UNIVERSIDADE
BEIRA INTERIOR

# Industrial Sensors Online Monitoring and Calibration Through Hidden Markov Models

**Alexandre Daniel Batista Martins**

Tese para obtenção do Grau de Doutor em
**Engenharia e Gestão Industrial**
(3º Ciclo de Estudos)

Orientador: Professor António João Marques Cardoso
Co-Orientador: Professor José Manuel Torres Farinha

**Janeiro de 2024**

UNIVERSIDADE
BEIRA INTERIOR

# Industrial Sensors Online Monitoring and Calibration Through Hidden Markov Models

**Alexandre Daniel Batista Martins**

Tese para obtenção do Grau de Doutor em
**Engenharia e Gestão Industrial**
(3º Ciclo de Estudos)

Orientador: Professor António João Marques Cardoso
Co-Orientador: Professor José Manuel Torres Farinha

Provas de Doutoramento: 22 de dezembro de 2023
**Júri:**
PRESIDENTE:
Prof. Doutor Joaquim Mateus Paulo Serra

VOGAIS:
Prof. Doutor António João Marques Cardoso
Prof. Doutor Diego Galar
Prof. Doutor Jorge Manuel dos Reis Gama
Prof. Doutora Isabel da Silva Lopes
Prof. Doutor Mateus Daniel Almeida Mendes

**Janeiro de 2024**

## Declaração de Integridade

Eu, Alexandre Daniel Batista Martins, que abaixo assino, estudante com o número de inscrição D2738 de/o Engenharia e Gestão Industrial da Faculdade da Engenharia, declaro ter desenvolvido o presente trabalho e elaborado o presente texto em total consonância com o Código de Integridades da Universidade da Beira Interior.

Mais concretamente afirmo não ter incorrido em qualquer das variedades de Fraude Académica, e que aqui declaro conhecer, que em particular atendi à exigida referenciação de frases, extratos, imagens e outras formas de trabalho intelectual, e assumindo assim na íntegra as responsabilidades da autoria.

Universidade da Beira Interior, Covilhã ____/____/2024

Assinado por: **Alexandre Daniel Batista Martins**
Num. de Identificação: 14258080
Data: 2024.01.18 10:45:36+00'00'


(assinatura conforme Cartão de Cidadão ou preferencialmente assinatura digital no documento original se naquele mesmo formato)

# Dedication

The completion of this PhD was one of the most challenging, remarkable, and rewarding stages in my life. As such, it is with great desire that I want to leave a few words of thanks to all those who participated in it.

I start by thanking my family, especially my parents and sister who were with me and supported me in my journey, allowing all this to be possible.

I thank the University of Beira Interior and my supervisors for the training and support provided during this journey.

First, my thanks to Professor António João Marques Cardoso, for accepting the challenge of guiding me in this academic journey. It is with enormous appreciation that I express my gratitude for his support and for all the help he gave me in my development as an academic.

A special thanks to my co-supervisor Professor José Torres Farinha, for believing in me and giving me the opportunity to perform this PhD. I thank him for all his unconditional support, which greatly contributed to my scientific and personal training. Through his optimism, it was possible to reinforce the stimulus to the achievement of the defined objectives.

I would also like to thank Professor Inácio Fonseca for his friendship, intelligence, encouragement, suggestions, and advice regarding the conceptualisation of the methodology developed during the course of this PhD.

I want to thank my doctoral colleagues who accompanied me on this journey, from which I can highlight Balduíno Mateus, Edmundo Pais, and João Rodrigues. The group's help and cooperation allowed this path to be followed in a safer way.

I also thank the Universidade Lusófona de Humanidades e Tecnologias, and especially Professor Oliveira Santos. Without his management capacity and trust in me, none of this would be accomplished. As such, a special thanks.

# Resumo

Esta tese visa demonstrar uma metodologia capaz de diagnosticar, através do Modelo Oculto de Markov (HMM, Hidden Markov Models), o estado de saúde de equipamento de produção, bem como o estado de calibração dos sensores acoplados ao equipamento. Através de uma metodologia bem definida, as observações recolhidas pelos sensores são otimizadas para dar entrada num HMM. Após o treino, as observações são traduzidas em estados ocultos, que representam o diagnóstico do equipamento em estudo, sendo: Estado 1 - "Bom funcionamento"; Estado 2 - "Alerta"; Estado 3 - "Avaria/Descalibrado".

Após a recolha dos dados, estes passam por um processo de limpeza que irá melhorar a sua qualidade e integridade. Em seguida, é realizada uma fase de geração de características. Esta fase é extremamente importante porque a informação pode ser gerida para o equipamento desejado. É através desta fase que se pode distinguir o diagnóstico entre o equipamento de produção e o dispositivo de leitura (sensores). Em seguida é realizada uma redução dimensional dos dados, através da Análise de Componentes Principais (PCA, Principal Components Analysis) e uma extração de novas características que, embora em quantidades menores, cada uma tem mais informação. Em seguida, a nova matriz de dados é aplicada a um Clustering, realizado por K-means, com o objetivo de agrupar dados semelhantes dentro do mesmo grupo. Isto fará com que dados de bom funcionamento fiquem num cluster e dados de mau funcionamento fiquem num cluster diferente. Estes clusters serão os estados observáveis otimizados. Subsequentemente, o HMM traduz os estados observáveis numa sequência de estados ocultos que representam o diagnóstico do equipamento.

Além da metodologia disponível para detetar diferentes tipos de informação do mesmo conjunto de dados, esta tem mais capacidades, tais como: imputar valores em séries temporais com poucas amostras através dos métodos da Rede Neural Profunda (DNN, Deep Neural Network), nomeadamente o modelo Multi-Layer Perceptron (MLP), bem como realizar o prognóstico do estado de saúde do equipamento através da DNN, Gated Recurrent Unit (GRU).

# Palavras-chave

Metrologia; OLM; Manutenção; HMM; Sensores; *Machine Learning*.

# Resumo Alargado

O fluxo contínuo de produção sem desperdícios e sem stock é o objetivo de qualquer empresa. Isto requer, entre outras coisas, um plano de manutenção eficaz do equipamento de produção, que atue antes que ocorra uma falha. Desta forma, é possível evitar paragens inesperadas e, consequentemente, aumentar a disponibilidade do equipamento. Quando se fala de equipamento de alta criticidade, a sua disponibilidade é essencial para o fluxo de produção. Em certas indústrias, as paragens inesperadas do equipamento de produção podem conduzir a perdas monetárias muito elevadas. Para este tipo de equipamento é vital utilizar manutenção baseada em condições (CBM, Condition Based Maintenance) e mesmo manutenção preditiva (PdM, Predictive Maintenance). Para estes tipos de manutenção é necessário ligar sensores ao equipamento de produção. Os sensores serão responsáveis pela tradução do estado de "saúde" (condição) do equipamento. Ou seja, os sensores são responsáveis pela tradução dos estados físicos em sinais elétricos mensuráveis. Desta forma, é possível obter dados que, se estudados e manipulados da forma correta, podem dar uma visão de apoio à manutenção CBM e PdM. Para extrair informação fiável é necessário confiar nos dados recolhidos. Para isso, é necessário que os sensores estejam calibrados. A calibração dos sensores é da responsabilidade da metrologia, nomeadamente da metrologia industrial. Através de sucessivas calibrações, efetuadas por uma sucessão hierárquica de padrões metrológicos, é possível garantir a fiabilidade de um sensor industrial. Portanto, a metrologia industrial tem um papel muito importante no que diz respeito à manutenção baseada nas condições do equipamento. Pode dizer-se que a metrologia industrial é a base da manutenção CBM e PdM. Assim, para ter uma boa estratégia de manutenção em equipamentos de alta criticidade, é também necessário ter uma boa estratégia de calibração de sensores. Para isso é necessário parar de efetuar calibrações de forma periódica. Este tipo de estratégia pode levar a um excesso de calibrações, a um elevado custo de mão-de-obra e, mesmo a não efetuar calibrações quando realmente necessário. Por conseguinte, é essencial adquirir uma estratégia de Monitorização de Calibração Online (OLM, Online Monitoring). Com esta abordagem, os sensores só são removidos do equipamento de produção para serem calibrados quando é realmente necessário. Além disso, este tipo de estratégia pode ser alargado para estudar as perspetivas de manutenção do equipamento de produção. Isto porque os dados utilizados para detetar avarias nos sensores são os mesmos que são utilizados para detetar avarias nos equipamentos de produção.

Esta tese visa desenvolver uma metodologia OLM capaz de detetar o estado de calibração dos sensores para assegurar a qualidade dos dados a serem utilizados nas estratégias de

manutenção CBM e PdM. Além disso, a metodologia desenvolvida é também utilizada para diagnosticar o estado de saúde dos equipamentos de produção. Uma vez que a metodologia utiliza uma fase de geração de características, é possível separar o tipo de informação desejada relativamente ao equipamento a ser estudado. Além disso, o objetivo desta tese foi desenvolver uma metodologia que utiliza o Modelo de Markov Oculto (HMM, Hidden Markov Model) como ferramenta para classificar o estado de saúde dos equipamentos.

Assim, a metodologia desenvolvida mostrou ser capaz de classificar, através do mesmo conjunto de dados, o estado de saúde do equipamento de produção, bem como o estado de calibração dos sensores. Além disso, é uma metodologia genérica que pode ser utilizada em diferentes tipos de equipamentos com diferentes tipos de sensores. Pode funcionar em modo online, permitindo insights em tempo real para evitar falhas inesperadas. É também uma metodologia que funciona sem informação prévia sobre o comportamento do equipamento e sem conhecimentos técnicos sobre o mesmo. A metodologia tem três estados de classificação do equipamento, provenientes do HMM: Estado 1 - "Bom Funcionamento"; Estado 2 - "Alerta"; Estado 3 - "Falha / Fora de Calibração". Através destes estados, a metodologia é capaz, de forma autónoma, diagnosticar o estado de saúde do equipamento, quer o de produção quer os dispositivos de leitura (sensores).

# Palavras-chave

Metrologia; OLM; Manutenção; HMM; Sensores; *Machine Learning*.

# Abstract

This thesis aims to demonstrate a methodology able to diagnosis, through the Hidden Markov Model (HMM), the health state of production equipment, as well as the calibration state of sensors reading equipment. Through a well-defined methodology, the observations collected by the sensors are optimised to give input into a HMM, that are translated into hidden states, which represent the diagnosis of the equipment under study, being: State 1 - "Good working"; State 2 - "Warning"; State 3 - "Fault/Uncalibrated".

After collecting the data, it goes through a cleaning process that will improve its quality and integrity. Then, a feature generation phase is performed. This phase is extremely important because the information can be managed for the desired equipment. It is through this stage that we can distinguish the diagnosis between the production equipment and the reading equipment. Next, a dimensional reduction of the data is performed, through Principal Component Analysis (PCA) and an extraction of new features that, although in smaller amounts, have more information each one. Then, the new data matrix is applied to a Clustering, performed by K-means, with the objective of grouping similar data within the same group. This will cause good working data to be in one cluster and bad working data to be in a different cluster. These clusters will be the optimized observable states that give input to the HMM. Subsequently, the HMM translates the observable states into a sequence of hidden states that represent the diagnosis of the equipment.

