible manner so that brainless IA has no choice to take wrong decisions. And that is exactly the topic of this thesis, which is not about the use of the latest and most powerful algorithms, rather than how to properly code the context and use it in the best possible manner to improve the performance of our decision taking army, our future Dragon warriors.

As presented in the previous chapter 1, huge advances in the theoretical aspects of Condition Based Maintenance have been done recently, nevertheless, industrial contexts have some complexities that are sometimes overlooked in research works.

The use of "Context-awareness" and "Context" is widely adopted by pervasive and mobile computing field [SGW16] but its definition is somehow fuzzy. In an attempt of clearing out the fuzziness of the term, a good definition was provided by [DAS01] which defines context as ***"Any information that can be used to characterise the situation of entities (i.e. whether a person, place or object) that are considered relevant to the interaction between a user and an application, including the user and the application themselves"***. As the concept of context has not a noteworthy presence in the field of maintenance [SW15], we have adopted the definition to the field adapting the previous definition. For us, context would be: ***"The information considered relevant for the characterisation of the interaction between an asset and its monitoring"***.

Following this definition and, according to our experience, industrial scenarios are governed by two key factors that influence the application of state-of-the-art monitoring algorithms in industrial applications and, at the same time, are interrelated:

- **Use case:** Use case is defined by the asset that is being monitored, the faults it is expected to have (if there is any knowledge about it), the sub-components, the number of units of the asset that are available (mass-produce non-mass-produced), also, and most importantly the operating conditions: steady, varying, commanded by demand or by external factors, repeatability, etc.

- **Data availability:** Data availability refers to the kind and type of data that is available for the development of the monitoring process. This includes, the sensors used for the monitoring with their sampling frequencies, the external conditions that are recorded (temperatures, other sensors reflecting operating condition but not related to health), the historic data of the machine available, the types of records stored in the historic data (are there any faults? Is everything nominal data?), etc.

The degree of complexity and the maturity of the monitoring systems will be directly affected by these factors that compose the context. To transfer a monitoring approximation to a final application it must have similar if not equal context. At the same time, the knowledge obtained in similar context is quite generalizable, and it might be transferred to a similar context.

This is the main point of this thesis work that identifies three industrial contexts from three particular use cases. In each particular case an approach that improves the current monitoring of the asset is presented, and, as the approaches are based in the concept of context and are use case agnostic, we support that the approach can be transferred to similar contexts.

The concept of context is leveraged by the proximity Tekniker has to real industrial applications, and drives this work to focus the research on a way that allows the transfer of the research outcomes to the industrial applications. In the current scenario we are in, this means working with non-mass-produced machines that are too complex to simulate, data that is difficult to obtain, lack of knowledge regarding most important indicators and varying operating conditions that are not commanded.

2.2 Research question

Based on the previous experience acquired in Tekniker and its closeness to real industrial applications a set of common scenarios is identified. These scenarios include:

- Non mass-produced machines operating under steady conditions: This scenario is given typically in machine tool. Generally, the number of machines is reduced, and the knowledge and understanding of the possible faults is scarce.
- Machines operating under varying operating conditions which is not commanded: This scenario represents machines where forcing certain operation is not possible and there is need to obtain a health indicator.
- Diagnosis of machines that have no fault related records: This scenario applies mostly to cases where the number of machines is reduced and, hence, the historic database is also reduced and has no records of faulty machines. Therefore, developing diagnosis algorithms under these conditions is complex.

These contexts identified by Tekniker are quite frequent. That is, clients and machines can vary, but the limitations and problematic are similar. For that reason, this project formulates and tries to answer the following Research Questions, that provide solutions to the monitoring problems that fall in either of the previous contexts:

1. What is the maximum information that can be obtained from a steady operation context?
2. How can varying operating context be analysed and stabilised to ease machine monitoring?
3. How can the diagnostic knowledge of a data-based algorithm be expanded in a context with no faulty data?

This thesis project is an attempt to provide answers to this set of Research Questions.

2.3 Research outcomes

Regarding the scientific contributions generated during the research, the following list displays the articles that have been published and include contributions of the author.

(a) López de Calle, K., Ferreiro, S., Konde, E., Bravo, I., Arnaiz, A., Sierra, B., 2017. Feature selection for remaining useful life prediction of spur- gears. 30th International Conference on Condition Monitoring and Diagnostic Engineering Management (COMADEM 2017), Jost Institute for Tribotechnology, University of Central Lancashire Preston PR1 2HE Lancashire, UK. 8.

(b) Imaz, I.B., López de Calle, K., Ferreiro, S., García-Arribas, A., Ardakani, H.D., Lee, J., 2018. Fingerprint analysis concept for gearbox health monitoring on speed transitory conditions using motor current signature analysis, in: PHM Society European Conference.

(c) López de Calle, K., Ferreiro, S., Arnaiz, A., Sierra, B., 2018. ***Comparison of Automated Feature Selection and Reduction methods on the Condition Monitoring issue***. Procedia Manufacturing 16, 2–9. https://doi.org/10.1016/j.promfg.2018.10.150

(d) López de Calle, K., Ferreiro, S., Arnaiz, A., Sierra, B., 2019. ***Dynamic condition monitoring method based on dimensionality reduction techniques for data-limited industrial environments***. Computers in Industry 112, 103114. https://doi.org/10.1016/ j.compind.2019.07.004

(e) López de Calle, K., Ferreiro, S., Roldán-Paraponiaris, C., Ulazia, A., 2019. ***A Context-Aware Oil Debris-Based Health Indicator for Wind Turbine Gearbox Condition Monitoring***. Energies 12, 3373. https://doi.org/10.3390/en12173373

Additionally, the following work is being considered for publication by the Journal of Intelligent manufacturing.

(f) López de Calle, K., Ruiz, C., Ferreiro, S., Arnaiz, A., Gómez, M., Sierra, B., Starr, A., 2020 **Hybrid modelling for linear actuator diagnosis in absence of faulty data records**.

From the previous, only those publications with major contributions of the author and answering the research questions proposed in this thesis are appended in Part II (titles in bold).

2.4 Structure of the thesis

Alongside this work three different cases of study are presented: rotating machinery, wind turbines and electromechanical actuators. In all of them our goal has been similar: to develop algorithms that help us knowing more about the current health status of the machines. At the same time, each case of study is immerse in a different context with different constraints, which has required the development of adapted solutions while allows to employ different advantages of the context.

Each context and case of study are described in separate chapter as a brief introduction to the articles, that are the major contribution of this work. Additionally, for simplicity, each chapter has been structured following the same pattern:

- Background: Describes the particularities of the studied context.
- Case study: Provides information of the machines.
- Innovation: The novelties presented in the project are described.
- Conclusions: Summarises the findings.
- Publications: The publications related to the research are listed.

Each chapter can be briefly summarised as follows:

Chapter 3: Steadily operating machines

This chapter describes the works related to rotating machinery carried out to obtain the maximum possible knowledge from machines that operate under steady conditions.

Chapter 4: Finding stable and repeatable contexts in real varying operating conditions

In this chapter the research done in the study of wind turbines is introduced, where the operating conditions are varying and it is not possible to directly compare measurements.

Chapter 5: Extending context of machine learning algorithms

In this chapter the limitations of contexts without faulty data are presented. In particular, a procedure to diagnose faults in a linear actuator which is based on hybrid modelling is studied.

Chapter 6: Conclusions

Finally, this chapter answers the research questions according to the findings of this thesis project and summarises the most important aspects of the research.

The following Table 2.1 contains the relation of each chapter with the research questions and the studied contexts.

Article	CS	RQ	Chapter	Context
(c)	Rotating machinery	1	3	Steady operation
(d)	Rotating machinery	1	3	Steady operation
(e)	Wind turbines	2	4	Varying operation
(f)	Linear actuators	3	5	Lack of faulty data

Tab. 2.1.: Relation of published works with research questions and contexts

For a better understanding of the thesis, readers are encouraged to read the related publications of the chapters 3, 4 and 5, once the respective chapter is read. After finishing the previous chapters and the related publications proceed to the general conclusions.

3 Steadily operating machines

> *The pessimist complains about the wind; the optimist waits to change; the realist adjusts the sails.*
>
> — **William Arthur Ward**
> (American writer)

Regardless of the attempts to reduce the failures of the machines, degradation and deterioration occur, and, in the long term, machines are condemned to deteriorate and fail. In an industrial context, there are various possibilities to deal with this fact, such as: apply corrective actions after the failure occurs to restart activity; try to detect the incipient fault to stop the machine before it fails catastrophically; or, besides detecting the incipient fault and stopping the machine, try to identify which indicators are more related to the fault, so that this fault can be recognised better in the future. Similarly to the dilemma presented by Arthur Ward about the different possible actuation to take under an apparently adverse context, of course, we opt to adapt our monitoring system to gain maximum benefit from the inevitable faults of the machines.

3.1 Background

As already detected by Casoetto et al. in [Cas+03], there are two main approaches for condition based assessment (nowadays better known as Condition Based Maintenance):

- The first approximation attempts to detect the condition of the assets by recognizing indicators of failure. For that purpose, this approach requires expert and/or a priori knowledge about the assessed machine or process, as the failure modes should be known in advance to be able to compare nominal indicators against current ones to characterize failures. This fact makes this approach very application specific.

- The alternative approach implies characterizing the condition of the machine by overlapping the latest system records to the ones that are known to be healthy, assuming the difference or drift between records is caused by the degradation. This approach has the advantage of not requiring neither faulty data nor large historical data.

Paying attention to the evolution of Condition Monitoring during the last years, it is possible to see that still today most of the research falls in either of the previous categories.

For example, the following works would fall in the first category: starting with the work of Djurdjanovic et al. [DNL02], where worn and sharp tools of CNC lathe machines were diagnosed by using time-frequency features. Later, in [SR09] a spur bevel gearbox test rig was used to extract vibrations under varying conditions, treat them with wavelets and, finally, use a decision tree in order to diagnose faults. Followed by the work of Li et al.[Li+11], where a two-stage feature selection algorithm was developed in order to classify among eight fault states, including combined faults. Going to the latest works, such as, [Bra+17], where the application of Wavelets and Double synchronous averaging techniques for motor current signature are compared in the diagnosis of faulty spur gears; or in [Zha+20], where various datasets are processed using state of the art Deep Learning techniques in order to assess the performance of the algorithms together with different pre-processing of the data.

At the same time, the presence of works using the second approach is slightly lower, but there are still cases. Such as [Cas+03], where the performance confidence value (CV) was used the visualize the evolution of wear in a welding machine in which Current, Voltage and Force were monitored. Or the research started by Ruiz-Carcel and Starr [RS15] were a physical model of an Electro-mechanical actuator was built and the evolution of the features was assessed by using a PCA and monitoring Q and T^2 statistics. Later in [RS18b] this approach was validated by replicating it with data created in a test rig. And in even more recent research [Maz+19], Mazzoleni, Previdi, Scandella and Pispola adopted similar procedures also for the monitoring of actuators, but added an additional layer of accuracy, as they use metrics to monitor threshold value violations.

The first approximation has some advantages, as having faulty records allows proving the effect of different signal processing algorithms as well as the accuracy of machine learning algorithms. This assumption is easy to reproduce in a test rig or a physical model (as in the majority of the works following this approximation), but this situation rarely occurs in industrial applications.

Although these approaches can be useful to gain insight in the adequateness of signal processing techniques or the possibility to discern between faulty and non-faulty data, in the real life, transferring the diagnostic models to the application is quite complex. Mainly, because the operating conditions are hardly the same as in the test rig, and, also, because real use cases tend to have more environmental noise that challenges the robustness of the algorithms. Additionally, it is not possible to directly put the diagnostic algorithms in production, and the faulty data records they need to be retrained are scarce in real life.

In that sense, the alternative approach is better suited to real industrial environments, where faulty data is scarce, and thresholds need to be learnt on the fly. Assuming steady operating context and assuming all the variability is produced by wear allow the second approach to be more appropriate for the industrial context. Nevertheless, the drawback of using a simpler procedure is that these models provide less information about the state of the machine. As only information regarding the detection of the fault is given, but not information about the type of fault.

The work presented in this research has tried to exploit those steady context in order to maximise the information obtained during the monitoring by using Dimensionality Reduction algorithms. In this way, relevant information related to the features of interest is obtained and, hence, the gap between the first and the second approximation can be reduced as more machines are monitored.

3.2 Case study: Rotating machinery

Rotating machines can be found in numerous industries, such as oil industry, aviation, mining and transportation, among others. Furthermore, they tend to operate under adverse conditions suffering high temperatures and high loads, hence, they suffer from performance degradation and mechanical failure [Li+17]. Rotating machines include pumps, motors, rolling bearings, gears, gearboxes, shafts, fans... and they are among the most important equipment in modern industrial applications [Liu+18]. For example, bearings, which are mechanical components used frequently in most rotating devices, can constitute as much as 44% of the total number of faults in some devices [Cer+18].

Due to the criticality of rotating machines, they have been widely addressed in the CM literature. Even though different sensors can be used for their monitoring such as acoustic emissions [Cer+18], temperature, pressure, oil analysis, noise [ASC11] or motor current signature [Bra+17], most of the works rely on the use of vibrations, as excess vibrations reflect unbalance, misalignment, worn gears or bearings, losseness among others as Vishwakarma [Vis+17] states.

Regarding the different signal processing methods used to work with vibration signals, three categories can be distinguished: time domain, frequency domain and time-frequency domain techniques; the first includes the use of statistical moments such as RMS, Kurtosis, Crest factor or noise filtering techniques such as Time Synchronous Averaging or the envelope; the second, includes the transformation of raw signals into frequency domain by means of Fourier Transform, in general, frequency domain features are better indicators of faults, as the resonance frequency component (or fault component) can be better detected in frequency domain when signals are stationary; lastly, for non-stationary applications, time-frequency techniques are used, such as the Short Time Fourier Transform (STFT) or the wavelets, that allow the use of long time intervals for more precise low-frequency information or short time intervals for better precision of high-frequencies.

In this research rolling bearings and gears working under steady conditions have been considered. In addition to stopping the machine before having catastrophic failure, obtaining as much knowledge as possible in relation to faults/degradation during the monitoring of the bearings and gears is attempted. For that reason, focusing much in the complexity of the features extracted from the vibration sensor is out of the scope of the research, hence, simple statistical descriptors and basic frequency domain descriptors are extracted.

3.3 Innovation

As discussed in the the previous section 3.1, the second category of monitoring works share a common denominator: constant operating conditions. This fact is not trivial, as this context can be exploited with algorithms based on statistical distributions, as any change in the distribution is given by the degradation of incipient faults in the system due to the lack of interference of varying operation. This simple assumption about the stability of the context has been widely exploited for failure detection, instead, this work goes a step further. In addition to answering the question "When should the machine stop before being too late?" the question "If the machine is stopping due to wear/fault, which indicators are reflecting this damage the most?" is also answered.

Two different datasets are employed in the research: the first one is generated in FZG test rigs used to test lubrication as well as other aspects of spur-gears; the second dataset is taken from an open repository, created for the PRONOSTIA challenge and consists of run to failure tests of bearings [Nec+12].

In both cases, features are extracted from the raw signals and the set of features is reduced with a dimensionality reduction technique. Four different techniques are compared: LDA, Relieff, Autoencoders and PCA. After applying the techniques and reducing the dimensionality of the initial dataset to a single dimension, a quality control chart is used to determine when the system is out of control. However, as the reader might have realised, some of the previous algorithms require of class values which is not something that it is available. And there is where advantage of the stable context is taken to go one step further and consider that the latest data taken from a machine has to be always in the same or worse condition than in previous instants (considering no maintenance actions). This way, it is possible to consider the problem as a classification problem and use supervised dimensionality reduction techniques.

In addition, the dimensionality reduction process is repeated periodically in order to determine which features are showing the greatest differences in comparison to the initial data-window in order to obtain extra information related to the best indicators of machine degradation.

Lastly, some considerations about the final application of the algorithms are included by measuring their performance and other aspects with a set of metrics designed with the industrial needs in mind: cost, interpretability, and effectiveness. This set of metrics helps to better understand other parameters that sometimes are overlooked, such as computational costs or the comprehensiveness of the feature reduction. The metrics here presented provide a fairer framework to compare four algorithms with monitoring purposes.

3.4 Conclusion

According to the findings, it is possible to utilise Dimensionality Reduction techniques to synthesise the information of large amounts of variables being faults detected at the same time. Regarding the comparison among algorithms, two algorithms obtained better results: Relief and LDA. The first one, because instead of feature projection makes ranks of features and, therefore, they are still understandable. The second, because its accuracy was greater than the rest of the algorithms. It is not coincidence that both best algorithms were supervised, as including class value benefits by giving extra knowledge. None of the previous could have been used without using domain knowledge (comparing nominal and latest observations), hence, it is clear that the application of ML algorithms benefits from domain knowledge. Or, in other words, better understanding of the problems to solve improves the chances of using tools better suited for its solution.

Regarding the metrics used in the work, it has to be considered that other works just focus on the accuracy, however, other aspects also affect in the selection of the "best" algorithm. Therefore, these metrics provide a more complete framework of comparison. Additionally, taking a look of the sub-dimensions the metrics that were developed were validated, as they reflected correctly their original purpose.

Finally, it is proven that obtaining additional information while the monitoring is carried out when the context is kept steady and all variance is attributed to degradation is possible. This has been achieved by using supervised dimensionality reduction algorithms that distinguish the features that change the most from the beginning to the stopping point.

3.5 Publications

The study carried out on this topic led to two publications:

The Tesconf conference that took place the 6th and 7th of November 2018, in Cranfield (UK), the work "Comparison of Automated Feature Selection and Reduction methods on the Condition Monitoring issue" was presented and later published in Procedia Manufacturing.

López de Calle, K., Ferreiro, S., Arnaiz, A., Sierra, B., 2018. Comparison of Automated Feature Selection and Reduction methods on the Condition Monitoring issue. Procedia Manufacturing 16, 2–9. https://doi.org/10.1016/ j.promfg.2018.10.150

Additionally, the article titled "Dynamic condition monitoring method based on dimensionality reduction techniques for data-limited industrial environments", authored by part of the Intelligent System unit in Tekniker (Kerman López de Calle, Susana Ferreiro and Aitor Arnaiz) together with the Robotics and autonomous systems group from the University of the Basque Country (UPV/EHU) represented by Basilio Sierra was published in the journal Computers in Industry.

López de Calle, K., Ferreiro, S., Arnaiz, A., Sierra, B., 2019. Dynamic condition monitoring method based on dimensionality reduction techniques for data-limited industrial environments. Computers in Industry 112, 103114. https://doi.org/10.1016/ j.compind.2019.07.004

Both articles are attached in PART II of the work.

4 Finding stable and repeatable contexts in varying operating conditions

> *Science is about sailing (and taming) uncertainty.*
>
> — **Fernando Blanco**
> @FBpsy on Twitter, Experimental Psychologist

Starting in late 2019 and during the first and second quarters of 2020 with a yet unclear expiration date, a previously unknown menace hangs over the world. This menace takes the name of Coronavirus (COVID-19) and with a high mortality and basic reproduction ratios, this virus spreads fast across the world causing devastation on its way. Most of the countries are forced to enact complete lock-downs of the citizens in their homes and only minimum services (first sector) are kept working in many countries. Despite the measures taken to avoid the spread of the virus, it has caused numerous fatalities as well as incipient worldwide economical crisis. In such a deplorable situation, many citizens and media do not understand why governments did not act faster. Meanwhile, governments hide behind the fact that they act following scientific criteria. Once again, science/scientist are to be blamed for the bad decisions.

In that context, Fernando Blanco (@FBpsy on twitter) warns with his tweet media and other displeased citizens about how science works: "If you expect science to provide certainty and assurance, if you think that (scientific) debate is symptom of ignorance, then you have now clue of what science is". The truth is that science is based on evidence, therefore, as new evidence arises previous assumptions are revised and, in some cases, they may be rejected and as new assumptions are validated.

The scope of this thesis is not related to the COVID-19 (despite great part of it being written during confinement). However, during the coverage of our field (machinery monitoring algorithms) we have had the chance to study new evidence that has allowed us to revise and validate hypothesis formulated by previous works, as well as to draw new hypothesis.

4.1 Background

Data-based algorithms used in condition monitoring are developed typically following this procedure:

1. The faults that need to be diagnosed are identified.
2. A test rig where these faults can be seeded is built.
3. Sensors/signals that could be helpful to diagnose the faults are recorded during the tests.
4. Meaningful features/transformations are used to diagnose or new features/-transformations are tested.
5. Data-based algorithms are tested to determine their diagnostic capabilities.

This procedure is vastly extended and has helped in the development of many signal transformations and improved both the extraction of diagnostic features and the use of diagnostic algorithms. Nevertheless, test rigs are built following certain assumptions, and, even if they are built to resemble as much as possible to the final application, there are some aspects that are difficult to reproduce and some external influences that are neglected in their development. Consequently, the evidence and conclusions drawn in a test rig (with typically non varying conditions) might not be accurate in a final application were operations are not held constant (they might not be conclusive under the new evidence). In other words, the operation context of test rigs is not totally accurate reproduction of the application context, which is hardly stationary.

Being aware of this fact, the applied research in Tekniker has lately focused on the development of repetitive tests (known as Fingerprints) in the applications that allow forcing machines to work under certain repetitive conditions. These Fingerprints are taken periodically and allow obtaining stable data from machines that work under very transient conditions. This way comparing measurements from one point in time to another is possible, as the measurements are taken in similar steady contexts. Fingerprint has been satisfactorily used for the monitoring of machine tool and other types of machinery, and it eases the knowledge transfer from test rigs to final applications, as the Fingerprints can be taken in conditions that are closer to the ones of the rigs. A representative example is the work presented by Ferreiro et al. in [Fer+16].

Nevertheless, not every machine can be forced to operate in a specific way, or the cost of doing it is too expensive for the sole purpose of monitoring. For example, the case study here presented, wind turbines (WT), can not be forced to operate in a desired way as the driving force is the wind, which can not be controlled. Therefore, alternatives to Fingerprint are required for their monitoring, which is what the the following sections present.

4.2 Case study: Wind turbine gearboxes

The increasing electrical energy needs together with the abundance and availability of wind energy are triggering the growth of wind turbine industry [Tch+14; Ham+09; NW13]. This surge in wind energy demand comes together with an increase of the size of the turbines that has been found correlated to higher failure rates [Ech+08; Su+17; Gar+12].

Considering Operation & Maintenance (O & M) of wind turbines can comprise from 10-20% of the total cost of energy (COE) according to Tchakoua et al. [Tch+14], and it can reach up to 30% in big of shore wind farms [CZH15], the use of Condition Monitoring Systems (CMS) is a must. CMS are proven techniques that successfully detect health status, detect faults and provide asset remaining life estimations [Tch+14; Ham+09; Gar+12]. The application of CMS in wind turbines supports the identification of the state of health of the turbines remotely and reduces the need of visual and on site inspection for that purpose, which is particularly cost saving in offshore wind farms[NB07].

The various works analysing failure statistics of gearboxes support the thesis of gearbox being one of the most delicate components in WTs [CMM16]. These works show some controversial conclusions regarding the tendency to failure that gearboxes have as Pfaffel, Faulstich and Rohrig recognise in [PFR17], as some of them find high failure rates [HDR06], whereas others have less gearbox related failures reported [PFR17; Su+17]. In any case, most of the studies relate the longest WT downtimes associated to gearbox failures [NW13; PFR17] and find it one of the costliest parts of the turbines [CMM16].

Gearbox monitoring is a widely studied topic in the literature. The prominent approaches in this field are based on vibration, oil debris, acoustic emissions and current signature analysis, among others. Even if there is a prevalence of vibration based monitoring works [Ham+09; NW13; Gar+12], oil debris monitoring (ODM) techniques have been found of interest for gearbox monitoring, because of the higher correlations they show with wear creation as demostrated by Kattelus, Miettinen and Lehtovaara in [KML17], and the additional capability of monitoring the oil quality and the state of other parts of the gearbox [Tch+14; Ham+09].

Nevertheless, the proven advantages of condition monitoring (CM) [NB07; Ham+09] are difficult to transfer to WT use cases, and they have a smaller presence in the literature than test rig based ones [Art+18]. This is because the variability of operation conditions of WTs affects the extraction of indicators while it specially damages the systems of WTs [Ham+09]. Most of the works presenting real in-service WT data are based on the use of SCADA (Supervisory Control and Data Acquisition) data, which is readily available in general. Typically, it is used to compare performances among WTs using power curves [Gon+19]. Additionally, temperatures from the SCADA have been modelled and compared over time to use differences as alarms as the different works reviewed by [TW17] show. However, the success of these techniques is limited [NW13]. Partly, because of external influences (such as the outside temperature) that require the alarms to be manually supervised by operators [TW17].

Consequently, the inclusion of additional CM sensors in operating WTs is flourishing [Gar+12]. And an increasing number of works present findings from real use cases, including some that show oil debris sensors (ODS) for the monitoring of in service WTs as the works of García Marquez et al. [Gar+12] and the one by Nie and Wang [NW13] state. As other works showing data from in service turbines [Dup10; Fen+13; KDT18] they rise similar conclusions: cumulative or averaged values are needed to avoid the noise caused by the varying operation, as the operation affects the sensed magnitude. These findings are also supported by the extensive work carried out by Sheng [She16], were a full-scale WT gearbox of 750 kW is tested with in-line and online sensors and samples taken during the tests. They mention the need of filtering influences caused by operational conditions; recommend to focus in trends instead of in absolute values, and suggest considering big particle size (>14 μm) indicators in particular. Also, they identify that damaged gearboxes have much higher debris generation rates than healthy ones.

Literature related to wind turbine monitoring being considered, the difficulties of gearbox diagnosis are well assessed. Engineering a solution that could cope with the complexities of the varying operating contexts was necessary, which is, exactly, where the contribution of this work resides.

4.3 Innovation

The work here presented adds new evidence to the literature by publishing in service studies of 3 WTs that are monitored with ODSs and shows an approximation used to deal with the varying conditions of the turbines with the aim of developing diagnostic algorithms. These are the two major innovations present in this work:

Firstly, how optical oil debris sensors behave in real operating conditions is studied. In addition to reproducing some monitoring techniques found in the literature, certain actions of the wind turbines (generator breaking and boosting) that might be particularly damaging the gearbox (as hypothesised by research in test benches) is analysed. For doing so, machine learning techniques are employed to avoid manual identification of these actions over the whole dataset. Later, the relation between these operations and the debris sensors is studied by analysing the distribution of the correlation.

Additionally, in an attempt to provide a health indicator of the gearbox, the concept of Fingerprint is extended to contexts with varying conditions. Basically, the premise is as follows: if it is not possible to force certain repetitive operation along the time to obtain comparable measurements, lets find which operations are better suited for taking measurements. In order to give answer to that premise, a segmentation is carried out. Several operating regions (OR) are defined and all operation sequences (OS) that belong to each of the 5 compared ORs are identified. Once all this sequences (or operation states) are identified, their appropriateness to be used as a basis for a Health Indicator (HI) of the gearbox is studied. The aspects considered are: the frequency, how many OSs occur weekly; the duration, mean time of OSs and; the stability or the variation of the different signals during the OS. Finally, once the best OR is identified, it is possible to develop a diagnostic algorithm that filters undesired context and only takes values of stable and repeatable contexts.

4.4 Conclusions

This work reveals interesting findings. Firstly, it is possible to validate some of the conclusions that were previously hypothesised in test rigs, such as the noise found in particle counting sensors [Fen+13; Dup10; She16] and the validity of using cumulative particle rates for reducing that noise [Fen+13; Dup10]. Also, greater differences in the ODS than in the rest of SCADA signals in damaged turbines were found, particularly in bigger particle sizes, indicating the sole use of SCADA might be better backed by additional sensors (as Feng et al. suggest in [Fen+13]) and remarking the use of bigger particles as earlier indicative of failure (as recognised by Sheng [She16]).

However, some discrepancies with previous works arose. For instance, the analyses of brakings and boosting shows an increase in particle creation during brakings whereas no increase is detected during the acceleration, in contraposition to the findings of [She16].

In addition, some inconsistencies among the ODS of the wind turbines were found, that were not justified according to our knowledge. This incongruence may be related to the nature of the machines, and could be caused by the natural variation of the systems, however, this variation is rarely considered in test rigs.

Finally, regarding the need of finding stable and repeatable contexts to implement algorithms, a procedure developed to extend the Fingerprint approach to varying operation environments that do not allow imposing the operation of the machine is validated. The approach, that consists on isolating measurement points that have been taken in similar conditions and are frequent, last long enough to avoid inertia of the machine and are stable, has served as basis to develop a Health Index that identifies the most damaged machines and allows tracing their evolution to particle threshold levels indicative of malfunctioning.

4.5 Publications

A conference and a prominent publication were developed throughout the study of this topic in the thesis project.

The DEEPWIND 2019 conference in Trondheim, Norway, that took place from the 16th to the 18th of January 2019. The slides that were presented are available in the final report of the conference, under the title: "Excluding context by means of fingerprint for wind turbine monitoring" available here [@Tan19].

The article titled "A context-aware oil debris-based health indicator for wind turbine gearbox condition monitoring", authored by part of the Intelligent System unit in Tekniker (Kerman López de Calle, Susana Ferreiro and Constantino Roldán) in collaboration with the Department of NE and Fluid Mechanics of the University of the Basque Country (UPV/EHU) was published in the special issue Maintenance management of wind turbines of the MDPI Energies journal. A copy of the article can be found in PART II.

López de Calle, K., Ferreiro, S., Roldán-Paraponiaris, C., Ulazia, A., 2019.A Context-Aware Oil Debris-Based Health Indicator for Wind Turbine Gearbox Condition Monitoring. Energies 12, 3373. https://doi.org/10.3390/ en12173373

5 Extending context of machine learning algorithms

> *Ez ikusi, ez ikasi - Do not see, do not learn.*
>
> — **Antoine Thomson d'Abbadie**
> (Explorer, geographer, ethnologist, linguist and astronomer)

The philanthropist Antoine Thomson d'Abbadie was, among other things, an eager astronomer. His passion for the astronomy was so deep that his castle in Hendaye had its own astronomical observatory. However, during his study of the refraction constant he could not see the peak of the mountain Larrun (the highest point surrounding the area) with his telescope, because the castle was between the mountain and the observatory. If he wanted to see the peak of the mountain, he had to remove the walls of the castle, and so he did. He took a cannon and made a hole by shooting in the direction that the telescope would see the peak of the mountain. The experiment did not end up successfully because the light could not pass through the thick walls due to diffraction. But, instead of being ashamed of his failure, he decided to keep the hole and wrote around it "Ez ikusi, ez ikasi." in other words "Do not see, do not learn".

Some centuries later, one of the problems we face in the monitoring of industrial assets is of similar nature, as the machines we attempt to monitor are not produced in a serialised way and have not had sensors before, therefore, we lack of faulty data records. In this context, the algorithms we attempt to train for failure detection find themselves with the same difficulties Antoine had, because, algorithm that does not see a failure, algorithm that will not be capable of diagnosing it. In other words, *Ez ikusi, ez ikasi*. Consequently, some techniques are needed to overcome the absence of faulty data, and cannons do not seem to be an appropriate option.

5.1 Background

As discussed in chapter 3, in the cases where the behaviours of defective machines are unknown, it is necessary to use methods based on the statistical properties of the sensor measurements to detect anomalous behaviours. This is the case of statistical process control (SPC) methods, which are the most common and most used among them. However, these methods do not have the capability of diagnosing faults, which is the next step in condition monitoring.

A typical approach used when no data is available is using physical models. These models, based on the mathematical formulations of first laws represent the approximate behaviour of the system, and can be used to learn how the time-varying systems might behave under different operating conditions, transients and environmental conditions [Gal14]. In some cases, they might even include the simulation of failures, which helps to understand how the machine could operate under faulty conditions.

Nevertheless, the development of physical models is subject to some limitations that harden their implementation, such as: the need of in depth understanding of the physics governing the system; the trade-off between the fidelity of the model and the computational cost of the simulation; the time needed to develop the model and the need of taking assumptions and tuning parameters in order to simplify the problem; as the assumptions and the parameters directly affect the fidelity of the physical model [Let+15; Gal14; LK14].

In contrast, purely data-based models do not rely on the understanding of the causality behind the signals of the model, as they infer these relations from the input data to the output. Instead of requiring an in depth understanding of the physics of the system to build a model, they use different statistical methods to model the input/output relation, relying for that on previously observed data [AKC15]. This methodology has the advantage of being faster and cheaper to develop and it is more appropriate when many records of machines are available. Additionally, it has shown better results than physics-based modelling in some complex engineered methods such as aircraft engines or/and wind turbines [Li+19].

Considering the pros and cons of each modelling technique and the limitations that each particular problem has, some authors have combined both of the aforementioned strategies in order to solve problems that are not well suited for either technique and require of the strengthens of both. Particularly for prognosis purposes, where the underlying relation governing the degradation of the system is unknown, plenty of works have relied on the hybridisation of different models. The reasons for combining these modelling approaches vary from one work to another, including: difficulties in the obtention of completely accurate physical models [Mat+15]; the high computational cost of running parallel models [Bal15; Nar+10]; the lack of understanding of all the physical effects governing the system [ACK13]; an improved pattern recognition by using data-driven methods [Bal15]; and, finally, the lack of data records (mostly related to failures) [**En2014**] or fault growth models [Bal15].

5.2 Case study: Linear actuators

Linear actuators are found in a variety of systems, such as valves, door openers, aircraft systems, machine tools, robot arms, etc. Lately, electromechanical actuators in particular are gaining importance in aeronautics, as they have some advantages over the typically used hydraulic actuators such as increased safety and reliability; easier and reduced maintenance; reduced weight, volume and complexity of transmission paths; and higher efficiency [Qia+17]. However, even if numerous advances have been carried out in their development over the last years this technology is not mature yet and needs further research.

In recent years, great part of the effort spent in the research of linear actuators has been divided in two main trends: The design of more robust or fault tolerant actuators, and the development of condition monitoring systems for already existing ones.

Essentially, the idea underlying fault tolerant actuator is dividing the actuator in smaller actuators that are connected in series and parallel so that in case any or some of them fail the actuator could still operate. The concept of this high redundancy actuators (HRA) was developed in inspiration of the human body, where a muscle is composed by many individual muscle cells and each one of them contributes to the force and travel of the muscle [Dav+08]. Most of the research conducted in that area discuss the different possible set-ups (series/parallel and number of actuators) and their tolerance to failure. Some examples are the works [ADW14; ADW16; MVS18] where different set-ups are tested in models and validated in rigs to asses their tolerance to failure.

On the other side, in addition to the research carried out in the development of more robust actuators, there is a second research area related to the condition monitoring of the actuators. Recent works in this area have focused in the detection of damages and health status of the actuators in order to provide better diagnostic abilities.

In the work carried out by Ehrmann, Isabey and Fleischer [EIF16], the possible sensors to monitor rack and pinion sensors are discussed. They research previous works related to the monitoring of rack and pinions, and, due to the lack of them, they compare rack and pinions to similar components. Finally, they determine which sensors could be better suited for their monitoring, concluding vibration analysis, position error and dynamic characteristics of the control loop and acoustic emission could be well suited for monitoring rack and pinions.

Ferreiro et al. [Sus+13] develop a physical model of an actuator where four different faults are seeded: degradation of the motor, increase in friction, backlash and external force. The actuation process is segmented and descriptors are extracted, using these descriptors different diangostic Machinel Learning algorihtms are trained and the best ones (ID3 classification tree and Bayes Net bayesian networks) are used as fault indicators. They face some difficulties when diagnosing mild faults but the overall classification accuracy is high for severe cases.

Voltage and current signals in time and time-frequency domain are used together with PCA, optimal transformation and a support vector machine algorithm to monitor an electromechanical actuator by [Kn+15]. Their algorithm is capable of detecting stroke related failures and reduced voltage.