Besides the methodology available to detect different types of information from the same data set, it has more capabilities, such as: imputing values in time series with few samples through Deep Neural Network (DNN) methods, namely the Multi-Layer Perceptron (MLP) model; performing the equipment health status prognosis through the Deep Neural Network (DNN), the Gated Recurrent Unit (GRU).

# Keywords

Metrology; OLM; Maintenance; HMM; Sensors; Machine Learning.

# Contents

# List of Figures

# List of Tables

# Acronyms

| | |
|---|---|
| AI | Artificial Intelligence |
| *BIPM* | *Bureau International des Poids et Measures* |
| CC | Calibration Certificate |
| CM | Condition Monitoring |
| CBM | Condition-Based Maintenance |
| CNN | Convolutional Neural Networks |
| DNN | Deep Neural Network |
| DBSCAN | Density-based Spatial Clustering of Applications with Noise |
| LDA | Discriminant Analysis |
| EURAMET | European Association of National Metrology Institutes |
| EFTA | European Free Trade Association |
| EU | European Union |
| FFT | Fast Fourier Transform |
| FG | Feature Generation |
| GRU | Gated Recurrent Unit |
| GMM | Gaussian Mixture Models |
| CGPM | General Conference on Weights and Measures |
| HMM | Hidden Markov Model |
| ICA | Independent Component Analysis |
| ICT | Information and Communication Technologies |
| IT | Information Technology |
| CIPM | International Committee for Weights and Measures |
| OIML | International Organization of Legal Metrology |
| VIM | International Vocabulary of Metrology |
| KPIs | Key Performance Indicators |
| LDA | Linear Discriminant Analysis |
| LSTM | Long Short-Term Memory |
| ML | Machine Learning |
| *MAE* | Mean Absolute Error |
| *MAPE* | Mean Absolute Percentage Error |
| NIST | National Institute of Standards and Technology |
| NML | National Metrology Laboratories |
| NN | Neural Networks |
| NPP | Nuclear Power Plants |

| | |
|---|---|
| NRC | Nuclear Regulatory Commission |
| OLM | Online Calibration Monitoring |
| IPQ | Portuguese Quality Institute |
| PdC | Predictive Calibration |
| PdM | Predictive Maintenance |
| PCA | Principal Analysis Components |
| PCs | Principal Components |
| RNN | Recurrent Neural Network |
| R$MSE$ | Root Mean Square Error |
| USA | United States of America |

# Chapter 1

# Introduction

## 1.1 Background

To maintain a good production flow in any industrial process, companies must achieve a good performance of their assets to reduce equipment wear and, consequently, avoid production stoppages and costs of economic development, inherent to non-production and loss of product quality. To achieve this goal, maintenance is a strategic tool. According to Campbell & Jardine "Smart organizations know they can no longer afford to see maintenance as just an expense. Used wisely, it provides essential support to sustain productivity and fuel growth, while driving down unneeded and unforeseen overall expenses" [1]. Maintenance evolved from a conventional idea, where it was seen as a necessary evil inducing high costs. Nowadays, it is different, being considered a strategic tool that allows the optimization of production and the business performance. Additionally, it is a tool that cannot be seen independently of the production area, as it generates value for manufacturers [2]. Better maintenance, aided by information systems, helps increase the life cycle of assets, reducing effluent emissions and energy consumption [3]. Therefore, it is essential to identify the best maintenance strategies that optimize the maintenance production, logistics and related areas, aiming to maximize the company's incomes.

As is well known, the industrial sector has developed considerably along time. As explained by Xu et al. [4] and Oztemel & Gursev [5], the industrial revolutions began at the end of the 17th century, with significant differentiation in the way of producing and operating in the industries, increasing productivity and, in turn, benefiting the entire value chain added to economic development. The first industrial revolution came from steam engines, which alleviated manual labour. In the 20th century, the second industrial revolution began, where electricity was introduced into industrial systems, allowing mass production and division of work. We are currently finishing the third industrial revolution that started in the 1970s. This revolution is characterized by electronics and Information Technology (IT), improving, and automating production [6]. We are now moving towards the evolution of the fourth industrial revolution, called Industry 4.0. It is like the third industrial revolution, but with a more significant extension of the digital paradigm, causing a wide range of economic and social

disruptions through digital transformation, such as new products and services and complex and unpredictable jobs and wellbeing [7]. We are evolved from an industrialized society to a post-industrialized community based on knowledge, services, and information. Industry 4.0 provides the connection of all means of production that allows real-time interaction between them, achieving greater operational efficiency, productivity gains, growth, and greater competitiveness, in addition to the development of new business models, services and products [6].

Then, these smart factories, made up of sensors, computers, and networked assets, can adapt to unexpected changes such as product characteristics or even asset failures. The factory's ability to communicate with its users in real-time makes production processes whole monitored, reducing breakage rates, and thus contributing to continuous improvement. According to Smit et al., "(...) the information thus generated can be analysed with big data and cloud computing processes which allows detecting and addressing invisible issues such as machine degradation, component wear, etc., in the factory floor" [8].

So, we are faced to another scenario in which every society depends on information everywhere and always. Therefore, new technologies, processes, and skills were added to the information and the people who use them. Market evolution requires organizations to find new ways to improve their products and services to satisfy their customers. To achieve these goals, it is of utmost importance to analyse colossal amounts of data. It is necessary to acquire the ability to transform the collected data into products to allow the industry's development and growth. The importance of data storage and its study emerged to provide reliable and well-founded answers to doubts about a given topic, based on information from the past and in real-time. The data itself is a raw element that can exist somewhere, whether useful or not, because it has no meaning. Data is simply a value, a measure, which only has some meaning that can be contextualized [9], [10]. Thus, when the relationship and connection between the data are present, information emerges, giving meaning to the data, and the information is structured following a cognitive process, and its validation generates knowledge. According to Wodecki, "(...) data is the perception of reality through signals from "sensors" (human senses as well as devices)" [10].

The world has become a digital place, with technological advances and the emergence of various types of sensors, the internet and Information and Communication Technologies (ICT), generating a large volume of information flow (Big Data) [11], [12]. According to Zhu [13], Big Data is defined as an information asset characterized by large volumes,

speeds, and varieties, striving for specific technologies and methods to transform this data into valuable information. Big Data works with heterogeneous data sets: structured, semi-structured, and unstructured. There is an opportunity to define ideas about new types of data and their content, presenting the challenge for companies of how to act with the increase in data volume, processing speed, and data diversity; storing and analysing this data requires new techniques tools. Thus, the business has the possibility to become more agile and enable conscious decision-making, which were not within the scope of the company [11], [14]–[21]

Due to this availability of data, greater computational power, and more robust algorithms, it becomes evident the interest around Artificial Intelligence (AI) [22], [23]. John McCarthy was the founder of the concept of Artificial Intelligence (AI) in mid-1955, describing it as a process where a machine behaves intelligently if it is equal to human behaviour. Currently, there are several proposals to define this term, each with its purpose. Still, most focus on the concept of creating computer programs or machines with behaviours considered intelligent and comparable to humans. That is, it is a branch of science, usually associated with informatics, which helps to find solutions to complex problems in a more humane way [22], [24], [25]. The engineering community has widely used computers to optimize and/or automate numerous tasks. With its development, the increase in data, and more accurate algorithms, assistance methods are improved and made AI capable of functions such as structured and unstructured data classification, pattern detection, optimization, and predictive modelling. According to Lee et al. [26], "(...) Industrial AI is a systematic discipline, which focuses on developing, validating and deploying various machine learning algorithms for industrial applications with sustainable performance."

Machine Learning (ML), a subarea of AI, discusses the study and construction of algorithms responsible for making the computer learn without being explicitly programmed. They can learn from their mistakes and make predictions based on the data. Rather than simply following programmed instructions, these algorithms can build a model from sample inputs to make predictions or decisions using available data. It allows machines to learn and make predictions by recognizing patterns.

Predictive Data Analytics covers business processes and data computational models that allow a company to make data-driven decisions [27]. ML can essentially represent or overlap with Predictive Modelling, a tool for Predictive Data Analysis (Figure 1).

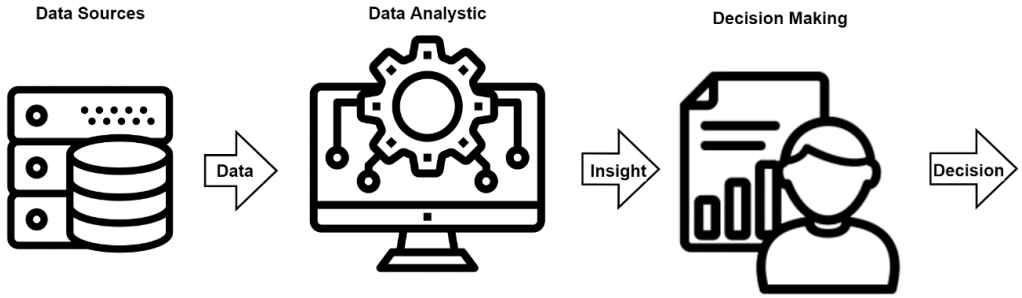



Figure 1 - Scheme of how a Machine Learning approach works.

With the development of ML methods, the detection and monitoring of shop floor processes can be simplified, making the decision driven by data augmentation easier. The possibility of real-time communication between humans, machines, and sensors, allows for fewer stops, detection of production defects, and improvement in the supply chain. This real-time communication allowed self-improvement and learning from experience [28], [29].

Through the data augmentation based on intelligent sensors, many opportunities for informed decision-making have been created. ML models have contributed to implementing Big Data technologies in several applications [30]. Based on this, ML has gained popularity, and companies have changed their decision-making processes in manufacturing applications such as predictive maintenance, fault diagnosis, schedule optimization, and product quality inspection. This introduction and implementation of ML in maintenance processes has changed the traditional predictive approaches to support decision-making processes [31], [32].

## 1.2 Problem description

When we talk about AI and Big Data, we must be aware that in the industrial environment industrial sensors are the largest source of data used in Condition-Based Maintenance (CBM), since this type of maintenance is based on the state of the equipment. According to Farinha [3], “Condition-Based maintenance is the maintenance planning that involves monitoring the equipment condition (Condition Monitoring - CM), usually predicting equipment failure”. It is necessary to make constant measurements to assess the health of the equipment, with the sensors being responsible for continuously interpreting the condition of the equipment. In this way, it is imperative to use reliable data, so that the information obtained is as accurate as possible with reality. This is guaranteed through the Science of Measurement, i.e., the Metrology.

Metrology is responsible for ensuring the reliability of the data provided by the sensors. The objective of Metrology is to guarantee the effectiveness and precision of the measurements made by the measuring instruments. Guedes [33] characterizes Metrology as the science of measurement that comprises all aspects, theoretical and practical, related to measurement. The metrology guarantees the traceability of reading equipment, and, in turn, the reliability of data collection is the calibration.

The calibration of reading instruments is an essential part of the asset management strategy. Companies invest a lot of money acquiring this type of equipment. For this reason, uncalibrated devices, providing wrong values, can imply that companies spend a lot of money, namely due to production stoppages, safety problems, production of inferior products, etc. [34]. By ensuring that reading instruments are calibrated, the company can confirm that the product meets customer requirements [35], [36]. Thus, calibration is one of the measurement instrument metrology services, which is a fundamental tool to ensure measurement traceability. According to Eren [34], calibration provides consistency in readings and reduces errors, validating measurements universally. According to Khosrow-Pour [16], "Data veracity is about the certainty of data meaning. This feature expresses whether data reflect properly the reality or not. It depends on the way in which data is collected. It is strongly linked to the credibility of sources. For example, the veracity of the data collected from sensors depends on the calibration of sensors". Then, the calibration can be considered the alkyl skin of the prediction [37].

For this reason, it is essential to address this topic, demonstrating the importance of calibration in supporting CBM and how it can be planned. Davies [38] refers a fundamental monitoring rule: "the instrumentation must be more reliable than the equipment being monitored". The same author also explains that the calibration must be valid for long periods or the integration of the calibration verification procedures about a reference standard as an integral element of the process because, although the sensors are pre-calibrated, some manipulations can affect their measurements [39].

Through frequent calibration, it is possible to minimize instrument error. Therefore, calibration shall be performed periodically according to the type of instrument and its stability characteristics, using adequate and traceable standards.

The importance of using calibrated equipment in a CBM strategy, according to Grous [40], is:

- To plan and trade with confidence;

- To optimize resources to be competitive;
- To ensure compatibility of measurements at different locations, at different times, but under the same conditions, thus justifying the adequacy of repeatability.

Through the industrial revolution and the emergence of I4.0., it began to increase the number of sensors in industrial processes, who could find a vast field of measurement for the equipment used. For this reason, it became necessary to acquire calibration management strategies. Shagluf et al. [41] explained that a complete calibration, performed on large or complex machines, can take several days. Therefore, it is necessary to define the best sensor calibration strategy depending on each situation.

There are two types of calibration strategies: Time-based preventive calibration strategies, which are performed at regular intervals, with calibration cycles not exceeding the time required for the sensor to exceed its tolerance; Predictive Calibration (PdC) strategy, that can have periodic checks of sensors, using relatively non-invasive methods, thus allowing calibration to be scheduled through a more informed process. Another form of PdC strategy is done by monitoring equipment conditions, called On-Line Monitoring (OLM). According to Hines et al. [42], it can be implemented for instrument monitoring, equipment monitoring, or operation monitoring; however, the acronym OLM is commonly used for extending sensor calibration intervals. OLMs are designed to monitor and diagnose sensors with online measurements during their measurements. According to Hoffmann [43], the OLM system offers an alternative approach to traditional time-directed calibration and opens the possibility of a calibration strategy based on sensor conditions. In this way, there may be an elimination or reduction of unnecessary field calibrations, reducing the associated labour costs and the health risks of personnel that may arise from the removal of the sensors (for example, if it is in oil or nuclear energy companies), and reduce the potential for calibration failure. Also, through this strategy, a sensor that requires calibration is not ignored, leading to unexpected downtime of production equipment and safety risks. Therefore, it is essential to define OLM strategies and models to support maintenance methodologies based on equipment conditions and sensor data collection.

## 1.3 Purpose and objectives

The objective of this research is to propose a modelling approach for a monitoring system that is generic to any type of equipment and sensors, without the need for prior information on equipment failures. The main tool used is the Hidden Markov Model (HMM), with the objective of classifying the state of a production equipment and the

need for calibration of a reading equipment, using only the data collected by the reading equipment.

The research has the following specific objectives:

1. To study modelling strategies, identifying possible improvements and propose a model capable of classifying the equipment health status;
2. To develop a model integrating several ML methods that optimize and improve the classification made by the HMM;
3. To define the performance improvement models to make predictions for the equipment health status in the time ahead;
4. To validate the modelling approach for classifying the equipment failure and the need for sensor calibration, as well as its prediction for the future, using real data.

## 1.4 Research questions

The following research questions (RQs) are answered in this work:

RQ 1: How important is metrology in supporting condition maintenance strategies?

RQ 2: How to diagnose the health status of a production equipment using the HMM?

RQ 3: How to use the health status diagnosis methodology of health equipment for a small data set?

RQ 4: How to predict the fault diagnosis for the time ahead?

RQ 5: How to diagnose the calibration status of a measuring equipment and the health state of a production equipment through the HMM?

The five attached papers address the five RQs, as is shown in Table 1.

TABLE 1 - RELATIONSHIP BETWEEN RQS AND APPENDED PAPERS

| *RQ* | *Paper I (Appendix A)* | *Paper II (Appendix B)* | *Paper III (Appendix C)* | *Paper IV (Appendix D)* | *Paper V (Appendix E)* |
|---|---|---|---|---|---|
| *RQ 1* | X | | | | |
| *RQ 2* | | X | X | X | X |
| *RQ 3* | | | X | | |
| *RQ 4* | | | | X | |
| *RQ 5* | | | | | X |

## 1.5 Scope and limitation

The research explores how, through data collected by sensors, it is possible to classify the state of health of a production equipment and the state of calibration of the reading

equipment, proposing a model of several ML tools and a classification made by the HMM.

It validates the modelling approach for two diagnostic processes, the health status of a production equipment and the calibration status of the reading equipment, as well as a diagnostic prediction for the future.

Research limitations include the following:

- Since the model predicts fault diagnosis and calibration status without previous information, there is no way to check the reliability of the developed models, since the study carried out in the company does not have enough time;
- The study is designed to use the classification made by HMM, thus restricting the diagnosis only to this method;
- The production equipment diagnosis mode indicates that there is a fault in the equipment but does not specify which type of fault or which component.

Future works will be oriented to solve these limitations.

## 1.6 Authorship of appended papers

Table 2 summarizes each author's contribution to the attached articles. The contribution is divided in the following tasks:

1. Conception and design study;
2. Data collection;
3. Data analysis and interpretation;
4. Writing the article;
5. Critical analysis of the article.

TABLE 2 - CONTRIBUTION OF EACH AUTHOR TO THE APPENDED PAPERS

| *Author* | *Paper I* | *Paper II* | *Paper III* | *Paper IV* | *Paper V* |
|---|---|---|---|---|---|
| *Alexandre Martins* | 1,4,5 | 1,3,4,5 | 1,3,4,5 | 1,3,4,5 | 1,3,4,5 |
| *Inácio Fonseca* | | 3,5 | 3,5 | 3,5 | 3,5 |
| *José Torres Farinha* | 5 | 2,5 | 2,5 | 2,5 | 2,5 |
| *Balduíno Mateus* | | | | 3 | |
| *João Reis* | | 5 | 5 | 5 | 5 |
| *António João Marques Cardoso* | 5 | 5 | 5 | 5 | 5 |

## 1.7 Outline of the thesis

The thesis consists of six chapters. The first chapter describes the problem and establishes the research objectives. Chapter 2 presents the theoretical background. Chapters 3 and 4 describe the research methodology and summarize the attached documents, respectively. Chapter 5 offers the results and discussion. Chapter 6 presents the conclusions and suggests future work. Finally, they are attached the papers referred in Chapter 1.

# Chapter 2

# Theoretical framework

## 2.1 Maintenance

Industrial maintenance appears to solve and minimize problems inherent to the performance of a company's production physical assets, aiming to reduce wear and tear and equipment failures and, consequently, production downtimes and the corresponding high costs to guarantee their competitiveness. Maintenance is, therefore, an integral part of a company and, according to Leturiondo [44], it is composed of a set of actions that are being developed to ensure the proper functioning of equipment and facilities. Pinto [45] adds that maintenance is a set of activities through a combination of management, technical and economic interventions. On the other hand, Ferreira [46] defines maintenance as a set of actions that preserve or restore equipment in a specific state, guaranteeing a particular service at a minimum global cost. Through the maintenance terminology standard, NP EN 13306:2017 [47], maintenance can be defined as "the combination of all technical, administrative and management actions, during the life cycle of an asset, aimed at maintaining it or restoring it to a state of in which it can perform the required function".

Maintenance is a set of procedures that aims to maximize equipment operation in its lifetime, preventing or correcting failures or wear and tear. To mention the difference between the useful life and economic life of equipment, Conceição [48] explains that the useful life of equipment refers to the time between the acquisition of the equipment until its end of life. The economic life is the period, from the entry of new equipment to the moment when the costs incurred in the use of that equipment are minimized.

Consequently, it is essential to identify the best maintenance strategies related to production logistics, aiming to maximize the assets' availability. How much more this organizational strategy is effective, more consolidates the maintenance policy adopted. Proper maintenance provides companies' rational budgets and improved benefits, improving the operational flow.

According to Conceição [48], Barbosa [49], Faria [50] and Calvão [51], maintenance has the following objectives: to guarantee optimal operating conditions; to maximize the

utilization of maintenance resources; to prolong the equipment life cycle; to minimize the parts inventory; to speed up the ability to react; to record and archive the historical data of each equipment to facilitate the detection of possible problems in the future; to increase safety and productive output; to reduce the number of emergencies and the number of breakdowns; to increase the reliability of equipment; to get maximum productivity with minimum costs; to have a good level of organization and operation from an integrated management perspective; to actively interact with other company functions; to lower manufacturing costs; to reach a better product quality; to increase flexibility; to speed up response to orders; to shorten delivery times; to increase clients satisfaction. So, maintenance is essential and plays a vital role in improving a company's entire manufacturing process.

The mission of a maintenance organization, according to Ferreira [46], both in industry and the services provision, is summarized in three points: to safety people and goods (namely the environment); to ensure the quality levels; to secure the cost of the product or service. “Effective maintenance aligns a company’s needs for increasingly efficient maintenance effort, realistic resources, doing the right thing by the employees and the environment” [52].

For cost-efficiency, they are necessary two components of the maintenance costs: the cost of maintenance resources and the cost of consequences related to the maintenance level [44]. If there is low maintenance, little money is spent in the short term; that is, the cost of maintenance resources is low. But, if this situation, in first time, reduces the average time between breakdowns, in times ahead, it will imply an increase in the average repair time, an increase in production losses, and, consequently, delays in deliveries, which causes customer dissatisfaction. In this way, the money saved, initially, on maintenance, will cost more for the company in the long run. By contrast, if maintenance is excessive, there will be few failures, but very high use of resources, implying a high cost. Thus, it becomes necessary to define the best maintenance strategy to be used depending on the company's needs and the criticality of the equipment (Figure 2).

In summary, and citing Pinto [45], we can say that companies that prioritize good management over the maintenance function are likely to be winning companies.

To explain the theory regarding maintenance, all types of existing maintenance will be briefly described, to later address in more depth the Condition Based Maintenance.

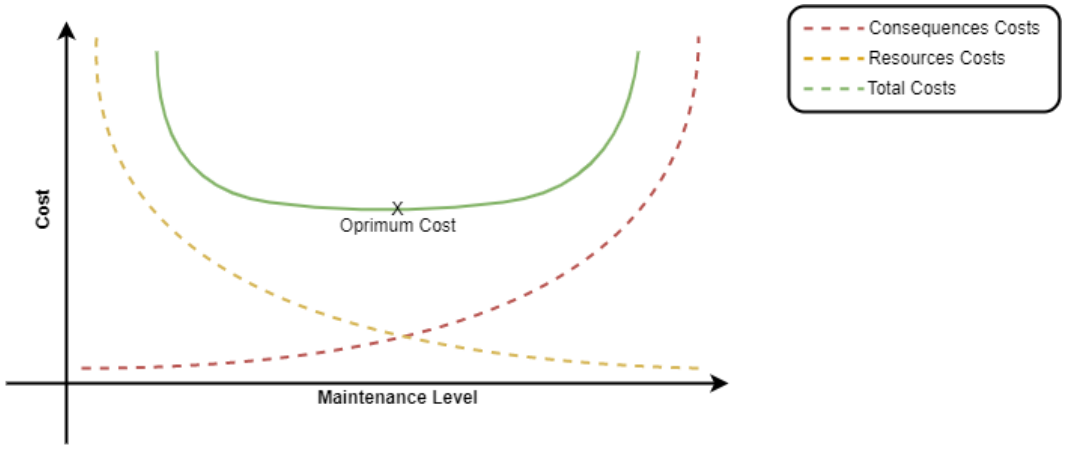



Figure 2 - Definition of maintenance needs and strategies to optimize cost.