Considering electrical failures, in [Cai+16] a data-driven bayesian network based algorithm with the purpose of detecting failures in the inverters of electromechanical actuators is developed. They extract some signal features using fast fourier transform and the dimensions of the samples are reduced using principal component analysis (PCA). A combination of both experimental and simulated data that includes short circuit and open circuit damages with different combinations of power switches of the SPWM is used to detect failures.

An electromechanical actuator used in an unmanned underwater vehicle is modelled in [KM18]. They also model two additional faults: load faults and coupling loss fault. The three models (nominal and faulty ones) are run in parallel and paying attention to the residuals they determine whether the actuator is operating correctly or it is damaged. The residuals they compare are: Current, position and angular speed. An electrohydrostatic actuator is studied in [GH14], they apply a state estimator referred to as SVSF-VBL which is based on the smooth variable structure filter (SVSF) and sliding mode concepts. They conclude that besides getting good estimates, SVSF-VBL approach is valid for detecting failures.

In the work carried out by [SI17] vibration sensors are used to distinguish faulty or worn ball screws from others in good shape. They design a device capable of harvesting energy from the actuator movement that measures vibrations at the same time. At the end they are capable or recognizing the worn actuators.

There are also some works that use the combination of models (hybridisation) for diagnosing and prognosing failures in linear actuators. This is the case of the research started Narasimhan [Nar+10] and continued by Balaban [Bal15]. These works develop a complete condition monitoring system for electromechanical actuators. They follow the TRASCEND diagnosis architecture, where a first diagnostic layer determines if the actuator is deviating from its common behavior, then a qualitative model is triggered in order to diagnose the failure and is supported by a data-driven model that improves diagnosis accuracy. After that, a Gaussian process regression (GPR) model (because there is not explicit fault degradation model) is used to estimate the Remaining Useful Life (RUL). In their in depth study they include various fault modes (jam fault, spall, motor failure, sensor fault), as well as different load profiles generated simulating real operating conditions.

5.3 Innovation

This work presents an approach that would like to cover those assets in which there is no faulty data available but diagnosing the assets is needed by means of condition monitoring algorithms. As the training context of diagnostic algorithm is limited by the lack of fault related data, it is necessary to expand that context to a wider one that includes faults. This is only possible by either damaging the machine in purpose (which implies downtime and highs costs); waiting until the machines suffers from all the possible faults (time consuming and undesirable) or by recreating the physical behaviour of the faults with a physical model, which is the path we have followed.

A physical model of an electro-mechanical actuator which is a twin of the real actuator is built. The most common faults (as indicated by the FMECA analysis) are seeded in the model and data records are created. Features are taken from both the real signals (the ones from the test-rig) and the synthetic signals (the ones created by the physical model). Because of the deterministic behaviour of the model, additional synthetic observations are generated by adding the noise from the real descriptors.

Due the difficulties faced during the combination of both data sources, a method for the deletion of features that are too different among data sources is developed. This way features that are inconsistent from one data source to the other are removed so that the dataset is consistent while the detectability of the faults is maintained with the remaining features. For that purpose, the method performs a classification of faults using only synthetic data and also classification of data between real or synthetic. This is repeated iteratively removing the feature that shows greatest differences among classes in the distinction between real/synthetic data. At the end, the accuracy of both classifications is analysed (real/synthetic and faults diagnosis) and the features left at the iteration with the worst real/synthetic results and the best fault diagnosis results are kept.

Once the data are combined, different scenarios are tested using Linear Discriminant Analysis (LDA) algorithm. In these scenarios LDA is trained using non-faulty data from the real actuator and the faulty and non-faulty synthetic data generated on the physical model. Among these scenarios single load cases, multiple load cases, class imbalance corrections, PCA projections, the effect of severity and chronological predictions are tested. The validation of the model is done by using real data which contains faults but has not been shown to the ML algorithms before.

Additionally, it needs special mention that the work follows the open research approach. The aim of this approach is to provide transparent, open and reproducible research outcomes that can be critically reviewed and reused by other researchers. Particularly, the data has been made open-access (it was actually open access thanks to Cranfield University [@RS18a]) and, also, the source code used for the analysis has been uploaded to an open access repository [@Lop+20].

5.4 Conclusions

In this work the possibility to augment the learning context of the algorithms by creating synthetic datasets with physical models is proved. Furthermore, algorithms are also used to detect previously unseen operating conditions. Adding to that the fact that better results are obtained when creating models that diagnose in a single load, the great potential hybrid modelling has for condition monitoring is highlighted.

Consequently, the use of digital-twins/hybrid models to avoid the limitations of industrial scenarios with lack of faults is validated. This is done by showing to data-based models how machines behave under unregistered operations or under unseen faults, which enables the algorithms to later detect these operations and faults even if there was no data obtained in these situations. This is the other interesting aspect of this work, that covers different load cases, which is more representative of real operation conditions.

Additionally, similarly to other works in the literature, difficulties have been encountered during the development of the physical model. As the model could just partially resemble the real behaviour and produce features that do not exactly resemble the real features. In relation to that, a method that allows removing the features that do not represent consistently the reality according the measured features (the ones from the rig) has been developed. The results suggest the method improves significantly the accuracy of the diagnosis algorithm. Furthermore, the method is generic enough to be transferable to other works attempting to hybridise data from physical models and real operation.

5.5 Publications

The research of limited contexts have been incorporated in the work "Hybrid modelling for linear actuator diagnosis in absence of faulty data records". The paper awaits the review by the Journal of Intelligent Manufacturing, the pre-print version of the work attached to PART II of the dissertation. In any case, readers are encouraged to access the analysis which is available in the repository [@Lop+20].

López de Calle, K., Ruiz, C., Ferreiro, S., Arnaiz, A., Gómez, M., Sierra, B., Starr, A., 2020 Hybrid modelling for linear actuator diagnosis in absence of faulty data records 20.

6 Conclusions

Here are the conclusions of the 3 year long adventure:

RQ 1: What is the maximum information that can be obtained from a steady operation context?

According to the findings, steady contexts that have been broadly addressed in the literature can be pushed a little bit further. Besides monitoring and detecting anomalies, dimensionality reduction algorithms can be used to gain extra knowledge so that the faults can be related with certain features. Particularly, supervised DR algorithms such as LDA are well suited for this task, as they can be used to distinguish among the early healthy measurements and the ones taken when anomalies are detected. This approximation settles the foundations to use process control algorithms in industrial environments to fill a database with information related to faults and fault indicators, so that future faulty observations can be compared to the historic database and they can be diagnosed. It is concluded that DR can be used to understand which features are related to the incoming fault while the system is monitored.

Steady operation context are quite infrequent in reality even if they are one of the scenarios more explored in the research literature. As mentioned in chapter 4, Tekniker has experience in the development of methods to turn varying operation into steady by meas of Fingerprint, but this approach has limitations, such as the cases where the operation is not commanded by the operator. In these cases steady operation regions need to be identified a posterior which is exactly what is attempted in the following Research Question 2.

RQ 2: How can varying operating contexts be analysed and stabilised to ease machine monitoring?

Chapter 4 has introduced the reader to a context with varying conditions. In this case, the varying operation is given by wind, which is stochastic and is used to command the operation of the turbine. In an attempt to obtain a Health Indicator of the turbine, the full operation of the turbine has been segmented into operation states considering different operation criteria. In order to determine which criteria is best, steadiness, duration, and frequency of occurrence of the operation states has been analysed. Once the best criteria is determined, only the operation states suggested by the best criteria are used to collect data from these operation states. Finally, after filtering the data, thresholds provided by laboratory studies are established to determine maximum bearable wear values.

Regardless of the use case, this procedure can be applied to similar machines where the operation is not steady but some operation patterns are repeated. Varying conditions can be cancelled by identifying and isolating data from repeatable operating conditions that occur during the varying process. At the same time this enables to use the approach proposed for steady conditions in varying conditions. This way even complex varying scenarios can be dealt as if they were steady. Nevertheless, variability of operation is not the only obstacle in the development of diagnostic algorithms, there is also the lack of faulty data, which is exactly what RQ 3 tries to answer.

RQ 3: How can the diagnostic knowledge of a data-based algorithm be expanded in a context with no faulty data?

In chapter 5 a scenario without faulty data is presented in which diagnosing faults is desired. In order to expand the available data for a data-driven algorithm, a physical model that reproduces the operation of the real machine is built and faults are seeded in the model. This way it is possible to use the data from the model (synthetic data) to train the data-based algorithm and, hence, diagnose faults from the machine before they ever occur.

This procedure can be used in other machines if they can be modelled and the most common faults are known. In conclusion, it is possible to expand the knowledge of the data-based algorithms by adding new faulty data records from physical models in contexts without real faulty data.

In addition to finding answer to the research questions, other findings that need to be mentioned are:

The implementation of algorithms in industrial scenarios faces additional external difficulties as the lack of mutual understanding between maintenance manager and monitoring algorithm developer. In this regard, the works related to chapter 3 present some metrics as an attempt to settle a common ground. These metrics reflect three aspects that concern managers: cost, effectiveness and interpretability. Each of these metrics summarises the information of other sub-dimensions, that provide detailed information of the behaviour of the DR algorithms under the hood: how much time they require to be computed; how deterministic the results are; various ways of measuring accuracy (true positive rate, positive predictive value, noise detection); the distribution of the weights of the features (how sparse features are); whether they are linear combinations or not; and the change in feature weights from the beginning to the end. This complete set of sub-dimensions can be used to better understand how the DR algorithms work. And also, to set a common language by using concepts understood by both the algorithm developer and maintenance manager that can ease taking the decision of choosing "the best" algorithm, or even simply trusting and relaying on an algorithm to take decisions. Which proves that understanding of the algorithms and the needs to be solved can be matched by a common group of metrics.

Another interesting finding is related to the appropriate use of machine learning algorithms and the benefits domain knowledge provides to ML and vice versa. Starting at chapter 3 where domain knowledge is used to enable classification algorithms (older samples must be in worst health condition); going to chapter 5 where subsequent instances are used for voting and obtaining better results (consecutive instants must be in the same damage stage); going through chapter 4, where, in order to understand whether WTs might be creating more debris during braking and boosting or not, a ML algorithm is trained to identify these instants in the database (improving the knowledge of the domain); one conclusion is clear: domain knowledge can be as beneficial to ML as ML can be to obtain more domain knowledge. The comparison of ML algorithms are quite frequent in the CM community, however, it seems like what makes the real difference is the proper integration of domain knowledge in the problem. Additionally, ML algorithms are vastly used for fault diagnosis providing them with different faults (a type of data which is difficult to obtain). Instead, here they are used to gain more insight from the data by identifying certain patterns occurring in a large database (a type of data commonly found in industry) which could be done manually, but would take insane amounts of time. According to the experiences here exposed, ML algorithms are extremely useful tools with great potential. And, in that sense, the challenge is finding real problems that can be solved with this tools instead of making up classification/regression problems so that the use of these tools is justified.

Also, during the development of the research and particularly in the last work presented in chapter 5, reproducible research practices have been adopted. This includes publishing data openly or using data from open repositories, as well as uploading the code used for the analysis to open access repositories. These practices are gaining popularity and interest among other scientific fields, as they allow a better review of the work, and research can be continued by other researchers. It is not common in condition monitoring field, however, some recent works that follow those practices have been identified. This is the case of the brilliant work presented in [Zha+20]. Publishing code and data is sometimes difficult when talking about projects related to industrial partners. Nevertheless, it is our sincere belief that it considerably raises research quality standards and allows cooperation among researchers.

All in all, the major contribution of this work resides in solving not the specific use cases, but the contexts of the problem. As explained in chapter 2, finding solutions for contexts instead than only for the use case implies that similar contexts could use the same approaches here presented. That is, the knowledge can be transferred. For instance, dimensionality reduction algorithms could be used to learn from the possible sources of failure of other machines than rotating machinery, regardless of not having vibration sensors as long as they operate in steady conditions. Similarly, the approach used to deal with varying conditions in wind turbines could be exported to other machines that also operate under varying conditions (the gearbox of a car for example) and be used to obtain a health indicator. And, the same happens for the hybrid approach used for electromechanical actuators, that could be used in any other machine that lacks faulty data and can be modelled with a physical model. The ability to transfer the knowledge is particularly interesting in the monitoring of industrial assets, as, many of the difficulties in the applied scenarios of this field are related to the acquisition of representative data with good enough quality standards and to deal with the variability of the operation. This work covers contexts with variability in the operation, the obtainment of knowledge in steady monitoring contexts and goes one step further by proposing a way the develop diagnosis algorithms for contexts with no faulty data. In short, this thesis project provides a road-map for the improvement of applied monitoring strategies from varying operating conditions to the diagnosis of machines.

6.1 Future Works

This thesis project is defined by the analysis of three different contexts occurring in three different applications, that, at the same time, are quite representative of many of the situations and cases that are given in machine condition monitoring. On the one hand, the fact that various cases are analysed leaves little chance to go in detail in each of them. On the other, confronting different scenarios provides a better overview of the current paradigm and challenges of condition monitoring.

In any case, some dead ends have been identified throughout the development of the thesis that could be addressed in following works or by other researchers.

Regarding the steady operating conditions, storing variables that might be of interest to distinguish a failure is a good starting point for diagnosis, but it still requires of manual or human decision to interpret incoming new faults. Automating this piece of work would be of great interest, and, for doing so, algorithms that use features instead of observations would be needed. In addition, algorithms that could learn new classes on-line when no similar faults are found in the database would be required.

For the case of varying conditions, the operation regimes were manually identified with the help of experts. However, automated time-series clustering methods could be used to segment the distinct regions in an autonomous way and later, apply the same procedure here presented to measure frequency, duration and stability of operation states. This clustering automation could extremely beneficial in scenarios with scarce experience in relation to the optimal operation regime.

Also, regarding hybrid modelling, this work has studied how to work without having faulty data. Nevertheless, this real faulty data might appear in the long term (even if only in the form of mild or low severity faults), and it could be integrated into the models to improve diagnosis. These new algorithms could probably be based on the use of incremental learning algorithms that allow partial retraining of the models when new observations are available.

Finally, a general lack of works presenting data from real industrial scenarios is observed. Exceptionally, data from wind turbines tends to be from real applications, but in these cases neither data nor code are disclosed. Further improvements in applying and presenting algorithms into operational environments are needed if condition monitoring research is going to go any further. For that purpose, stronger cooperation between researchers and industries is needed, as well as the will to disclose and present the findings to the research community.

Bibliography

[ACK13] Dawn An, Joo Ho Choi, and Nam Ho Kim. "Options for Prognostics Methods: A review of data-driven and physics-based prognostics". In: *54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. 2013, p. 1940 (cit. on p. 35).

[AKC15] Dawn An, Nam H Kim, and Joo-ho Choi. "Practical options for selecting data-driven or physics-based prognostics algorithms with reviews". In: *Reliability Engineering and System Safety* 133 (2015), pp. 223–236 (cit. on p. 34).

[ADW14] Hasmawati Antong, Roger Dixon, and Christopher Ward. "Modelling and Building of Experimental Rig for High Redundancy Actuator". In: *UKACC International Conference on Control (CONTROL* July (2014) (cit. on p. 35).

[ADW16] Hasmawati Antong, Roger Dixon, and Christopher P Ward. "High Redundancy Actuator with 12 Elements: Open-and Closed-loop Model Validation". In: *IFAC-PapersOnLine* 49.21 (2016), pp. 254–259 (cit. on p. 35).

[Art+18] Estefania Artigao, Sofia Koukoura, Andrés Honrubia-Escribano, et al. "Current signature and vibration analyses to diagnose an in-service wind turbine drive train". In: *Energies* (2018) (cit. on p. 28).

[ASC11] F. Al-Badour, M. Sunar, and L. Cheded. "Vibration analysis of rotating machinery using time–frequency analysis and wavelet techniques". en. In: *Mechanical Systems and Signal Processing* 25.6 (Aug. 2011), pp. 2083–2101 (cit. on p. 20).

[Bal15] Edward Balaban. "Prognostic Health-Management System Development for Electromechanical Actuators". In: *Journal of Aerospace Information Systems* 12.3 (2015) (cit. on pp. 35, 37).

[Bra+17] Inaki Bravo-Imaz, Hossein Davari Ardakani, Zongchang Liu, et al. "Motor current signature analysis for gearbox condition monitoring under transient speeds using wavelet analysis and dual-level time synchronous averaging". In: *Mechanical Systems and Signal Processing* 94 (2017). Publisher: Elsevier, pp. 73–84 (cit. on pp. 18, 20).

[Cai+16] Baoping Cai, Yubin Zhao, Hanlin Liu, and Min Xie. "A data-driven fault diagnosis methodology in three-phase inverters for PMSM drive systems". In: *IEEE Transactions on Power Electronics* 32.7 (2016), pp. 5590–5600 (cit. on p. 36).

[CMM16] James Carroll, Alasdair McDonald, and David McMillan. "Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines". In: *Wind Energy* (2016) (cit. on p. 27).

[Cas+03] Nicolas Casoetto, Dragan Djurdjanovic, Rhett Mayor, and Jun Ni. "Multisensor Process Performance Assessment Through Use of Autoregressive Modeling and Feature Maps". en. In: *Journal of manufacturing systems* (2003), p. 9 (cit. on pp. 17, 18).

[Cer+18] Mariela Cerrada, René-Vinicio Sánchez, Chuan Li, et al. "A review on data-driven fault severity assessment in rolling bearings". en. In: *Mechanical Systems and Signal Processing* 99 (Jan. 2018), pp. 169–196 (cit. on p. 20).

[CZH15] Christopher J Crabtree, Donatella Zappalá, and Simon I Hogg. "Wind energy: UK experiences and offshore operational challenges". In: *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy* 229.7 (2015), pp. 727–746 (cit. on p. 27).

[Dav+08] J Davies, T Steffen, R Dixon, et al. *Modelling of High Redundancy Actuation Utilising Multiple Moving Coil Actuators*. Vol. 41. 2. IFAC, 2008, pp. 3228–3233 (cit. on p. 35).

[DAS01] Anind K. Dey, Gregory D. Abowd, and Daniel Salber. "A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications". en. In: *Human–Computer Interaction* 16.2-4 (Dec. 2001), pp. 97–166 (cit. on p. 12).

[DNL02] Dragan Djurdjanovic, Jun Ni, and Jay Lee. "Time-Frequency Based Sensor Fusion in the Assessment and Monitoring of Machine Performance Degradation". en. In: *Dynamic Systems and Control*. New Orleans, Louisiana, USA: ASMEDC, Jan. 2002, pp. 15–22 (cit. on p. 18).

[Dup10] Richard Dupuis. "Application of Oil Debris Monitoring For Wind Turbine Gearbox Prognostics and Health Management". In: *Annual Conference of the Prognostics and Health Management Society, 2010 1* (2010) (cit. on pp. 28, 29).

[Ech+08] E. Echavarria, B. Hahn, G. J. W. van Bussel, and T. Tomiyama. "Reliability of Wind Turbine Technology Through Time". In: *Journal of Solar Energy Engineering* 130.3 (2008), p. 031005 (cit. on p. 27).

[EIF16] Christopher Ehrmann, Philippe Isabey, and Jürgen Fleischer. "Condition monitoring of rack and pinion drive systems : necessity and challenges in production environments". In: *Procedia CIRP* 40 (2016), pp. 197–201 (cit. on p. 36).

[Fen+13] Yanhui Feng, Yingning Qiu, Christopher J. Crabtree, Hui Long, and Peter J. Tavner. *Monitoring wind turbine gearboxes*. 2013 (cit. on pp. 28–30).

[Fer+16] Susana Ferreiro, Egoitz Konde, Santiago Fernández, and Agustín Prado. "INDUSTRY 4.0: Predictive Intelligent Maintenance for Production Equipment". In: *European Conference of the Prognostics and Health Management Society* (2016), p. 9 (cit. on p. 27).

[GH14] SA Gadsden and SR Habibi. “State Estimation and Fault Detection of an Electrohydrostatic Actuator”. In: *ASME/BATH 2014 Symposium on Fluid Power and Motion Control*. American Society of Mechanical Engineers Digital Collection. 2014 (cit. on p. 36).

[Gal14] Diego Galar. *Hybrid Models for Rotating Machinery Diagnosis and Prognosis : Estimation of Remaining Useful Life*. May. 2014 (cit. on p. 34).

[Gar+12] Fausto Pedro García Márquez, Andrew Mark Tobias, Jesús María Pinar Pérez, and Mayorkinos Papaelias. “Condition monitoring of wind turbines: Techniques and methods”. In: *Renewable Energy* 46 (2012), pp. 169–178 (cit. on pp. 27, 28).

[Gon+19] Elena Gonzalez, Bruce Stephen, David Infield, and Julio J. Melero. “Using high-frequency SCADA data for wind turbine performance monitoring: A sensitivity study”. In: *Renewable Energy* 131 (2019), pp. 841–853 (cit. on p. 28).

[HDR06] Berthold Hahn, Michael Durstewitz, and Kurt Rohrig. “Reliability of Wind Turbine–Experiences of 15 years with 1,500 WTs”. In: *Wind Energy* (2006) (cit. on p. 27).

[Ham+09] Z. Hameed, Y. S. Hong, Y. M. Cho, S. H. Ahn, and C. K. Song. “Condition monitoring and fault detection of wind turbines and related algorithms: A review”. In: *Renewable and Sustainable Energy Reviews* 13.1 (2009), pp. 1–39 (cit. on pp. 27, 28).

[KML17] J Kattelus, J Miettinen, and A Lehtovaara. “Detection of gear pitting failure progression with on-line particle monitoring”. In: *Tribiology International* September 2016 (2017), pp. 1–7 (cit. on p. 28).

[KM18] Matt Kemp and Eric J Martin. “Fault Isolation of an Electro-mechanical Linear Actuator”. In: *Proceedings of Prognostics and Health Management Society conference 2018*. 2018 (cit. on p. 36).

[Kn+15] Christian Kn, Zakaria Marsil, Markus Rekla, Johannes Reuter, and G Clemens. “Fault Detection in Linear Electromagnetic Actuators Using Time and Time-Frequency-Domain Features Based on Current and Voltage Measurements”. In: *20th International Conference on Methods and Models in Automation and Robotics (MMAR)* (2015), pp. 547–552 (cit. on p. 36).

[KDT18] Alexios Koltsidopoulos Papatzimos, Tariq Dawood, and Philipp R. Thies. “Data Insights from an Offshore Wind Turbine Gearbox Replacement”. In: *Journal of Physics: Conference Series* 1104.1 (2018) (cit. on p. 28).

[Let+15] U Leturiondo, M Mishra, D Galar, and O Salgado. “Synthetic data generation in hybrid modelling of rolling element bearings”. In: *Insight-Non-Destructive Testing and Condition Monitoring* 57.7 (2015), pp. 395–400 (cit. on p. 34).

[Li+11] Bing Li, Pei-lin Zhang, Hao Tian, et al. “A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox”. en. In: *Expert Systems with Applications* 38.8 (Aug. 2011), pp. 10000–10009 (cit. on p. 18).

[Li+17] Xiaochuan Li, Fang Duan, David Mba, and Ian Bennett. “Multidimensional prognostics for rotating machinery: A review”. en. In: *Advances in Mechanical Engineering* 9.2 (Feb. 2017), p. 168781401668500 (cit. on p. 20).

[Li+19] Zhixiong Li, Dazhong Wu, Chao Hu, and Janis Terpenny. “An ensemble learning-based prognostic approach with degradation-dependent weights for remaining useful life prediction”. In: 184.December 2017 (2019), pp. 110–122 (cit. on p. 34).

[LK14] Linxia Liao and Felix Köttig. “Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction”. In: *IEEE Transactions on Reliability* 63.1 (2014), pp. 191–207 (cit. on p. 34).

[Liu+18] Ruonan Liu, Boyuan Yang, Enrico Zio, and Xuefeng Chen. “Artificial intelligence for fault diagnosis of rotating machinery: A review”. en. In: *Mechanical Systems and Signal Processing* 108 (Aug. 2018), pp. 33–47 (cit. on p. 20).

[MVS18] G Arun Manohar, V Vasu, and K Srikanth. “Development of a high redundancy actuator with direct driven linear electromechanical actuators for fault-tolerance”. In: *Procedia computer science* 133 (2018), pp. 932–939 (cit. on p. 35).

[Man15] Asociación Española de Mantenimiento. *El mantenimiento en españa: Encuesta sobre su situación en las empresas españolas*. Asociación Española de Mantenimiento, 2015 (cit. on p. 8).

[Mat+15] Ion Matei, Anurag Ganguli, Tomonori Honda, and Johan De Kleer. “The case for a hybrid approach to diagnosis: A railway switch”. In: *CEUR Workshop Proceedings* 1507 (2015), pp. 225–234 (cit. on p. 35).

[Maz+19] Mirko Mazzoleni, Fabio Previdi, Matteo Scandella, and Giulio Pispola. “Experimental Development of a Health Monitoring Method for Electro-Mechanical Actuators of Flight Control Primary Surfaces in More Electric Aircrafts”. en. In: *IEEE Access* 7 (2019), pp. 153618–153634 (cit. on p. 18).

[Nar+10] Sriram Narasimhan, Indranil Roychoudhury, Edward Balaban, and Abhinav Saxena. “Combining Model-Based and Feature-Driven Diagnosis Approaches – A Case”. In: (2010) (cit. on pp. 35, 37).

[Nec+12] Patrick Nectoux, Rafael Gouriveau, Kamal Medjaher, et al. “PRONOSTIA: An experimental platform for bearings accelerated degradation tests.” In: 2012 (cit. on p. 21).

[NW13] Mengyan Nie and Ling Wang. “Review of condition monitoring and fault diagnosis technologies for wind turbine gearbox”. In: *Procedia CIRP* 11.Cm (2013), pp. 287–290 (cit. on pp. 27, 28).

[NB07] Julia Nilsson and Lina Bertling. “Maintenance management of wind power systems using condition monitoring systems - Life cycle cost analysis for two case studies”. In: *IEEE Transactions on Energy Conversion* 22.1 (2007), pp. 223–229 (cit. on pp. 27, 28).

[PFR17] Sebastian Pfaffel, Stefan Faulstich, and Kurt Rohrig. “Performance and reliability of wind turbines: A review”. In: *Energies* 10.11 (2017) (cit. on p. 27).

[Qia+17] Guan Qiao, Geng Liu, Zhenghong Shi, and Yawen Wang. “A review of electromechanical actuators for More / All Electric aircraft systems”. In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 232.February 2018 (2017) (cit. on p. 35).

[RS15] C Ruiz-Carcel and A Starr. “Development of a novel condition monitoring tool for linear actuators”. In: *The twelfth international conference on condition monitoring and machinery failure prevention technologies*. 2015, pp. 1–12 (cit. on p. 18).

[RS18b] Cristobal Ruiz-Carcel and Andrew Starr. “Data-based detection and diagnosis of faults in linear actuators”. In: *IEEE Transactions on Instrumentation and Measurement* 67.9 (2018). tex.publisher: IEEE, pp. 2035–2047 (cit. on p. 18).

[SR09] N. Saravanan and K.I. Ramachandran. “Fault diagnosis of spur bevel gear box using discrete wavelet features and Decision Tree classification”. en. In: *Expert Systems with Applications* 36.5 (July 2009), pp. 9564–9573 (cit. on p. 18).

[SGW16] Bernard Schmidt, Diego Galar, and Lihui Wang. “Context Awareness in Predictive Maintenance”. en. In: *Current Trends in Reliability, Availability, Maintainability and Safety*. Ed. by Uday Kumar, Alireza Ahmadi, Ajit Kumar Verma, and Prabhakar Varde. Cham: Springer International Publishing, 2016, pp. 197–211 (cit. on p. 12).

[SW15] Bernard Schmidt and Lihui Wang. “Predictive maintenance: literature review and future trends”. In: *The International Conference on Flexible Automation and Intelligent Manufacturing (FAIM), 23-26 June 2015, University of Wolverhampton, UK*. Vol. 1. The Choir Press. 2015, pp. 232–239 (cit. on p. 12).

[She16] Shuangwen Sheng. “Monitoring of Wind Turbine Gearbox Condition through Oil and Wear Debris Analysis: A Full-Scale Testing Perspective”. In: *Tribology Transactions* (2016) (cit. on pp. 28–30).

[Su+17] Chun Su, Yang Yang, Xiaolin Wang, and Zhaoyong Hu. “Failures analysis of wind turbines: Case study of a Chinese wind farm”. In: *Proceedings of 2016 Prognostics and System Health Management Conference, PHM-Chengdu 2016* December (2017) (cit. on p. 27).

[SI17] Thapanun Sudhawiyangkul and Don Isarakorn. “Design and realization of an energy autonomous wireless sensor system for ball screw fault diagnosis”. In: *Sensors and Actuators A: Physical* 258 (2017), pp. 49–58 (cit. on p. 37).

[Sus+13] Susana Ferreiro, Andres Jimenez, Jon Madariaga, et al. “Health monitoring for electro-mechanical nose landing gear door actuator of a uav, based on simulation modelling and data-driven techniques”. In: *Chemical Engineering Transactions* 33 (July 2013), pp. 655–660 (cit. on p. 36).

[TW17] Jannis Tautz-Weinert and Simon J. Watson. "Using SCADA data for wind turbine condition monitoring – a review". In: *IET Renewable Power Generation* 11.4 (2017), pp. 382–394 (cit. on p. 28).

[Tch+14] Pierre Tchakoua, René Wamkeue, Mohand Ouhrouche, et al. "Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges". In: *Energies* 7.4 (2014), pp. 2595–2630 (cit. on pp. 27, 28).

[Vis+17] Manish Vishwakarma, Rajesh Purohit, V. Harshlata, and P. Rajput. "Vibration Analysis & Condition Monitoring for Rotating Machines: A Review". en. In: *Materials Today: Proceedings* 4.2 (2017), pp. 2659–2664 (cit. on p. 20).

[Zha+20] Zhibin Zhao, Tianfu Li, Jingyao Wu, et al. "Deep Learning Algorithms for Rotating Machinery Intelligent Diagnosis: An Open Source Benchmark Study". en. In: *arXiv:2003.03315 [cs, eess]* (Mar. 2020). arXiv: 2003.03315 (cit. on pp. 18, 44).

Webpages

[@Gru17] SPRI Grupo. *Invest in Basque Country - Strategic sectors*. en-US. tex.ids: StrategicSectors2017a libraryCatalog: www.spri.eus. May 2017. URL: `https://www.spri.eus/invest-in-basque-country/en/invest-basque-country/sectors/` (visited on June 9, 2020) (cit. on p. 7).

[@Lop+20] Kerman Lopez de Calle – Etxabe, Cristobal Ruiz – Carcel, Andrew Starr, et al. *Hybrid Modelling for Linear Actuator Diagnosis in Absence of Faulty Data Records*. 2020. URL: `https://github.com/klopetx/Hybrid_EMA_Diagnosis` (cit. on pp. 38, 39).

[@Pre08] General Secretariat for Communication Presidency of the Government. *Maquina Herramienta*. en. Library Catalog: www.basquecountry.eus. Apr. 2008. URL: `https://www.basquecountry.eus/t32-sectores/en/contenidos/noticia/maquina_herramienta_08/en_ma_he/maquina_herramienta.html` (visited on June 9, 2020) (cit. on p. 7).

[@RS18a] Cristobal Ruiz-Carcel and Andrew Starr. *Data Set for "Data-Based Detection and Diagnosis of Faults in Linear Actuators"*. Mar. 7, 2018. URL: `https://cord.cranfield.ac.uk/articles/Data_set_for_Data-based_Detection_and_Diagnosis_of_Faults_in_Linear_Actuators_/5097649` (cit. on p. 38).

[@Tan19] John Olav Tande. *EERADeepWind'2019*. en. 2019. URL: `https://www.sintef.no/globalassets/project/eera-deepwind-2019/tr-2019_00247_eera-deepwind.pdf/` (cit. on p. 30).

[@Vav18] Bob Vavra. *2018 Maintenance Survey: Playing Offense and Defense*. Mar. 15, 2018. URL: `https://www.plantengineering.com/articles/2018-maintenance-survey-playing-offense-and-defense/` (visited on July 28, 2020) (cit. on p. 8).

Part II

Part II: The research

7 Comparison of Automated Feature Selection and Reduction methods on the Condition Monitoring issue

- Authors: López de Calle, Kerman; Ferreiro, Susana; Arnaiz, Aitor; Sierra, Basilio.
- Publisher: Elsevier
- Journal: Procedia Manufacturing
- Year: 2018
- Quartile (Scimago/WoS): Q2/-
- DOI: 10.1016/j.promfg.2018.10.150

Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 00 (2018) 000–000

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

7th International Conference on Through-life Engineering Services

Comparison of Automated Feature Selection and Reduction methods on the Condition Monitoring issue

Kerman López de Calle*[a,b], Susana Ferreiro[a], Aitor Arnaiz[a], Basilio Sierra[b]

[a]IK4-TEKNIKER, Iñaki Goenaga street, 5, 20600 Eibar, Gipuzkoa (SPAIN)
[b]University of the Basque Country (EHU-UPV), Faculty of Informatics, Manuel Lardizabal Ibilbidea, 1, 20018 Donostia, Gipuzkoa (SPAIN)

Abstract

Condition Monitoring is a key task of condition-based maintenance (CBM). The economic efficiency together with the reliability of CBM is expanding its use to areas where it was not used before. However, this expansion of CBM systems is ballasted by two main obstacles: the maintenance related knowledge of recently monitored assets, which is usually scarce and not structured; and the lack of failure related data, which is a major obstacle of data-driven methods. With the aim of developing a method to automatically select and/or reduce features, this work compares widely used dimensionality reduction and feature selection methods, which are capable of automatically obtaining knowledge while the monitoring is carried out. Those methods can later be used to help operators decide which features could be more related to the degradation of the machine. The purpose of this work is to determine which of those compared methods could be more appropriate for detecting anomalies in order to develop monitoring systems.

Keywords: Condition Monitoring; Data Analytics; Autonomous Maintenance; Automatic Feature Selection; Dimensionality Reduction;

* Corresponding author. Tel.: +34-943-206-744-9641;
E-mail address: Kerman.lopezdecalle@tekniker.es

1. Background

Condition Monitoring is essentially the design of a system which is able to monitor the condition of an asset through sensors. Once the system detects the condition of the asset, it is possible to: first, detect anomalies; secondly, diagnose these anomalies; and, in some cases, even predict them before they occur.

Some authors recognise that, for a condition monitoring process to be reliable, it is required to fuse more than a single signal [1]. Other researches have shown that apparently less interesting signals could be more appropriate for certain purposes; that is the case of [2], in which it is stated that the direction with lowest vibration resulted more interesting for detecting wear. At the same time, the range of different machines operating in industry and the required specific knowledge for monitoring those machines obstructs the monitoring task. On the one hand, some have been vastly studied (such as most rotating machinery), whereas, on the other hand, there are other machines which have been less studied (such as clutch-breaks and other machine tools).

An ideal data-driven procedure of condition monitoring would be based on the following schema *Fig. 1* (a):

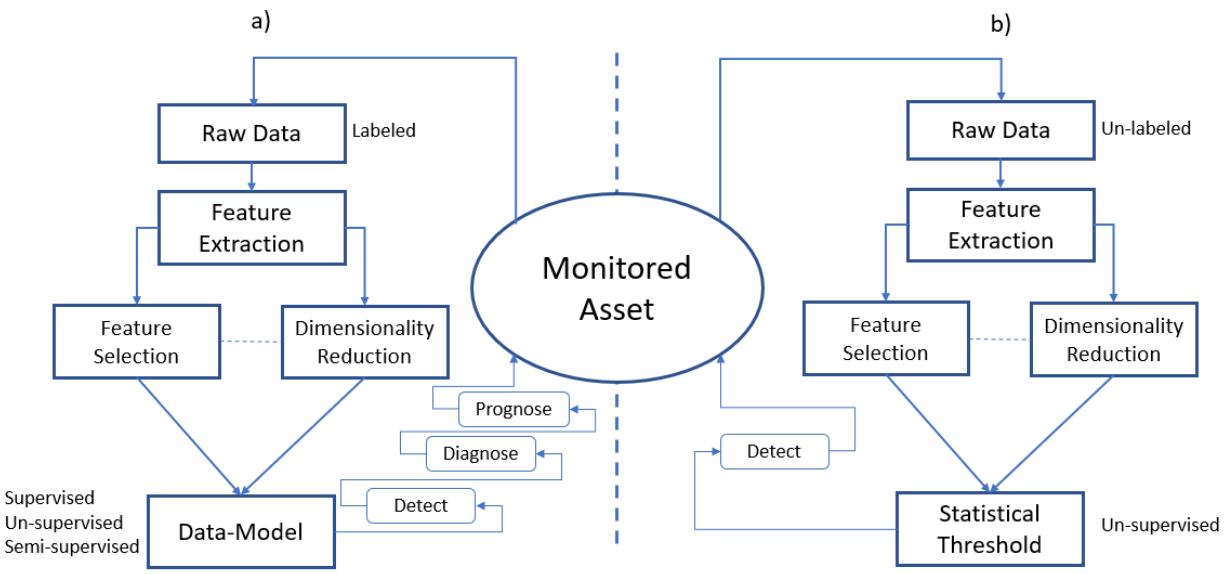

Fig. 1. (a) Ideal data-driven monitoring approach. (b) Approach in recently monitored assets.

However, most of the times there are not enough labelled data, or the quality is not good enough, which is a very common phenomenon in assets which have been recently monitored. For that reason, it is sometimes necessary to learn 'on the fly' and decide which feature is more representative of the condition of the asset, as shown in *Fig. 1* b). This work focuses on the cases in which data becomes available along the time and there is no way to train a regular classifier due to the lack of labelled data.

In order to determine which extracted features are more related to the asset condition (Feature Selection) or how to combine them (Feature Extraction), some authors have constructed Health Indicators from features to make validations [3]. Nevertheless, those indicators are extracted after analysing the whole evolution of the feature from run-to-failure, which is not applicable in the scope of this study.

In this work the reality shown in *Fig. 1* b) is simulated. The simulation confronts the difficulty of being an un-supervised problem, as there is no label determining the real condition of the asset. For that reason, the evaluation of different methods is not trivial and requires properly developed metrics.

The aim of feature selection and reduction techniques is double in these frameworks. First, those methods allow determining which features are more relevant for the process according to some criteria; besides, the dimensionality reduction eases the monitoring task, as monitoring a single dimension is less complicated than tracking several features at a time.

2. Methodology

As the purpose of this work is to identify a technique capable of monitoring an asset and to detect anomalies while more knowledge is extracted during the monitoring process, several methods have been compared in order to evaluate which one performs better and under which circumstances.

2.1. Simulation:

With this aim in mind, a simulation was carried out. The simulation assumed a newly monitored device starting to feed a data repository. After an initial tracking period of time, the stored data was analysed periodically by different algorithms and a new tracking dimension was obtained according to some criteria. The new dimension was then evaluated by process control limits and it was decided whether the system was working under normal or abnormal/damaged conditions.

As previously explained, the issue could be considered unsupervised due to the lack of labels. However, the addition of a simple assumption allowed the use of supervised methods. The assumption consisted of considering that degradation level cannot be reduced with the time (unless maintenance actions are carried out) and, therefore, comparing the most recent instances with the initial ones would be similar to having "degraded" and "non-degraded" classes.

2.2. Data:

In order to obtain robust results, the original idea was to compare the algorithms in different scenarios, this work shows, however, how the algorithms behaved in the monitoring of bearings. The bearing dataset provided by [4] contained run-to-failure tests of bearings with two vibration channels monitored throughout the process.

Three different strategies were analysed in relation to the data feeding of the algorithms. The first one consisted of feeding all the data currently available. The second one used the initial 20 instances and the last 20 instances every time the algorithms were refreshed. And the third one used 20 instances from the end and the remaining 20 instances sampled from different instants along the process.

The following Table 1 contains a detailed explanation of the data-partitioning strategies:

Table 1. Details of data feeding strategies.

Data splitting method	Takes all data	Deterministic	Creates class	Explanation
First-last-n40	False	True	LDA and Relief	Takes first n and last n values in each refreshing.
Sampled-split	False	False	LDA and Relief	The 'recent' n/2 instances are taken from the data belonging to the last 10% of the data. The rest is randomly sampled from the remaining data following a probability.
First-last-all	True	True	LDA and Relief	Splits all available data in two, considering the first half 'old/undamaged' and the second half 'recent' instances.

The recalculations of the algorithms took place every time 5 new instances entered the system, and the refreshing process began once 50 measurements were stored in the simulated data-repository.

2.3. Descriptor extraction:

Raw data usually consists of vast amounts of measurements taken in very short time intervals. But, for our monitoring purpose, that data needed to be somehow compressed in descriptors, which are values synthetizing raw data by retaining certain information.

This work is intended to develop a generic monitoring technique. For that reason, the descriptors used assume no complex understanding of the assets, even if it is not the case, as bearings have been studied in depth and very specific descriptors have been identified in the literature. In order to simulate that naivety, simple time domain descriptors have been used as proposed by [5][6][7]. The extracted descriptors were: Mean, Standard Deviation, Clearance, Impulse Factor, Crest Factor, Kurtosis, Peak value, Root Mean Square, Root, Shape factor and Skewness.

2.4. Dimensionality reduction:

Despite the possibility of monitoring the evolution of each of the descriptors independently, due to the complexity of this approach in systems/assets with various signal sources, it is usually decided first to reduce the dimensionality to a single dimension. This dimension should represent the state of the asset and it is originated by combining or pruning the previously extracted descriptors in various ways by following some criteria.

The dimensionality reduction/feature selection methods studied in this work are the following:

- PCA: Principal Component Analysis is used to represent the input features with new dimensions, with the aim of finding new linearly un-correlated dimensions explaining as much variance as possible. The implementation used can be found in [8].
- Autoencoders: Autoencoders are Neural Networks used to extract patterns from input values. Their mid layers are constrained with fewer neurons than input and output layers so that, when they are trained with the same inputs as class values, they need to obtain a compressed pattern in their hidden layers. This compressed pattern can be used to generate a new dimension. Implemented in R [8] with the autoencoder package provided by [9].
- LDA: Linear discriminant analysis uses class values. It measures the mean and standard deviation values of each class in order to later calculate the probability of new instances to belong to one class or another. It creates new projections with C-1 new different dimensions, where C is the number of classes. Implemented with R [8] and package MASS [10].
- Relief: This supervised feature selection algorithm gives weights to features in relation to how well they distinguish the class from similar (close in distance) instances belonging to other classes, and, how different the feature is in neighbours of the same class. Implemented in R in the package FSelector [11].

Prior to being fed into the algorithms, the data was first pre-processed by extracting the mean value and dividing it by the standard deviation to compensate the scale effect.

It could be noticed that most of the aforementioned algorithms work in an unsupervised manner. However, it is not the case of the LDA and Relief, which require labels to work correctly. In order to provide labels, some domain related bias was introduced, which in this case meant adding a different label to the oldest and another label to most recent samples. In this way, LDA and Relief would rely on those descriptors showing greater differences from the beginning to the end of the process.

2.5. Monitoring:

Once the data have been processed and the dimensionality reduced, it can be tracked in order to monitor our system. Monitoring consists of tracking the evolution of process-related features and detecting possible deviations from normality. Typical monitoring methods calculate the mean of the feature and establish upper and lower boundaries. Different strategies are used to determine when the process is out of control.

In our simulation, upper and lower boundaries were recalculated each time the dimensions were refreshed. In the calculation of those boundaries, the instances considered out-of-control by the last run were not taken into account. The process was defined to be damaged or out-of-control when two consecutive out-of-control instances were given.

2.6. Evaluation criteria:

Evaluating the performance of distinct algorithms is difficult in CBM monitoring duties. Even if there is a relation between the measured signal and the health of the asset, that relation is complex and not always empirically defined. This means that, the real health state of the machine is unknown. Therefore, it is difficult to establish an objective criterion that defines how well an algorithm detects anomalies or/and damages due to the lack of real labels.

Researchers agree with the need of feature selection as a way of improving model performance by means of: eliminating noise; providing faster and more cost-effective models; and, helping to better understand the process that generated the data [12][13]. Nevertheless, different researchers have used different criteria to define the quality of the features as discussed by [14], with some more generic methods as proposed by [13][15][16] and some other very case-specific methods such as [17][18]. However, it was decided to utilize other criteria which would better match the monitoring purpose of the algorithms and the demands of the industry.

After some consideration, and, taking into account that the objective evaluation of the dimensionality reduction/feature selection is complex, the following criteria were considered for evaluating the feature selection and dimensionality reduction algorithms.

- Computational cost: time required to compute each transformation/reduction. Note that it might vary using different implementations of the algorithms.
- Understandability and traceability: interpretability of the output feature by the operator, and capability to trace the features which constructed the new dimension.
- Efficacy: defined as the distance to the limit given by the expert, to the limit given by the algorithms. In this case of bearing monitoring, the expert decision was to use the root mean square (RMS) value of the vibration signal. The score is a division of the distance to class divided by total asset life (in instances).
- Determinateness: measures the rate of change in the rank of weights in each refreshment, and the change compared to initial weight set. For that purpose, the distance metric for ranks proposed in [19] is used, which ranges from 0 to 1, being 1 absolute equality and 0 absolute disparity.

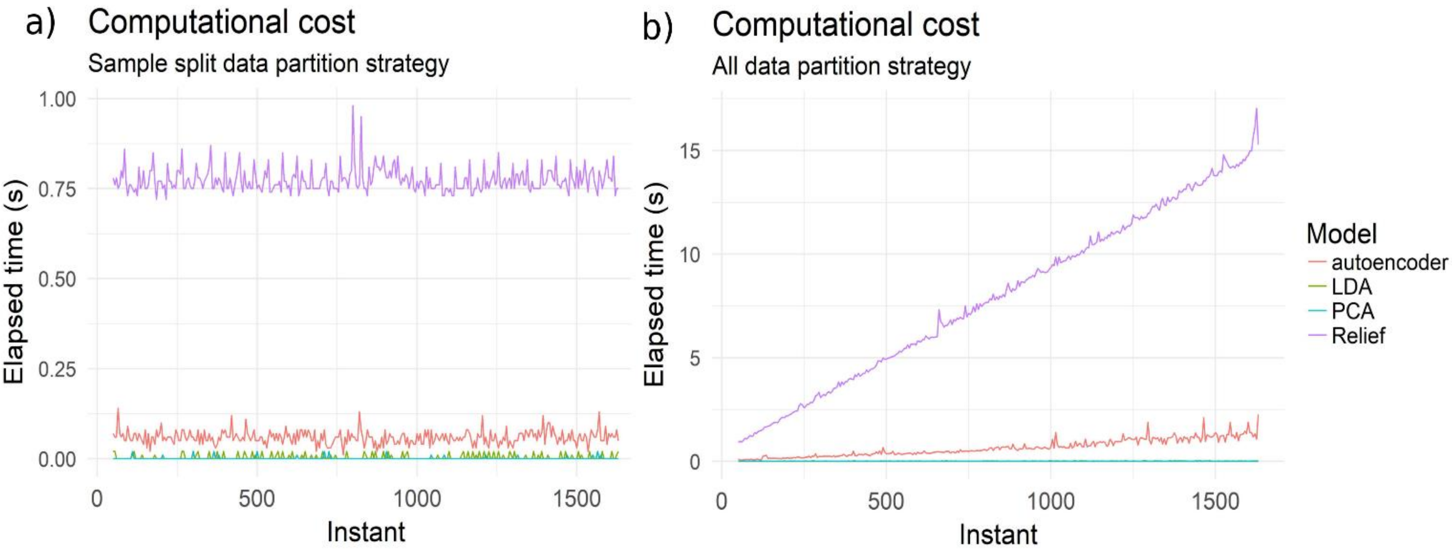

Fig. 2. a) Computational cost with random sampling partitioning strategy. b) Computational cost with all-data partitioning strategy.

3. Results

Once the simulations took place, it was possible to compare the algorithms as well as the effect of different partitions for each of the bearings in the dataset.

Regarding the partitioning method, evidence has shown that as the amount of data fed into the algorithms increases, the computational requirements also increase. As a result, methods which used a fixed quantity of data in each iteration are computationally more efficient. As seen in Fig. 2, using all the available samples requires linearly increasing the amount of time (b), which can be a problem for systems having big data volumes to process. This situation does not happen in strategies were the amount of data fed is fixed (a).

Table 2. Effectiveness of algorithms with different data-feeding strategies by bearing.

	first-last n40				first-last all				sampled split			
Bearing	PCA	LDA	autoenc.	Relief	PCA	LDA	autoenc.	Relief	PCA	LDA	autoenc.	Relief
Bearing1_1	0.75	0.80	-0.16	-0.04	NA	0.79	NA	0.00	0.79	-0.27	-0.30	0.00
Bearing1_2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Bearing2_1	0.06	0.14	0.35	0.14	NA	0.14	NA	0.14	0.06	0.13	0.35	0.14
Bearing2_2	0.16	0.22	0.18	0.97	0.20	0.17	0.16	0.97	0.97	0.21	0.16	0.77
Bearing3_1	NA	0.00	NA	0.00	NA	0.09	NA	0.00	NA	NA	NA	0.00
Bearing3_2	-0.15	0.68	-0.27	0.92	-11.31	-0.02	-0.27	0.92	-0.15	-0.09	-0.15	0.92
Mean	0.20	0.37	0.03	0.40	-5.55	0.23	-0.05	0.41	0.42	0.00	0.01	0.37

In relation to the algorithms, the previously mentioned metrics were used in order to measure their differences.

Regarding effectiveness, it is worth noting that the PCA algorithm together with the autoencoder sometimes find difficulties in trespassing the limits in some cases, even if the class descriptor (RMS value) does find out-of-control process. This finding is interesting, as many works have shown similar behaviors between PCA and autoencoders in relation to the creation of the new dimensions. Relief algorithm shows very irregular behavior, this may be caused by the fact that only a single feature has been chosen by Relief. Which means that in some cases the error is 0 because Relief has chosen the same feature as the class (RMS). In other cases, however, the features chosen by Relief are not that related to the RMS value, and therefore, the score is worse.

Overall, it seems that autoencoder is the dimensionality reduction method which has the closest predictive values to the class value, as the mean values show.

With respect to the computational cost, as the figure Fig. 2 shows, LDA and PCA seem to need similar computational times for the calculations, whereas autoencoder needs slightly higher ones and Relief requires quite longer times. These differences could be caused by the different levels of optimization in the implementation, or also due to the values of the hyperparameters used during the calculations.

However, the use of all the data requires longer computational times, which has to be taken into consideration if the volumes of data which are going to be treated will keep increasing in the future.

Regarding interpretability of the new dimensions, it is difficult to define how comprehensible the new dimensions are. The combinations proposed by algorithms, which compress various dimensions into a single one, hinder the tracing of the feature which is originally out-of-control. Paying attention to the weights might be an option, however, the high compression level (as many descriptors are reduced to a single one) makes the interpretation of weights risky. Because, even if some weights have extreme values in comparison to others, the sum of the rest of the weights might be compensating that unbalance. In that sense, the Relief algorithm, which has been used as a feature selector instead of a dimensionality reducer, can be considered the best choice, as it just selects the best feature for the monitoring. Nevertheless, even if it is possible to trace the feature which is out-of-control, it is difficult to obtain any information related to damage origination, even less in the case of recently monitored assets, where the knowledge is limited.

Lastly, in relation to the determinateness of the algorithms, as explained in the previous chapter, a ranking distance metric has been used with the weights obtained in each of the iterations. The distance metric has been used to measure the distance of the new weights compared to the ones calculated in the previous iteration and, also, to measure the distance from the initial weights to the last ones. Table 3 displays the distance values obtained during the simulation:

Table 3. Determinateness of algorithms by bearing and by feeding strategy.

		First-Last n40		First-Last all		Sample split	
Bearing	Model	from Previous	from Initial	from Previous	from Initial	from Previous	from Initial
Bearing1_1	autoencoder	0.769 +- 21.2 %	0.642 +- 26.5 %	0.936 +- 8.4 %	0.753 +- 12.62 %	0.683 +- 21.2 %	0.67 +- 17.9 %
	LDA	0.787 +- 11.6 %	0.676 +- 10.8 %	0.879 +- 10.2 %	0.821 +- 9.01 %	0.729 +- 12.9 %	0.713 +- 12.9 %
	PCA	0.83 +- 19.2 %	0.508 +- 37.4 %	0.976 +- 5.0 %	0.77 +- 14.81 %	0.629 +- 34.2 %	0.352 +- 42.6 %
	Relief	0.814 +- 10.9 %	0.58 +- 21.0 %	0.783 +- 13.4 %	0.659 +- 14.57 %	0.701 +- 19.5 %	0.659 +- 15.6 %
Bearing1_2	autoencoder	0.752 +- 18.5 %	0.704 +- 11.4 %	0.761 +- 20.2 %	0.716 +- 16.06 %	0.635 +- 29.0 %	0.73 +- 19.3 %
	LDA	0.818 +- 12.5 %	0.767 +- 10.2 %	0.891 +- 10.0 %	0.769 +- 5.07 %	0.724 +- 10.2 %	0.693 +- 10.7 %
	PCA	0.907 +- 9.0 %	0.71 +- 17.6 %	0.964 +- 9.1 %	0.618 +- 15.21 %	0.804 +- 12.9 %	0.63 +- 13.2 %
	Relief	0.655 +- 23.7 %	0.611 +- 21.3 %	0.555 +- 32.6 %	0.433 +- 26.79 %	0.526 +- 28.9 %	0.383 +- 46.5 %
Bearing2_1	autoencoder	0.862 +- 10.4 %	0.744 +- 8.1 %	0.939 +- 7.7 %	0.846 +- 4.37 %	0.736 +- 20.0 %	0.674 +- 8.0 %
	LDA	0.774 +- 10.9 %	0.747 +- 9.0 %	0.84 +- 11.1 %	0.743 +- 8.61 %	0.74 +- 14.9 %	0.595 +- 12.1 %
	PCA	0.898 +- 10.8 %	0.601 +- 11.6 %	0.979 +- 4.1 %	0.799 +- 5.88 %	0.714 +- 18.1 %	0.485 +- 30.7 %
	Relief	0.603 +- 33.3 %	0.594 +- 27.8 %	0.666 +- 22.4 %	0.546 +- 34.25 %	0.52 +- 30.6 %	0.596 +- 32.4 %
Bearing2_2	autoencoder	0.81 +- 14.0 %	0.41 +- 38.0 %	0.866 +- 13.9 %	0.437 +- 43.48 %	0.71 +- 17.5 %	0.547 +- 28.0 %
	LDA	0.798 +- 11.8 %	0.694 +- 10.5 %	0.888 +- 8.0 %	0.697 +- 12.63 %	0.743 +- 11.4 %	0.759 +- 9.6 %
	PCA	0.923 +- 7.4 %	0.628 +- 12.7 %	0.983 +- 4.5 %	0.694 +- 12.68 %	0.744 +- 23.1 %	0.636 +- 25.9 %
	Relief	0.846 +- 16.1 %	0.493 +- 11.0 %	0.764 +- 23.6 %	0.545 +- 14.86 %	0.696 +- 23.6 %	0.313 +- 28.8 %
Bearing3_1	autoencoder	0.647 +- 29.2 %	0.421 +- 31.4 %	0.912 +- 9.1 %	0.795 +- 10.94 %	0.583 +- 18.4 %	0.553 +- 34.7 %
	LDA	0.766 +- 13.6 %	0.687 +- 6.4 %	0.823 +- 8.7 %	0.739 +- 14.48 %	0.839 +- 6.7 %	0.826 +- 6.5 %
	PCA	0.824 +- 20.5 %	0.409 +- 38.1 %	0.994 +- 1.1 %	0.973 +- 0.21 %	0.632 +- 47.2 %	0.277 +- 68.6 %
	Relief	0.568 +- 38.7 %	0.444 +- 32.7 %	0.579 +- 34.7 %	0.532 +- 38.91 %	0.417 +- 23.3 %	0.42 +- 38.8 %
Bearing3_2	autoencoder	0.704 +- 21.4 %	0.525 +- 19.8 %	0.917 +- 12.0 %	0.436 +- 16.06 %	0.649 +- 22.2 %	0.521 +- 18.8 %
	LDA	0.806 +- 10.4 %	0.774 +- 8.9 %	0.945 +- 6.3 %	0.791 +- 9.36 %	0.763 +- 11.3 %	0.789 +- 8.2 %
	PCA	0.853 +- 12.2 %	0.42 +- 28.8 %	0.989 +- 4.3 %	0.597 +- 11.22 %	0.695 +- 23.6 %	0.515 +- 26.4 %
	Relief	0.791 +- 13.0 %	0.418 +- 30.4 %	0.639 +- 25.5 %	0.48 +- 26.04 %	0.489 +- 31.7 %	0.341 +- 33.4 %
	Mean value	0.783 +- 12.4 %	0.592 +- 10.79 %	0.853 +- 9.8 %	0.674 +- 9.44 %	0.671 +- 13.8 %	0.570 +- 12.1 %

Results have shown that the partitioning affects the mean determinateness of the models. That is because the number of new instances used for recalculations is bigger in the sample split in comparison to the other two cases. When all the samples are used for recalculations, only the last samples are new, therefore the change in weights is not considerable. When first-last 40n is used, as 20 samples are always the same, the change in weights is smaller than in random sampling, where half of the instances are randomly chosen in each iteration.

The difference from the initial weights to the weights calculated in each refreshing follows a similar trend. However, the weights are less similar in average and it has to be mentioned that even if the distance in the sample split is bigger than in the first-last 40 instances, that difference is really small.

Relief algorithms performs worse than the rest of the dimensionality reduction techniques.

PCA seems to change the least in each iteration, however, that change is big when the distance is considered in relation to the initial weights. Similarly, this happens to autoencoder, which tends to obtain a small rate in change in each iteration, but that distance increases when comparing it to initial weights. In that sense, LDA has de advantage of having more similar weights (compared to initials) and a considerably small rate of change in each iteration.

4. Conclusions:

Several dimensionality reduction and feature selection techniques used for the purpose of monitoring have been compared in this work. A specific evaluation criteria has been established and the different algorithms have been tested under the bearing dataset [4], with different data feeding strategies.

Different partitioning methods have shown different performances, however, it is clear that the less data is used, the more computationally efficient the algorithms are. Besides, the adaptability of sampling randomly could be an interesting phenomenon to study, as it could be used for by the operators to "teach" the algorithms new limits.

Furthermore, in relation to the evaluation of the algorithms, it must be mentioned that Relief algorithm performed much worse than the rest of the algorithms, after the autoencoder, and, lastly, LDA and PCA which had similar computational needs. However, those computational requirements might change with different hyperparameters (in the of Relief and autoencoder algorithms) or implementations of the algorithms.

Due to the need of establishing a class value for comparison among algorithms and the bias introduced by choosing a single feature as Health Indicator, results could vary greatly in the evaluation of efficacy. In our tests, autoencoder and LDA perform best, but autoencoder tends to create a dimension where no anomalies are detected.

Tracing weights in order to understand the anomalous features does not seem to be applicable, as the information found in the single dimension is highly compressed. In that sense, the use of Relief algorithm as feature selector eases the task, but, as drawback, it keeps less information.

Highest determinateness ratios are reached by PCA and LDA. However, PCA tends to have bigger changes in the weights from initial values, whereas LDA, scores worse in each refreshment but keeps weights closer from initials.

Lastly, the inclusion of "degraded" and "non-degraded" labels to allow the use of supervised algorithms (LDA/Relief) work has given satisfactory results. This concept could be interesting to develop in future works.

As stated in *Fig. 1* b), there are more steps involved in the monitoring task which are beyond the scope of this study but are interesting to assess. Future researches could include those other steps as well as other aspects of interest, such as the influence of the inclusion of domain knowledge in the feature extraction process and the capability to detect and avoid noisy features.

5. References:

[1] R. Teti, K. Jemielniak, G. O'Donnell, and D. Dornfeld, "Advanced monitoring of machining operations," *CIRP Ann. - Manuf. Technol.*, vol. 59, no. 2, pp. 717–739, 2010.

[2] J. C. Chen, L. H. Huang, M. A. X. Lan, and M. S. Lee, "Analysis of an Effective Sensing Location for an In-process Surface Recognition System in Turning Operations," *J. Ind. Technol.*, vol. 15, no. 3, pp. 1–6, 1999.

[3] G. Vachtsevanos, F. L. Lewis, M. . Roemer, A. Hess, and B. Wu, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. 2006.

[4] P. Nectoux *et al.*, "PRONOSTIA : An experimental platform for bearings accelerated degradation tests.," in *IEEE International Conference on Prognostics and Health Management, PHM'12*, 2012, pp. 1–8.

[5] Y. Lei, *Intelligent fault diagnosis and remaining useful life prediction of rotating machinery*. 2016.

[6] J. V Abellan-Nebot and F. R. Subrión, "A review of machining monitoring systems based on artificial intelligence process models," *Int. J. Adv. Manuf. Technol.*, vol. 47, no. 1–4, pp. 1–21, 2009.

[7] C. Hu *et al.*, "Options for Prognostics Methods: A review of data-driven and physics-based prognostics," *54th AIAA/ASME/ASCE/AHS/ASC Struct. Struct. Dyn. Mater. Conf.*, vol. 14, no. 1, pp. 1–14, 2013.

[8] R. R Development Core Team, *R: A Language and Environment for Statistical Computing*, vol. 1, no. 2.11.1. 2011.

[9] E. Dubossarsky and Y. Tyshetskiy, "autoencoder: Sparse Autoencoder for Automatic Learning of Representative Features from Unlabeled Data." 2015.

[10] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, Fourth. New York: Springer, 2002.

[11] P. Romanski and L. Kotthoff, "FSelector: Selecting Attributes." 2016.

[12] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[13] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *J. Mach. Learn. Res.*, vol. 3, no. 3, pp. 1157–1182, 2003.

[14] S. Jie, G. S. Hong, M. Rahman, and Y. S. Wong, "Feature Extraction and Selection in Tool Condition Monitoring System," pp. 487–497, 2002.

[15] G. Wang and Q. Song, "Selecting feature subset via constraint association rules," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7301 LNAI, no. PART 2, pp. 304–321, 2012.

[16] J. V. Abellan-Nebot and F. Romero Subirón, "A review of machining monitoring systems based on artificial intelligence process models," *Int. J. Adv. Manuf. Technol.*, vol. 47, no. 1–4, pp. 237–257, 2010.

[17] A. Ginart, I. Barlas, J. Goldin, and J. L. Dorrity, "Automated feature selection for embeddable Prognostic and Health Monitoring (PHM) architectures," *AUTOTESTCON (Proceedings)*, no. 404, pp. 195–201, 2007.

[18] M. Gerdes, D. Galar, and D. Scholz, "Automated parameter optimization for feature extraction for condition monitoring," *14th IMEKO TC10 Work. Tech. Diagnostics 2016 New Perspect. Meas. Tools Tech. Syst. Reliab. Maintainab. Saf.*, pp. 452–457, 2016.

[19] W. Webber, A. Moffat, and J. Zobel, "A similarity measure for indefinite rankings," *ACM Trans. Inf. Syst.*, vol. 28, no. 4, pp. 1–38, 2010.

8

Dynamic condition monitoring method based on dimensionality reduction techniques for data-limited industrial environments

- Authors: López de Calle, Kerman; Ferreiro, Susana; Arnaiz, Aitor; Sierra, Basilio.
- Publisher: Elsevier
- Journal: Computers in Industry
- Year: 2019
- Quartile (Scimago/WoS): Q1/Q1
- DOI: 10.1016/j.compind.2019.07.004

Computers in Industry 112 (2019) 103114

Contents lists available at ScienceDirect

Computers in Industry

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

Dynamic condition monitoring method based on dimensionality reduction techniques for data-limited industrial environments

Kerman López de Calle[a,b,*], Susana Ferreiro[a], Aitor Arnaiz[a], Basilio Sierra[b]

[a] IK4-TEKNIKER, Iñaki Goenaga Street, 5, 20600 Eibar, Spain
[b] Faculty of Informatics UPV/EHU, Manuel Lardizabal, 1, 20018 Donostia-San Sebastián, Spain

ARTICLE INFO

Article history:
Received 14 February 2019
Received in revised form 10 July 2019
Accepted 15 July 2019
Available online xxx

Keywords:
Condition monitoring
Condition-based maintenance
Dimensionality reduction
Feature selection
Feature projection

ABSTRACT

In recent years, the integration of intelligent condition monitoring systems has become crucial for the Operation and Maintenance in the industry. Condition-based maintenance strategies allow the identification of the current health status of industrial assets. Consequently, downtimes and maintenance costs are reduced by the efficient planning of corrective actions and the early detection of costly damages. However, the implementation of data-based monitoring techniques is often hampered by two main difficulties: Firstly, the lack of understanding of the monitored equipment; and secondly, the lack of data related to failures and their evolution, which is one of the major obstacles for the implementation.

This article proposes a method based on the use of dimensionality reduction (DR) techniques to perform real-time monitoring of the state of the machine. The performance of the approach is studied with four different DR techniques: PCA, LDA, Relief, and Autoencoders. The algorithms are tested in a simulation of various asset lifetimes in which they determine when the assets are working under abnormal conditions. Next, the results of the simulation are evaluated under a specifically designed evaluation criterion based on three key performance indicators; cost, interpretability, and effectiveness. In this way, a common situation in industrial field applications is imitated, where the information is gathered in real-time and there is not enough prior knowledge of the machine nor data about its degradation or failures.

1. Introduction

Condition monitoring (CM) is a concept totally related to the permanent availability of machinery, with minimal interruption or even without interruptions. In practice, this translates into the need to guarantee a fast and efficient maintenance that resolves or, ideally, prevents any defect in a timely manner. This process aims to monitor and detect failures in machines so that maintenance can be scheduled proactively when necessary, and not before. The condition of the asset, inferred by the indicators that show signs of loss of performance or future failures, must trigger maintenance within a specified period of time before the failure, so work can be completed before the asset fails or the performance falls below the optimum level. These indicators are the basis to detect the real condition of the component, detect anomalies in an early state, provide a diagnosis to determine what has happened and, finally, in some situations, even predict the failure before it occurs by forecasting the remaining useful life (RUL).

However, it is sometimes complex to fulfill the data requirements needed to develop algorithms. This means: having records of anomalies or failures (recognizing the type of failure and providing a diagnosis); and, having long enough historical data of the evolution of the degradation (being able to establish trends and to estimate the remaining useful life). These data requirements are only met when the machine is monitored with sensors and it works until the anomaly or failure occurs, which is undesired by the industry. In addition, the diversity of the types of failure and operating conditions of the machine hinder gathering fully representative datasets for the analysis. Against this background, it is usual to use test benches in order to simulate anomalies and failures; and even then, this is only possible with some components, such as the case of some mechanical components (i.e., bearings or gears), which are widely studied and where there is a wide knowledge of the field.

* Corresponding author at: IK4-TEKNIKER, Iñaki Goenaga Street, 5, 20600 Eibar, Spain.
E-mail addresses: kerman.lopezdecalle@tekniker.es (K. López de Calle), susana.ferreiro@tekniker.es (S. Ferreiro), aitor.arnaiz@tekniker.es (A. Arnaiz), b.sierra@ehu.eus (B. Sierra).

https://doi.org/10.1016/j.compind.2019.07.004
0166-3615/© 2019 Elsevier B.V. All rights reserved.

In summary, although the inclusion of sensor devices is a growing trend in the industry, data that represent anomaly or failure situations are rarely obtained. Furthermore, the analysis and knowledge extracted from the use of experimental test benches or other approaches is difficult to extrapolate to the actual use case of systems, where other aspects are involved, such as the complexity and the dynamic operating conditions among others. Therefore, it is necessary to implement other monitoring strategies that allow, without prior knowledge in the domain or operation of the equipment, to start monitoring and extracting relevant information during its life.

This document proposes an alternative procedure for the cases in which the analysis and the implementation of the algorithms collide with the previously mentioned limitations. Fig. 1 represents the two points of view: Fig. 1(a) represents the approach used when there is a representative dataset of potential failure modes or anomalies, and Fig. 1(b) represents the approach detailed in this work, to be used when there is no a priori knowledge about the most representative features of health, nor a representative set of data defining its normality and no-normality (anomalies or failures).

In both of the approaches presented in Fig. 1, the monitored asset is working while the sensors are generating data along its lifetime. This data is processed, some features are extracted and, later, dimensionality reduction (DR) techniques are used. Lastly, data models are used in order to detect, diagnose and predict possible failures. The main difference of the approaches is the fact that our approach does not assume having faulty data for DR nor for the building of data-based models, and, therefore, it is limited to detecting anomalies. Furthermore, the DR is carried out considering other criteria that are not related to the optimization of failure detection.

The rest of the article is organized as follows. Section 2 presents a review of previous works related to Condition monitoring (CM) and the position of the present work in that context. Section 3 describes the proposed approach: the simulation, carried out using data from real use-cases and emulating an operating machine that must be stopped before the failure; the Feature Extraction techniques, used to describe the signals; the basis of the DR algorithms, utilized to reduce the dimension and gain insight of the process; the monitoring techniques that define when the process is out of control; and lastly, the definition of the evaluation criteria. Section 4 briefly describes the experimental setup used for gathering the datasets. Next, Section 5 explains the results of the analysis and the evaluation. And finally, Section 6 closes the article with the most important conclusions and future works.

2. Literature review

2.1. Condition monitoring

Condition-based maintenance (CBM) is widely extended due to its cost effectiveness and its increasing capabilities of improving productivity and maintenance planning. The growth in the use of CM is so widespread that providers also seek to modernize existing equipment with CM capability. There is a greater number of machines where the data is already available and, therefore, intelligent algorithms are needed to carry out the analysis while considering the operation mode of the equipment.

In reference to the development of algorithms, there are mainly two methodologies: data-driven approaches and mathematical or physical modeling. Data-driven approaches collect information from the sensors in order to have real life data during long periods of time while waiting for the failure to occur and using this data to identify trends or features which correspond and characterize faults. In some cases, such as some mechanical components or systems (i.e., gears, bearings, specific types of gearboxes, etc.), real experimentation can be replaced by experimentation in the test benches (which can be relatively similar to the machine and can mimic its behavior or performance in real operation). These test benches allow the simulation of specific types of failures and algorithms for their detection to be developed. Mathematical or physical modeling allows the algorithm to be developed from the first principles and can be used to predict failure modes and their effect on measured parameters.