There are several strategies for determining when maintenance should occur on a given asset. However, each strategy has associated shortcomings and advantages that make it suitable for every use [53].

Cabral [54] refers that there are two essential types of maintenance, namely, planned maintenance and unplanned maintenance (Figure 3).

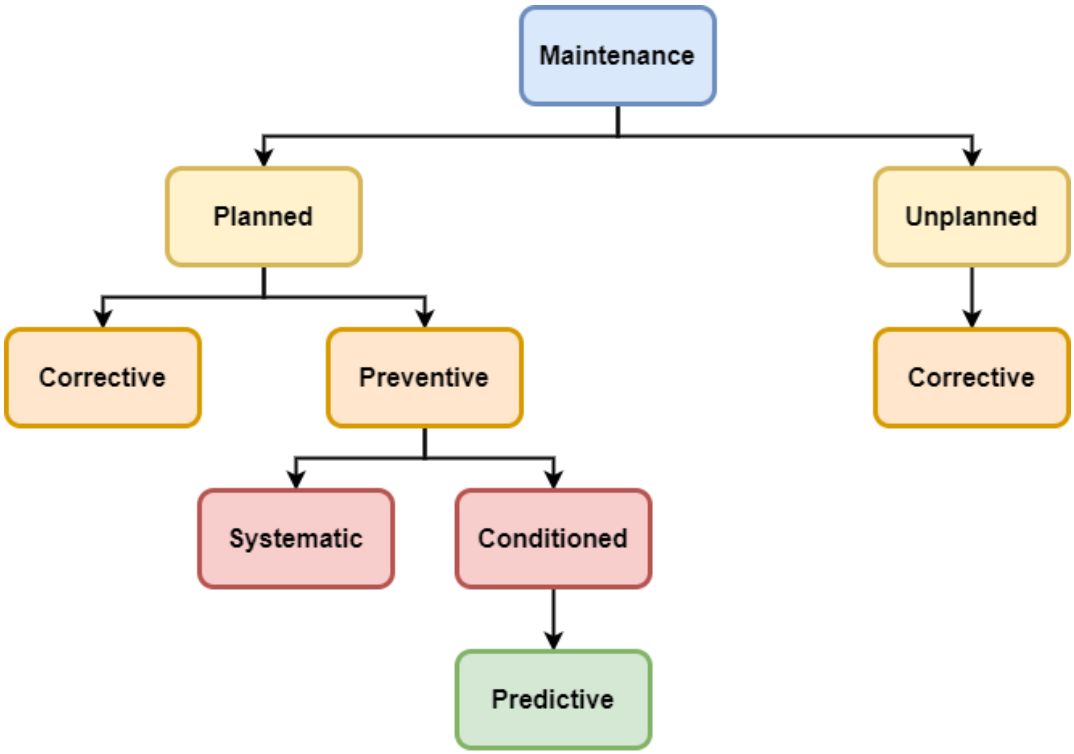



Figure 3 - Maintenance strategies.

### 2.1.1 Unplanned Maintenance (Corrective Maintenance)

This type of maintenance is done, as the name implies, without any planning. When there is a failure of equipment that leads to its stopping without any previous warning, it becomes necessary to carry out an emergency maintenance intervention (unplanned corrective maintenance). There is no plan, analysis, or preparation of any kind.

As Raposo [55] explains, unplanned maintenance corresponds to a policy in which interventions seek to recover the equipment to reach its normal operating state after the failure has occurred. Unplanned maintenance is the simplest and most primary form of intervention, being a management option, usually costly when evaluated economically from the point of view of the system operation.

Corrective maintenance includes identifying a broken component and repairing or replacing it to bring it to a state where the item can perform its required function, thus restoring equipment reliability [44]. It is reactive maintenance, where the consequences of its use can be: the reduction of the useful life of the equipment and, consequently, the production flow; the maintenance stop at unexpected and inopportune time, which may result in high costs to the company; the increased number of parts in stock for maintenance; to oblige frequent recourse to overtime; the lengthening downtime and also preventing the planning of machine downtime; to decrease the useful life of equipment; the high labour, parts, and service costs; the potential accidents or damages to the environment [54]–[57].

Since this maintenance is carried out, only when there is a failure in the equipment, there are no associated costs inherent to the planned maintenance. So, this type of maintenance can, and should be, reserved for equipment of minor importance for production (where the equipment availability is of little significance) and whose annual repair cost, as well as unexpected breakdowns, are acceptable. The choice of the corrective maintenance method is justified by the fact that the indirect costs of the damage are minimal, there are no safety problems, and there is equipment in which eventual failures do not critically affect the production or service [57], [58].

Regardless of the nature and level of maintenance adopted, there will always be residual the failures/breakdowns where corrective measures will be necessary. Thus, no matter how well planned and scheduled the maintenance is, it is impossible to have complete control over the operation of the machine; so, it is necessary, in certain situations, to take emergency action with a corrective nature [49].

### 2.1.2 Planned Maintenance

This type of maintenance is made in advance regarding the interventions to be carried out, considering the spare parts stocks, spare parts delivery times, and equipment availability.

It is maintenance that brings more significant economic benefits to the company and prolongs the equipment useful life and its availability for production. There is greater control over the equipment. It is a proactive maintenance policy that tries to avoid failures, increasing the safety and reliability of critical components. This approach requires an investment of time, technology, resources, and personnel [53].

According to Farinha [3], in planned maintenance the interventions follow an well-established program:

- To prevent failures or malfunctions and balance the maintenance workload;
- To tailor interventions to the asset production schedule;
- To prepare resources in advance to make interventions more cost-effective.

Planned maintenance, in turn, can be divided into two essential aspects, namely, corrective planned maintenance and preventive planned maintenance, which will be explained.

### 2.1.3 Corrective Planned Maintenance

This type of maintenance, where the intervention is carried out in a planned way, is used in two situations: the first, when there is a failure in the equipment and maintenance is not urgent and can be scheduled; the second is when the failure does not affect the operation of the equipment and, therefore, can be planned. This is only when there are no security problems, and production is not affected [56], [59].

### 2.1.4 Preventive Maintenance

According to NP EN 13306:2017[47], the preventive maintenance strategy is carried out at predetermined intervals or according to previously established criteria and is intended to reduce the probability of failure or degradation of the functioning of an asset.

As the chance of unexpected failures is reduced with preventive maintenance and, consequently, interruptions to the production program are minimized, the cost can be lower than corrective maintenance [44]. This maintenance carries out all actions performed before an equipment failure occurs. It is done in a revised, prepared and

planned way, where knowledge of the likely failure modes is necessary, being the key to this strategy to select the intervention period [44]. Its function is to prevent unplanned stops and premature damage to equipment, which, if not repaired, would result in corrective activities. This maintenance is based on the state of the equipment, installation location, data provided by the manufacturer, among others [48]. It is used to avoid serious consequences caused by asset failures, such as preventive maintenance policies, and conduct maintenance activities before an asset fails [60].

The objectives of this type of maintenance are: the reduction of corrective interventions; the increase of equipment reliability; the increase of the useful life and the effective life of the equipment; the improvement of work planning; the regularization of workloads; the improvement of the safety (fewer unplanned stops); the increasing of the production rate; the increasing of the equipment availability; the better management of spare parts stocks; the reduction of the general production costs; the better quality of the final product; the better conservation and durability of the equipment; the reduction of failure rate and unexpected failures; the decrease of secondary failures, reducing the cost of repairs; the ease of establishing a maintenance plan; the reduction of inventory costs [56], [58], [59], [61].

Its disadvantages include: excessive and unnecessary maintenance; an overly optimistic estimate implies a risk of system failure; some failures may occur despite the implementation of this strategy; maintenance is usually performed when the machine is still operating in acceptable condition; variations in operating conditions must be considered due to their influence on the estimation of the component failure time [44]. According to Higgins & Mobley [57], "a literal interpretation of the term is a maintenance program that is committed to the elimination or prevention of corrective and breakdown maintenance tasks".

### 2.1.5 Systematic Preventive Maintenance

This type of maintenance is carried out in a planned manner and works according to a defined time interval or other control variable according to the equipment specificity. For every kind of equipment, the types of inspections that must be carried out are defined, thus designing the operations to be carried out in each scheduled visit [54]. Raposo [55] explained that periodic preventive maintenance consists of a methodology for preventing defects subject to stoppage or low performance of equipment or production systems. The existing interventions run a systematic program, with intervals measured in a given unit of time or using another parameter that reflects the equipment operation [3].

This periodicity is defined based on data provided by the manufacturer (optimal operating conditions, lubrication points, periodicity, etc.) [62], [63]. It is a type of maintenance that has higher costs due to the several interruptions of the equipment and, sometimes, for the unnecessary replacement of components in an excessive way. Also, in this type of maintenance strategy performed at predetermined intervals, unnecessary replacement of components can expose the asset to a fault classification (inadequate assembly of components, improper installation, malfunction of components, improper storage procedures of components, etc.) [59].

### 2.1.6 Preventive Condition Based Maintenance

Regarding preventive Condition-Based Maintenance (CBM), one of the focuses of the doctoral program study analyses the operating condition of the equipment through non-destructive tests or condition evaluations carried out periodically, aiming to reduce the probability of failure. It evaluates several performance parameters of the equipment, indicating the need for intervention, thus allowing more extended machine operating time without maintenance interventions [56], [63], [64], [65]. The real health state of the systems is analysed, made through measurements of analysed physical parameters (vibration, noise, temperature, …), to detect failures early and to support decisions based on real data. This leads to problems being detected, analysed, and corrected before they lead to equipment failure [66]. In CBM, the useful life of the equipment is monitored through its operating condition, with 99% of equipment failures being preceded by certain signs, conditions, or indications that a failure may occur [61], making the maintenance model necessary for better equipment integrity management, lower life cycle cost, prevention of catastrophic failures, etc. In general, the main objective of CBM is to make a real-time assessment of the condition of the equipment to support maintenance decisions, thus reducing unnecessary maintenance and associated costs [67]. This type of maintenance leads to low expenses when already implemented, given the reduced need for permanent labour, the increase in the facilities' availability and productivity, and the reduction of fixed assets in spare parts and other materials. However, implementing this method requires a well-organized maintenance system, which implies to have a specific measuring equipment, sometimes expensive, and specialized training to carry out the diagnosis [54]. CBM is typically used to reduce maintenance costs by avoiding unnecessary maintenance activities based on condition monitoring information, requiring maintenance action on an asset when the evidence of abnormality is represented in the condition monitored data [60]. Quatrini *et al.* [68] stated that a CBM strategy is considered optimal to be adopted when a failure or degradation process can cause critical economic losses.

According to several authors, from which may be highlighted Lin & Banjevic [62], Goyal & Pabla [67], Prajapati *et al.* [69], Florian *et al.* [70], there are two types of approaches in CBM: Diagnostics, where further analysis of the event, deals with fault detection, isolation and identification when an abnormality occurs and prognostics; Prognosis, which deals with the prediction of failures and degradation before they occur, that is, it is the process of predicting the future failure, analysing the current and previous history of the operating conditions. The prognosis is seen as predictive Condition Based Maintenance. As the NP EN 13306:2017 [47] indicates, predictive maintenance is seen as conditioned maintenance performed according to the extrapolated forecasts of the analysis and evaluation of significant degradation parameters of the equipment. Farinha [3] explains that, it is part of the conditional maintenance "carried out in accordance with the extrapolated forecasts of the analysis and evaluation of significant parameters of the asset degradation". Goyal & Pabla [67] state that prognostics are much more efficient than diagnostics to achieve zero downtime performance. Diagnosis, however, is required when the prediction of prognostic failure fails, and a failure occurs [62]. In this way, equipment monitoring requires tools based on historical data and statistical inference methods to define the system's health status and early detection of pending failures with timely pre-failure actions [70]. So, CBM has its main process in the Condition Monitoring (CM), where signals are continuously monitored using certain types of sensors and/or other indicators, causing maintenance activities to be carried out only when necessary, before a failure occurs [71], [72]. So, CM is responsible for observing the condition of the equipment. Guo *et al.* [60], Ahmad & Kamaruddin [61] explained that the CM process has several ways to act. It has two objectives: to collect equipment condition data; to increase knowledge about the causes and effects of equipment failures and deterioration patterns. It can be performed in two ways: online, performed during the equipment's operating state; offline, performed when the equipment is not operating. It can also be carried out: periodically, which is carried out at specific intervals, such as an observation collection every hour or at the end of each work shift, with the aid of portable indicators; continuously, where the collection of observations is, as the name implies, carried out continuously and automatically based on measuring equipment, such as sensors. Therefore, as explained by Allenby [64], it is necessary to select the most suitable measurement equipment to configure systems and routines to respond to the individual needs of the equipment to be monitored in CBM. Technical advances in Condition Monitoring techniques have provided a way to achieve high availability and reducing downtime [65]; as the author says, there is a wide and growing variety of condition monitoring techniques for machine CBM and fault diagnosis. Understanding

the nature of each CM technique and the type of information measured will undoubtedly support a decision model.

Mobley [73] says that CM is much more than a maintenance scheduling tool and should not be restricted to maintenance management. It can be used to improve production capacity, product quality, and the overall effectiveness of factories. In this way, the author considers the CM as a management technique that uses the regular assessment of the actual operating condition of the plant's equipment, production systems, and plant management functions to optimize the total plant operation, being able to identify the most, if not all, of the factors that limit the effectiveness and efficiency of the entire plant. The working principle of condition monitoring makes observations available, which need to be interpreted to take appropriate actions accordingly. There are, generally, two stages in Condition-Based Maintenance [65]: the first one is related to the acquisition of condition monitoring data and their technical interpretations; the second stage is the maintenance decision making, that is, what to do now, once information data about conditions and their interpretations are available. According to Jardine [62], a CBM program consists of three main steps, which can be seen in Figure 4:

1. Data acquisition to obtain relevant observations to take information about the health equipment;
2. Data processing, to process and analyse the data or signals collected in step 1 to better understand and interpret the data. It is possible to forecast anomalies and prognostics for the future, that is, the diagnosis of the health status of the equipment;
3. Decision-making to support efficient maintenance policies.

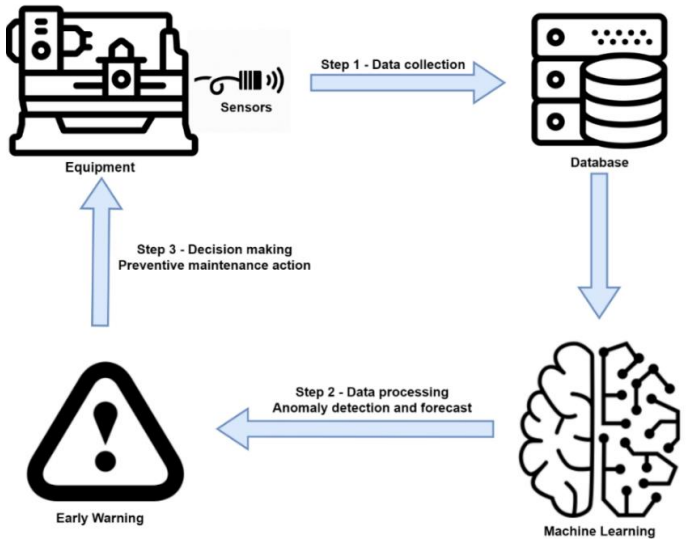



Figure 4 - Condition-Based Maintenance steps.

Through artificial intelligence processes, there is a development of powerful reasoning algorithms and forecasting techniques, which play an essential role in predicting the equipment life and to become one of the critical components for the success of CBM [69]. The same author also says that AI also focuses on machine learning algorithms. The purpose of these algorithms is to identify the complexity pattern of sensory data received from the operation or system under test and then create the smart decisions to be supported based on these data sets. This has a direct relationship with the CBM system because the purpose of this is to make the decision based on the data received from the system throughout its life cycle.

To understand the advantages of a CBM strategy over other maintenance methods, Leturiondo [44] highlights the following: lower failure rate and, consequently, more excellent reliability and machine safety; increased availability of equipment; repairs can be better planned; stocks can be reduced; provides enough data to improve knowledge about the equipment. The author also highlights the disadvantages: the need for necessary knowledge and the cost of resources; CBM requires knowledge of CM techniques, and high investment is required for data measurement tools and databases.

## 2.2 Sensors

Industrial sensors are the most significant source of data used in Condition Maintenance since this type of maintenance is done through the evaluation of the equipment health state made by this reading equipment. It is necessary to make periodic measurements, being this data systematically analysed to evaluate the condition of the equipment under study. The sensors are the responsible devices for translating the condition of the equipment at each moment. This is a topic that will be developed in a more detail in the next sections, including its operation and characteristics.

### 2.2.1 Predictive maintenance and sensors

Through the measurements performed by the sensors, where there is a collection of reliable and repeatable data, it is possible to monitor the condition of equipment. The information obtained from these data allows effective maintenance to be carried out and, consequently, to optimize machine performance. This type of maintenance is only possible through sensors, which are integrated into a machine or manufacturing equipment to provide the necessary information about the current condition of the system. This maintenance method has some advantages over others that are also used. According to Wild [71], using sensors, it is possible to evaluate the equipment even when they are running without having to stop production. The same author also says that

maintenance actions can be conveniently planned for the equipment and the production team. It is a maintenance method that allows for informed decision-making.

The equipment health conditions can be monitored and predicted through information acquired by the sensors. Optimal maintenance actions can be scheduled to avoid equipment breakdown, minimizing the total costs of operation and maintenance. CBM is a type of maintenance that, after its implementation, it is easy to follow. It is complex to implement since this type of maintenance requires the implementation of technological devices and tools to monitor production equipment [74]. Emmanouilidis & Pistofidis [72] refer that equipment condition monitoring involves analysing and processing samples from high-resolution sensors, using intelligent, statistical, and pattern recognition techniques to estimate a component's condition state machine or system. Then, a condition monitoring system will allow implementing CBM strategies, with decision supported on real condition of the monitored assets.

To perform a momentary diagnosis of the machine, keeping the necessary information for maintenance up to date, several types of sensors and analysis help in this data collection process. Wild [71] presents some of the forms of diagnosis:

- Vibration analysis - probably the most common and universally acceptable technique for monitoring condition in rotating machinery. Martins *et al.* [75] refer that vibration analysis is one of the most used techniques, adapting to both rotating and alternative equipment, allowing the identification of a high number of anomalies and influencing equipment availability;
- Temperature monitoring - by taking regular readings of the temperature of the machine components, it is possible to plan for deterioration due to temperature rise;
- Flow and Pressure Analysis of Hydraulic and Coolant Circuits - to perform condition monitoring on some hydraulic and refrigeration equipment, it is necessary to monitor flow and pressure performance. These two variables need carefully be monitored by sensors from which all flow and pressure readings can be taken over a period. Decreased flow indicates that the health of the system is in trouble - for example, worn impellers in pumps or leaks or broken seals;
- Oil analysis - is one of the most popular and straightforward methods for condition monitoring in the industry. The main objective is to detect early signs of debris in the oil. According to Smith & Mobley [76], the oil spectrometric analysis process is helpful for any mechanical device that uses oil for lubrication. It tests for the

presence of metals, water, glycol, fuel dilution, viscosity, solid particles, among other variables;

- Acoustic emission – it is a well-used technique for pressure vessels gaining popularity in other industrial areas;
- Thermography – it has been used to measure hot spots on electrical panels without removing covers or disconnecting power.

### 2.2.2 Industrial Sensors

According to Iniewski [77], sensor technologies have proliferated in several areas such as: science; product design; electronics; photonics; mechanics; chemistry, and biology. Sensors are used in the person's daily life, as well as in companies, and are responsible for detecting various signals such as sound, movement, optical or magnetic signals. The ability to have many small devices that transmit real-time data physically distributed close to detected objects brings new opportunities to observe and act in the world, bringing significant benefits to humanity [77].

The words sensor and transducer are both used in the context of measurement systems and often interchangeably. McGrath & Scanaill [78] refer that, there are no differences between them, and that the transducer is more used in the United States, while the sensor is more prevalent in Europe.

So, the sensors are responsible for directly connect a physical phenomenon and the data acquisition system, converting signals from mechanical quantities into electrical signals. They can convert the information received from a mechanical phenomenon, transforming it to be possible to read it (Figure 5).

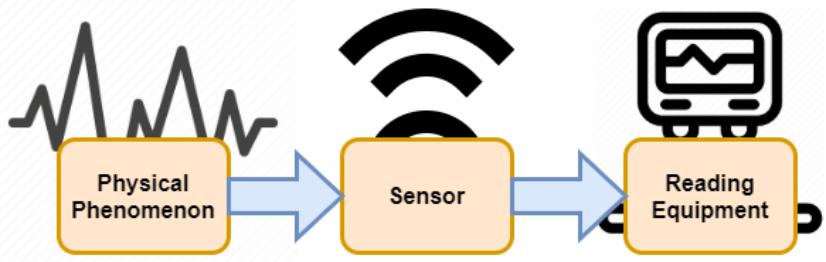



Figure 5 - Basic diagram of the operation of a transducer.

So, setting aside the various nuances of domain and application, a sensor measures something of interest and provides an output that can do something useful [78].

According to McGrath & Scanaill [78], the first sensor appeared in 1883 was a thermostat. From that date, several other sensors emerged; the first were simple devices, measuring a quantity of interest and producing some form of the mechanical, electrical, or optical output signal. Then, the advancement of technology, namely computing, widespread communications, connectivity through the Web, smart mobile devices, and cloud integration, led to an increase in sensors. This evolution is illustrated in Figure 6.