Various works in the literature have followed one or the other methodology aiming the monitoring of the condition of mechanical components. Regarding the data-driven ones, in [1], a systematic methodology for gearbox monitoring and fault classification is developed and evaluated for a dataset of gearbox vibration data. Moreover, Bechhoefer and He [2] describe a process for data driven prognostics and, as an example, uses a gear fault run to failure test. Their approach fuses the set of features into a health indicator through a statistical process. Saravanan and Ramachandran [3,4] studied gear box fault diagnosis using decision tree classification and artificial neural networks. The vibration signals of a spur bevel gear box in different conditions are used to demonstrate the application of various wavelets in feature

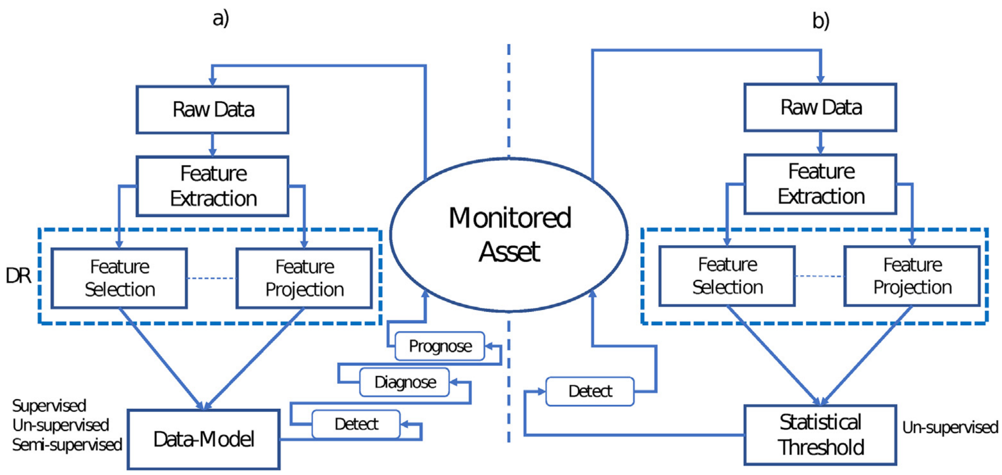

Fig. 1. (a) Typical approach to condition monitoring. (b) Approach of this article.

extraction. Del Río et al. [5] presents experimental testing and data collection of healthy and faulty gears to detect and predict early stages of failures in a gearbox, considering vibration signals, which can be obtained from EMA (electromechanical actuators) by means of triaxial accelerometer in on-ground testing. [6] proposes bearing monitoring (the most vulnerable part of induction motors) based on stator current signals processed with a deep learning architecture from an experimental campaign that artificially simulates bearing damage. Furthermore, Bravo-imaz et al. [7] focused on the analysis of motor current signature for fault diagnosis of gearboxes operating under transient speed regimes. A specifically designed and thoroughly monitored test bench was used to fully characterize the experiments, in which gears in different health status were tested. There are other works that adopt physics based models, as in [8], where a simple dynamic model of a single stage gearbox is used to show accurately the effect of the rotating input, output and mesh frequency components in the stator current signature for induction machine-based electromechanical systems. In order to validate the proposed model, a test-bed based on a wound-rotor 4 kW three-phase induction machine connected to a gearbox was used. Furthermore, other authors have fused both techniques, Li et al. [9] performs a hybrid fault diagnosis of a gearbox using four classifiers. Eight fault states, including gear defects, bearing defects and combination of gear and bearing defects, are simulated on a single-stage gearbox to evaluate the proposed feature extraction and selection scheme.

Nevertheless, the use of condition monitoring is not limited to mechanical components. Due to its proven benefits, it has been extended to a wide variety of systems and final applications as detailed below. For example, Djurdjanovic et al. [10] demonstrates the ability to use information from multiple sensors (from a CNC lathe machine) to assess the drift away from the sharp tool machine operation and to distinguish between a sharp and a worn tool through experiments. Next, in [11], the algorithm based on the fusion of multiple sensor inputs and data collected from experimentation on the welding machine is presented. It matches observed system signatures against those observed during its normal behaviour. Later, Djurdjanovic et al. [12] showed the development of an Agent to enable multi-sensor assessment and prediction of performance of products and machines considering some signal processing techniques and feature extraction (such as frequency-bands energies, wavelets, principal components, etc.), combined with a health assessment (based on Gaussian functions, Hidden Markov Models, Particle filter, etc.) from a web-enabled E-manufacturing test-bed realized by a car manufacturer. More recently, Bleakie and Djurdjanovic [13] applied condition monitoring and fault modeling to a set of standard built-in sensors on a modern 300-mm technology industrial plasma enhanced chemical vapor deposition (PECVD) tool. In another field of application, Ferreiro et al. [14] presents a development of health monitoring system for an electro-mechanical nose-landing gear door actuator of an unmanned aerial vehicle, based on a combination of simulation modeling and data-driven techniques. The aim of the work is to detect some failures at early stages to avoid a catastrophic fault that may cause serious damage to the unmanned aerial vehicle (UAV). Also, later, Ruiz-Carcel and Starr [15] presents a data-based condition monitoring tool for linear actuators tested successfully using computational simulations and then they propose in [16] a set of algorithms that makes use of features extracted from the controller through the analysis and experimentation with a specially designed test rig to recreate fault scenarios under different operation conditions.

However, for several reasons, both methodologies have disadvantages that complicate the transfer of the algorithms to the final applications. On the one hand, the typical data-based methodology is time consuming and the data capture of all failure and operation modes is needed but not guaranteed. On the other hand, mathematical and physical modeling must be validated through the combination of real-life data and/or the deliberate introduction of faults into the system; however, this task becomes expensive and cumbersome for large machines with complex dynamics. Furthermore, it requires a lot of expert knowledge about the machine and processes.

In addition, the real problems of the industry go further and are difficult to tackle, especially in small and medium-sized companies in manufacturing. These companies have new equipment that needs to be monitored. The performance of these machines is unknown even by the manufacturer and there is neither time nor resources for experimentation. Each machine is individual with its particularities and there is no guarantee that the knowledge extracted from one machine can be transferred to another due to contextual factors.

2.2. Dimensionality reduction

Regarding the use of DR techniques, their use is well established in CM. DR techniques are recognized tools for improving model performance by means of: the elimination or reduction of noise, providing faster and more cost-effective models, and helping to better understand the process that generated the data as explained in [17,18]. Also, applied to CM, the fusion of features has provided better results as described in [19]; and some apparently less interesting features have been demonstrated more appropriate for certain purposes as revealed in [20]. They have also been used in CM to construct Health Indicators as presented in [21,22]. Furthermore, it is recognized that the reduction of features used for monitoring purposes can be translated in cost reduction, as it reduces computational requirements as well as the number of sensors required for obtaining raw measurements as shown in [23]. For these reasons, the use of DR techniques is widely extended in CM, as detailed in [16,23].

2.3. Evaluation criteria

The evaluation of the performance of distinct algorithms is a difficult matter in CM. Even if there is a relation between the measured signal and the health of the asset, that relation is complex and not always empirically defined. Consequently, the real health state of the machine might be unknown. Therefore, it is difficult to establish an objective criterion that defines how well an algorithm detects anomalies or/and damages due to the lack of real labels.

DR algorithm performances are typically evaluated and compared by training a classifier and measuring its accuracy or related metrics (false positive rate, error rate, etc.) as explained in [18]. Some works go one step further and include additional metrics such as in [24], where runtime, proportion of selected features and a sensitivity, and confidence thresholds are included. In CM applications that include DR, accuracy and runtime related metrics prevail [23,25–27]. However, some works mention the need to consider other additional indicators such as power consumption in [23]. In any case, accuracy and runtime are not the only metrics that reflect how well algorithms are suited for the final implementation in CM systems. The evidence has shown that interpretable models are more accepted by the users according to [28]. Furthermore, there is an increasing general reluctance to implement non-interpretable algorithms (also called black box algorithms) reflected in the new regulations (GDPR)[29]. That is why the evaluation of CM algorithms must include additional and specific metrics tailored for its final application.

Focusing on the development of early stage monitoring algorithms for the cases of new or recently monitored and less known assets, this work tests a simulation based on the procedure showed in Fig. 1(b) and evaluates the dimensionality reduction algorithms under a specific evaluation criteria based on industrial needs.

3. Proposed approach

The approach presented below consists of two main parts detailed in depth in the following sections: Simulation and evaluation criteria. First, the simulation based on the diagram of Fig. 1(b) is explained, from which the performance of the DR algorithms is measured under the reproduction of a real scenario. Next, the evaluation criteria are shown, i.e., the metrics based on the most relevant implementation factors.

3.1. Simulation

Essentially, the simulation consists of taking run-to-failure data from the different data sources (described in Section 4) and emulating a recently monitored device with this data. The approach this work follows is presented in Fig. 2.

Firstly, an initial tracking period of time with no monitoring but just tracking is assumed. After that period of time, the data stored until $t_{current}$ is analyzed through Loop 1 and Loop 2 and, if there is no anomaly detected, $t_{current}$ is moved 10 instances (simulating new data generated by the machine). This process is repeated until an anomaly is detected in Loop 2 or the whole dataset has been monitored with the system without finding anomalies. In more detail, the asset life simulation follows these steps:

- Step (A) *Process simulation*: The data until $t_{current}$ is split into two windows (first and last window) that keep equal size and include the first 25 and the last (until $t_{current}$) 25 instances of the process.
- Step (B) *Label construction*: Labels are given to the windows coming from process simulation. This is done by, assuming, from the domain knowledge, that the measurements taken in posterior instants must be equally or more degraded, but never less than the ones taken at the beginning. Therefore, non-degraded class is given to initial records (first window) and degraded label to most recent records (last window). This approximation allows the use of supervised methods for dimensionality reduction.
- Step (C) *Dimensionality reduction*: DR algorithms are used with the data from the previous step. As a result, a feature transformation and weights related to feature values are generated. Weights are stored (see Step D) and the parameters used in the DR are transferred to the next step. This phase is explained in more detail in Dimensionality reduction algorithms (DR) section.
- Step (D) *Weight storing*: The weights obtained in the training phase of the algorithms are stored in order to study them with the metrics explained in Section 3.2.
- Step (E) *Signal reconstruction*: Using the parameters from the DR step and the whole signal data until $t_{current}$, the signals are projected in the new and single dimension (in the case of Feature Projection algorithms) or just the best feature is left (in the case of Feature Selection algorithms).
- Steps (F)–(I) *Monitoring(Loop 2)*: The projected feature is tracked and it is decided whether the system is under control or it is out of control. More details of these steps are given in the section Monitoring.

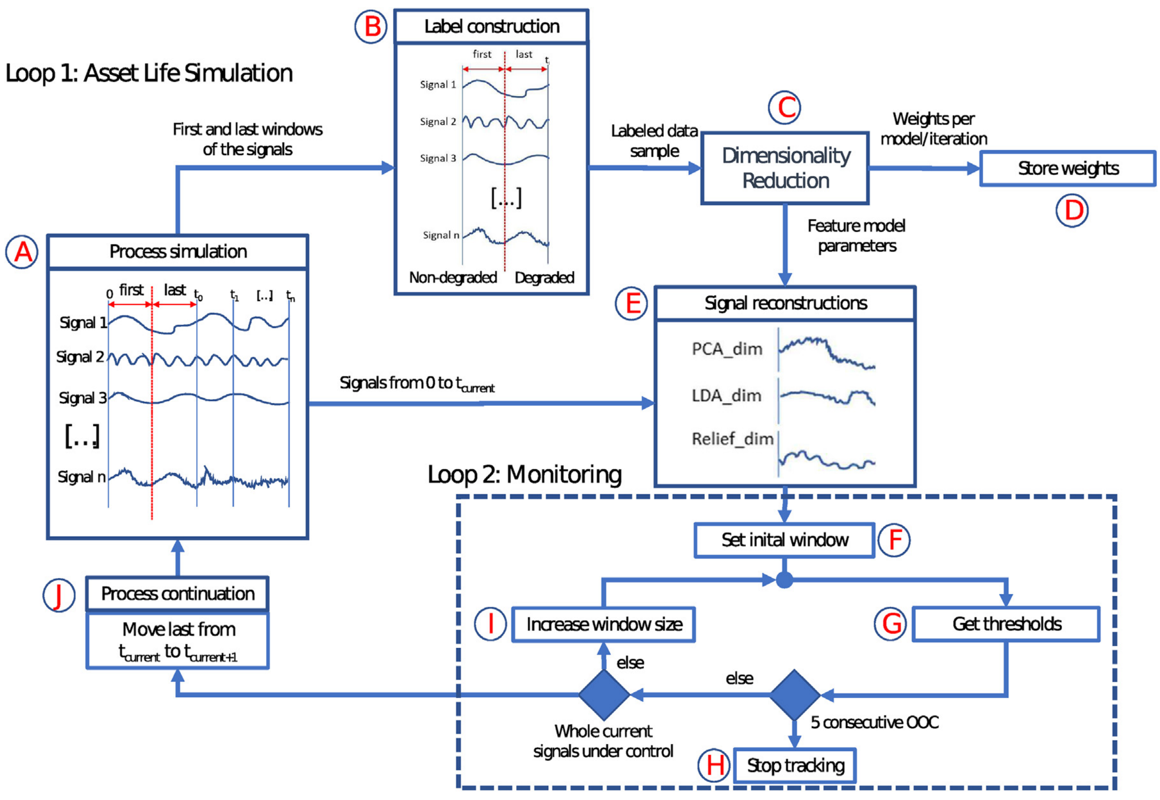

Fig. 2. Schema of the simulation.

- Step (J) *Process continuation*: If the process is not out of control, the simulated machine should keep working, $t_{current}$ is increased by 10 data instances and Loop 1 is triggered again.

Each algorithm has been tested separately, storing all the weights and projections created along the asset life simulation. Note that as different algorithms are tested, the projections suggested by each algorithm can lead to disparate results regarding when the machine should stop working.

3.1.1. Feature extraction (FE)

Instead of using raw data measurements for the monitoring, some statistical descriptors are typically used. These descriptors are generated as a result of a feature extraction process which tends to be specific according to the sensor producing the data stream. In our particular datasets, consisting of vibration, microphone, current intensity and encoder measurements; both simple statistical descriptors as well as more specific ones were extracted.

Regarding the simple time domain statistical descriptors, the ones proposed in [21,30,31] were extracted in the time domain raw data. The extracted descriptors were: Mean, Standard Deviation, Clearance, Impulse Factor, Crest Factor, Kurtosis, Peak value, Root Mean Square, Root, Shape factor and Skewness.

Besides, some more specific descriptors were extracted for the vibrations and the microphones. In the case of vibrations, frequency domain statistics were used. From the first five harmonics of the gear mesh frequency 10 Hz and 30 Hz windows were taken and the maximum and the RMS values were extracted similarly to [32]. Also, for the microphone, the energy index (EI) proposed in [33] was calculated using four different power values (1, 3, 5, 10). Detailed explanation of the features with more complexity is given in [34].

Not all descriptors were extracted in both data sources, more details regarding the data sources and descriptors can be found in Section 4.

Lastly, before using the descriptors with the DR algorithms, the descriptors were pre-processed by extracting the mean value and dividing them by the standard deviation to compensate the scale effect.

3.1.2. Dimensionality reduction algorithms (DR)

Assuming that basing a CM system on a single feature in unfeasible [19], the role of DR techniques is twofold in this work. First, these methods determine which features are more relevant for the process according to some criteria in a on-line way; besides, they ease the monitoring task, as monitoring a single dimension is less complicated than tracking several features at a time. Two families of DR algorithms are used in this work: feature selection and feature projection.

Feature selection (FS) algorithms identify the features that maximize the relation among the input variables and the target feature (a.k.a class) or just reduce the number of features by reducing the amount of redundancy among the variables. Feature projection (FP) algorithms, instead of choosing among the input variables, create new mappings of the features. Each algorithm has its own criterion to optimize the variable selection/reduction.

In addition to the previous classification of DR techniques, these techniques can also be classified regarding the use they make of the class variable. When they need a variable of interest, they are called supervised. Contrarily, if they do not require the class variable for the mapping or selection, they are known as unsupervised.

This work has utilized the following algorithms, all of them implemented in the R software [35]:

- *PCA*: Principal Component Analysis is used to represent the input features with new dimensions. These new dimensions are linearly un-correlated and explain as much variance as possible. The implementation used can be found in [35] based on the description of [36].
- *Autoencoders (AE)*: Autoencoders are Artificial Neural Networks (ANN) used to extract patterns from input values. Their mid-layers are constrained with fewer neurons than input and output layers so that, when they are trained with the same inputs as class values, they need to obtain a compressed pattern in their hidden layers. This compressed pattern can be used to generate a new dimension. The technique is well explained in [37], it is implemented with the autoencoder package provided by Dubossarsky and Tyshetskiy [38].
- *LDA*: Linear discriminant analysis uses class values. It measures the mean and standard deviation values of each class in order to later calculate the probability of new instances belonging to one class or another. It creates new projections with C-1 new different dimensions, where C is the number of classes. The proposed new dimensions try to maximize the difference between classes. It is implemented using the MASS package [36] with the description of [36].
- *Relief*: This supervised feature selection algorithm gives weights to features in relation to how well they distinguish the class from similar (close in distance) instances belonging to other classes, and how different the feature is in neighbours of the same class. It is implemented using the FSelector package [39] based on the description of Kononenko [40].

It should be noted that some of the aforementioned algorithms work in an unsupervised manner. However, it is not the case of the LDA and Relief, which require labels to work correctly. In order to provide labels, some domain related bias was introduced, as explained in Simulation section. In addition, it must be mentioned that neither PCA nor Relief reduce dimensions by themselves, what they do is give a ranking of feature importance from which only the best feature was later used.

Table 1 classifies the algorithms used in this article according to the classification explained previously in this section.

Note that each of the previous algorithms uses weights in order to assess the importance of the feature for the corresponding model. The weights do not have equal meaning from one algorithm to other, however, they do represent the importance the algorithm gives to each particular feature in comparison to the others. In the particular case of autoencoders, that do not assign importance to the variables, the weights are equivalent to the weights of the first layer, as it is the only connection of the variables to the layer with the single neuron.

For the training of the algorithms the default parameters of the implementation are used unless for the for the case of autoencoder, in which the mid layer has single neuron with tanh activation function, lambda (weight decay) is set to 0.002, beta (sparsity penalty) is equal to 6, rho (sparsity parameter) 0.01 and epsilon (weight initialization) 0.001 are used.

3.1.3. Monitoring

Once the data is processed and the dimensionality is reduced, the new dimensions are used in order to monitor the system as

Table 1
Classification of the algorithms.

	Feature projection	Feature selection
Supervised	LDA	Relief
Unsupervised	PCA, AE	

displayed in Fig. 2 Loop 2. Monitoring consists of tracking the evolution of process-related features and detecting possible deviations from normality. As this step is carried out after the DR step, and only a single feature is kept, traditional Statistical Process Control systems can be used.

Monitoring is carried out by using a Shewhart chart [41]. In step F, a window equal to the initial tracking period of time (as in step A) is taken. The mean (μ) and threshold values ($3 \cdot \sigma$) of the Shewhart chart are calculated (step G). If any five consecutive out-of-control (OOC) points are found, an anomaly is diagnosed, and the systems stops the process (step H). Otherwise, if the system is under control, Loop 2 continues by adding 1 instance to the initial window (step I) until either the system is stopped or $t_{current}$ is reached. If $t_{current}$ is reached and no anomaly is detected, the whole process continues in step J by adding more data point to the process simulation.

3.2. Evaluation criteria

The simulation confronts the difficulty of being an unsupervised problem, as there is no label determining the real condition of the asset. For that reason, the evaluation of different methods is not trivial and requires properly developed metrics.

With the intention of designing a generic comparison framework for industrial application cases, the proposed solution has taken into account the following three factors:

- *Efficacy*: Satisfies the desired results (detects the anomalies).
- *Interpretability*: The outputs are understandable.
- *Efficiency*: With the focus on optimizing resource needs.

The rest of the section is dedicated to explaining in detail the exact metrics used for each dimension.

3.2.1. Efficiency

Translating cost to industrial needs is synonymous with of money and time. In the particular case of software products, the computation time itself is related to both, as it determines the hardware requirements and the highest frequency rate of decision taking. This dimension has been represented by two sub-dimensions:

- *Computation time*: Time required by the algorithm to obtain the new dimensions. In order to force 0–1 range, the inverse values of the minimum and maximum times required by the compared algorithms are used.
- *Determinateness from previous*: Determinateness is defined in this work as the difference of the weights from one instant to another. In order to measure that change, a distance metric for ranks proposed in the work [42] and implemented in the gespeR package [43] has been used. This metric also ranges from 0 to 1, where 1 means absolute equality and 0 absolute disparity. Determinateness has been used in two different sub-dimensions, in the case of determinateness from previous, it stands as the average distance of the weights in respect to the weights of the previous instants for each of the recalculations. The idea of including this metric in the cost resides in the fact that the slower (higher determinateness from previous) the weights change from the previous iteration, the easier it is to reduce the calculation frequency without losing information.

3.2.2. Interpretability

This second dimension takes into account the amount of variables used by the models, whether or not the models require a mathematical mapping and how much the models change the values of the weights:

- *Sparsity*: This stands for the fraction of sparse weights of an algorithm. As the whole feature space has been reduced to the first selected feature or component, a sparse weight (and therefore feature not contributing to that feature/component) is considered a weight which is 10 times smaller than the maximum weight value of that algorithm.
- *Comprehensiveness*: Sparsity considers the whole output weights of the dimensionality reduction weights. However, the dimension used for tracking is a single feature in the case of FS algorithms, in contrast to dimensionality reduction techniques that fuse various signals into a single one. For that reason, a binary metric has been created giving 1 values to complete comprehensiveness and 0 for those methods using mathematical mappings (Feature Projection methods), as their traceability is more complex.
- *Average determinateness*: This metric averages the determinateness of weights in the first instant to the weights in the out-of-control instant(referred to determinateness from initial) with the determinateness from previous (the same metric as the previous dimension). This averaged metric shows, on the one hand, how much the transformation has evolved from the beginning (0 meaning totally different rank and 1 meaning no evolution) and, on the other hand, how erratic the transformations are (if determinateness from previous is high, it means there is a small change, whereas if it is big, it stands for erratic behaviour).

3.2.3. Efficacy

Considering a condition monitoring framework, efficacy has been considered as the capability of an algorithm to: detect anomalies; giving the smallest number of false positives; and, at the same time, extract as much valuable information as possible. This dimension has been disseminated in three different sub-dimensions that measure how accurate predictions are and how well noise is detected and eliminated. Due to the difficulties to assess the real health state of an asset, expert knowledge was considered for defining the ground truth feature for each dataset (see Section 4) and compared to the results provided by the algorithms.

- *True positive rate (TPR)*: This metric represents the fraction of cases that are abnormal (positive) according to the expert feature and are also considered abnormal by the algorithm, divided by the total number of real positive cases.
- *Positive predictive value (PPV)*: This metric assesses which fraction of the positive predictions is really positive. That is, from all the data instants that are found out of control by the tested algorithms, how many are really out of control according to the expert feature.
- *Noise detection (ND)*: In order to estimate the denoising capability of the dimensionality reduction algorithms, four synthetic variables have been created and added to the dataset. For the creation of each synthetic variable, the range of a randomly chosen variable is measured, and then this range is used to create random uniform noise. Noise detection metric takes 0–1 values, 0 being none of the synthetic variables are identified as noise (meaning weight was not sparse) and 1 being the four synthetic variables are identified as noise (had sparse weights).

Note that all metrics have 0 to 1 range and therefore can be easily averaged or/and compared. Each dimension is computed by averaging the corresponding sub-dimensions.

Table 2
Health indicators by gear.

Gear	Class feature
Test 5	Accelerometer X H2 Max
Test 14	Accelerometer Y H1 RMS 10 Hz
Test 18	Accelerometer Y Crest

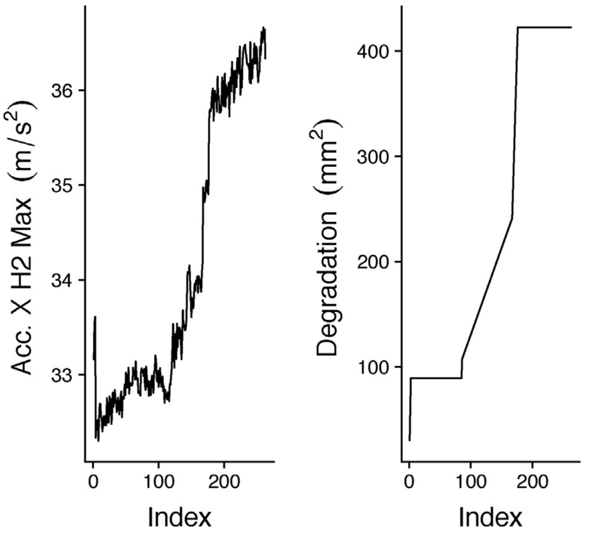

Fig. 3. Left side: Example of best feature according to the expert in Gear/Test 5. Right side: RMS of wheel and pinion degraded surface values with linear interpolation in the gaps.

Table 3
Characteristics of the different datasets.

Dataset	Features	Length
Gear 5	141	263
Gear 14	141	558
Gear 18	141	543
Bearing 1.1	22	466
Bearing 1.2	22	144
Bearing 2.1	22	151
Bearing 2.2	22	797
Bearing 3.1	22	85
Bearing 3.2	22	1637

4. Experimental setup

For the sake of robustness, two different data sources have been used for the simulations and evaluations. Both sources were originated in test rigs, but with the purpose of testing different components. As the simulations are based on these datasets that are taken under steady operating conditions, the load and speed variations (or any other factors influencing signals during operation) are not considered.

- Dataset 1: Bearings
  This dataset provided by Nectoux et al. [44] is open an contains run-to-failure tests of bearings with two vibration channels monitored throughout the process.
  For the monitoring purpose, some descriptors are extracted from the raw data. The set of time domain statistical descriptors explained in Feature extraction (FE) section has been used. In this dataset, root mean square (RMS) values of accelerometers have been used as the real health indicators for evaluating the efficacy.
- Dataset 2: Gears
  This dataset is private and has been generated in IK4-TEKNIKER [34]. It was generated in a FZG test rig, which is a standardized work bench well known for spur gear testing. The original purpose of the test was to assess the evolution of micro-pitting and pitting, which are kinds of wear generated on the surface of the gears. To this effect, the test had to be stopped periodically in order to measure the wear. Therefore, in addition to the signals monitored by the sensors, some surface degradation measures were made available.
  The test rig had different data sources integrated: 3 axial accelerometers, a microphone, current intensity meters and an encoder. For each one of the sensors, the same time domain statistical descriptors that were taken for the bearings dataset were extracted. Additionally, the more specific descriptors described in Section 3.1.1 (FE) were also taken for vibrations (harmonics) and the microphone (energy index). In the case of gears dataset, as there are some micro-pitting measurements in addition to the recorded signals, an indicator was created using the measurements of micro-pitted surfaces of both wheel and pinion gears. The measurements of both surfaces was averaged with a RMS value and, as there are less degradation records than sensor measurements, linearity has been assumed (from each measured micro-pitted instant to the next one) to impute degradation values for the rest of the sensor measurements. With the help of a correlation analysis and an expert, the most appropriate features were chosen. The features shown in Table 2 have been used in each case as health indicators. Fig. 3 shows an example of best feature used as health indicator together with the interpolated degradation variable.

Table 4
Average of sub-dimension results by data source.

Data source	Model	Efficiency		Interpretability			Efficacy		
		Comp. Eff.	Det. prev.	Sparsity	Compr.	Avg Det.	TPR	PPV	ND
Bearings	Autoencoder	0.937	0.716	0.583	0	0.638	0.233	0.243	0.667
	LDA	0.995	0.815	0.885	0	0.780	0.615	0.648	1.000
	PCA	1.000	0.877	0.359	0	0.727	0.558	0.508	0.667
	Relief	0.000	0.667	0.295	1	0.559	0.792	0.778	0.625
Gears	Autoencoder	0.987	0.695	0.870	0	0.580	0.000	0.000	0.833
	LDA	0.999	0.646	0.956	0	0.598	0.333	0.333	0.917
	PCA	1.000	0.839	0.475	0	0.549	0.333	0.333	0.833
	Relief	0.000	0.721	0.275	1	0.495	0.333	0.333	0.833

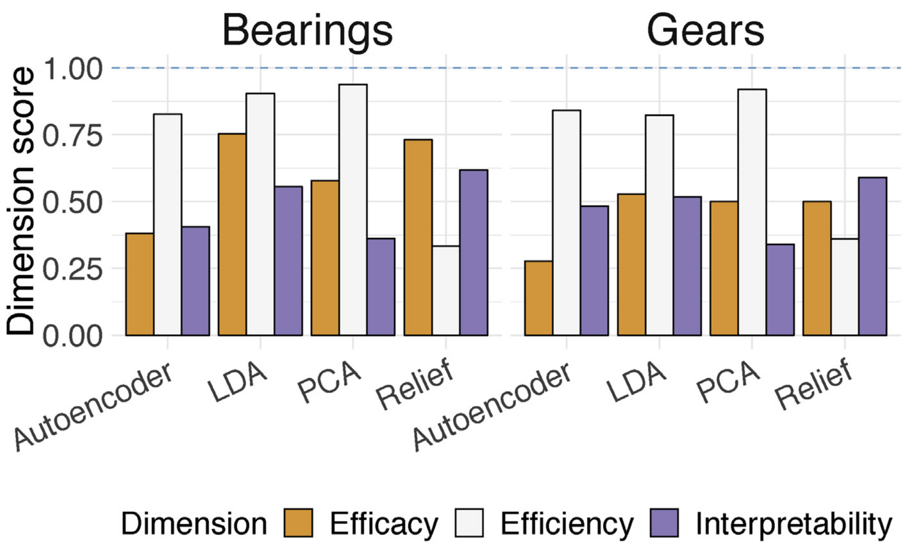

Fig. 4. Averaged main dimension scores by algorithm and case of use.

As the duration of the tests varied in each particular case because of the dependence on how long the bearings/gears needed to brake/get pitting, the datasets have different lengths. The final dimensions of the datasets are displayed in Table 3.

5. Results and discussion

After carrying out the simulations, the results have been evaluated according to the sub-dimensions and dimensions explained in Section 3.2. Complete results and aggregated tables for each case are displayed in the appendix (Tables A.5–D.8).

Table 4 shows the mean values of the different algorithms in each data source.

Overall, the algorithms tend to behave similarly from one case of use case to another, showing greater differences in Efficacy metrics. TPR and PPV are the most changing scores, and in both cases they obtain low values, penalizing specifically Autoencoders. Other interesting points are the huge differences that Relief has with the rest of the algorithms regarding Computational Efficiency. This happens because of the obtention of the metric, as the metrics has 0–1 range which is achieved by scaling elapsed time to that range using the minimum

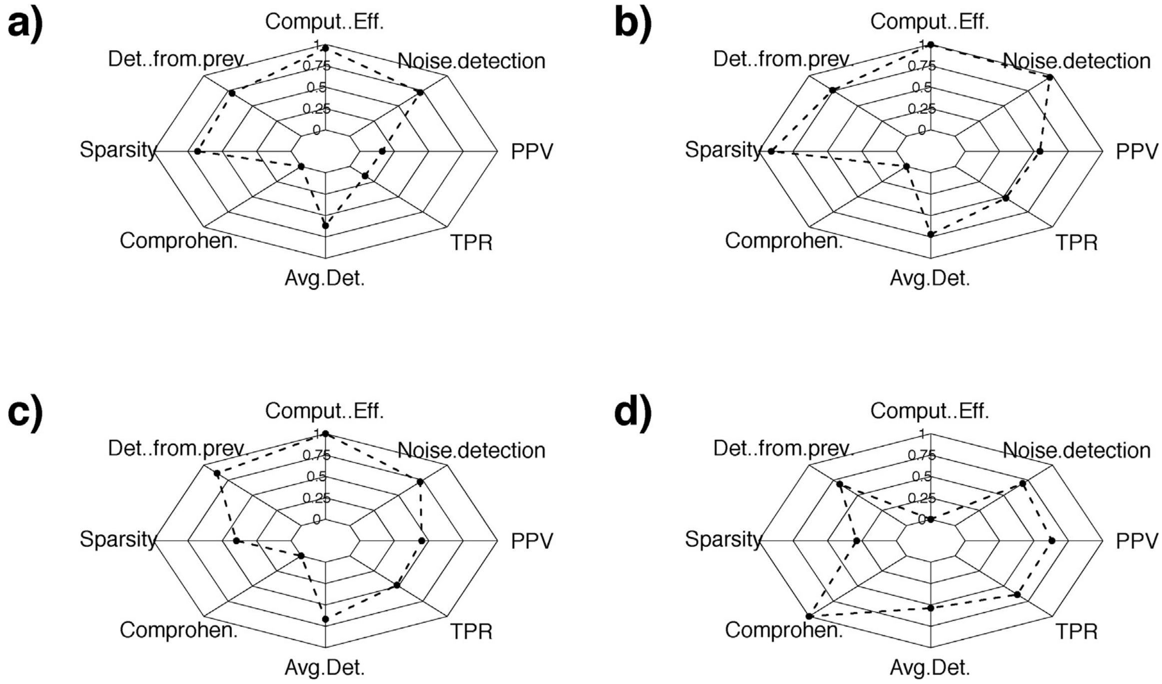

Fig. 5. Radar charts representing averaged sub-dimension scores of all cases of use for each algorithm. (a) Autoencoder, (b) LDA, (c) PCA, (d) Relief.

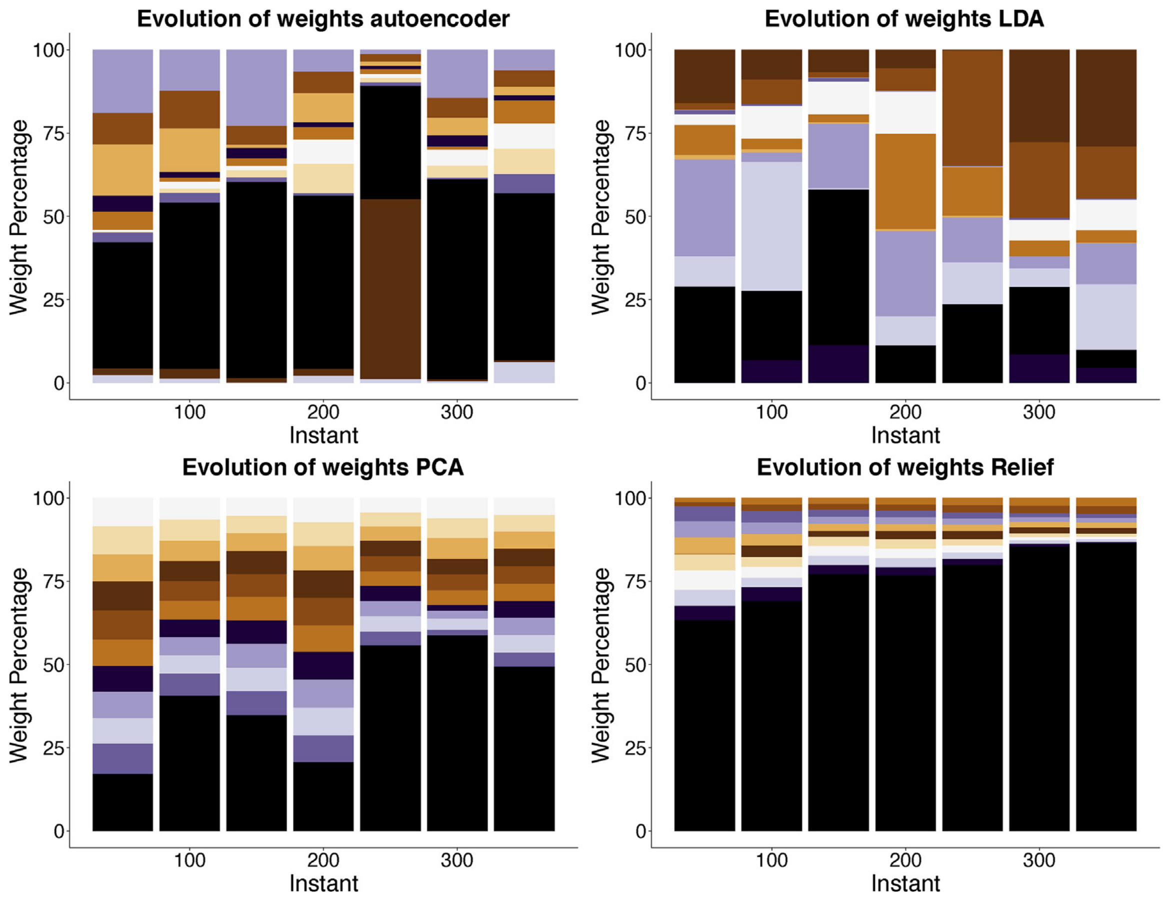

Fig. 6. Evolution of the weights of Gear Test 14. In all cases, the colored proportions represent top 10 variables according to each algorithm (they are different) while black color represents the sum of the rest of the variables.

and maximum values. Relief algorithm has been slow (possibly due to reasons related to implementation) and it has taken considerably longer times than the rest of the algorithms. Adding the scaling to that, it always obtains the lowest values (0) achieving very similar scores to the rest of the algorithms. This nuance of the scale should be taken into account, also the real elapsed time in addition to the sub-dimension should be considered. It is also interesting to note the capability of detecting noise that algorithms have demonstrated, in which LDA clearly excels. As the noise generated were uniformly distributed variables, one could expect to obtain better noise detection scores by supervised algorithms. However, possibly because it does not tend to generate sparse weights (see Sparsity), it is not the case of Relief.