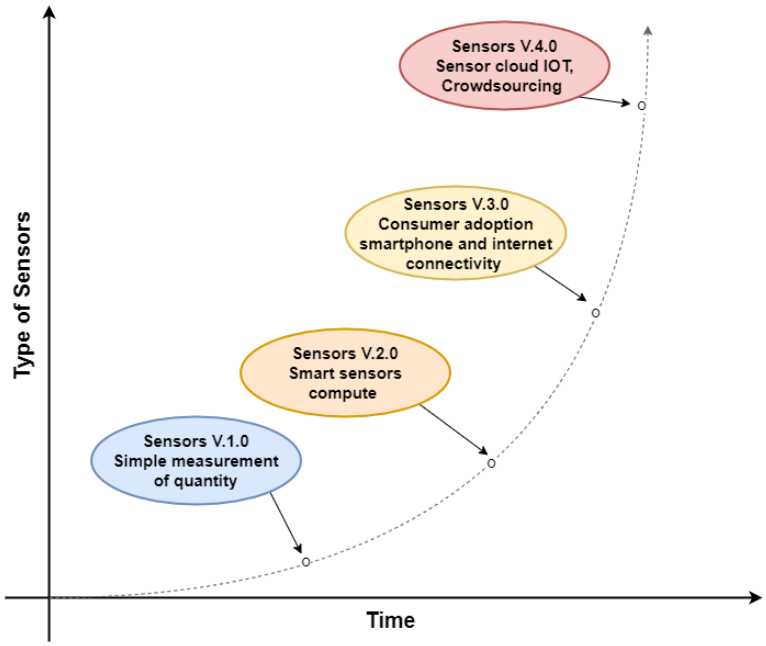



Figure 6 - Time evolution of sensors.

Sensors convert a mechanical phenomenon into a measurable electrical signal. Some sensors do not respond naturally to changes in mechanical phenomena, requiring signal conditioning. Before digitizing the sensor output, the signal may need other components and circuits to produce a signal that can take full advantage of the measurement hardware to reduce the noise effects from external interference [79]. Wilson [80] adds that sensors do not operate alone; they require a more extensive system consisting of signal conditioners and various analogue or digital signal processing circuits. For example, the system can be a measurement system, a data acquisition system, or a process control system. The same author also refers that sensors and their associated circuits measure various physical properties, such as: temperature, force, pressure, flow, position, light intensity, etc. These properties act as a stimulus to the sensor, and the sensor output is conditioned and processed to provide the corresponding measure of the physical property.

Wild [71] refers that the sensors can be used alone or together to monitor a specific situation, such as:

- Manufacturing operations;
- Health equipment condition;
- Inventory control;
- Working in progress;
- Identification of parts, tools, pallets, etc.

Sensors are a vital instrument in this subject of study because, through the data collected continuously from the manufacturing system, it is possible to analyse and determine the condition of the systems and the trends of how the conditions have changed since the previous analysis. This, along with technical training of parts and machines, can be used to "predict" when equipment may fail [71]. Today, the sensing of a factory can be considered more efficient, done with less supervision, and in unmanned manufacturing. Sensors can be used to control the manufacturing process, and their measurements allow a better understanding of the process, generating improvements [81]. The connection between profit and process measurement is illustrated in Figure 7.

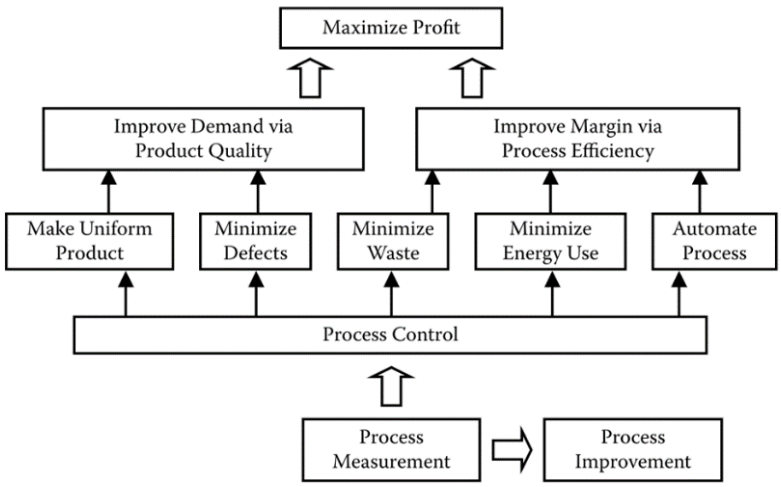



Figure 7 - Process measurement benefits for manufacturing operation and profits [81].

### 2.2.3 Type of sensors

According to Farinha [3], “there are many types of sensors, according to each type of condition variable. The reasons to choose a sensor are diverse, as are the type of output signal, the range of output values, the environmental conditions, the physical dimensions, and so on”. The same author also states that the sensor signals can be digital or analogue, and the digital sensors can have an interface to communicate with other

devices. In turn, analogue sensors no longer have this capability, so an analogue to digital converter is needed to enable communication with other devices.

Another way for characterizing the sensors is by measuring the factors of interest. We can describe the sensors as active (with contact) and passive (without contact). According to Wilson [80], McGrath & Scanaill [82] and Keysight Technologies [83], active sensors need an external source of excitation. They generate an electric current when the external physical environment changes. It requires physical contact with the factor to be measured, disturbing its state. The same authors explain that passive sensors generate their electrical output signal without external voltages or currents. They change their resistive, capacitive, or inductive characteristics along with their physical parameters. An external power source is required to induce an electrical output. It does not require direct contact with the factor to be measured, which will not imply a disturbance.

In Table 3, can be seen some examples of sensors and their characteristics.

TABLE 3 - TYPICAL SENSOR AND THEIR OUTPUTS [80]

| ***Property*** | ***Sensor*** | ***Active/Passive*** | ***Output*** |
|---|---|---|---|
| ***Temperature*** | Thermocouple | Passive | Voltage |
| | Silicon | Active | Voltage/Current |
| | RTD | Active | Resistance |
| | Thermistor | Active | Resistance |
| ***Force/Pressure*** | Strain Gage | Active | Resistance |
| | Piezoelectric | Passive | Voltage |
| ***Acceleration*** | Accelerometer | Active | Capacitance |
| ***Position*** | LVDT | Active | AC Voltage |
| ***Light Intensity*** | Photodiode | Passive | Current |

In both active and passive sensors, most reading outputs are made at relatively small voltages, currents, or resistance changes, requiring some signal conditioning. For this reason, the entire class of signal conditioning circuits has evolved. Amplification, level translation, galvanic isolation, impedance transformation, linearization, and filtering are fundamental signal conditioning functions that may be required [80].

Currently, there are sensors capable of making measurements of practically all existing physical quantities. For example, for temperature measurement, there are thermocouples, resistance thermometers, thermistors, and semiconductor junctions, which convert the temperature of the medium that are in contact with into a proportional analogue signal; for flow measurement, there are, among others, turbine flow meters, which generate a rectangular wave signal whose frequency depends on the flow velocity;

load cells are available for measuring through voltages; for pressure measurement, there are several types of pressure transducers; etc. [84]–[86]

### 2.2.4 Smart Sensors

The increase in industrial data has led to the increase use of data collectors. Currently, smart sensors are widely used in industrial production and, because of it, they arise many opportunities and challenges for supported decision-making [30]. The task of conventional sensors for measuring the equipment condition is to measure parameters and transmit, usually by wire, the measurements to the data acquisition unit, which are then treated and processed. Nowadays, it becomes possible to "transfer" intelligent behaviour to the sensor, being possible to process the data signal before transmitting. This is possible through hardware and software components (microcontrollers, memory, and basic processing software) added to the sensor [72].

Bolton [87] says that smart sensors can be programmed for specific requirements, such as processing raw input data, correcting items such as nonlinearities, and then sending the processed data to a base station; sending a warning signal when the measured parameter reaches some critical value. A smart sensor is a concept that represents sophisticated sensor functionality. They are defined in the IEEE 1451 family of standards, which describes a set of open communication interfaces, where smart transducers have features for self-identification, self-description, self-diagnosis, self-calibration, location recognition, time recognition, data processing, reasoning, data fusion, alert notification (report signal), standard-based data formats, and communication protocols, supported by the so-called TEDS (Transducer Electronic Data Sheet) [81], [87], [88]. The key feature of these standards is the definition of TEDS, corresponding to storage, identify the transducer, calibration, correction data, measurement range, manufacturing-related information, etc. These sensors moved from sensing endpoints to "sensor nodes", and the device-level architecture evolved rapidly [72]. Zanjireh & Larijani [84] explain that a sensor node corresponds a sensor network capable of performing some processing, collecting sensory information, and communicating with other nodes connected in the network. This makes it possible to support sensor intelligence for condition monitoring systems. Smart sensor behaviour can range from simple signal amplification to advanced data modelling techniques for condition monitoring. The characteristics and functionalities of an intelligent sensor, according to Emmanouilidis & Pistofidis [72], are:

- Processing capacity and software routines suitable for processing data locally;

- Make efficient use of network infrastructure through complex protocols and distributed communication standards, implementing policies that improve network robustness and flexibility and lessen the load on centralized nodes;
- Can support the execution of advanced distributed processes such as collective decisions, node task allocation, and network-wide workflow management;
- Classify data according to its criticality to avoid unnecessary data processing during a critical stage of the monitoring item;
- Self-diagnosis and self-calibration capability and periodically request coordination sensors to collect and process network statistics;
- Can be reprogrammed, facilitating the network to receive remote software updates. Additional processing techniques can be downloaded to a smart sensor.

The same authors also report that typical sensor node nowadays designs provide software capable of responding to a specific monitoring need, thus allowing wireless modules to perform:

- Smart Detection – able to decide when to send machine parameters, locally assessing their novelty, significance, and confidence to get the first decision;
- Smart sustainable monitoring tools - constant monitoring of the sensor network and adapting its topology and operation to new parameters, conditions and events (i.e., node failure). Network self-diagnosis and calibration can now be placed inside the tool (sensor nodes) and is no longer the responsibility of the instrumentation technician;
- Smart Diagnostics - Critical/Warning level measurements automatically invoke the local execution of simple or complex diagnostic procedures.

It is possible to fulfil different condition monitoring tasks through intelligent sensor networks. In this way, it is possible to create a flexible, programmable, customized tool that supports sustainable machine operation. According to Emmanouilidis & Pistofidis [72], "a collective of smart sensors can more than adequately fulfil the processing needs of most condition monitoring processes. The need for pre-scheduled visual inspections, maintenance actions and data readings is reduced when a group of smart sensors is able to identify critical events from fused novel data and automatically produce reports that can drive the scheduling of maintenance tasks."

The integration of smart sensors allows for more excellent reliability and performance and can reduce production test costs. But it is also necessary for companies to measure the initial development costs, which can be significant [78].

Another important concept about smart sensors is the Internet of Things (IoT). It is a network of electronic devices of low cost and low energy consumption, where communications occur without human intervention (in this case, we are talking about smart sensors). All devices are connected, and each device acts as a "smart node" in the network, gathering information and performing low-level signal processing to filter out noise signals and reduce the bandwidth required for communications. The nodes are connected to the internet and send information to the "cloud", securely to protect, store and process data [87], [89]–[94]

According to Addepalli *et al.* [90], IoT promises an increase in smart wireless sensors, cheap and low power consumption, increasing the volume of data. In this way, it is possible to understand the product life cycles detailed in a few minutes. The same author also refers that with smart sensors and IoT technologies, such as Radio Frequency Identification (RFID) and MTConnect, data acquisition becomes increasingly economical and widespread. He also explains the importance of connecting sensors to machines or connecting machines to each other for quick decision making.