Visualization of the three main dimensions also presents interesting additional information. Fig. 4 suggests that LDA should be selected for these cases of use followed by PCA, then Relief and lastly Autoencoder. However, this visualization also shows that, for applications aiming to have interpretable results, Relief would be a better option due to its high scores in Interpretability and Efficacy.

The radar-charts displayed in Fig. 5 represent the scores obtained in each of the sub-dimensions by each algorithm and show two different patterns: one for FP algorithms and another for the FS ones. This difference is mostly caused by both the comprehensiveness and computation efficiency scores. The first because it only belongs to FS algorithms; and, the latter, because of the aforementioned reasons related to scale.

Taking some of the weights of the 10 most valuable variables for each algorithm along the simulation process (Step D Weight storing in Fig. 2), Fig. 6 is created.

The evolution of the weights presented in the graph shows how recalculating the models has detected changes in the importance of the features as a consequence of changes in the input data, as the evolution in fractions along the time show. Note that each algorithm tends to behave in a deterministic way throughout the simulation process, either they assign big values to a small number of variables or they tend to give homogeneous values to each variable. This effect has been correctly reflected by the sparsity dimension, as it obtains small values for the case of Relief (see Table C.7 Gear Test 14) whereas LDA obtains a quite high sparsity value (meaning it tends to focus on a small group of variables).

In relation to the evolution of the weights over time, the different determinateness values have demonstrated their use for detecting both the change from initial values and the change in each DR recalculation step. As an example, in the evolution of the weights shown in Fig. 6, it is possible to see that the algorithm that has the lowest average determinateness value, Autoencoder, is clearly the algorithm with the greatest change in weights from the beginning to the end. Meanwhile, PCA, which has the highest determinateness from previous score, shows the smoothest changes in each iteration. Notice that, even if the algorithms should select the most interesting variables for the process, in this particular case of gear test 14, there is no clear agreement between

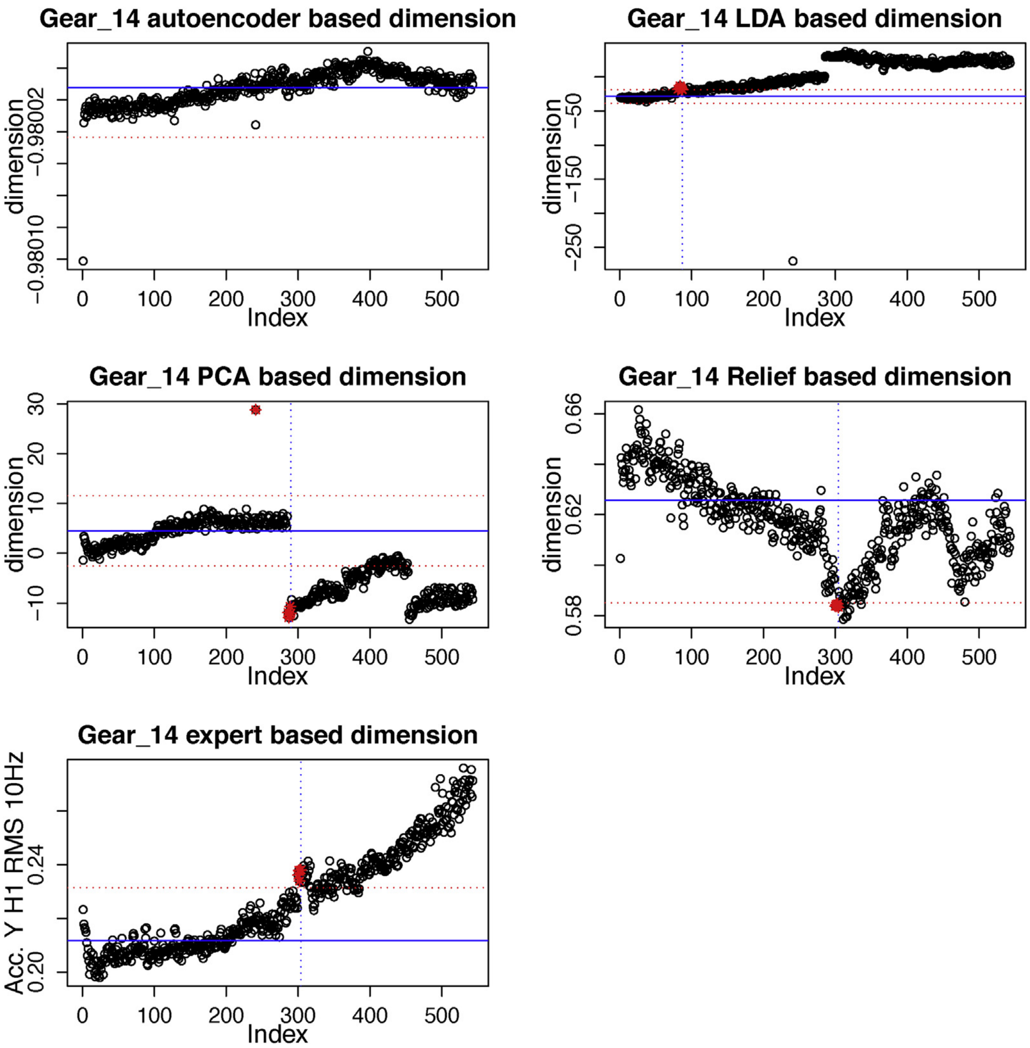

Fig. 7. Dimensions created by algorithms in Gear Test 14. The visible dimensions correspond to the Signal reconstruction step E in the simulation. Solid blue line corresponds to the mean value of the Shewhart chart and the horizontal discontinued lines symbolize the thresholds. The stopping point is represented as the vertical discontinued line. The points with red stars are out of control.

algorithms. This can be caused by different reasons: due to strong correlations in the variables (as they are originated from same or similar sensors); or, because of the different criteria used by DR algorithms to reduce dimensionality (variables maximizing the separation between degraded and non-degraded instances or combinations of variables retaining as much information as possible).

Fig. 7 depicts the new dimension created by each DR algorithm in Step (E) Signal reconstruction. This figure raises some contradictions. The stopping points found by most of the algorithms (vertical line in the graph) are really close to the stopping point obtained by the expert based dimension. However, this is not reflected by Efficacy metrics Table C.7 in Gear 14, just because the points found out of control by the algorithms (marked with a star in the graph) do not coincide with the points found out of control by the expert feature, therefore, the Efficacy score (without considering ND) is null for all the algorithms, when this should not be the case.

Evaluating the metrics globally, the following aspects are noteworthy:

Efficacy shows some limitations. On the one hand, it displays accurately the capability of detecting noise. On the other hand, considering DR methods can have more information than the best feature, a 100% agreement between expert and algorithms would be complicated to reach. Furthermore, the aforementioned differences between stopping points have not been considered, their inclusion as sub-dimension could affect the Efficacy score greatly. Additionally, it has to be taken into account that the expert criteria is used as base-line. Due to the large amount of features available in the Gears dataset, the expert was helped by a correlation analysis, which helped to find variables highly correlated with the degradation. Consequently, the expert criteria (probably trying to find monotonic and linearly correlated features) could have seen its prior judgments reinforced even if the features suggested by the correlation were not the best. Therefore, the base-line should be questioned in all the comparisons of features for monitoring purposes, where the real class value is hardly known. As a result, basing two thirds of the Efficacy only on TPR and PPV is not accurate, and this metric should be readjusted to reflect reality in a more detailed way.

Interpretability metric scores live up to the expectations. Sparsity is a good measure of the tendency to focus on small subsets of variables, nevertheless, it does not consider that FS algorithms use a single dimension for representing the process regardless of how the rest of the weights are distributed. However, this fact is corrected by the comprehensiveness sub-dimension, which penalizes DR techniques using mathematical mappings. Regarding average determinateness, it is not very clear whether it produces satisfactory results, as it might have been penalizing models that have encountered big changes in features from the initial calculations, which is not intrinsically bad.

In relation to the Efficiency, ignoring the limitations of the scaling, it represents the costs required for implementation. It considers the time needed for each recalculation, and the possibility of reducing the amount of recalculations without losing information.

All in all, representing the evaluation with three main dimensions seems a good intermediary in order to reduce the gap between the user and the implementer, as these metrics could be used as a common language between both.

Regarding the best DR algorithm for the monitoring purposes, LDA shows the best potential in both data sources, with good Efficacy and Efficiency but with worse results than Relief for Interpretability. The operator should consider including either one algorithm or the other. However, these results do not imply these algorithms are always the best for the cases of use (bearings/gears), and simulations and evaluations should be done in similar or different applications.

Finally, it is important to note that domain knowledge is essential for the development of CM techniques. Without domain knowledge it would not be possible to define appropriate features nor to use the small nuance of the degradation in order to use surpervised algorithms. In this particular case, that would imply not being capable of using supervised algorithms (LDA, Relief), which are among the best candidates. Therefore, it is important to note that the inclusion of domain knowledge is decisive for the implementation of any data-based model. This is because, if the knowledge is correctly translated into the data driven model, the performance will increase sharply.

6. Conclusions

This work presents an innovative approach based on the use of dimensionality reduction algorithms for the early detection of anomalies and the obtention of process related knowledge in scenarios with limited data. Four different DR algorithms have been compared for monitoring purposes in two different case scenarios: bearing monitoring and gears monitoring. Finally, the results of the simulation have been evaluated with a set of metrics created ad hoc considering the requirements found in the industry.

DR algorithms have demonstrated the capability of detecting changing variables in a system, as the evolution of the weights has shown. Additionally, that change has been monitored in order to determine acceptable/unacceptable thresholds. Furthermore, as this approach can easily work with large sets of variables, it becomes a good starting point for recommending variables in highly multivariate datasets, to better understand and detect changing variables and to initiate CM systems with little data.

Some specifically designed metrics have been used to evaluate the algorithms. These metrics are based on the industrial implementation needs, more precisely on the evaluation of the Interpretability, Efficacy and Efficiency. This inclusion of additional dimensions to the Efficacy takes other important factors into consideration. These other factors are neglected in other works, but directly affect the final chance of implementing the algorithms. The set of proposed sub-dimensions correctly reflects its fundamental purpose, with some need of adjustment for Efficacy, that should be improved by adding more informative sub-dimensions in the future.

There is no single and best algorithm for CM. Nevertheless, in the case scenarios presented, two algorithms stand out: LDA, which scores the best average efficiency and efficacy; and Relief, which provides the most understandable features, and scores well in the other dimensions. The final decision should be left to domain experts who can make the best decision considering the proposed metrics and the requirements of the final application.

This work focuses on two, often unattended, paradigms: the application of CM in systems with data limitations; and the need to include additional metrics besides performance for the analysis of the implementation of algorithms. However, some issues are not sufficiently addressed here and it would be interesting to explore and analyze them in future works. For instance, the inaccuracies of Efficacy metrics, which do not consider the distance between stopping points should be added as a sub-dimension to Efficacy; the study of the agreement between the important features according to different DR algorithms; the possibility of extending the algorithm, by incrementally learning from failures from one test to another; or, lastly, the extension of the approach to other industrial cases of use not related to rotating machinery and/or involving varying operations.

Declaration of interest

The authors declare no competing interests.

Appendix A. Model aggregated scores in sub-dimensions

Table A.5
Mean of algorithm scores by model.

Model	Comp. Eff.	Det. prev.	Sparsity	Compre.	Avg Det.	TPR	PPV	ND
Autoencoder	0.954	0.709	0.679	0	0.618	0.156	0.162	0.722
LDA	0.997	0.759	0.908	0	0.720	0.521	0.543	0.972
PCA	1.000	0.864	0.397	0	0.667	0.483	0.449	0.722
Relief	0.000	0.685	0.288	1	0.537	0.639	0.630	0.694

Appendix B. Model aggregated scores in main dimensions by data source

Table B.6
Scores by data source and model.

Data source	Model	Efficiency	Interpretability	Efficacy
Bearings	Autoencoder	0.827	0.407	0.381
	LDA	0.905	0.555	0.754
	PCA	0.938	0.362	0.578
	Relief	0.334	0.618	0.731
Gears	Autoencoder	0.841	0.483	0.278
	LDA	0.823	0.518	0.528
	PCA	0.920	0.341	0.500
	Relief	0.360	0.590	0.500

Appendix C. Scores in sub-dimensions for each test

Table C.7
Sub-dimension scores for all cases.

Gear	Model	Comp. Eff.	Det. prev.	Sparsity	Compre.	Avg Det.	TPR	PPV	ND
Bearing 1 1	Autoencoder	0.914	0.673	0.538	0.000	0.617	1.000	0.889	0.500
Bearing 1 1	LDA	0.990	0.818	0.962	0.000	0.787	0.625	0.556	1.000
Bearing 1 1	PCA	1.000	0.866	0.269	0.000	0.710	0.750	0.545	1.000
Bearing 1 1	Relief	0.000	0.784	0.231	1.000	0.637	0.750	1.000	0.750
Bearing 1 2	Autoencoder	0.925	0.763	0.846	0.000	0.685	0.000	0.000	0.750
Bearing 1 2	LDA	0.993	0.797	0.923	0.000	0.756	0.400	1.000	1.000
Bearing 1 2	PCA	1.000	0.904	0.115	0.000	0.809	0.600	1.000	0.500
Bearing 1 2	Relief	0.000	0.571	0.308	1.000	0.511	1.000	1.000	0.750
Bearing 2 1	Autoencoder	0.933	0.847	0.577	0.000	0.791	0.000	0.000	1.000
Bearing 2 1	LDA	0.993	0.796	0.923	0.000	0.795	1.000	0.667	1.000
Bearing 2 1	PCA	1.000	0.900	0.885	0.000	0.831	1.000	0.500	0.750
Bearing 2 1	Relief	0.000	0.559	0.308	1.000	0.494	1.000	0.667	0.250
Bearing 2 2	Autoencoder	0.954	0.683	0.577	0.000	0.663	0.400	0.571	0.750
Bearing 2 2	LDA	0.997	0.826	0.808	0.000	0.794	0.000	0.000	1.000
Bearing 2 2	PCA	1.000	0.941	0.269	0.000	0.791	0.000	0.000	0.750
Bearing 2 2	Relief	0.000	0.805	0.192	1.000	0.640	0.000	0.000	0.750
Bearing 3 1	Autoencoder	0.939	0.651	0.308	0.000	0.521	0.000	0.000	0.500
Bearing 3 1	LDA	1.000	0.842	0.885	0.000	0.801	0.667	0.667	1.000
Bearing 3 1	PCA	0.997	0.768	0.385	0.000	0.553	1.000	1.000	0.500
Bearing 3 1	Relief	0.000	0.497	0.500	1.000	0.495	1.000	1.000	0.750
Bearing 3 2	Autoencoder	0.957	0.678	0.654	0.000	0.548	0.000	0.000	0.500
Bearing 3 2	LDA	0.996	0.808	0.808	0.000	0.749	1.000	1.000	1.000
Bearing 3 2	PCA	1.000	0.881	0.231	0.000	0.665	0.000	0.000	0.500
Bearing 3 2	Relief	0.000	0.789	0.231	1.000	0.573	1.000	1.000	0.500
Gear 14	Autoencoder	0.985	0.627	0.722	0.000	0.473	0.000	0.000	0.500
Gear 14	LDA	0.999	0.657	0.958	0.000	0.597	0.000	0.000	1.000
Gear 14	PCA	1.000	0.848	0.569	0.000	0.515	0.000	0.000	0.750
Gear 14	Relief	0.000	0.732	0.396	1.000	0.518	0.000	0.000	0.750
Gear 18	Autoencoder	0.989	0.794	0.979	0.000	0.685	0.000	0.000	1.000
Gear 18	LDA	0.999	0.625	0.958	0.000	0.604	1.000	1.000	1.000
Gear 18	PCA	1.000	0.866	0.500	0.000	0.601	1.000	1.000	1.000
Gear 18	Relief	0.000	0.750	0.111	1.000	0.496	1.000	1.000	0.750
Gear 5	Autoencoder	0.986	0.662	0.910	0.000	0.582	0.000	0.000	1.000
Gear 5	LDA	1.000	0.657	0.951	0.000	0.594	0.000	0.000	0.750
Gear 5	PCA	1.000	0.804	0.354	0.000	0.531	0.000	0.000	0.750
Gear 5	Relief	0.000	0.679	0.319	1.000	0.472	0.000	0.000	1.000

Appendix D. Scores in main dimensions for each test

Table D.8
Main dimensions scores for all cases.

Gear	Model	Efficiency	Interpretability	Efficacy
Bearing 1 1	Autoencoder	0.794	0.385	0.796
Bearing 1 1	LDA	0.904	0.583	0.727
Bearing 1 1	PCA	0.933	0.326	0.765
Bearing 1 1	Relief	0.392	0.623	0.833
Bearing 1 2	Autoencoder	0.844	0.510	0.250
Bearing 1 2	LDA	0.895	0.560	0.800
Bearing 1 2	PCA	0.952	0.308	0.700
Bearing 1 2	Relief	0.286	0.606	0.917

Bearing 2 1	Autoencoder	0.890	0.456	0.333
Bearing 2 1	LDA	0.895	0.573	0.889
Bearing 2 1	PCA	0.950	0.572	0.750
Bearing 2 1	Relief	0.279	0.601	0.639
Bearing 2 2	Autoencoder	0.819	0.413	0.574
Bearing 2 2	LDA	0.912	0.534	0.333
Bearing 2 2	PCA	0.970	0.354	0.250
Bearing 2 2	Relief	0.402	0.611	0.250
Bearing 3 1	Autoencoder	0.795	0.276	0.167
Bearing 3 1	LDA	0.921	0.562	0.778
Bearing 3 1	PCA	0.883	0.312	0.833
Bearing 3 1	Relief	0.249	0.665	0.917
Bearing 3 2	Autoencoder	0.818	0.401	0.167
Bearing 3 2	LDA	0.902	0.519	1.000
Bearing 3 2	PCA	0.941	0.299	0.167
Bearing 3 2	Relief	0.394	0.601	0.833
Gear 14	Autoencoder	0.806	0.398	0.167
Gear 14	LDA	0.828	0.518	0.333
Gear 14	PCA	0.924	0.362	0.250
Gear 14	Relief	0.366	0.638	0.250
Gear 18	Autoencoder	0.892	0.555	0.333
Gear 18	LDA	0.812	0.521	1.000
Gear 18	PCA	0.933	0.367	1.000
Gear 18	Relief	0.375	0.536	0.917
Gear 5	Autoencoder	0.824	0.497	0.333
Gear 5	LDA	0.829	0.515	0.250
Gear 5	PCA	0.902	0.295	0.250
Gear 5	Relief	0.339	0.597	0.333

References

Al-Atat, H., Siegel, D., Lee, J., [1]. A systematic methodology for gearbox health assessment and fault classification. Int. J. Progn. Health Manage. 2, 1–16.

Bechhoefer, E., He, D., [2]. A process for data driven prognostics. Proceedings of the Prognostics, Health Management Solutions, Conference 24–26.

Saravanan, N., Ramachandran, K.I., [3]. Fault diagnosis of spur bevel gear box using discrete wavelet features and Decision Tree classification. Expert Syst. Appl. 36 (5), 9564–9573.

Saravanan, N., Ramachandran, K.I., [4]. Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). Expert Syst. Appl. 37 (6), 4168–4181.

Del Río, S., Fernández, S., Bravo-Imaz, I., Konde, E., Irigaray, A., [5]. AI and statistical technologies for manufacturing and maintenance strategies improvement: Health monitoring for electromechanical actuators. IGI Global 215–233.

Cipollini, F., Oneto, L., Coraddu, A., Savio, S., [6]. Unsupervised deep learning for induction motor bearings monitoring. Data-Enabled Discovery Appl. 3 (1), 1.

Bravo-imaz, I., Davari, H., Liu, Z., García-arribas, A., Arnaiz, A., Lee, J., [7]. Motor current signature analysis for gearbox condition monitoring under transient speeds using wavelet analysis and dual-level time synchronous averaging. Mech. Syst. Signal Process. 94, 73–84.

Kia, S.H., Henao, H., Capolino, G.A., [8]. A modeling approach for gearbox monitoring using stator current signature in induction machines. Conference Record - IAS Annual Meeting (IEEE Industry Applications Society) 1–6.

Li, B., Zhang, P.L., Tian, H., Mi, S.S., Liu, D.S., Ren, G.Q., [9]. A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox. Expert Syst. Appl. 38 (8), 10000–10009.

Djurdjanovic, D., Ni, J., Lee, J., [10]. Time–frequency based sensor fusion in the assessment and monitoring of machine performance degradation. Dyn. Syst. Control 15–22.

Casoetto, N., Djurdjanovic, D., Mayor, R., Ni, J., Lee, J., [11]. Multisensor process performance assessment through use of autoregressive modeling and feature maps. J. Manuf. Syst. 22 (1), 64.

Djurdjanovic, D., Lee, J., Ni, J., [12]. Watchdog agent – An infotronics-based prognostics approach for product performance degradation assessment and prediction. Adv. Eng. Inf. 17 (3–4), 109–125.

Bleakie, A., Djurdjanovic, D., [13]. Feature extraction, condition monitoring, and fault modeling in semiconductor manufacturing systems. Comput. Ind. 64 (3), 203–213.

Ferreiro, S., Jiménez, A., Madariaga, J., Novillo, E., Fernández, S., Alia, A., Morante, E., [14]. Health monitoring for electro-mechanical nose landing gear door actuator of a UAV, based on simulation modelling and data-driven techniques. Chem. Eng. Trans. 33.

Ruiz-Carcel, C., Starr, A., [15]. Development of a novel condition monitoring tool for linear actuators. The Twelfth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies 1–12.

Ruiz-Carcel, C., Starr, A., [16]. Data-based detection and diagnosis of faults in linear actuators. IEEE Transactions on Instrumentation and Measurement 67 (9), 2035–2047.

Saeys, Y., Inza, I., Larrañaga, P., [17]. A review of feature selection techniques in bioinformatics. Bioinformatics 23 (19), 2507–2517.

Guyon, I., Elisseeff, A., [18]. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Teti, R., Jemielniak, K., O'Donnell, G., Dornfeld, D., [19]. Advanced monitoring of machining operations. CIRP Ann. Manuf. Technol. 59 (12), 717–739.

Chen, J.C., Huang, L.H., Lan, M.A.X., Lee, M.S., [20]. Analysis of an effective sensing location for an in-process surface recognition system in turning operations. J. Ind. Technol. 15 (3), 1–6.

Lei, Y., [21]. Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery. Butterworth-Heinemann.

Vachtsevanos, G., Lewis, F.L., Roemer, M., Hess, A., Wu, B., [22]. Intelligent Fault Diagnosis and Prognosis for Engineering Systems, 456. Wiley Hoboken.

Adams, S., Meekins, R., Beling, P.A., Farinholt, K., Brown, N., Polter, S., Dong, Q., [23]. A comparison of feature selection and feature extraction techniques for condition monitoring of a hydraulic actuator. Annual Conference of the Prognostics and Health Management Society 311–321.

Wang, G., Song, Q., [24]. Selecting feature subset via constraint association rules. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7301 LNAI 304–321.

Jie, S., Hong, G., Rahman, M., Wong, Y., [25]. Feature extraction and selection in tool condition monitoring system. In: McKay, B., Slaney, J. (Eds.), AI 2002: Advances in Artificial Intelligence. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 487–497.

Ginart, A., Barlas, I., Goldin, J., Dorrity, J.L., [26]. Automated feature selection for embeddable Prognostic and Health Monitoring (PHM) architectures. AUTOTESTCON (Proceedings) 195–201.

Gerdes, M., Galar, D., Scholz, D., [27]. Automated parameter optimization for feature extraction for condition monitoring. 14th IMEKO TC10 Workshop on Technical Diagnostics 2016: New Perspectives in Measurements, Tools and Techniques for Systems Reliability, Maintainability and Safety 452–457.

Green, L.A., Mehr, D.R., [28]. What alters physicians' decisions to admit to the coronary care unit? J. Family Pract. 45 (3), 219–226.

Centre of the European Union, [29]. EU General Data Protection Regulation. Off. J. Eur. Union .

Abellan-Nebot, J.V., Romero Subirón, F., [30]. A review of machining monitoring systems based on artificial intelligence process models. Int. J. Adv. Manuf. Technol. 47, 237–257.

Sikorska, J.Z., Hodkiewicz, M., Ma, L., [31]. Prognostic modelling options for remaining useful life estimation by industry. Mech. Syst. Signal Proces. 25, 1803–1836.

Hu, C., Smith, W.A., Randall, R.B., Peng, Z., [32]. Development of a gear vibration indicator and its application in gear wear monitoring. Mech. Syst. Signal Process. 76-77, 319–336.

Al-Balushi, K.R., Addali, A., Charnley, B., Mba, D., [33]. Energy index technique for detection of acoustic emissions associated with incipient bearing failures. Appl. Acoustics 71, 812–821.

Lopez de Calle, K., Ferreiro, S., Konde, E., Bravo, I., Arnaiz, A., [34]. Feature selection for remaining useful life prediction of spur-gears. 30th International Conference on Condition Monitoring and Diagnostic Engineering Management (COMADEM

2017), Jost Institute for Tribotechnology, University of Central Lancashire Preston PR1 2HE Lancashire, UK. 98–105.
R Development Core Team, R., [35]. R: A Language and Environment for Statistical Computing, volume 1. .
Venables, W.N., Ripley, B.D., [36]. Modern Applied Statistics with S, fourth edition Springer, New York.
Baldi, P., [37]. Autoencoders, unsupervised learning, and deep architectures. J. Mach. Learn. Res. 37–49.
Dubossarsky, E., Tyshetskiy, Y., [38]. autoencoder: Sparse Autoencoder for Automatic Learning of Representative Features from Unlabeled Data. R package version 1.1..
Romanski, P., Kotthoff, L., [39]. FSelector: Selecting Attributes. R package version 0.21.. .
Kononenko, I., [40]. Estimating attributes: Analysis and extensions of RELIEF. European Conference on Machine Learning 171–182.
Montgomery, D.C., [41]. Introduction to Statistical Quality Control. John Wiley & Sons.
Webber, W., Moffat, A., Zobel, J., [42]. A similarity measure for indefinite rankings. ACM Transactions on Information Systems 28, 1–38.
Schmich, F., [43]. gespeR: Gene-Specific Phenotype EstimatoR. R package version 1.10.1.. .
Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., Varnier, C., [44]. PRONOSTIA: An experimental platform for bearings accelerated degradation tests. IEEE International Conference on Prognostics and Health Management, PHM'12 1–8.

9

A Context-Aware Oil Debris-Based Health Indicator for Wind Turbine Gearbox Condition Monitoring

- Authors: López de Calle, Kerman; Ferreiro, Susana; Roldán, Constantino; Ulazia, Alain.
- Publisher: MDPI
- Journal: Energies
- Year: 2019
- Quartile (Scimago/WoS): Q2/Q3
- DOI: 10.3390/en12173373

energies

MDPI

Article

A Context-Aware Oil Debris-Based Health Indicator for Wind Turbine Gearbox Condition Monitoring

Kerman López de Calle [1,2,*], Susana Ferreiro [1], Constantino Roldán-Paraponiaris [1] and Alain Ulazia [3]

[1] Intelligent Information Systems unit, IK4-TEKNIKER, Iñaki Goenaga street 5, 20600 Eibar, Spain
[2] Department of Computer Science and Artificial Intelligence, University of the Basque Country (UPV/EHU), Faculty of Informatics 649, 20080 Donostia, Spain
[3] Department of NE and Fluid Mechanics, University of the Basque Country (UPV/EHU), Otaola 29, 20600 Eibar, Spain
* Correspondence: kerman.lopezdecalle@tekniker.es

Received: 16 July 2019; Accepted: 27 August 2019; Published: 2 September 2019

check for updates

Abstract: One of the greatest challenges of optimising the correct operation of wind turbines is detecting the health status of their core components, such as gearboxes in particular. Gearbox monitoring is a widely studied topic in the literature, nevertheless, studies showing data of in-service wind turbines are less frequent and tend to present difficulties that are otherwise overlooked in test rig based works. This work presents the data of three wind turbines that have gearboxes in different damage stages. Besides including the data of the SCADA (Supervisory Control And Signal Acquisition) system, additional measurements of online optical oil debris sensors are also included. In addition to an analysis of the behaviour of particle generation in the turbines, a methodology to identify regimes of operation with lower variation is presented. These regimes are later utilised to develop a health index that considers operation states and provides valuable information regarding the state of the gearboxes. The proposed health index allows distinguishing damage severity between wind turbines as well as tracking the evolution of the damage over time.

Keywords: condition monitoring; condition based maintenance; wind turbine; oil debris monitoring; gearbox

1. Introduction

In a world with an ever-increasing electric energy demand, wind energy is getting attention, and has become the fastest growing renewable energy source because of its availability and abundance [1,2]. In this way, the global wind turbine installed power capacity is an increasing trend [3], and two phenomenons are arising: wind turbine (WT) hub size is increasing; and, the fast expansion of wind farms (WF) requires finding better settlements. Consequently, industries are expanding to inhospitable locations, such as offshore hard-to-reach places [4] in a pursuit of better wind resources, as typically, these locations provide higher wind power resources with less turbulence [5].

One of the biggest burdens wind farms face are the Operation & Maintenance (O&M) costs that, as the authors of [1] state, can comprise 10–20% of the total cost of energy (COE) for wind project, and reach up to 35% for a WT at the end of life, a figure that goes up to 30% in the case of big offshore wind farms [6]. Furthermore, various works have related higher failure rates in bigger WTs as compared to smaller ones [7–9]. Finding a positive correlation between average wind speed and failure rate that is reinforced in offshore sites [8,10]. Therefore, it is necessary to assess the health state of the systems and subsystems of WTs, in order to organise maintenance actions and reduce downtimes and losses due to unforeseen stops. In this field, condition monitoring systems (CMS) are well known technology with proven success for health status detection, fault identification and prediction [1,2,9]. Such technology allows identifying the state of the assets remotely, reducing considerably the need of visual inspections on site, which is a costly matter mostly for offshore farms [11].

Particularly, gearboxes represent a delicate component of WTs [10]. The various failure statistic analysis carried out lead to some controversy on its tendency to failure [12], as some of them have less failures reported [8,12], whereas others find high failure rates related to gearboxes [13]. Anyway, most of the studies associate the longest WT downtimes to this component [3,12] and emphasise the need of proper monitoring techniques to avoid them. Additionally, it is considered one of the costliest parts of the WTs [10] and although there are attempts to replace them with direct drives to reduce costs, studies question this assumption, and still the majority of offshore wind turbines rely on gearboxes [14]. Consequently, their monitoring is of vital importance.

Gearbox and gearbox related subsystem condition monitoring (CM) has been broadly addressed on the literature. For that purpose a wide variety of approaches have been tested, such as vibration, oil debris monitoring (ODM), acoustic emissions and current signature among others. Vibration-based CMS prevail over other kinds of sensors [2,9], with much research carried out on the field of signal processing of vibration signals in time, frequency, time/frequency and order domains [3]. However, some works find ODM techniques also interesting, for their higher correlation with wear creation [15], or for the added value they have for monitoring both the oil quality and the state of the gearbox parts [1,2].

Nevertheless, even if the advantages of CM are proven [2,11], its transfer from experimental tests to real WT use cases is less known in the literature [16]. This is because the variability of operation conditions of WTs affects the extraction of indicators while it specially damages the systems of WTs [2]. Most of the works presenting real in-service WT data are based on the use of SCADA (Supervisory Control and Data Acquisition) data, which is readily available in general. Typically, it is used to compare performances among WTs using power curves [17]. These benchmarking procedures are usual for other O&M issues such as pitch misalignment correction [18,19] and the identification of defective anemometers [20]. Additionally, temperatures from the SCADA have been modelled and compared over time to use differences as alarms as the different works reviewed by the authors of [21] show. However, the success of these techniques is limited [3]. Partly, because of external influences (such as the outside temperature) that require the alarms to be manually supervised by operators [21]. Consequently, the inclusion of additional CM sensors in operating WTs is flourishing [9], and an increasing number of works present findings from real cases of use:

- In the work by the authors of [22], two case studies are analysed: in the first one, physical principles that relate the difference in temperatures with the efficiency, rotational speed and power output are used, and the approach is validated by using the deviation of the temperature with respect to power in order to foresee a failure; in the second one, vibration and particle counter sensors are used and the evolution of the signals is studied before and after the replacement of a bearing. They suggest using cumulative particle counts to better detect failures instead of direct particle creation measurements and to combine various sensors in order to improve confidence in the diagnosis.
- In the work by the authors of [23], they create a health indicator based on the centroids proposed by a Self Organising Map in order to group WTs according to health status using SCADA data. This way operators are given additional information regarding the health state of the WTs, and can plan consequently.
- In the work by the authors of [16], current and vibration analyses are used to diagnose a turbine drive train, they emphasise on the difficulty of using signal processing techniques that are only proven at laboratory scale, and recognise the complexity of calculating the remaining useful life (RUL) and establishing damage/healthy thresholds, especially with a lack of available historical data.

- In the work by the authors of [24],the vibration signature of a sample of healthy wind turbines is shown, most relevant indicators are identified on averaged power spectra and the dependence of amplitudes on the operation is studied. They conclude that the high impact of wind speed on vibration amplitudes has to be taken into account to develop CMS.
- In the work by the authors of [14], the data gathered before and after a planetary gear was changed due to spalling is examined. They integrate temperature, vibration and particle counter signals in order to reduce false alarms, and prove the ability to distinguish healthy and warning states.
- In the work by the authors of [25], they suggest the use of moving averages (of both short and long term trends) of ODM to generate a count rate propagation model. Then, they establish an acceptance threshold based on the equivalent maximum angle of spall which is related to bearing geometry; and, lastly, they estimate remaining useful life (RUL).

Most of the works related to on service WT utilise vibration and/or oil debris sensors [3,9]. The works based on ODM from the previously mentioned ones [14,22,25] agree on the same difficulties for the development of ODM systems: the need of averaging or using cumulative values instead of using directly particle generation rates; and the tendency of particle creation rate to vary with operation. These findings are supported by the extensive work of the authors of [26], in which a full-scale WT gearbox of 750 kW is tested with in-line and online sensors and samples taken along the time. In their findings, the need of filtering influences caused by operational conditions is remarked; they recommend to focus in trends instead of in absolute values, and suggest considering big particle size (>14 µm) indicators in particular; also, they identify that damaged gearboxes have much higher debris generation rates than healthy ones.

Taking into consideration the interest of having real on-service WT operation data analysed, and that some of the limitations of ODM of WT are already identified on the literature, this work aims to provide a better insight for the development of ODMs. For that purpose, the data obtained in three WTs monitored with oil debris sensors are studied for a period of six months; the readings of the sensor are compared to other traditional SCADA based monitoring techniques; and, lastly, a study of the different operation states is carried out to determine which filtering criteria is better to develop an health index that considers operating conditions.