Through the progress of sensors and IoT technology, it is possible to acquire condition data much more efficiently and cost-effectively, allowing CBM to become a more attractive aspect. Lai *et al.* [91] enunciate the advantages of adopting IoT for CBM: an opportunity for more cost-effective remote monitoring in real-time; rapid data analysis capability, which enables faster and better maintenance decisions based on real-time information on equipment conditions provided by advanced cloud computing; allows companies to promote integrated operations of products and services, such as WSNs or remote monitoring systems, that consist of distributed autonomous devices equipped with sensors that monitor physical or environmental conditions and can cooperate with internal systems to monitor the status in real-time of a product or equipment.

At least, an in-depth analysis is required, using artificial intelligence methods, to gain insights and competitive advantages in manufacturing systems.

## 2.3 Metrology

As referred above, smart sensors can perform various tasks, but their acquisition is still expensive. As such, it may be more advantageous to carry out in-depth analyses using conventional sensor data, using artificial intelligence methods to obtain insights and competitive advantages in manufacturing systems. Furthermore, no matter how expensive or sophisticated, an instrument should never be trusted 100%. An anomalous reading should not be disregarded, and costly actions should not be taken based on

unconfirmed evidence from a single instrument. Another aspect to consider concerns the reliability of the data collected by the sensors, which is only guaranteed through metrology, and the calibration of the sensors carried out by accredited laboratories. At this stage, it makes sense to introduce the term metrology and explain its role in ensuring the reliability of the data provided by the sensors, which are then worked on and studied for use in CBM. As Shagluf *et al.* [41] refer, modern metrology introduces vital factors such as downtime and product quality when measuring maintenance effectiveness. According to Rabinovich [92], "the true value of a measurand is the value of the measured physical quantity, which, being known, would ideally reflect, both qualitatively and quantitatively, the corresponding property of the object. Measuring instruments are created by humans, and every measurement on the whole is an experimental procedure. Therefore, results of measurements cannot be accurate. This unavoidable imperfection of measurements is expressed in their inaccuracy. Quantitatively the measurement inaccuracy is characterized by the notion of either limits of error or uncertainty".

Grous [40] refers that, in applied science, measurements are not accurate, as they are always subject to errors due to various human and material causes. Qualifying an error to subsequently quantify an uncertainty proves that the validity of the measurement result is in doubt. Therefore, evaluating measurement uncertainties generating errors is a very complex task.

For these metrics to exist, measuring instruments must prove that measurements are always the same, at any time and in any part of the world.

Therefore, the objective of Metrology is to ensure the effectiveness and accuracy of measurements performed by these measuring instruments [79]. According to the Portuguese Quality Institute (IPQ)[1] website, Metrology, as a science of measurement, provides reliable material support for the measurement system, essential in the sectors of the economy, health, safety, and the environment, constituting a critical infrastructure technology in modern societies. As in practically everything in people's daily lives, especially in the field of Engineering, there must be a metric; good use of Metrology is essential to ensure universal uniformity.

The Meter Convention, the treaty that created the International Bureau of Weights and Measures (BIPM), an intergovernmental organization under the authority of the General Conference on Weights and Measures (CGPM), with oversight by the International Committee for Weights and Measures (CIPM), was signed in Paris on May 20, 1875, by

[1] http://www1.ipq.pt/pt/metrologia/apresentacao/Pages/Metrologia.aspx (Accessed on 14/01/2022)

representatives of seventeen nations. In addition to founding the BIPM and establishing how the activities of the BIPM are to be financed and administered, the Meter Convention established a permanent organizational structure for member governments to act in concert on all matters relating to units of measurement [95], [96].

One of the CGPM's competencies is to discuss and examine the work carried out by the National Metrology Laboratories (NML); in turn, the BIPM makes recommendations on new determinations of fundamental Metrology in the remaining fields of Metrology. Currently, the BIPM comprises 56 Member States, including the most industrialized countries [97].

In the European panorama of the quality and sustainable development system, we can see, through Figure 8, its regional organizations.

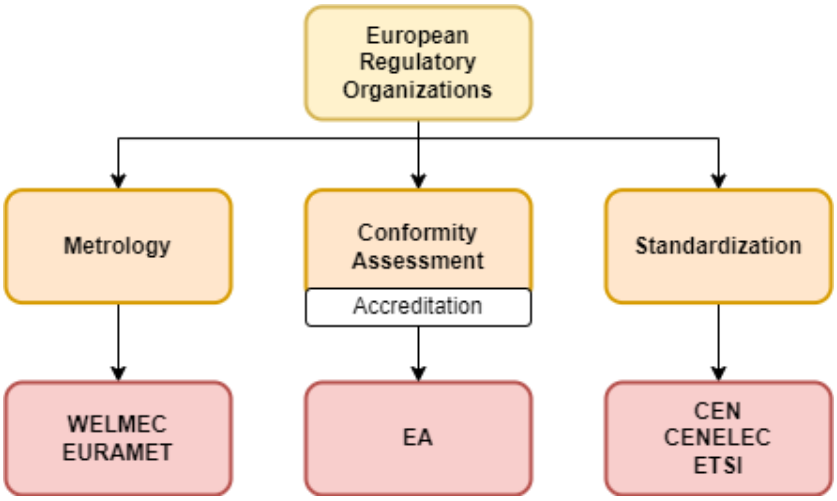



Figure 8 - European Regional Organizations of the quality system and sustainable development.

Regarding the European metrology regulators, according to the official website of the European Union[2], the global harmonization of measuring instruments is very advanced and, to achieve greater efficiency in this aspect, the European Association of National Metrology Institutes (EURAMET) was created. This is the Regional Metrological Organization that represents the national metrological organizations of the European Union (EU) and European Free Trade Association (EFTA) countries, coordinating the activities of its members. In the case of legal metrology, the European regional metrology organization is WELMEC. The members of this organization are the national authorities responsible for legal metrology in EU and EFTA countries:

[2] https://ec.europa.eu/growth/single-market/goods/building-blocks/legal-metrology/support-organisations_en [Accessed on 14/01/2022]

- EURAMET[3] - is the Regional Organization for Standardization (ROS) in Europe, responsible for the coordination and cooperation of the National Institutes of Metrology (NIM) in several areas, such as: metrology research; traceability of measurements to IS units; international recognition of national measurement standards; related calibration and measurement features. EURAMET facilitates the development of national metrological infrastructures through knowledge transfer and cooperation between members. Its mission is to develop and disseminate an integrated, economic, and internationally competitive measurement infrastructure for Europe, always considering the needs of industry, companies, and governments. One of the organization's main objectives is to increase the benefits of metrology to society. There are two main tools to achieve these goals: The European Metrology Research Program (EMRP) and the European Metrology Program for Innovation and Research (EMPIR), with more than 100 joint research projects to date.
- WELMEC[4] - WELMEC is the European Cooperation in Legal Metrology. It was founded on June 8, 1990, and its committee comprises three categories of entities: Members; Associate Members; and Observer Organizations. It also corresponds to various European trade associations and other regional metrology organizations. The main objective of WELMEC is to determine a harmonized and firm approach to European legal metrology. It is concerned with establishing, maintaining, and improving communication channels among its members and associate members, aiming to develop mutual trust through participation in shared activities. In addition, it also actively develops links with other regional and international bodies that have an interest in European metrology. WELMEC's work is carried out by their working groups, which later make guidance documents (WELMEC guides) available on its website.

Portugal can now be framed in this context once the European regulatory framework in the scope of quality and sustainable development has been defined.

Portugal has the *Sistema Português da Qualidade* (SPQ), which, according to Decree-Law No. Nº140/2004, assumes itself as a structure of national scope, which includes, in an integrated way, the entities that bring together efforts for the promotion of quality in Portugal and that ensures the coordination of the three subsystems: standardization; qualification; and metrology. In this ambit, the SPQ assumes the objective of

[3] https://www.euramet.org/about-euramet/ [Accessed on 14/01/2022]
[4] https://www.welmec.org/welmec/welmec-tour/welmec-information/ [Accessed on 06/07/2020]

guaranteeing and developing quality through all entities that, voluntarily or inherently, intervene in the various sectors of society. There are the three pillars of quality represented by Portuguese national bodies within this system, which will be mentioned later.

According to the CEN[5] website, a National Standardization Organization (NSO) is a one-stop-shop for all interested parties and is the main focal point of access to the concerted system, which comprises regional and international standardization. It has the responsibility of the national members of CEN to implement European standards as national standards.

Through the CEN-CENELEC[6] website, it can be verified that there are 34 NSOs, where Portugal is included through the *Instituto Português da Qualidade (IPQ)*. Figure 9 illustrates some of the logos used by the NSOs in Europe.


Figure 9 - National Standardization Organizations of Europe belonging to CEN/CENELEC[7].

IPQ, in addition to other activities related to quality and sustainable development, manages the Central Metrology Laboratory and the Legal Metrology Service. It coordinates other primary laboratories, looks after the hierarchical chains of standards, ensures national scientific and industrial metrology representation, and installs the metrology museum.

[5] https://standards.cen.eu/dyn/www/f?p=CENWEB:5 [Accessed on 08/07/2020]
[6] https://www.cencenelec.eu/aboutus/Communities/Pages/default.aspx [Accessed on 08/07/2020]
[7] https://www.anacom.pt/streaming/Modulo2_Normalizacao_Portugal.pdf?contentId=1347372&field=ATTACHED_FILE [Accessed on 27/01/2022]

### 2.3.1 Concepts

Metrology is the science of measurement, whose main objective is to ensure that measurement equipment reproduces values of specific reproducible and similar quantities internationally [79]. According to Sousa [98], Metrology is responsible for everything related to practical and theoretical aspects of measurement. Guedes [33] characterizes Metrology as the science of measurement that comprises all aspects, both theoretical and practical, related to measurement, whatever their uncertainty and the domain of science and technology to which they refer, relating processes to each other, the instruments, the place, the metrologist, etc. The main goal of metrologists is to find ways to measure various physical quantities as accurately as possible. Atom sensors, such as atom interferometers and atomic clocks, are examples of a class of instruments that can measure these quantities with very high precision [99]. In practical terms, Metrology comprises diverse aspects (for example, procedures, uncertainties, errors, standards, etc.) [95]. It includes all aspects, theoretical and practical, related to measurement, regardless of their uncertainty and the domain of science and technology to which they refer [100], [101]. Grous [40] explains that uncertainty reflects how a quantity is measured and the confidence given to a result. Instruments in measurement involve calibrations and manipulations, requiring appropriate procedures and calculations, a subject that will be discussed later.

To frame the different sets of terms related to Metrology, the International Vocabulary of Metrology (VIM)[8] was taken as a reference, which appears in the context of world metrology and seeks international harmonization of terminologies and definitions used in the fields of Metrology and instrumentation. This document is considered as a reference for metrologists, standardizing the terminologies, being divided into five parts:

1. Quantities and units;
2. Measurements;
3. Measuring devices;
4. Properties of measuring devices;
5. Measurement standards.