2. Data and Methodology

2.1. Wind Farm and Turbines

This study analyses the data produced by 3 WTs which are located in the wind farms at Bayo and Monteros, in Zaragoza (Spain). Both wind farms are close one another and undergo similar influences of the wind. The natural barriers of the Iberian System mountain ranges in the south and the Pyrenees mountain ranges in the north constitute a funnel effect that creates the meteorological occurrence known as cierzo; a dry, usually cold and accelerated flux of air intensified by the natural funnel going through the Ebro valley. Cierzo is more frequent during winter and the beginning of spring, and is compensated by the antagonistic phenomenon known as bochorno, that goes in the opposite direction to cierzo and tends to be softer. Additionally, these oposing phenomena provide the wind with copious kinetic energy and make the region an interesting location for the exploitation of wind energy [27].

The WTs have a 58 m diameter rotor and three blades. Their rated power is 850 kW and cut-in and cut-out wind speeds are 3 m/s and 20 m/s, respectively. They have planetary gearboxes with 1/62 transmission ratios coupled with asynchronous generators. The mineral lubricant is cleaned by offline oil filters and the online oil debris optical sensors is installed in a bypass of the lubrication system.

Regarding the health status of the gearboxes, visual and endoscopic inspections carried out on-site reveal different levels of damage. Two of the gearboxes show medium wear levels (WT 1 and WT 2) with micropitting present in most of the gears, whereas the last one is diagnosed with medium–high wear level showing greater surfaces damaged by micropitting in some gears and pitting in the sun gear. However, no corrective actions have been recommended yet.

2.2. Optical Oil Debris Sensor

Oil samples can be taken and analysed offline in laboratories, however, this procedure delays the decision making process and requires to access the WTs. Therefore, online oil debris sensors are an attractive way of determining the quality of the lubricant and safeguard the components of the gearbox.

In particular, this work uses a optical oil debris sensor. This kind of sensors monitor the fluid condition and contamination using optical technology by capturing high-resolution images of the moving fluid, and later applying advanced processes of image digitisation and spectral analysis. They detect, quantify and classify the particles bigger than 4 microns by size and/or shape, in addition of distinguishing these particles from air bubbles [28]. Besides wind turbine lubrication system monitoring, this kind of technology is well-suited for other industrial applications such as automotive, steel sector, wastewater treatment or cement industries [29] as all of the previous use lubrication systems.

2.3. Dataset

The study is based on a dataset consisting of six months long records of 3 WTs. The data records are taken with one minute frequency from the SCADA. At the same time, additional measurements provided by online optical oil debris sensors are taken. Variables from the SCADA represent the operation of the WTs, whereas the ones provided by the sensor indicate the amount of particles of size greater than 4, 6 and 14 micrometers (ISO.4, ISO.6 and ISO.14, respectively) present on the lubricating oil according to the ISO 4406 standard [30]. These values of the oil sensor represent the particle generation rate, as the oil is being continuously filtered. Details of the variables of the SCADA and the oil debris sensor with the units of measurement are presented in the Table 1.

Table 1. Variables available in the dataset and measurement units by data source.

Source	System	Units of Measurement
SCADA	Pitch angle	°
	Gearbox temperature	°C
	Wind speed	m/s
	Generator speed	rpm
	Active power	kW
Oil debris sensor	ISO.4	scale
	ISO.6	scale
	ISO.14	scale

For privacy reasons the data is shown in a normalised way along this work within a 0 to 1 range corresponding to minimum and maximum values of each of the variables in the dataset.

2.4. Methodology

In order to gain better insight on the use of oil debris sensors to obtain health indicators, the study has two parts: an exploration and correlation analysis stage, in which an overview of the data is presented and some methods of the literature contrasted; and the comparison of operation regions and health index (HI) development, where different operating regimes are compared and the most appropriate one is chosen as the basis to develop a HI. The methods used in each of the parts are presented below.

2.4.1. Exploration and Correlation Analysis

In an initial stage, various visualisation and correlation techniques are used:

- Pearson correlation: Coefficient used to measure the degree of linear association between two variables, presented in the work by the authors of [31].
- Spearman's correlation: Nonparametric coefficient that reflects the degree of monotonicity between two variables explained in the work by the authors of [32].
- Principal component analysis (PCA): PCA is a orthogonal transformation that turns a set of variables into a set of linearly uncorrelated variables. This technique is widely used for visualisation purposes in order to reduce multidimensional spaces to lower dimensional representations with the minimum information loss [33].
- Exponential moving average: In contrast to regular moving average where the average of a window of values in taken, this type of moving average is used when latest values are need to have more importance. The implementation used in this work can be found on the work by the authors of [34].
- Local regression (LOESS): This nonparametric method is used on local subsets of data. The implementation is based on the work by the authors of [35].
- Decision trees: Decision trees are machine learning (ML) algorithms used for classification. Their goal is to predict values of a target variable based on the inputs. Trees are built by splitting input variables with criteria that maximise the probability of having instances of certain group in each partition. The trees used in this work used Gini impurity index for partitioning and are implemented in the work by the authors of [36], which is based on the work by the authors of [37].

2.4.2. Comparison of Operation Regions and Health Index (HI) Development

During the initial exploration, the influence of the operation in the particle creation rates is detected; nevertheless, as there is no clear correlation identified between operation variables and particle creation, it is decided to consider only the measurements that are taken under the same operation conditions. Furthermore, a methodology is used to define which operating conditions are the most appropriate for monitoring purposes. The following techniques were used.

- Operation region (OR) definition: In order to find the optimal instants for taking measurements, several operating regions (OR) are explored. Each operating region is defined by a set of rules/criteria, such as wind speed in range (x m/s, y m/s), active power equals nominal power, etc. The ORs analysed in this work are suggested by experts in the domain, and are depicted in the following Figure 1 with a short explanation added in Table 2.
- Operation states (OS): Instead of considering each time instant individually as a data point with certain associated variable values (pitch angle, active power, ISO.4… etc.) groupings of data points have been studied. These groups or operation states are generated considering the different operation regions previously presented, and correspond to a set of chronologically continuous points over time fulfilling the rules proposed by the OR. Every time the machine works under the criteria of an OR we say it has entered in a new OS related to that OR, that lasts as long as the WT keeps working under the constraints of the OR.

Table 2. Descriptions of the different operating regions.

Operating Region	Characteristics
Nominal	Stable power generation. Varying pitch
N. & low pitch	Similar to nominal, but more restrictive and not including high wind speeds, delimited using pitch values.
Ramp to nominal	Ranges from about the middle of the power curve to beginning of nominal operation.
Ramp	Values taken only during the power ramp.
Pre-ramp	Values taken before the generator speed ramp starts.

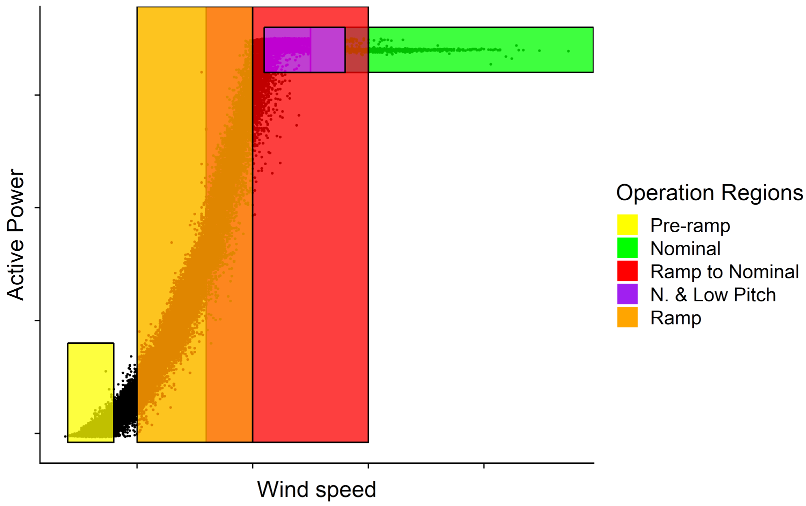

Figure 1. Operation regions over active power against wind speed plot.

- Operation clustering: With the purpose of analysing the steadiness of the different OR the following procedure was used in order to generate data cluster representing the variability of the operation in each OS according to the different ORs.

  1. Scale all the variables between 0 and 1 corresponding to the maximum and minimum values of each variable.
  2. Taking an OR (Example:Nominal) find the respective number of OS occurrences in the dataset $\{OS\}_{i=1}^{m}$, where m is the number of occurrences.
  3. Create a matrix for each OS_i where $i = 1, 2, \ldots, m$:

$$OS_i = \begin{pmatrix} a^i_{11} & a^i_{12} & \cdots & a^i_{1p} \\ a^i_{21} & a^i_{21} & \cdots & a^i_{2p} \\ \vdots & & \ddots & \vdots \\ a^i_{n_i 1} & \cdots & \cdots & a^i_{n_i p} \end{pmatrix} \tag{1}$$

     where p is equal to the number of sensors considered and n_i is the length of the i-th OS; therefore, these matrices contain the values of the p operation variables along the OS.
  4. Then, the difference vector of each variable is calculated by OS. This vectors represent the variability of the operation during the OS and give as a result the new matrix D:

$$D_i = \begin{pmatrix} d^i_{11} & d^i_{12} & \cdots & d^i_{1p} \\ d^i_{21} & d^i_{22} & \cdots & d^i_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ d^i_{n_i-1,1} & d^i_{n_i-1,2} & \cdots & d^i_{n-1,p} \end{pmatrix}, \quad i = 1, 2, \ldots, m \tag{2}$$

     where $d_{jk} = a^i_{j+1,k} - a^i_{jk}$,that is, each element of the difference vector is the difference between the measurement in that instant (j) and the following measurement $(j+1)$, for each $j = 1, 2, \ldots, (n-1), k = 1, 2, \ldots, p$.

5. Then matrix R is computed.

$$R = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \dots & r_{mp} \end{pmatrix} \tag{3}$$

R is the result of computing the columnwise quadratic mean of the D_i matrices, and represents the average values of the variability considering both negative and positive values. They are computed in the following way.

$$r_{ik} = \sqrt{\frac{\sum_{j=1}^{n_i-1} d_{jk}}{n_i - 1}} \tag{4}$$

6. From the R matrix, two metrics are obtained:
   (a) Centroid: The average position of the points contained in W. Computed as follows.

$$Centroid = (\mu_1, \dots, \mu_p) \tag{5}$$

   where for all $k = 1, \dots, p$ the average μ_k is calculated as follows.

$$\mu_k = \frac{1}{m} \sum_{i=1}^{m} r_{ik} \tag{6}$$

   (b) Cluster dispersion: Mean of the variable variance value that represents how disperse the cluster is; it is calculated as follows.

$$Dispersion = \sum_{k=1}^{p} \sigma_k \tag{7}$$

   Each σ_k for $k = 1, 2, \dots, p$ is the standard deviation computed in the following way.

$$\sigma_k = \frac{1}{m} \sum_{i=1}^{m} (r_{ik} - \mu_k)^2$$

This procedure is repeated separately for each WT and OR. Therefore, there are five ORs by three WTs, a total of 15 data clusters.

- Operation state and cluster metrics: From the clusters of data and the OS some metrics are calculated that help identifying the most interesting OR. These metrics are as follows.
  - Weekly occurrence ratio: Average number of times per week the WT enters in an OS as defined in the OR.
  - Steadiness: The euclidean distance from the centroid (or mean point) of a cluster to the total steadiness (no variation) point.
  - Dispersion: Indicates how spread the data points within a cluster are. Defined previously in Cluster dispersion.

3. Results

As in Section 2.4, results chapter is divided in two parts. The first part, Section 3.1, explains the exploratory analysis that is carried out over the dataset and the relations found between the variables. The second part, Section 3.2, shows the steps that were taken in order to identify the best conditions for obtaining measurements along the time in order to obtain a health index of the gearboxes.

3.1. Exploration and Correlation Analysis

Taken a sample of the whole dataset, Pearsons and Spearmans correlations are studied. In order to identify any possible difference between power generation and during no generation, correlation is also measured in separate samples. However, no significant correlations (neither Pearsons nor Spearmans) are found between operation variables and particle generation data. Regarding the operational variables, some show high degree of association because of the control system. Furthermore, the association of the same variables among WTs over overlapped time spans yields high correlation which means they face similar environmental conditions (wind). However, this is not the case of ODSs, that do not correlate from one WT to another. Nevertheless, in light of the strong correlations between particle indicators (ISO.4, ISO.6 and ISO.14), and following the advice provided by the authors of [26], it is decided to follow the study using ISO.14 indicator as only indicator for particles in order to simplify the study.

After the brute correlation study, the variables are visually studied against the wind speed in Figure 2. The different variables of the SCADA data plotted against the wind speed show the typical patterns that can be found in wind turbines, and are extremely similar one to another. The greatest (but yet small) differences are found in gearbox temperature, suggesting there could be some differences in the cooling system or on the efficiency of the gearboxes.

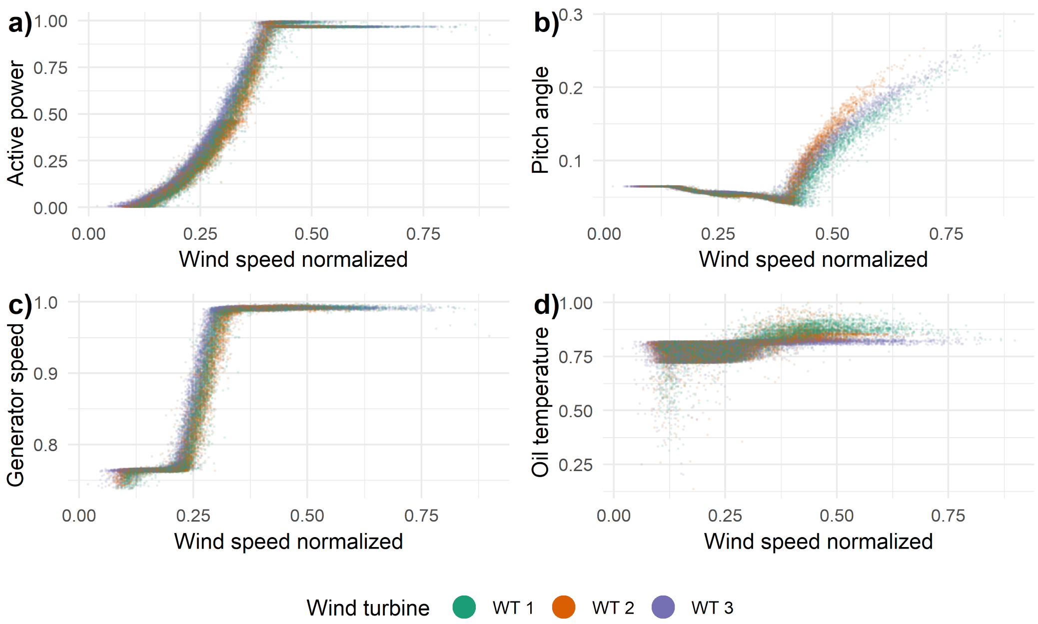

Figure 2. SCADA variables against wind speed by WT. (**a**) Active power against wind speed, (**b**) pitch angle against wind speed, (**c**) generator speed against wind speed and (**d**) oil temperature against wind speed.

As the signals of the ODSs are discrete, much noisier and is almost impossible to visualise anything in the raw measurements, the measurements are given some pretreatments by averaging the values in 0.33 m/s wind speed bins, which creates the pattern visible in Figure 3c.

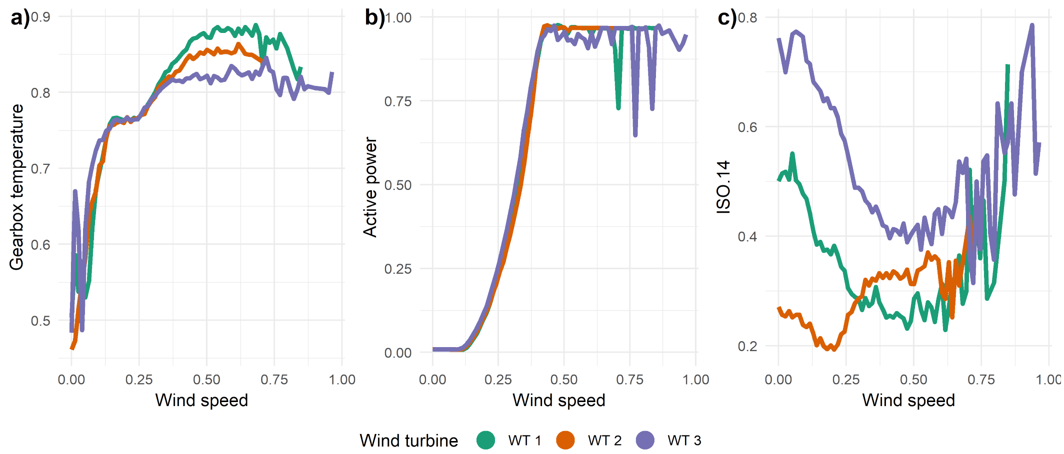

Figure 3. Averaged values of gearbox temperature, active power and ISO.14 in 0.33 m/s wind speed bins. (**a**) Gearbox temperature. (**b**) Active power. (**c**) ISO.14.

Averaged values show great differences between the particle creation rates among WTs. At the same time, the influence of operation over particle generation is visible. Interestingly, the behaviours do not coincide exactly between WTs: WT 1 and WT 3 show big similarities, with high wear creation with low wind speeds and lower wear creation at medium speed or nominal operation; meanwhile, WT 2 shows a different pattern, as its wear creation increases proportionally with wind speed. These differences in the behaviours of the WTs regarding particle creation and operation are also present in the averaged values of wear generation during power production and no power production (including: idling, generator turning without active power generation and idling because of overload) as Figure 4 demonstrates.

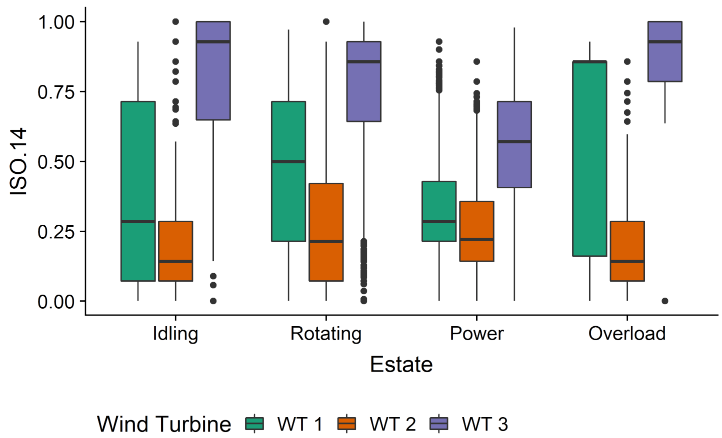

Figure 4. Boxplots of ISO.14 particle creation rates during different operations.

Again, WT 2 does not act as the other WTs. In any case, the previous figures suggest the particle creation is greater during no power generation, meaning braking and acceleration could be causing higher particle creations. Furthermore, there is a clear distinction in the mean level of particle creation rates that match the visual diagnostics of the gearboxes, showing higher values in WT 3, and lower values for WT 1 and WT 2. This variation of particle levels that indicates disparate damage severity, is complex to detect by just paying attention to the the SCADA variables. As Figure 3a,b shows, the same binned in variables typically used for condition monitoring by benchmarking (Active power and gearbox temperature) are not sufficiently different in order to make comparisons between turbines and determine whether WTs could be damaged. Whereas these differences are clearly visible in the binned ISO.14 values (Figure 3c).

This fact is clearer when cumulative particle creations are used. Instead of using raw signals, using cumulative particle rates provides a better insight of the degradation process, as it allows us distinguishing changes in the slopes. In Figure 5, we see clear and increasing differences between WTs in the trends generated when plotting cumulative particle creations against cumulative power generation. If cumulative temperature is observed, the differences among WTs, even if existent, are quite small which reduces the possibility to correctly diagnose failures using only SCADA data. Additionally, the presence of similar shapes for the three turbines in both temperature and particle creation and considering the same periods of time are studied indicate some common factor could be causing the sharp increase in the middle of the curves, which is visible in both variables.

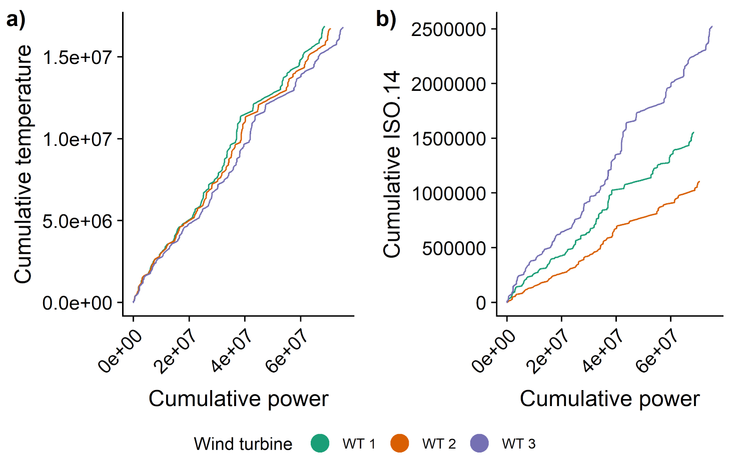

Figure 5. Cumulative values against cumulative active power. (**a**) Temperature. (**b**) ISO.14.

In order to reproduce previous findings in the literature, it is decided to analyse braking and acceleration registered in the SCADA data. For doing so, the Generator speed of five days of operation is taken and it is manually labelled adding "braking", "boosting" or "other" labels. Then, five lagged variables of the generator speed, a exponentially smoothed generator speed (using a bin size of 15) and a difference vector are created. With this data a decision tree is trained using the default parameters for classification cases and it is used to segment the remaining generator speed data in brake/boost/other. With the data split in these groups, it is possible to study the sequences occurring in the data. Boosting and braking sequences are studied by measuring the spearmans correlations of the ISO.14 variables with the smoothed generator speed. In Figure 6 the distribution of the correlations obtained by WT is presented.

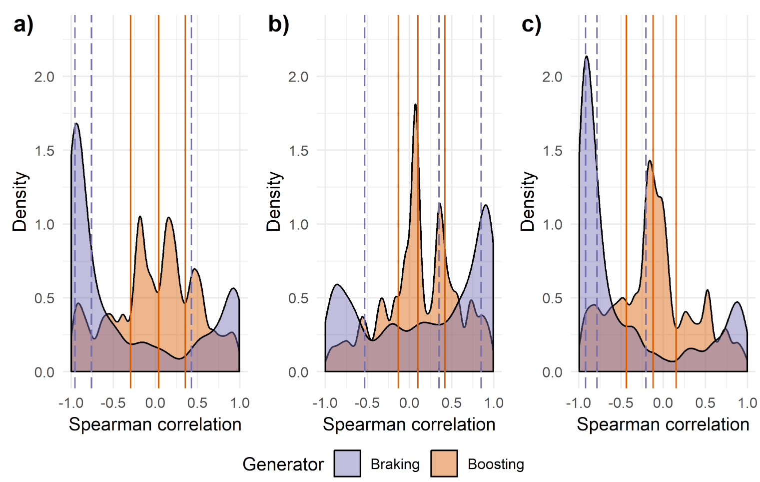

Figure 6. Density plot of the spearman correlation of ISO.14 in respect to the exponentially smoothed generator speed by wind turbine, vertical lines represent quartiles. (**a**) WT 1, (**b**) WT 2 and (**c**) WT 3.

The distribution of the correlation shows here different behaviours. In the braking sequences, WT 1 and WT 3 (Figure 7a,c) have bimodal distributions with a minor mode in strong positive correlation values and the major mode in very strong negative correlation values. This implies that there is a predominant tendency to create more particles during stops (speed decreases and particle generation increases), but is not always the case, as in some cases the correlation is positive (speed decrease with particle decrease). In WT 2 the opposite behaviour is identified, even if the correlation distribution is also bimodal, the major mode is on positively correlated values, meaning in this WT there is a tendency to decrease particle generation when the generator is stopping. Regarding the boosting sequences, the overall correlation values are quite low, which implies there is no clear relation between the increasing speed and the particle creation. The predominance of the major mode in very negative correlation values together with the quartile lines so far from the 0 value indicates braking generate an increase in particle creation, at least for turbines WT 1 and WT 3.

Taking WT 3, boosting and braking sequences were analysed in depth. The following Figure 7 presents two examples of sequences with strong spearman correlation values with negative and positive correlation.

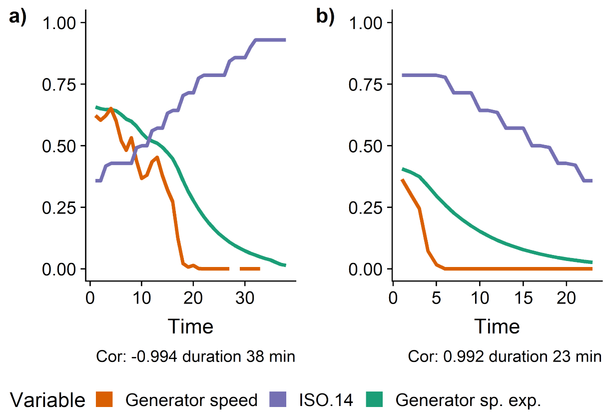

Figure 7. Examples of OSs showing high spearman correlation between ISO.14 and smoothed generator speed. (**a**) Positive correlation. (**b**) Negative correlation.

Interestingly, despite the there is a clear unbalance in the number occurrences, stopping can lead to both an increase or a decrease in particle generation. Note that the generator speed decreases faster than the ISO.14 level, and the nature of the exponentially smoothed speed is more similar to the one of the ISO.14 variable that is more influenced by the inertia of the system than generator speed.

The same procedure is followed for boosting, in this case, considering predominant correlation is near 0 (meaning there is no monotonicity) examples with low correlation are also studied. Figure 8 displays occurrences with high positive correlations (a), highly negative correlations (b) and no correlation (c).

With the uniform distribution of correlation for boosting cases and the different cases shown in Figure 8 there is no way of identifying an expected behaviour of the particle generation during boosting sequences. Furthermore, Figure 8c reveals an unexpected behaviour during idling. As the sensor is giving high ISO.14 particle levels. This fact occurs mostly in WT 3 but is also reported in WT 1, but with a lower frequency. Off-line oil filters should operate continuously regardless of the operation of the machine, but this finding suggest the filter could be stopping in certain situations, which explains also the big difference of particle generation found in Figure 4.

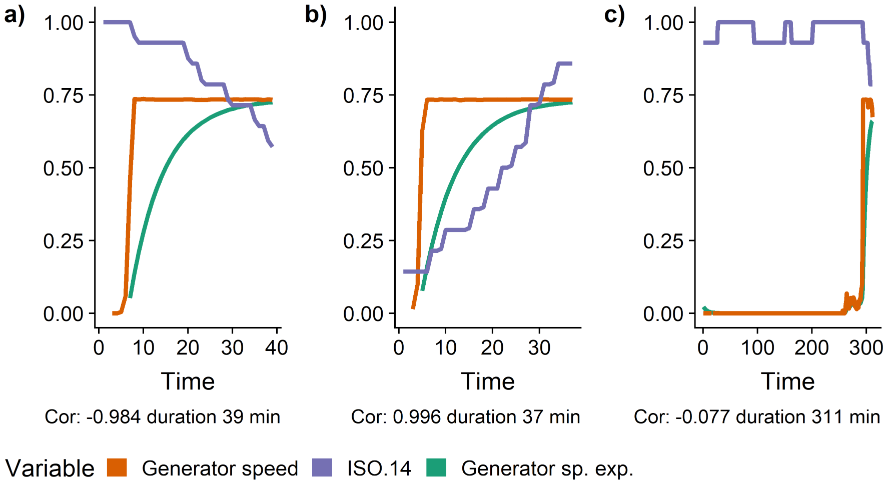

Figure 8. Boosting examples with disparate spearman correlation values between ISO.14 and smoothed generator speed. (**a**) Positive correlation. (**b**) Negative correlation. (**c**) No correlation.

3.2. Comparison of Operation Regions and Health Index (HI) Development

At this point the influence of operation over the particle creation is evident; therefore, it is decided to isolate measurements taken under similar conditions to compare them along the time and use these filtered measures to build a HT. For this purpose, the procedure explained in Section 2.4.2 is carried out. The operational data is taken, different operation regions are defined one by one as explained in Figure 1 , the operation states produced in each turbine are generated and once all WT have been processed, the OSs are studied. Considering that there could be a delay between operation conditions and the effect of those conditions on the oil debris content, it is decided to remove occurrences (Operation States) shorter than a minimum length. In order to define the most appropriate minimum required duration, the following table, Table 3, is created, where the effect of filtering with different duration is presented.

Table 3. Weekly OS frequency of different minimum times by each wind turbine (WT) and operation region (OR).

Turbine	OR	>5	>10	>15	>30	>45
WT 1	N. & pitch	179	64	30	8	3
WT 1	Nominal	1073	532	357	189	113
WT 1	Ramp-to-nominal	2204	1159	794	417	294
WT 1	Pre-ramp	1119	367	155	25	9
WT 1	Ramp	3451	1906	1380	757	481
WT 2	N. & pitch	44	15	6	2	2
WT 2	Nominal	985	507	339	183	128
WT 2	Ramp-to-nominal	1833	1057	758	427	297
WT 2	Pre-ramp	1040	325	140	29	5
WT 2	Ramp	3316	1893	1340	744	502
WT 3	N. & pitch	262	103	50	21	11
WT 3	Nominal	1144	600	410	206	127
WT 3	Ramp-to-nominal	2191	1159	820	453	320
WT 3	Pre-ramp	1299	403	174	19	3
WT 3	Ramp	3282	1840	1339	754	511

The time filter reveals that most occurrences have very short duration, as moving the filter from 5 to 10 min reduces the number of occurrences to a half in most of the ORs. Furthermore, very restrictive ORs, such as N. & pitch, are less present in the database, and the ones with wider limits (that also coincide with the most frequent wind speeds) are more present in the dataset. Considering longer OSs should reduce the amount of noise created by previous operation regimes, while there should be a sufficient week rate in order to obtain enough indicators over time, it is decided to keep OSs that last longer than 10 min.

Following with the procedure, once data matrix R is obtained for each WT, it is possible to see the different clusters that are created and represent the variability of the measurements. Figure 9 represents the two principal components of the operational variables in the dataset (pitch angle, gearbox temperature, wind speed, generator speed and active power) that are generated in WT 1 using PCA algorithm. The variability retained by each dimension is displayed in the axes. It is interesting to see the representations that the different ORs take. The first principal component (Dim 1) mostly contains pitch difference, generator speed and gearbox temperature. Most of the clusters have great part of the variation related to this feature, whereas in the second component (Dim 2), wind speed, active power and generator speed are causing most of the variation, and this dimension affects mainly Ramp and Ramp-to-nominal clusters that show very high dispersion.

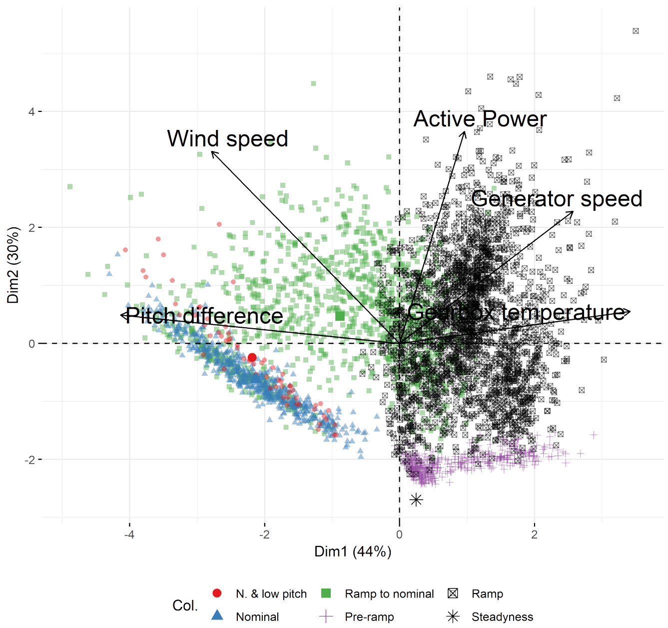

Figure 9. Example of the PCA obtained from the different OR and RMS (root mean square) values of the OS generated in WT 1.

The steadiness point, that is, the point with no variation is also represented on the graphs as a star. The euclidean distance to that point (the steadiness) is measured and the results presented in the following Table 4:

Table 4. Steadiness of the different OR by WT.

	N. & Pitch	Nominal	Ramp to Nominal	Pre-Ramp	Ramp
WT 1	2.00×10^{-3}	1.50×10^{-3}	1.50×10^{-3}	3.00×10^{-4}	1.10×10^{-3}
WT 2	1.40×10^{-3}	1.00×10^{-3}	1.10×10^{-3}	4.00×10^{-4}	1.00×10^{-3}
WT 3	2.00×10^{-3}	1.50×10^{-3}	1.50×10^{-3}	4.00×10^{-4}	1.10×10^{-3}
Average	1.80×10^{-3}	1.30×10^{-3}	1.40×10^{-3}	4.00×10^{-4}	1.10×10^{-3}

There are extreme differences regarding how much operation variables vary during the OSs. From the least steady OR (Ramp-to-nominal) to the steadiest one (Pre-ramp) the distance is ten times bigger, meaning the Pre-ramp operation regime is much steadier than the Ramp-to-nominal OR.

Regarding the variation of the clusters, that is, how close from the centroid the data-points are, the results displayed in Table 5 are obtained.

Table 5. Variation of the ORs by WT.

	N. & Pitch	Nominal	Ramp to Nominal	Pre-Ramp	Ramp
WT 1	7.56×10^{-5}	4.64×10^{-5}	7.58×10^{-5}	6.96×10^{-6}	5.76×10^{-4}
WT 2	1.98×10^{-5}	2.78×10^{-5}	6.19×10^{-4}	9.03×10^{-6}	5.36×10^{-4}
WT 3	7.31×10^{-5}	5.51×10^{-5}	6.80×10^{-4}	9.32×10^{-6}	6.34×10^{-4}
Average	5.62×10^{-5}	4.31×10^{-5}	6.86×10^{-4}	8.44×10^{-6}	5.82×10^{-4}

The smallest variation values are obtained by Pre-ramp OR, with clear difference to with the rest of the ORs.

Considering the information provided by the frequency study, the PCA visualisation, distance to steadiness and centroid variation, which OR should be chosen for HI purposes is determined. N. & pitch is discarded because of it low occurrence frequency, Ramp and Ramp-to-nominal show high dispersion, which means there are operational fluctuations in the ORs they delimit. Between Nominal and Pre-ramp ORs, according to steadiness and dispersion criteria Pre-ramp should be chosen, therefore, it is decided to consider the measurements taken under Pre-ramp OR.

Taking the Pre-ramp OR, the ODM variables as well as the other SCADA variables are averaged with the RMS (root mean square) value of each OS. Figures 10 and 11 display the RMS values obtained from the OSs generated using these points. Over the points, the coloured curves represent the fitting provided by the LOESS algorithm using a high span fraction (0.75) in order to retain the trend instead of local variations. The grey shade represents the 95% confidence interval of the fitted curve.

Note that power generation stays under very strict limits, shows almost no variation along the time and, besides the latest trend values (that have less data points), almost no difference between turbines. Similarly, temperature shows higher variation but the trend keeps very stable along the time. Interestingly, the differences in temperature that are visible do not match the expectations: WT 1 and WT 2 (diagnosed with medium wear wear level) have higher temperature values than WT 3 (diagnosed with medium–high wear level). Regarding particle generation, it is possible to see the sudden increase WT 2 and WT 3 have could be related to the same increase of OSs with higher Active power production (Figure 10b). The differences in trend values do correspond to the damage levels of the gearboxes, showing that WT 3 is in a worse condition than the rest of the turbines.

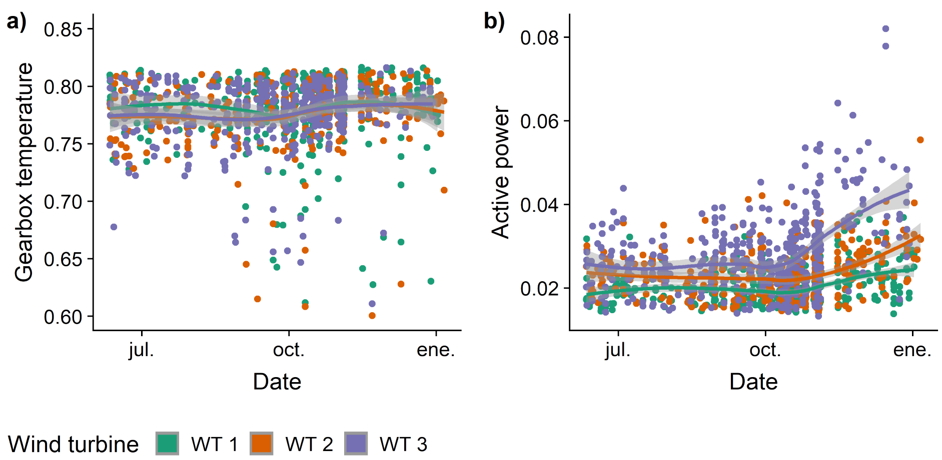

Figure 10. RMS values of different variable OSs throughout the time, generated using type Nominal OR. Curve is obtained by local regression (LOESS) smoothing. (**a**) Gearbox temperature. (**b**) Active power.

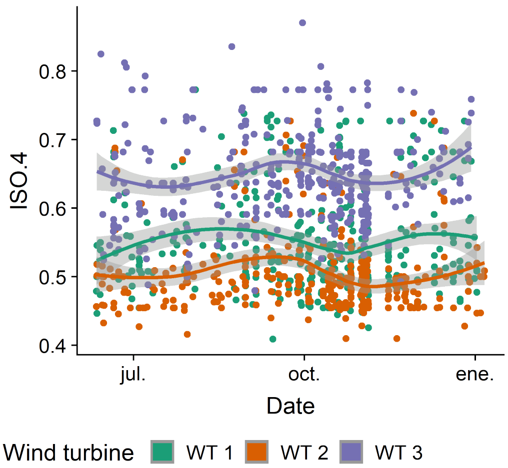

Figure 11. RMS values of ISO.14 variable and smoothed trend fitted with LOESS.

Lastly, the comparison of the evolution of the particles in different ISO particle size (Figure 12) demonstrates the correlations among ISOs that are found in the correlation analysis. ISO.14 shows the greatest difference between WTs, which means it could reflect the damage status in a more accurate way, and therefore it is better suited for comparison purposes. However, the contrast between WTs is clearly visible also in the rest of the ISOs. The scale of the figure is ranged between 0 and 25 ISO values and thresholds proposed by laboratory experience are included for both warnings (20/18/15) and danger (21/19/16) for ISO 4/6/14, respectively. Note that, even if the trends are far from reaching the thresholds, the real-time measurements surpass the thresholds more than once. This means that operation variation can cause great spikes in the particle generation rate and, in order to obtain an overview of the condition of the gearbox, it is required to focus on the trends instead of in instantaneous values. Considering ISO.14 keeps under very low values, the state of the gearbox could be considered yet to be healthy. Either if the smoothed trend would reach values close to the thresholds or it would increase sharply, gearbox should be considered in danger.

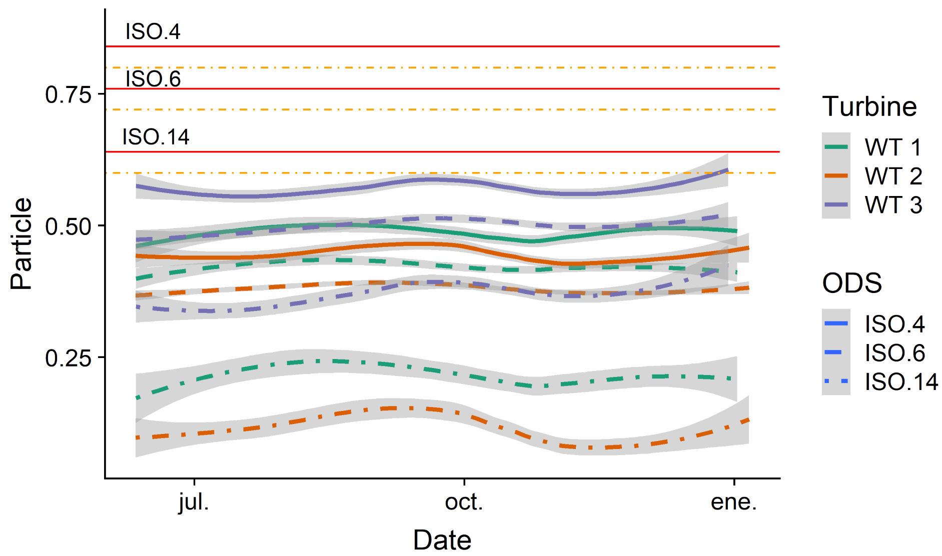

Figure 12. LOESS smoothed particle generation rates (ISO 4, 6 and 14) of Pre-ramp ORs by WT. The horizontal lines at the top represent warning and danger thresholds for the different particle sizes.

4. Discussion and Future Outlook

This study presents the data obtained through the monitoring of three WTs with oil debris optical sensors during six months. In this way, turbines with gearboxes in different stages of deterioration are presented and compared. Initially, an attempt to correlate the operation of the WTs and ODSs has been carried out. After that, a way to identify repeatable and steady operation regimes has been used as a basis for developing a health indicator.

Different works have shown in-service oil debris monitoring in the literature [22,25] or have studied the relation of the operation in behaviour of ODSs in full scale test rigs [26]. Nevertheless, this work is particular as it presents both: first an study of the influence of the operation on debris generation; and then, it proposes a method for the identification of the optimal instants to obtain measurements considering operation.

The number of studied turbines is reduced and the installation of the lubrication systems is equal according to our knowledge. However, two clear behaviours have been discovered during the exploratory phase: two of the turbines show similar trends whereas another one seems to behave in a completely different way. Therefore, the results here displayed should be understood in this context.

We have faced difficulties when working with the noise of raw ODS's measurements that are also reported in other works [22,25,26], and using the techniques already present in the literature (cumulative particle rates [22,25]) has been useful to reduce the noise. In comparison to the sole use of variables that are directly obtained from the SCADA (active power and generator temperature), significant improvement has been detected when using ODSs, as they show greater differences between the health status of the gearbox. This fact validates the thesis of the need of including additional sensors for defining with higher accuracy the damage levels of the systems [22].

Due to the varying operation and the difference behaviours of WTs, it has been difficult to find clear correlations between operation and particle creation. ISO measurements are highly correlated among themselves, but it is difficult to find association to other operation variables. Only in a detailed inspection of braking and acceleration periods, contrarily to what is reported in the literature [26], a general tendency to increase particle generation has been detected when generator is braking, whereas no increase has been detected during acceleration. Furthermore, some periods of high particle creation have been identified when the generator is idling, this phenomenon could be caused by different behaviours of the oil filtering system, but this assumption has not been proved.

In concordance to the findings of the authors of [26], gearboxes that are more deteriorated (the case of WT 3) have shown a tendency to generate more particles rates than gearboxes in better condition (WT 1 and WT 2). This fact is clearly visible in the cumulative particle creation or the binned ISO against wind speed plots. Also, the pattern of particle generation that is visible in the binned wind speed of WT 1 and 3 reminds the Stribeck curve, which could explain the high particle creation rates at low speeds and the high rates at higher speeds. Furthermore, the differences in particle creation rates are more evident when considering bigger particle size (ISO.14), as the authors of [26] stated. Nevertheless, the sensor in WT 2 provides patterns that differ from the sensors in WT 1 and WT 2. Considering how close the patterns are in WT 1 and WT 3, two hypothesis could be possible: either sensor is not working correctly; or the lubrication system is affected by factors not included in the SCADA.

Regarding the development of the HI, in contrast to the proposals of other works in the literature that make trends over the whole data [25], our method considers the operation regime in which the measurements are taken and uses only measurements that are obtained under the same circumstances. On the one hand, this leads to have periods of time without indicators, but this issue has been considered by choosing ORs with high number of occurrences. On the other hand, an analysis of the operation has been carried out to identify instants with lesser variation in the operation, which should provide more stable measurements and less influenced by the operation. However, using smoothing techniques has still been necessary in order to make trends visible.

According to the analysis of the operation, the WTs tend to move fastly from one OR to another, as the high number of short OSs reveals. Regarding the steadiness of the different ORs analysed, in the pre-ramp zone the operational variables remain more stable than in the rest of ORs that are in the power ramp, as they show a bigger distance to steadiness point.

As the authors of [16] recognise, establishing limits for admissible and nonadmissible damage is one of the biggest challenges in in-service machinery. The limits proposed in this work are based on laboratory experience but might need to be readjusted by interacting with bigger WT databases, as there is no total failure record in the dataset under study.

Lastly, even if ODSs have been demonstrated capable of detecting diverse levels of damage in gearboxes, with the current analysis it is not possible to determine which component of the gearbox is really damaged, which is possibly to do with other sensors such as vibration sensors. In order to detect the root cause of the damage with ODS, a characterisation of the kind of particles would be needed, including shape and elemental composition in addition to the particle count and size. Without these requirements, visual inspection will be needed to determine which component is exactly damaged.

The findings of this work suggest a promising future for optical oil debris sensors in the field of WT monitoring. At the same time, the need of being aware of all the details of the case study is also concluded, as there are some inconsistencies that are not explainable by the sole analysis of the SCADA and ODS data, but might be explainable if more details of the installation of the sensors and the WT itself were made available. In this regard, improving the cooperation and trust between WT owner and researchers would be a key factor for doing better analyses.

The addition of more monitored turbines to the study as well as prolonging the studied period of time would validate the results and determine whether one of the groups is just anomalous for external reasons (such as unreported differences in the systems), or there are really other turbines in which the oil debris follows the same behaviours. In this same line, it would be interesting to keep observing the differences in the HI that are proposed while visual inspections are done periodically in order to prove the validity of the approach for diagnosing the state of the turbine and also to learn to adjust the limits of admissible/nonadmissible thresholds of particle generation before having severe damages.

Furthermore, using the characterisation of the generated debris in order to related the visual inspections with the results provided by online sensor would give and additional value to the monitoring, as the root cause of failure could be identified. Also, adding vibration sensors in order to determine which source of data could provide better insight, more valuable information or identifying possibilities of synergy would be of interest.

Author Contributions: Conceptualisation, all the authors; Methodology, all the authors; Original draft preparation, K.L.d.C.; Writing—Review & Editing, C.R.-P., A.U. and K.L.d.C.; Software, K.L.d.C., C.R.-P. and S.F.; Supervision, S.F. and A.U.

Funding: This work was performed with the financial support of the FRONTIERS IV (ELKARTEK KK-2018/00096) Project financed by Eusko Jaurlaritza.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Tchakoua, P.; Wamkeue, R.; Ouhrouche, M.; Slaoui-Hasnaoui, F.; Tameghe, T.A.; Ekemb, G. Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges. *Energies* **2014**, *7*, 2595–2630. [CrossRef]
2. Hameed, Z.; Hong, Y.S.; Cho, Y.M.; Ahn, S.H.; Song, C.K. Condition monitoring and fault detection of wind turbines and related algorithms: A review. *Renew. Sustain. Energy Rev.* **2009**, *13*, 1–39. [CrossRef]
3. Nie, M.; Wang, L. Review of condition monitoring and fault diagnosis technologies for wind turbine gearbox. *Procedia CIRP* **2013**, *11*, 287–290. [CrossRef]

4. Marti-Puig, P.; Blanco, A.M.; Cárdenas, J.J.; Cusidó, J.; Solé-Casals, J. Feature selection algorithms for wind turbine failure prediction. *Energies* **2019**, *12*, 453. [CrossRef]
5. Manwell, J.F.; McGowan, J.G.; Rogers, A.L. *Wind Energy Explained: Theory, Design and Application*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2010. [CrossRef]
6. Crabtree, C.J.; Zappalá, D.; Hogg, S.I. Wind energy: UK experiences and offshore operational challenges. *Proc. Inst. Mech. Eng. Part A J. Power Energy* **2015**, *229*, 727–746. [CrossRef]
7. Echavarria, E.; Hahn, B.; van Bussel, G.J.W.; Tomiyama, T. Reliability of Wind Turbine Technology Through Time. *J. Sol. Energy Eng.* **2008**, *130*, 031005. [CrossRef]
8. Su, C.; Yang, Y.; Wang, X.; Hu, Z. Failures analysis of wind turbines: Case study of a Chinese wind farm. In Proceedings of the 2016 Prognostics and System Health Management Conference (PHM-Chengdu 2016), Chengdu, China, 19–21 October 2016. [CrossRef]
9. García Márquez, F.P.; Tobias, A.M.; Pinar Pérez, J.M.; Papaelias, M. Condition monitoring of wind turbines: Techniques and methods. *Renew. Energy* **2012**, *46*, 169–178. [CrossRef]
10. Carroll, J.; McDonald, A.; McMillan, D. Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines. *Wind Energy* **2016**. [CrossRef]
11. Nilsson, J.; Bertling, L. Maintenance management of wind power systems using condition monitoring systems—Life cycle cost analysis for two case studies. *IEEE Trans. Energy Convers.* **2007**, *22*, 223–229. [CrossRef]
12. Pfaffel, S.; Faulstich, S.; Rohrig, K. Performance and reliability of wind turbines: A review. *Energies* **2017**, *10*, 1904. [CrossRef]
13. Hahn, B.; Durstewitz, M.; Rohrig, K. Reliability of Wind Turbine–Experiences of 15 years with 1500 WTs. In *Wind Energy*; Peinke, J., Schaumann, P., Barth, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2006. [CrossRef]
14. Koltsidopoulos Papatzimos, A.; Dawood, T.; Thies, P.R. Data Insights from an Offshore Wind Turbine Gearbox Replacement. *J. Phys. Conf. Ser.* **2018**, *1104*. [CrossRef]
15. Kattelus, J.; Miettinen, J.; Lehtovaara, A. Detection of gear pitting failure progression with on-line particle monitoring. *Tribiol. Int.* **2018**, *118*, 458–464. [CrossRef]
16. Artigao, E.; Koukoura, S.; Honrubia-Escribano, A.; Carroll, J.; McDonald, A.; Gómez-Lázaro, E. Current signature and vibration analyses to diagnose an in-service wind turbine drive train. *Energies* **2018**, *11*, 960. [CrossRef]
17. Gonzalez, E.; Stephen, B.; Infield, D.; Melero, J.J. Using high-frequency SCADA data for wind turbine performance monitoring: A sensitivity study. *Renew. Energy* **2019**, *131*, 841–853. [CrossRef]
18. Elosegui, U.; Egana, I.; Ulazia, A.; Ibarra-Berastegi, G. Pitch angle misalignment correction based on benchmarking and laser scanner measurement in wind farms. *Energies* **2018**, *11*, 3357. [CrossRef]
19. Astolfi, D. A Study of the Impact of Pitch Misalignment on Wind Turbine Performance. *Machines* **2019**, *7*, 8. [CrossRef]
20. Rabanal, A.; Ulazia, A.; Ibarra-Berastegi, G.; Sáenz, J.; Elosegui, U. MIDAS: A Benchmarking Multi-Criteria Method for the Identification of Defective Anemometers in Wind Farms. *Energies* **2019**, *12*, 28. [CrossRef]
21. Tautz-Weinert, J.; Watson, S.J. Using SCADA data for wind turbine condition monitoring—A review. *IET Renew. Power Gener.* **2017**, *11*, 382–394. [CrossRef]
22. Feng, Y.; Qiu, Y.; Crabtree, C.J.; Long, H.; Tavner, P.J. Monitoring wind turbine gearboxes. *Wind Energy* **2013**, *16*, 728–740. [CrossRef]
23. Blanco, M.A.; Gibert, K.; Marti-Puig, P.; Cusidó, J.; Solé-Casals, J. Identifying health status of wind turbines by using self organizing maps and interpretation-oriented post-processing tools. *Energies* **2018**, *11*, 723. [CrossRef]
24. Escaler, X.; Mebarki, T. Full-Scale Wind Turbine Vibration Signature Analysis. *Machines* **2018**, *6*, 63. [CrossRef]
25. Dupuis, R. Application of Oil Debris Monitoring For Wind Turbine Gearbox Prognostics and Health Management. In Proceedings of the Annual Conference of the Prognostics and Health Management Society, Portland, OR, USA, 10–16 October 2010.
26. Sheng, S. Monitoring of Wind Turbine Gearbox Condition through Oil and Wear Debris Analysis: A Full-Scale Testing Perspective. *Tribol. Trans.* **2016**. [CrossRef]

27. Cuadrat Prats, J. El clima de Aragón. *Geografía Física de Aragón. Aspectos Generales y Temáticos*; Asociación de Geógrafos Español: Murcia, Spain, 2004; pp. 15–26.
28. Mabe, J.; Zubia, J.; Gorritxategi, E. Photonic low cost micro-sensor for in-line wear particle detection in flowing lube oils. *Sensors* **2017**, *17*, 586. [CrossRef] [PubMed]
29. Lopez, P.; Mabe, J.; Miró, G.; Etxeberria, L. Low cost photonic sensor for in-line oil quality monitoring: Methodological development process towards uncertainty mitigation. *Sensors* **2018**, *18*, 2015. [CrossRef] [PubMed]
30. British Standard. *BS ISO 4406:1999 Hydraulic Fluid Power—Fluids—Method for Coding the Level of Contamination by Solid Particles*; International Organization for Standardization: Geneva, Switzerland, 1999. [CrossRef]
31. Pearson, K. VII. Note on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* **1895**, *58*, 240–242.
32. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **1904**, *15*, 72–101. [CrossRef]
33. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417–441. [CrossRef]
34. Ulrich, J. TTR: Technical Trading Rules. R Package Version 0.23-4. 2018. Available online: https://cran.r-project.org/web/packages/TTR/TTR.pdf (accessed on 15 July 2019).
35. Cleveland, W.; Grosse, E.; Shyu, W. Local regression models. In *Statistical Models in S*; Chambers, J.M., Hastie, T.J., Eds.; Software Pacific Grove: Wadsworth, OH, USA, 1992; pp. 309–376.
36. Therneau, T.; Atkinson, B. rpart: Recursive Partitioning and Regression Trees. R Package Version 4.1-13. 2018. Available online: https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf (accessed on 15 July 2019).
37. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*; Wadsworth Int. Group: Wadsworth, OH, USA, 1984; Volume 37, pp. 237–251.

10 Hybrid modelling for linear actuator diagnosis in absence of faulty data records

- Authors: Lopez de Calle – Etxabe, Kerman; Ruiz – Carcel, Cristobal; Starr, Andrew; Ferreiro, Susana; Arnaiz, Aitor; Gomez, Meritxell; Sierra, Basilio;
- Publisher: Submitted to Springer
- Journal: Submitted to Journal of Intelligent Manufacturing
- Year: 2020
- Quartile (Scimago/WoS): -
- DOI: -

Hybrid modelling for linear actuator diagnosis in absence of faulty data records

Authors: Kerman López de Calle – Etxabe[12*], Cristobal Ruiz – Carcel [3], Andrew Starr[3], Susana Ferreiro[1], Aitor Arnaiz[1], Meritxell Gómez[12], Basilio Sierra[2]

1: TEKNIKER. Iñaki Goenaga street, 5 - 20600 Eibar, Gipuzkoa, Spain

2: UNIVERSITY OF THE BASQUE COUNTRY (UPV/EHU). Manuel Lardizabal Pasealekua, 1 - 20018 Donostia-San Sebastián, Gipuzkoa, Spain

3: CRANFIELD UNIVERSITY. College Rd, MK43 0AL, Cranfield, Bedford, UK

***: Corresponding author.**

Keywords: Condition based maintenance, fault diagnosis, linear actuators, hybrid modelling.

Abstract:

The advantages of Condition Based Maintenance over alternative maintenance strategies have been widely proven. The use of detection, diagnosis and prognostic algorithms allows operators to adjust schedules to optimise the cost of maintaining the machinery while avoiding breakdowns and downtimes. However, the previous techniques are not always applicable to the whole range of machines we find in industry. Particularly in cases where machines are unique or not-mass-produced, it is difficult to obtain faulty data records, and this limitation complicates the development of diagnostic monitoring algorithms.

This work proposes an approach that combines the data from physical models with data-based models (also known as hybrid modelling) to sort out the lack of faulty data records. Such approach facilitates the development of diagnosis algorithms that are given in some industrial scenarios. For doing so, faulty data records are generated in a physical model and used to augment the real non-faulty data records so that diagnostic algorithms can be trained. This procedure allows using data-based algorithms to detect real faulty data and to analyse their diagnostic potential.

Our approach is tested in a condition monitoring case study applied to linear actuators. A specific test rig was built and used to collect data from healthy and faulty cases (the later only used for validation purposes). Additionally, a physical model which simulates nominal (healthy) and faulty conditions was used to generate synthetic data. Synthetic and real measured data were combined to develop a diagnostic model and the algorithm was validated in the detection of real faulty cases. Besides the detection of certain faults, this approach has been proved to be valuable to detect also unseen operating conditions.

The results obtained in this work prove the validity of hybrid models for those cases in the industry where there are physical or economical limitations to obtain certain data records that, therefore, difficult the implementation of diagnostic algorithms.

Introduction:

The prevalence of condition-based maintenance (CBM) over preventive and corrective maintenance is a proven fact. Furthermore, good CBM strategies should be based on the analysis and predictions of historical and real-time data, and should be available of updating recommendations as soon as new data exists (Bousdekis et al., 2018). However, there are some cases where certain data types (faulty records) might not be available in the historical data. Consequently, the development of diagnostic algorithms is obstructed, as it is necessary to possess records obtained during faulty condition operations to develop them. At the same time, having faulty machines is, obviously, undesired and avoided by the industry. And only when many machines are operating, there might be the chance to obtain some fault related data. This scenario of lack of faulty data is particularly common in non-mass-produced machines. Therefore, designing diagnostic algorithms under these circumstances is complex.

The alternative approach here proposed consists in the combination of physics-based and data-based models. This method also known as hybrid modeling, has been previously suggested by other works in the literature and tries to benefit from the advantages of both modelling techniques precision and applicability (Medjaher and Zerhouni, 2013; Mishra et al., 2015). Some of the reasons behind the use of hybrid models are: the difficulty of obtaining completely accurate physical models (Matei et al., 2015); the cost of running parallel models (Balaban et al., 2015; Narasimhan et al., 2010); the complexity and lack of understanding of all the physics governing the system (An et al., 2013; Benkedjouh et al., 2015); improved pattern recognition of data-driven models (Balaban et al., 2015); and not having enough data records (mostly related to failures) (Mishra et al., 2015) or lack of fault growth models (Balaban et al., 2015).

Regarding the field of application, hybridization has been utilized in a range of different applications: rotating machinery (An et al., 2013; Leturiondo et al., 2017; Li et al., 2019; Qian et al., 2017), railway switches (Matei et al., 2015), battery life prediction (Liao and Köttig, 2014), mechatronic systems (Medjaher and Zerhouni, 2013) and, also, in electro-mechanical actuators (Balaban et al., 2015; Narasimhan et al., 2010).

Particularly, this work focuses on the monitoring of linear actuators, a widely extended mechanism for the generation of linear motion. In the past, this linear motion profiles were produced with cam-follower and crank-rod mechanisms. Later on, these techniques evolved to hydraulic and pneumatic actuation systems, which have been extensively used for applications requiring motion control or high forces. In recent times, linear actuators are found in a variety of systems, such as valves, door openers, aircraft systems, machine tools, robot arms, etc. Furthermore, electro-mechanical actuators (EMA) in particular are gaining popularity in aeronautics, as they have some advantages over the typically used hydraulic actuators, such as increased safety and reliability, easier and reduced maintenance, reduced weight, volume and complexity of transmission paths, and a higher efficiency (Qiao et al., 2018). However, this technology is not mature yet, and needs further research.

Currently, there are two main trends in the latest research regarding linear actuators.

First, some authors have focused on the design and development of fault tolerant actuators, a.k.a. high redundancy actuators. The underlaying idea behind fault tolerant actuators is designing actuators that are composed by several small actuation elements coupled, so that they will be able to work even under faulty conditions, this duplicity in actuators gives the name high redundancy actuators (HRA). HRA was inspired by human musculature. They mimic human muscles, that are composed by many individual muscle cells each of them contributing to the force and travel of the muscle (Davies et al., 2008). Current research focuses on the different types of architectures these actuators have, and in the effect these architectures have in the resilience of the actuator. For example, in the work by (Antong et al., 2014) a 3 serial element in 4 parallel structure is modelled, later the model is validated with a real rig in (Antong et al., 2016) considering open-loop and close-loop controls. Other works such as the one by (Manohar et al., 2018) model a 3×3 series-in-parallel architecture HRA and seed faults, they study the degradation and the possibility to keep actuating even under 3 faulty actuators.

In relation to the latest works in the field of condition monitoring of linear actuators, recent research focuses on the detection and diagnosis of faults and the remaining useful life estimation of the actuators. For that purpose monitoring systems are typically based on position error and current (Ruiz-Carcel and Starr, 2015, 2018a). Also, in lesser extent vibration (Balaban et al., 2015; Sudhawiyangkul and Isarakorn, 2017) when rotating elements are involved (balls crew/rack and pinion), temperature (Balaban et al., 2015) and also acoustic emissions could be used as Ehrman collates (Ehrmann et al., 2016). The typical faults that are considered are: lack of lubrication, spalling, backlash (Ruiz-Carcel and Starr, 2018a), external forces (jamming)(Susana Ferreiro et al., 2013), including some others more related to the current supply of EMAs, such as open circuit, close circuit (Cai et al., 2017); according to Narasimhan (Narasimhan et al., 2010) and Balaban (Balaban et al., 2015) in addition to the previous faults, winding shorts, and common sensor faults like bias, drift and scaling should be considered. Linear actuator monitoring systems have also considered variety of operating conditions, which includes, various loads and motion profiles as sinusoidal, trapezoidal (Ruiz-Carcel and Starr, 2018a), sine sweep and triangular (Narasimhan et al., 2010). It is also common to find some pre-treatment of the signals before plugin them into the diagnosis or prognosis models. In that sense, they can be as simple as statistical descriptors taken from time domain (Ruiz-Carcel and Starr, 2018a) going through the use of fast fourier transform (FFT) for vibrations (Sudhawiyangkul and Isarakorn, 2017) or currents (Cai et al., 2017) for frequency domain going to the more complex time-frequency domain techniques such as Discrete Wavelet Transform (DWT), Wigner-Ville-Distribution (WVD) or the Mel-Frequency-Cepstral-Coefficient (MFCC) that Knöbel employs (Knöbel et al., 2015). And, for the reduction of the dimensionality of the data, principal component analysis (PCA) must be noted, for its use is widely extended as the amount of works using it suggests (Cai et al., 2017; Knöbel et al., 2015; Mazzoleni et al., 2019; Ruiz-Carcel and Starr, 2018a). Regarding the final diagnosis models, the works focusing in fault diagnosis have used different algorithms for that purpose, for example, (Cai et al., 2017) employs Bayesian networks, (Knöbel et al., 2015) uses support vector machines and Neural Networks are used by (Balaban et al., 2015).

Actuator monitoring systems have also benefited from the use of physical models. In that regard, some different trends haven been identified. The works by (Balaban et al., 2015) or (Kemp and Martin, 2018) run physical models and compare the model data to the one generated in reality, furthermore, (Kemp and Martin, 2018) runs additional faulty models to detect which fault shows smaller residuals and is therefore the one the happening in the rig. Ruiz-Carcel (Ruiz-Carcel and Starr, 2015) uses the model to validate a statistical process control (SPC) based monitoring technique that is later validated in a real rig in (Ruiz-Carcel and Starr, 2018a). Similarly, (Susana Ferreiro et al., 2013) and (Cai et al., 2017) and (Knöbel et al., 2015) use data from the physical model to train machine learning diagnostic algorithms. Ferreiro uses a data set fully generated with physical model data from the design stage of the actuator, Knöbel uses nominal data from the rig and enriches the dataset with faults from the physical model and, lastly, Cai combines the data generated in the rig together with the one from the physical model. Note that the diagnostic models used in the previous works are trained with fully synthetic (generated in the physical model) or data combined from the model and the real use case. In contrast, the scenario here presented is one in which obtaining real faulty data is not possible.

In this scenario, the work here presented continues with the research initiated in (Ruiz-Carcel and Starr, 2015) where a data-based anomaly detection algorithm was developed for an EMA and tested in a physical model. Later, in (Ruiz-Carcel and Starr, 2018a), this approach was improved and validated in a real actuator. This second work showed the capability of detecting failures and to give additional indications of which features were more distant from normality. The research work presented in this paper deals with the automated diagnosis of the faults and copes with a scenario that is quite frequent in the industry: Not having faulty records. The final aim of the work is to prove that, without having real faulty records, it is possible to learn to distinguish and identify failures without seeing them in the real life in advance. That is, we attempt to anticipate to real failures by using synthetic faulty data produced by the physical model. For that reason, real faulty data is only used for validation purposes along the work.

Methodology:

Essentially, to enrich the real data with unseen faults, it is necessary to have a reliable physical model that, besides contemplating the same operating conditions as the real use case, must represent also faulty conditions. For doing so the schema displayed in the following Figure 1 has been followed:

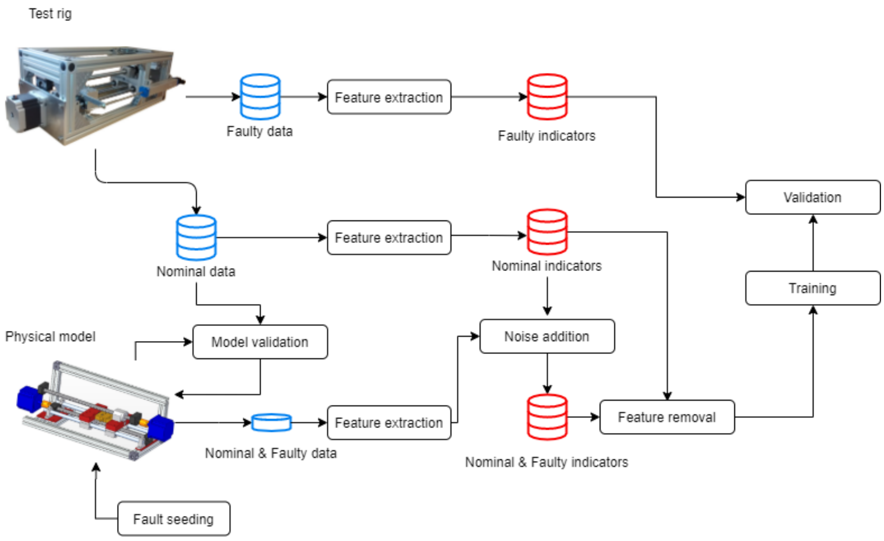

Figure 1: Schema of data pipeline.

The test rig is used to generate data in various operating conditions. Furthermore, some mechanical faults can be seeded in the test rig. In parallel, a first approximation of the physical model is carried out, and, once the real nominal data is available, it is possible to adjust the model to behave as the real use case. Then, the expected faults are seeded in the physical model. Later, features are extracted from both real nominal and synthetic data and, using the noise originally found in nominal data, more synthetic data samples are created and the features that do not resemble similarity between real and synthetic data are removed. Finally, the synthetic data augmentation can be validated by training a Machine Learning algorithm that is tested afterwards on the detection of real faults.

The following subsections describe to process in more detail:

1. Test rig:

The main element of the test rig consists of a ball-screw mechanism with a threaded shaft containing a helical raceway for the displacement of the bearing balls housed inside the nut. Varying loads are generated by attaching a secondary actuator. The actuators are connected through a load cell, so that the cell provides feed back to the controller of the secondary actuator. With this controller, different operating conditions can be represented by changing the load setpoint. The commanded load setpoints for the second actuator are: 20 Kgf, 40 Kgf and -40 Kgf.

The signal provided by a linear transducer is used by the position controller to command the main actuator to produce two motion profiles: trapezoidal (constant speed) and sinusoidal (smooth acceleration and deceleration). However, only trapezoidal profiles are considered in this work. These trapezoidal profiles are defined to complete a 120mm extension in 5 seconds, stop for 3 seconds and retract in another 5 seconds. Motor current in this actuator is also measured for monitoring purposes.

In addition, 3 types of faults are seeded with increasing severities: the bolt in the holding the seal of the is tightened to increase friction simulating lack of lubrication; surfaces of the screw are damaged starting with 1 mm diameter defects and gradually increasing up to 4 mm diameter defects including the removal of the sidewall of neighboring channels; lastly, backlash defect is obtained by substituting the balls (originally 3.15 mm of diameter) by smaller ones (3 mm and 2.5 mm diameter).

Further details regarding the test rig and the seeded faults can be found in (Ruiz-Carcel and Starr, 2018a). Each test in the final dataset contains 5 motion cycles, and each test is repeated 10 times. The full data set contains tests of both motion profiles (sinusoidal and trapezoidal), the varying load conditions (20 Kgf, 40 Kgf and -40 Kgf), and the normal and faulty conditions. The whole dataset is accessible at (Ruiz-Carcel and Starr, 2018b).

2. Physical model:

The model of the actuator follows the model developed in (Ruiz-Carcel and Starr, 2018a) and was developed in Matlab using Simscape toolbox. The model contains an electrical motor which is controlled by a PID using the position measurement and desired setpoint, a rack and pinion block that transfers the rotational motion to linear one and, for simplicity sake, an equivalent force in substitution of the secondary actuator. In addition, the model also includes an equivalent rotational inertia, a mass, and linear friction block. The physical model allows varying the intensity of the force to simulate the same operating conditions that are tested in the test rig.

2.1. Model validation

For the combination of data from the physical model and the rig it is necessary that they coincide. Therefore, the physical model must be validated and adjusted if necessary. The Matlab Optimization toolbox was used for the adjustment of the physical model using nominal displacement and current measurements from the test rig for the validation and adjustment.

The final signals from the test rig and the signals from the physical model are displayed in the following Figure 2:

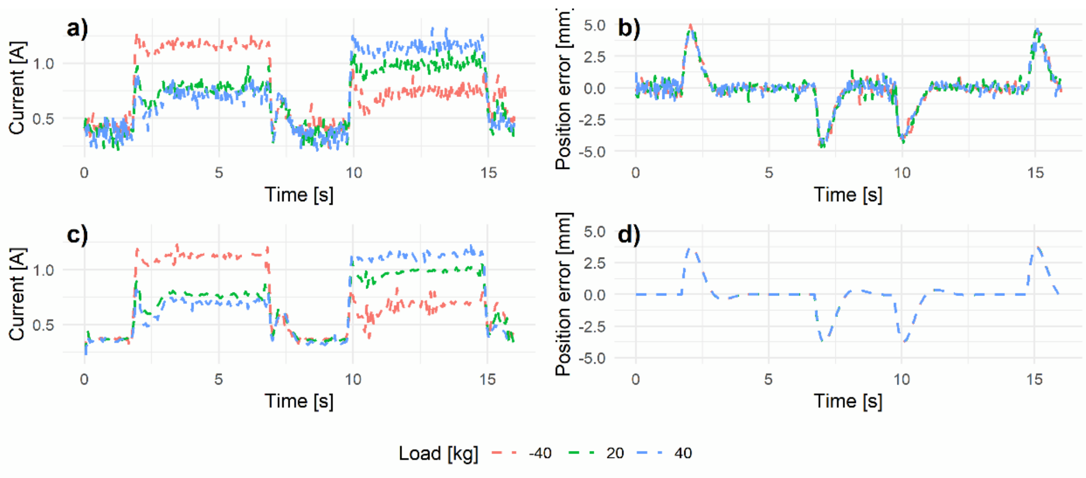

Figure 2: Real and synthetic signals. a) Real current b) Real position error c) Synthetic current d) Synthetic position error

2.2. Fault seeding

According to our approach, for the purpose of foreseeing faults it is necessary to seed them first in the physical model. This work focuses on the diagnostic of lack of lubrication and spalling faults. Lack of lubrication is introduced in the physical model by increasing the translational friction coefficient during the whole translation of the actuator, which is done with increasing friction values to study gradually increasing severities. Regarding the spalling fault, it is simulated by a severe increase of friction in the particular point of the motion path where the fault is located on the screw. This test is also done with varied friction values corresponding to more severe spalls.

3. Feature extraction:

Raw signals contain information that can be exploited to develop diagnostic algorithms. However, redundant data is removed and key indicators are typically extracted from the raw signal. This process is known as feature extraction. The feature extraction schema adopted in this work follows the one in (Ruiz-Carcel and Starr, 2018a). However, due to the added difficulty of manipulating two distinct data sources (synthetic and real) only the simplest features are kept reducing the differences between data sources.

Setpoint, position, position error and current are the signals available at the test rig of this work. From that set of signals, the ones used for the monitoring are the position error and the current.

The feature extraction process for the signals follows these steps:

1) Segment each cycle using the trapezoidal motion into: extension, retraction, extended and idling segments.
2) Detect tail of the extension/retraction segments (last two thirds of the segment) and get descriptors (maximum, minimum peak to peak and mean values).
3) Detect the head of the extension/retraction segments (first third of extension/retraction segments) and obtain overshoot (absolute maximum value).

The following Figure 3 displays a visual explanation of the extracted points.

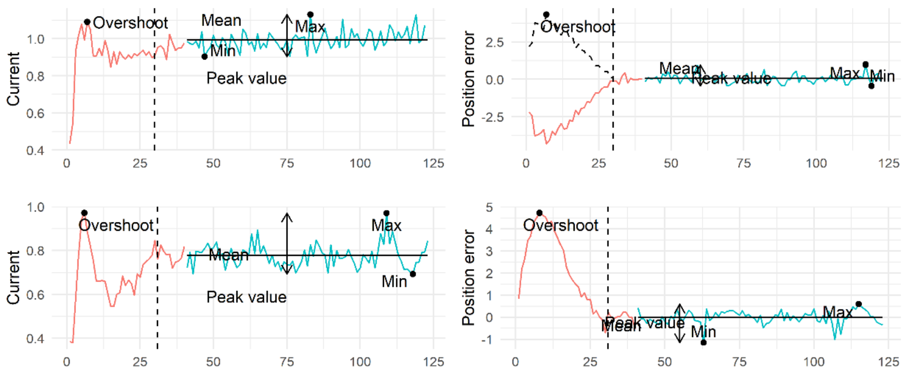

Figure 3: Descriptors taken from current and position error. Top: Retraction. Bottom: Extension.

As consequence of the feature extraction process, signals are synthetized into indicators or descriptors.

4. Augmentation by noise addition:

The data produced by the physical model (the synthetic data) is limited to a single case for each load and fault condition and it is not subjected to any process variability. That is, repeating the simulation will not alter the resulting signals. However, the development of machine learning (ML) algorithms requires of bigger data samples for robustness sake. Additionally, as the final goal is to diagnose data originated in the real use case, it is necessary to introduce the variability of the process in the synthetic data.

For the calculation of new observations, first, the standard deviation value of the nominal cases of the test rig per variable is computed, after, random samples are taken from a uniform distribution that is centered in the single synthetic values and has the same variability as the real data.

5. Feature removal:

Physical modelling tends to be based in some simplifications. As side-effect of the simplifications, the data generated by the model is not exactly equal to the one generated in the test rig. After feature extraction, it is necessary to reduce the differences between the data that is generated by the model that has been augmented with the noise addition (synthetic data) from the data that belongs to the test rig (real data). Otherwise, it would not be possible to train with data from one source to test on data from the other. At the same time, it is necessary to keep the input features with good enough diagnostic capability, or it will not be possible to diagnose any failure.

The procedure utilized to reduce the number of features from the dataset is as follows:

1. Create two datasets:

a. Nominal: Contains healthy data from both sources (Real/Synthetic).
b. Damaged: Only includes data from the physical model (Synthetic data), includes nominal cases and faults.

2. For each dataset:
    a. Split in train/test taking datasets randomly containing 66% and 33% of the observations respectively.
    b. Scale and center train/test datasets using mean and standard deviation from the train set.

3. Use LDA algorithm over Nominal dataset in a binary classification problem (Real/Synthetic) compute accuracy, kappa and store ranking of variables that separate both classes the most.

4. Use the LDA for classification of faults in damaged dataset (Spalling/Lubrication/Nominal), store accuracy and kappa.

5. Remove the most valuable feature (from step 3).

6. Iterate steps 3 to 5 until only two features are left.

The aim of this procedure is to determine which features can be removed to make the data from the different sources more alike without losing too much diagnostic capability. For this purpose, in each iteration the performance metrics and remaining features are kept for studying later the evolution of both metrics in relation to the remaining features.

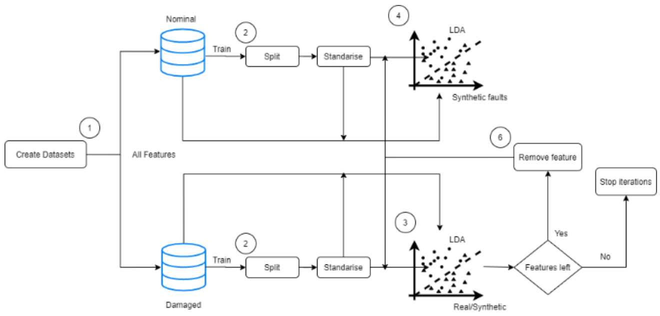

Figure 4: Feature removal process schema. Numbers correspond to the steps in the description.

Note that faulty data from the real dataset is not used, as the aim is to reproduce a scenario where this type of data is not available, which is common in the industry.

6. Fault diagnosis:

The bulk of the work consists in the use of a classification algorithm to diagnose some real faults that the algorithm has not seeing before. In addition, some pre-processing techniques have been analyzed, to determine which combination of pre-processing could be better suited to this type of problems.

6.1. Pre-processing:

The performance of statistical algorithms can be typically improved by some data pre-processing. Typically, standardization (dividing by the standard deviation after mean subtraction) is used to improve comparison of data of different magnitudes. In addition, alternative pre-processing techniques can also be used with the aim of improving algorithm performance.

6.1.1. Principal Component Analysis (PCA):

Principal component analysis is a dimensionality reduction technique that tries to find linear combinations (or Principal Components) of the variables that maximize the variance while they are uncorrelated between each other. PCA is used as pre-processing algorithm to reduce the amount of predictor variables used by the classification algorithm without causing major losses of information. For that purpose, a reduced set of principal components that contains as much variance as possible is used instead of the original features.

6.2. Classification algorithms:

Classification algorithms are used when a set of features can be used as predictors of a target class, which can be binary (taking to values) or multiclass (having more than two classes). Classification algorithms try to find the relation between the features and the target class value in order to predict the class value when only features are available.

6.2.1. Linear Discriminant Analysis (LDA)

LDA finds the directions that maximize the separation between classes to use these directions later in the prediction of the class of the individuals. The directions, or linear discriminants, are linear combinations of predictor variables.

When predictors used by the LDA algorithm are standardized, the discriminant weights reflect the importance of each variable in class separability according to the algorithm.

6.3. Metrics for fault diagnosis:

An interesting aspect of fault classification or fault diagnosis is the metrics used to evaluate the performance of the algorithms. The metrics used in this work to evaluate the performance are introduced below:

6.3.1. Accuracy:

Accuracy measures the number of hits of an algorithm. That is, the percentage of times the algorithm correctly predicts the target class out of the total number of cases.

6.3.2. Class wise accuracy:

Class wise accuracy measures the accuracy considering each class independently. Using class wise accuracy is particularly interesting with imbalanced data (big differences in the number of observations per class).

6.3.3. Cohens Kappa:

In this application, Cohens Kappa is used to measure the degree of concordance between the real and the predicted class values. Besides considering the agreement among raters (as accuracy), it also considers the probabilities of hypothetical probability of chance agreement, which is calculated from the observed data. It takes values from 1 to -1 being 1 complete agreement, 0 non greater than chance agreement and -1, worse than random agreement.

7. Good practices in scientific computing

One aspect considered during the development of this work is the reproducibility of the experiments. Awareness has risen in the scientific community regarding the serious problems in the reproduction of scientific works. Consequently, some authors have developed certain workflows that allow to better audit scientific works as well as to reduce the possibilities of making mistakes and/or the chance to not finding them (Wilson et al., 2017). Following these recommendations, the code used for the processing of the data and the analysis has been uploaded to an open repository (Lopez de Calle – Etxabe et al., 2020). As well as the raw data, that was made accessible for previous publications (Ruiz-Carcel and Starr, 2018b). The code is written in R language, and we recommend using Rstudio IDE for its reproduction. Further details for reproducing the results of this work can be found in the code repository (Lopez de Calle – Etxabe et al., 2020).

Results:

Once the physical model was validated with nominal data from the rig (real nominal), it was possible to generate synthetic data representing faulty states of the actuator. Then, the features were extracted from the signals and, to increase the size of synthetic data, the dataset was enriched by adding more samples. For that purpose, the augmentation by noise addition explained in the previous section Methodology was used. That way, a dataset with 2850 observations and 20 features (see Table 2 below) per observation was created. Each observation had a load condition, different health status (nominal, lubrication fault or spalling fault) and different origin (Synthetic or Real) as summarized in the Table 1. Note that each type of fault (including nominal) had different severities and 50 observations were generated for each load case and severity, consequently the number of observations for lubrication and spalling fault was greater than the one of nominal observations.

Table 1: Number of observations per data source, load and fault type.

Real		Fault		Synthetic		Fault	
Load	Lubrication	Spalling	Nominal	Load	Lubrication	Spalling	Nominal
-40	100	400	50	-40	200	150	50
20	100	400	50	20	200	150	50
40	100	400	50	40	200	150	50

In sight of the differences among the features obtained from each data source (real/synthetic), feature removal process was applied. During the process, the modelling accuracies and Kappa metrics were stored while the set of input features was reduced by one feature (the one separating data source the most) in each iteration. The evolution of the accuracy and the kappa in the diagnosis and in the separability of real and synthetic data for each iteration is displayed in the following Figure 5.

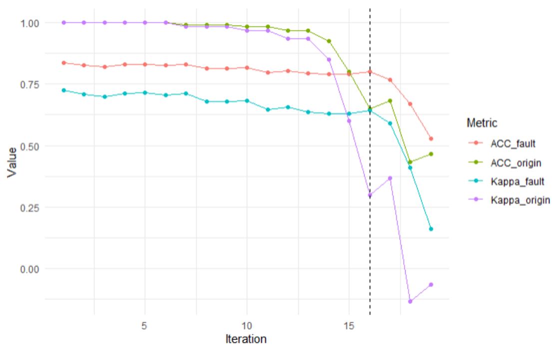

Figure 5: Feature removal Kappa and Accuracies per iteration.

According to the evolution, it was decided to keep the feature set of the 16th iteration (represented by a vertical line in Figure 5) because the feature set had already lost the separation capability of the data sources while the diagnostic capability was still high. As a consequence of the feature removal, the dimensionality of the dataset was reduced to the features shown in the following Table 2:

Table 2: Extracted features per signal and motion type. Grey shaded features represent the ones left after the feature removal process.

Current		Position error	
Extension	Retraction	Extension	Retraction
Mean	Mean	Mean	Mean
Max	Max	Max	Max
Min	Min	Min	Min
Peak	Peak	Peak	Peak
Overshoot	Overshoot	Overshoot	Overshoot

Diagnosis under different contexts

Initially, diagnosis was tested by taking together data from all load conditions and comparing different pre-treatments. That way the effect of dimensionality reduction techniques such as PCA, the use of different initial feature sets (with/without feature removal process), the augmentation of observations by repetition in order to balance classes, and the use of a majority vote technique with sets of cycles to give a single diagnostic (adding chronological information) were tested.

The diagnosis consisted in the creation of a training dataset that was fed to the LDA model and a testing dataset where the predictions of the model were compared to the real labels. For the validation of the tests 75% of the training data was randomly chosen per test and it was validated in another random sample containing 75% of the test set. It must be kept in mind that the training and testing datasets did not contain the same data, the cases were pre-filtered in purpose to test our hypothesis, that is, training data contained faults from synthetic data and nominal data from both the rig and the physical model while testing dataset contained only the data from the test rig (faulty and nominal).

The test name represented the features that were used as input as well as the additional treatments that data had. The tests named "All Features" had all the features that were extracted initially. The rest of the tests, named "Final Features" contained the 5 features chosen in the feature removal process. The tests containing severe faults took only the most severe cases for each type (lubrication and spalling) and nominal data. Regarding the cases that had a PCA pre-processing stage, the following table displays the number of Principal Components that were retained in each case and the amount of cumulative variance they retained:

Table 3: Number of retained principal components and cumulative variance for the tests with PCA pre-processing.

Test	Retained PCs	Cumulative variance
Final Features PCA	3	86.67 %
All Features PCA	10	89.43 %

In the test that had data augmentation, the training samples were repeated to have approximately the same number of samples per class. In that way, the initial cases (450 observation of lubrication fault, 338 of spalling fault and 225 of nominal) turned to be 900 observations of lubrication fault, 1014 of spalling fault and 900 nominal cases.

For robustness sake each test was repeated 10 times, each time the data was randomly taken from the pre-filtered cases. The results are showed in the following Figure 6 by means of boxplots, that depict the results obtained in diagnosis with all the loads. For a better understanding and interpretation of the results, not only Accuracy metric was considered, the Kappa coefficient was also obtained in each test.

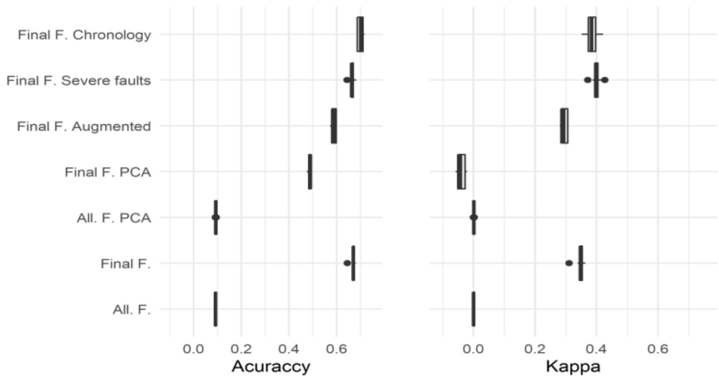

Figure 6: Diagnosis performance of different data pre-treatments modelling observations of all load cases together.

For a deeper insight of the accuracy regarding the different diagnosis of classes (Nominal/Lubrication fault/Spalling fault), the class wise accuracy was measured for the tests using the final set of features the results are presented in the following Figure 7.

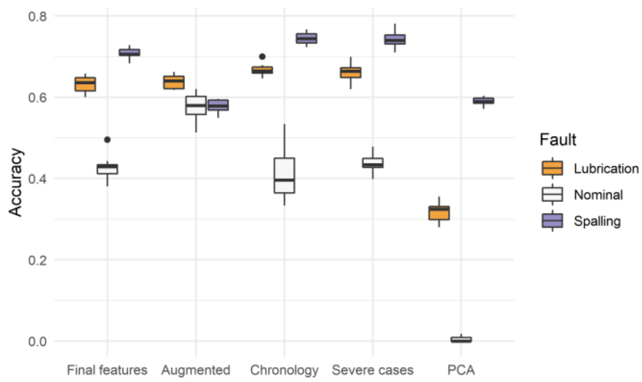

Figure 7: Class wise accuracy of test using the final set of features.

Finally, after comparing the models that included the whole range of loads some of the tests were repeated by modelling data belonging to each load separately to analyse whether diagnostic could be improved by modelling each load case with an ad hoc model. These results are presented in the following Figure 8:

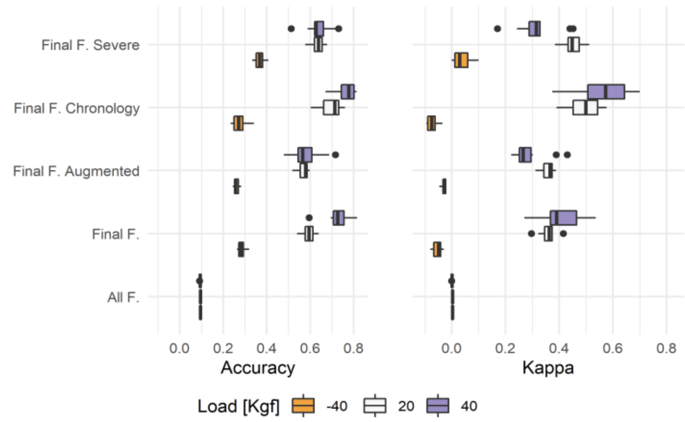

Figure 8: Different performance metrics for the diagnosis using a model per load case.

Detecting operational context

In sight of the interest in the identification of the operational load due to the considerable performance variation in the diagnosis, the capability to detect the operational context was tested. For that purpose, four different tests were carried out, in which the LDA algorithm was trained and the predictions were analysed. The first two tests contained data from all the synthetic cases (3 loads, 2 fault types and nominal) together with nominal data from the real dataset under 3 different loads. For the third and fourth tests only data from synthetic cases was used. The difference between the 1st and the 2nd tests and the 3rd and the 4th test was that the 2nd and the 4th had an extra dimensionality reduction step carried out by applying a PCA analysis and keeping only the 4 most important principal dimensions. In both cases (tests 3 and 4), the 4 Principal Components accounted for approximately 95 % of the variability. The following figure shows the performance in load detection:

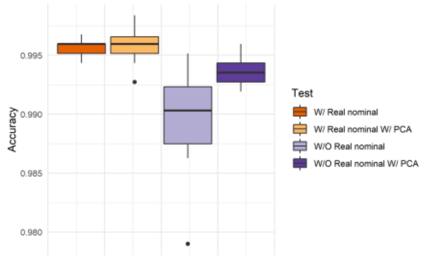

Figure 9: Context detection with and without PCA pre-treatment and with and without using real data.

Best-case scenario

Lastly, in order to have a baseline for comparison, the best cases were compared to the hypothetical of having the real faulty data representing the best-case scenario, in which faulty records are available and therefore, there is no limitation nor need of using a physical model to generate synthetic data of faults.

Due to its low performance, Load condition -40 was left aside during this comparison, consequently, load conditions 40 and 20 were combined. As in previous tests, synthetic data from faulty and nominal cases was enriched with nominal real data, in this test, however, it is compared to a model trained over data fully originated in the test rig. The set of features was reduced to the final features presented above and the effect of chronology was compared against individual prediction. Additionally, considering the particularity of the dataset that, only Kappa statistic was measured, and simple Kappa baselines were stablished for both multiclass (considering nominal, spalling fault and lubrication fault) and binary (spalling fault and lubrication fault) cases. For the multiclass case, the baseline prediction considered that every observation belonged to Spalling fault (class with greatest number of observations), for binary, several lubrication fault misclassification rates were included.

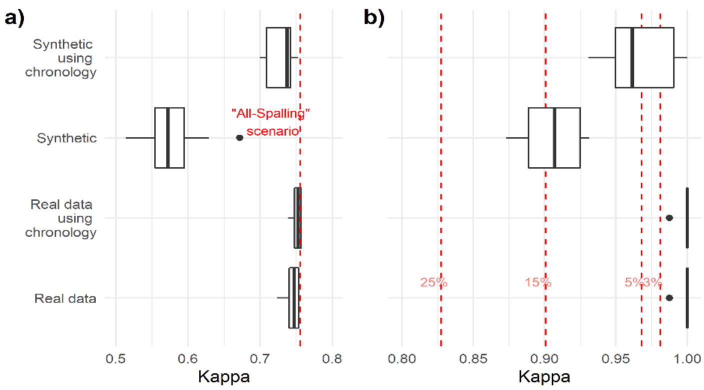

Figure 10: Comparison of results against the best-case scenario (having real data) vertical red lines represent additional baseline scenarios. a) Multiclass problem (nominal/spalling/lubrication). b) Binary class problem (lubrication/spalling).

Discussion:

Difficulties have arisen during the development of the tests, such as the difficulty to obtain a physical model with enough accuracy to reproduce faulty data with realistic current and position error signals. Although this has been partially accomplished (see the small differences in fig 2), the method proposed for dimensionality reduction has reduced the mismatch between the observations produced in the model and the real ones, which is proven by the huge performance disagreement between the diagnosis results between “All F.” and “Final F.” that Figure 6 shows. Furthermore, when PCA is used naively over all features, the results is still much worse than our proposed method (see “ALL F. PCA” in the same figure).

Regarding the ability to diagnose faults, besides the already mentioned prevalence of the reduced set of features the first diagnosis iteration (Figure 6, Figure 7 and Figure 8) reveals some interesting findings. On the one hand, using chronology is the best strategy in comparison to the others (using severe cases or using augmentation), in any case, none of the previous have an extreme contrast in comparison to directly using the final set of features (using augmentation is even worse than just using the features). On the other, same major concerns arise: firstly, when modelling each load separately load -40 generates catastrophic results (which is possible reducing the overall accuracy of the previous models) in comparison to the rest of loads (Figure 8); and, there are quite big differences from accuracy to kappa metrics, leading to think there is a class which is not being detected properly, which is obvious when looking to the Figure 7 as the class wise accuracy of nominal cases is close to null. This lack of sensitivity with nominal label (which is already reflected by the contrast between accuracy and kappa metrics) might be caused by the combination of the following factors:

- Severity: Initial degradation stages of the spalling fault do not have much impact nor in the operation nor in the signal, therefore, distinguishing them from nominal cases is non-trivial (or even not possible).
- Intermittence: Spalling fault does not manifest in all the cycles. Sometimes, even if the spall is in its biggest size, the balls managed to run smoothly through the bolt, which leads to an incorrect labelling of spalling fault.
- Imbalance: As there is a fixed number of observations per severity case, the number of observations of the dataset related to spalling fault is greater than the one of nominal cases (single severity case considered, see Table 1). This class imbalance combined to the previous factors greatly affects the generalization capability of the algorithms. Note that in figure Figure 7 the only test with balanced class wise accuracies is the augmented case which improves the nominal case detection at the expense of reducing overall accuracy.

This difficulty to discern nominal cases and spalling faults is present in even more favorable scenarios as figure Figure 10 suggest. Regardless of no considering load case -40, the overall diagnosis of multiclass case (Figure 10 a)) is lower than the simple baseline of considering everything as spalling, due to the aforementioned reasons. Nevertheless, the binary classification problem (Figure 10 b)) shows far better results. Firstly, it demonstrates that the use of chronology improves kappa, as in both cases baseline and our approach, results using chronology have higher accuracies. Secondly, it proves that the faults can be discerned under no real data conditions with a hybrid model with little chance for error. As the baselines suggest only about 5% of the lubrication faults would be mismatched as spalling, possibly corresponding to cases with very mild lubrication fault.

However, for making the previous scenario possible, it is necessary to detect the loads under which the algorithm is working, as some load conditions offer very low accuracies (L -40). This fact is probably related to the physical model being better adapted for some load cases than others. In any case, detecting the operating context is not very challenging as the Figure 9 shows that extremely high accuracies are reached in load detection even when no real data is used. This has very interesting implications, as it shows that a physical model can reproduce unseen scenarios that can be later learned by machine learning models.

Conclusions:

Hybrid models and digital twins have recently gained attention in the literature. This work demonstrates the value of having a physical model, twin of the monitored asset, for the improvement of the diagnosis and, also, the detection of new operation contexts. In other words, this work shows how data limitations (one of the major drawbacks of data-based models) can be partially overcome with the data created from the physical model.

This study has tried to gain insight in the expansion of the boundaries diagnostic algorithms have when dealing with scenarios with no faulty records. For that purpose, data from a test rig has been used to represent a real use case and a physical model has been used to create data of faulty and nominal/healthy cases. A real scenario with no faulty data has been simulated by using only synthetic data to train the diagnostic algorithms, later, the performance of our approach has been evaluated by testing the predictions with the real faulty data.

According to our results, using physical models to augment the input data pool for diagnostic algorithms seems a viable way of providing a good starting point for diagnostic algorithms. The results we have obtained show a strong capability to differentiate among faults, with a worse performance in the distinction among nominal and spalling faults. However, we have proved that, even if the real data was available for the modelling, the same problem would have occurred due to the intermittent behavior, the imbalance among classes and the soft failure severities. Hence, faults are not trivial to detect even with the real data.

Additionally, besides the use in fault detection, the possibility to detect different context in an extremely accurate way is proven, even if they were previously unseen. This implies that algorithms' horizons can be expanded to develop normality models in operation regions previously unobserved or that diagnostic algorithms that only make predictions in the regions they perform optimally can be developed.

As suggested by (Cai et al., 2017), the obvious differences among real and synthetic data have led to difficulties during the integration of the data. The model was validated with real nominal data, yet the slight differences in the nature of the signals has caused the data to not match exactly. As countermeasure, a feature removal method has been proposed. The method is based in the recursive elimination of features by monitoring the diagnostic capability and their capability to be used to detect the origin of the data source and finding a point with an acceptable trade-off. Our results show that the method has made a significant difference in the final diagnosis, and we believe this method is generic enough to be transferred to other works that try to combine physical model data and real data. Furthermore, our results show significant improvement in comparison to the PCA, which is the preferred tool according to our review.

Also, the added value of domain knowledge has been manifested. The better understanding of the problem (faults should no fix by themselves in subsequent cycles) has allowed to make use of chronological information by adding a majority voting layer to the diagnosis algorithm, which has shown the best accuracy compared to the rest of diagnosis algorithms.

Lastly, for the sake of reproducibility and scientific rigor, we have given open access to the procedure and results we have obtained by publishing the code used for the pre-processing and the analysis, as well as to the data sets used. We have the strong belief that this practice will ease the reproduction of our results and gives other researchers the opportunity reuse part of our work or to improve it.

In respect to the loose ends left in the research, two interesting research fields have been identified: the poor results in the distinction of healthy and spalling faults suggest the need of methods to improve the detection of intermittent faults and deal with imbalance; and, secondly, the possibility to study how the diagnosis ability could be improved as soon as faulty data points appear. We believe the former could be addressed by using alternative policies to majority voting, policies that could explore different ways of combing diagnosis predictions to increase the detection of those intermittent kind of faults. Regarding the latter, a possible direction could be the use of online machine learning algorithms, a branch of machine learning that deals with algorithms that are updated as soon as new data becomes available

Bibliography:

An, D., Choi, J.H., Kim, N.H., 2013. Options for Prognostics Methods: A review of data-driven and physics-based prognostics, in: 54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. p. 1940.

Antong, H., Dixon, R., Ward, C., 2014. Modelling and building of experimental rig for high redundancy actuator, in: 2014 UKACC International Conference on Control (CONTROL). pp. 384–388.

Antong, H., Dixon, R., Ward, C.P., Ward, P., Antong, H., Dixon, R., Christopher, P., Ward, P., 2016. ScienceDirect Elements : Closed-loop Elements : Closed-loop Elements : Open- and Closed-loop Model. IFAC-PapersOnLine 49, 254–259. https://doi.org/10.1016/j.ifacol.2016.10.563

Balaban, E., Saxena, A., Narasimhan, S., Roychoudhury, I., Koopmans, M., Ott, C., Goebel, K., 2015. Prognostic health-management system development for electromechanical actuators. Journal of Aerospace Information Systems 12, 329–344.

Benkedjouh, T., Medjaher, K., Zerhouni, N., Rechak, S., 2015. Health assessment and life prediction of cutting tools based on support vector regression. Journal of Intelligent Manufacturing 26, 213–223.

Bousdekis, A., Magoutas, B., Apostolou, D., Mentzas, G., 2018. Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance. Journal of Intelligent Manufacturing 29, 1303–1316.

Cai, B., Zhao, Y., Liu, H., Xie, M., 2017. A Data-Driven Fault Diagnosis Methodology in Three-Phase Inverters for PMSM Drive Systems. IEEE Transactions on Power Electronics 32, 5590–5600. https://doi.org/10.1109/TPEL.2016.2608842

Davies, J., Steffen, T., Dixon, R., Goodall, R.M., Zolotas, A.C., Pearson, J., 2008. Modelling of high redundancy actuation utilising multiple moving coil actuators. IFAC Proceedings Volumes 41, 3228–3233.

Ehrmann, C., Isabey, P., Fleischer, J., 2016. Condition monitoring of rack and pinion drive systems: necessity and challenges in production environments. Procedia CIRP 40, 197–201.

Kemp, M., Martin, E.J., 2018. Fault isolation of an electro-mechanical linear actuator, in: Proceedings of Prognostics and Health Management Society Conference, 1. Presented at the Annual Conference of the PHM Society. https://doi.org/10.36001/phmconf.2018.v10i1.539

Knöbel, C., Marsil, Z., Rekla, M., Reuter, J., Gühmann, C., 2015. Fault detection in linear electromagnetic actuators using time and time-frequency-domain features based on current and voltage measurements, in: 2015 20th International Conference on Methods and Models in Automation and Robotics (MMAR). pp. 547–552.

Leturiondo, U., Salgado, O., Galar, D., 2017. Validation of a physics-based model of a rotating machine for synthetic data generation in hybrid diagnosis. Structural Health Monitoring 16, 458–470. https://doi.org/10.1177/1475921716676053

Li, Z., Wu, D., Hu, C., Terpenny, J., 2019. An ensemble learning-based prognostic approach with degradation-dependent weights for remaining useful life prediction. Reliability Engineering & System Safety 184, 110–122.

Liao, L., Köttig, F., 2014. Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. IEEE Transactions on Reliability 63, 191–207.

Lopez de Calle – Etxabe, K., Ruiz – Carcel, C., Starr, A., Ferreiro, S., Arnaiz, A., Gomez, M., 2020. Hybrid modelling for linear actuator diagnosis in absence of faulty data records. GitHub repository. https://doi.org/10.5281/zenodo.3961536

Manohar, G.A., Vasu, V., Srikanth, K., 2018. ScienceDirect Modeling and simulation of high redundancy actuator for fault- tolerance. Materials Today: Proceedings 5, 18867–18873. https://doi.org/10.1016/j.matpr.2018.06.234

Matei, I., Ganguli, A., Honda, T., de Kleer, J., 2015. The case for a hybrid approach to diagnosis: A railway switch., in: DX@ Safeprocess. pp. 225–234.

Mazzoleni, M., Previdi, F., Scandella, M., Pispola, G., 2019. Experimental Development of a Health Monitoring Method for Electro-Mechanical Actuators of Flight Control Primary Surfaces in More Electric Aircrafts. IEEE Access 7, 153618–153634. https://doi.org/10.1109/ACCESS.2019.2948781

Medjaher, K., Zerhouni, N., 2013. Hybrid prognostic method applied to mechatronic systems. The International Journal of Advanced Manufacturing Technology 69, 823–834.

Mishra, M., Leturiondo, U., Salgado, O., Galar, D., 2015. Hybrid modelling for failure diagnosis and prognosis in the transport sector: Acquired data and synthetic data. Dyna 90, 139–145.

Narasimhan, S., Roychoudhury, I., Balaban, E., Saxena, A., 2010. Combining model-based and feature-driven diagnosis approaches-a case study on electromechanical actuators.

Qian, P., Ma, X., Cross, P., 2017. Integrated data-driven model-based approach to condition monitoring of the wind turbine gearbox. IET Renewable Power Generation 11, 1177–1185.

Qiao, G., Liu, G., Shi, Z., Wang, Y., Ma, S., Lim, T.C., 2018. A review of electromechanical actuators for More/All Electric aircraft systems. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 232, 4128–4151.

Ruiz-Carcel, C., Starr, A., 2018a. Data-based detection and diagnosis of faults in linear actuators. IEEE Transactions on Instrumentation and Measurement 67, 2035–2047.

Ruiz-Carcel, C., Starr, A., 2018b. Data set for "Data-based Detection and Diagnosis of Faults in Linear Actuators." https://doi.org/10.17862/cranfield.rd.5097649

Ruiz-Carcel, C., Starr, A., 2015. Development of a novel condition monitoring tool for linear actuators, in: The Twelfth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies. pp. 1–12.

Sudhawiyangkul, T., Isarakorn, D., 2017. Design and realization of an energy autonomous wireless sensor system for ball screw fault diagnosis. Sensors and Actuators A: Physical 258, 49–58.

Susana Ferreiro, Andres Jimenez, Jon Madariaga, Eva Novillo, Santiago Fernandez, Andres Alia, Esteban Morante, 2013. Health monitoring for electro-mechanical nose landing gear door actuator of a uav, based on simulation modelling and data-driven techniques. Chemical Engineering Transactions 33, 655–660. https://doi.org/10.3303/CET1333110

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., Teal, T.K., 2017. Good enough practices in scientific computing. PLoS Comput Biol 13, e1005510. https://doi.org/10.1371/journal.pcbi.1005510

Declaration:

All manuscripts must contain the following sections under the heading 'Declarations'.

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

To be used for non-life science journals

Funding (information that explains whether and by whom the research was supported)

Conflicts of interest/Competing interests (include appropriate disclosures)

Authors declare no conflicts of interest.

Availability of data and material (data transparency)

Code availability (software application or custom code)

Authors' contributions (optional: please review the submission guidelines from the journal whether statements are mandatory)

Not applicable

To be used for life science journals + articles with biological applications

Ethics approval (include appropriate approvals or waivers)

Not applicable

Consent to participate (include appropriate statements)

Not applicable

Consent for publication (include appropriate statements)

Not applicable

11 Some final thoughts

> *Asi behar da ta jarraitu, aurreratuko bada. Asieran ez dira gauzac osotuac ecusten. Gueldica eguiten dira aurrerapenac*
>
> — **Ipui oncak (1804) by Vicenta Mogel**
> (First Basque female writer)

Antonia Vicenta Moguel lived during a time when Basque language was considered to be a language of uneducated villagers, furthermore, writing was the kind of duty a lady could not do. The brave Antonia proved those beliefs wrong when she wrote a book in Basque adapting some fables from Aesop. The book she wrote, Ipui onac, included the quote that begins this chapter, which translates to something like "You have to start and continue if you gonna ever move forward. At first you won´t see matters finished. Progress is made 'step by step' ".

Two centuries later, my grandmother Agustina Solozabal, who lives in the village besides Antonia's (and shares Antonia's bravery as well), found the way to synthesise the whole quote in to a short yet precise expression, and managed to transfer this pill of knowledge to her descendants. Her philosophy, "Txikirri-txikirri", represents the tireless effort that keeps driving humanity one step beyond. The kind of effort that draws a smile in the traveller's face when he/she looks back to the now distant starting point.

Enrolling on a PhD is saying yes to a roller-coaster journey where emotions have constant ups and downs (as life itself) due to paper rejections, paper publication acceptances, faulty experiments (do experiments ever go as expected?) and other vicissitudes that one faces during this complex journey. However, I would like to give a piece of advice to those who doubt whether to start the journey or not, or those other whose energies have been drained by the slope of the journey: Txikirri-txikirri.

I would like to thank you, for you who have read so far (even if you might have jumped some chapters or pages) and desire good luck in your own journey.