### 2.3.2 Practice areas

Metrology can be divided into three fundamental areas of activity [33], [79], [92], [94], [100], [102], [103], as can be seen in Figure 10.

[8] http://www1.ipq.pt/pt/metrologia/documents/vim_ipq_inmetro_2012.pdf (Accessed on 14/01/2022)

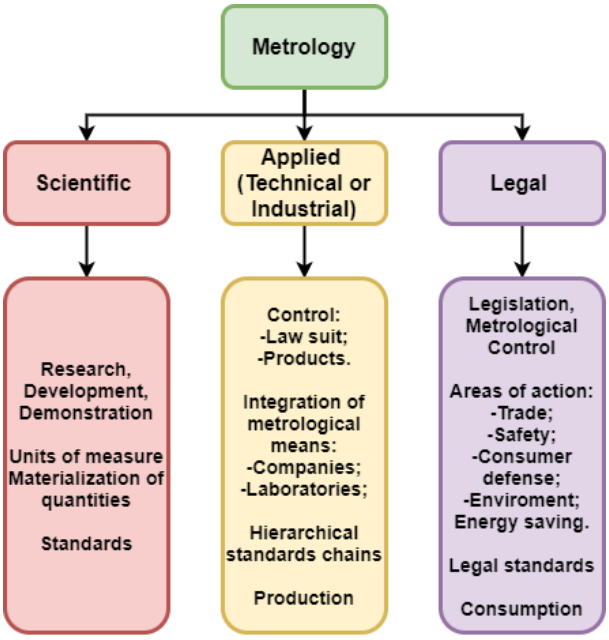



Figure 10 - Fundamental areas of metrology.

In the area of knowledge, we can say that metrology has the following objectives:

- Definition and maintenance of measurement units;
- Definition and maintenance of measurement standards at the highest metrological quality.

It is related to measurement units and international standards to maintain metrological quality. It is an area that operates with the highest accuracy and uncertainty, independent of other entities regarding traceability. Santos [103] refers that, in Portugal, Scientific Metrology is the responsibility of the IPQ, which intervenes through the Metrology Services Directorate, being its main objectives the following ones:

- The realization and maintenance of national standards;
- Participate in critical BIPM comparisons;
- Traceability of reference standards;
- Calibration of standards and measuring instruments;
- Organization of interlaboratory comparisons;
- Participation in the national accreditation system.

Regarding the Legal aspect, it is essential to emphasize the following:

- Commercial, fiscal, environmental protection, energy conservation, health and safety applications regulated by each State;

- Responsible for acting in the field of Legislation, with metrological control regarding the most diverse existing measuring instruments, concerning technical and legal regulatory requirements.

Neto [101] explains that it is the part of Metrology related to activities resulting from mandatory requirements, referring to measurements, measurement units, measuring instruments, and measurement methods developed by competent bodies.

According to Grous [40], it "is the intervention of the government to ensure the quality of measuring instruments or measuring operations affecting the public interest: safety of people, health and environmental protection, and fair transactions. The International Organization of Legal Metrology (OIML), gathering together government agencies analogous to the Sub-Directorate of Metrology, is responsible for establishing international guidelines on measuring instruments."

Concerning the technical or industrial aspect, the following stands out:

- Scientific and Industrial Metrology is a fundamental tool for growth and technological innovation, promoting competitiveness and creating a favourable environment for scientific and industrial development in any country. It is responsible for Proficiency Tests, a set of technical procedures for determining the performance of calibration or testing laboratories, through interlaboratory comparisons [101]. It is therefore essentially related to productive activity, exerting a specific control over processes and products that require an integration of metrological means at the level of companies and laboratories; it considers a hierarchical chain of standards, existing in laboratories or companies/organizations, which are also traceable by primary standards that can be national or international [104], [105].

### 2.3.3 Traceability

To have international metric coherence, there must be a traceability chain. That is an uninterrupted set of comparisons that ensure the result of a measurement or the value of a standard. It relates to the highest-level references, ending at the lowest level. Thus, it is guaranteed that the importance of the standard, even being of lower level, will agree with the higher ones. The traceability aspect operates in the applied area.

According to the VIM, traceability is the property of a measurement result whereby that result can be related to a reference through an uninterrupted and documented chain of calibrations, each contributing to the measurement uncertainty.

In the context of Metrology, metrological traceability is represented by a pyramidal level [33], [101], [106], as can be seen in Figure 11.

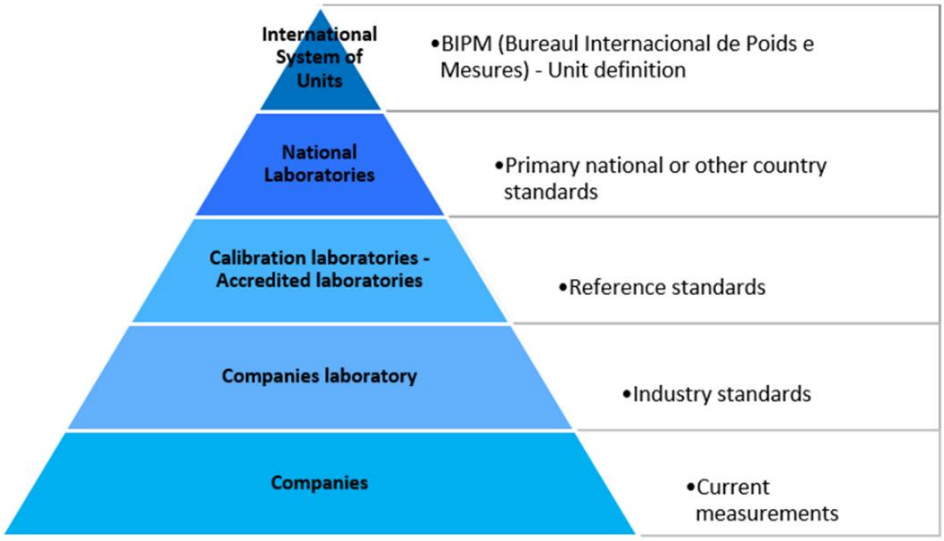



Figure 11 - Pyramidal traceability.

Wiemann *et al.* [106] argue that quality reliability is closely linked to the traceability chain and the availability of measurement standards. It also says that measurement standards are needed to establish an appropriate traceability chain traced back to the meter. Through traceability, it is possible to know the changes in the equipment, facilitating the management of several aspects [40], such as:

- Varied use and proper adjustment of equipment in the workplace;
- Selection of equipment among others offered by different suppliers;
- More or less accurate detection (based on records).

The author also adds that good traceability is necessary for a good analysis, defining the periodicity of traceability, classifying, archiving, and writing a procedure that describes the details of the instrument's way of life and keeps it up to date.

At the manufacturing level, Skliarov *et al.* [107] show that ensuring traceability of measurement results during the production process guarantees quality improvement. Pires [94] adds that the traceability of measurement must be guaranteed, allowing the necessary conclusions to be drawn about its metrological quality. Likewise, it is essential to carry out the most relevant actions to ensure the correct indication of the measuring instruments, which are the regular calibration of the measuring instruments (industrial metrology) and the periodic verification of the measuring instruments following legal

regulations (legal metrology). Verification is legal, based on regulations, standards, or decrees.

Any customer can require traceability in a contract or by standards such as ISO 9000-9004 and ISO 14253. This requirement is made to guarantee the quality of the measurement and for the protection of the buyer. Consequently, any tool or equipment used in production must ensure its measuring track [100].

In Europe, accredited European laboratories ensure traceability at the highest level. In the United States of America (USA), traceability is provided at the highest level directly by the National Institute of Standards and Technology (NIST) [98].

### 2.3.4 Calibration

According to Singhal *et al.* [108], "the veracity of big data requires careful design of data-acquisition and calibration strategies and of feature-extraction and selection strategies so that decision makers have clean, valid, and reliable inputs to use in making decisions." According to Sousa [98], one of the services of applied metrology is the calibration of measuring instruments, an essential tool to ensure the traceability of measurement. To establish a traceability chain, measurement standards traced back to the meter are used. This traceability is done through calibrations. The calibrated values and related measurement uncertainties of these high-level standards make that accredited calibration laboratories offer calibration services to work standards [106].

Calibration is the process of comparing an "unknown" element to an equivalent or better standard. It involves the determination of the metrological characteristics of an instrument, which is achieved through a direct comparison with standards, from which we can obtain the calibration curve. The calibration curve is specific for each device and converts the raw measurement into the corrected measurement [40]. It is obtained by subjecting the instrument to a real value of the quantity to be measured, which is supplied with a standard instrument and accurately reads the raw measurement provided.

Finally, a calibration corresponds to an issuing calibration certificate and placing a label. Based on this information, a user can decide whether the instrument is suitable for the application in question.

According to Smith & Mobley [76], and Mobley [104], the objective of a calibration system is to avoid the unreliability of the measurement tool through an immediate detection of deficiencies.

The main reasons for using calibrated instruments, according to Martins [79], are:

- Ensuring that an instrument's readings are consistent with other measurements;
- Determine the accuracy of instrument readings;
- Establish the feasibility of the instrument.

Through calibration, it is possible [95]:

- To assign the measured values to the indications;
- To determine the corrections related to indications;
- To determine other metrological properties.

Calibration is responsible for comparing a measuring instrument with a higher-level standard to determine the measurement error. Knowing this measurement error, it can evaluate the exact value of the metric to be measured.

### 2.3.5 Calibration certificate

Calibration is responsible for establishing the relationship between the value of the quantity produced and the value applied, and, through this process, its results are documented in a Calibration Certificate (CC). A current CC is a historical document, which only should be considered a baseline control, not guaranteeing that the meter is accurate at measurement. This is just a statement at the moment, under certain conditions, about the deviation between device indications and a reference standard [38], [40].

Therefore, before making potentially costly decisions based on a transducer or instrument reading, it is recommended to obtain a calibration check before and after the critical measurement. In this way, it will be possible to avoid stopping an equipment in good condition based on a sudden drop in performance, but conversely, it will support the decision to react if circumstances require [38].

ISO 17025:2005[9], General requirements for the competence of testing and calibration laboratories, describes the specific points that CC for measuring instruments must contain measurement results, including measurement uncertainty and/or a statement of compliance with a specification of identified metrology. Therefore, an indication of submission of the calibrated measuring instrument with established metrological requirements or different metrological characteristics for the CC is required. It is necessary to assess the conformity of the object with the specified requirements; the

[9] https://www.iso.org/obp/ui/#iso:std:iso-iec:17025:ed-3:v1:en [Accessed in 14/01/2022]

main requirement of these documents need to consider the uncertainty of the measurements during the conformity assessment [109].

According to Grous [40] and Zakharov & Neyezhmakov [109], the CC of an instrument provides the deviation, being the uncertainty in this deviation called calibration uncertainty. The sensor must consider the calculation parameters measurement uncertainty, which includes:

- uncertainty about the calibration performed during traceability;
- uncertainty due to device accuracy if not corrected;
- uncertainty related to instrument deviation (fatigue) between two calibrations;
- uncertainty linked to instrument characteristics (readability, repeatability, etc.);
- uncertainty linked to the environment if conditions are different during calibration;
- instrumental uncertainty of the standard, its instability, changes in operating conditions, mutual influence of the standard and the sensor to be calibrated;
- the observed variation in the calibrated sensor readings.

Grous [40] explains that the measurement uncertainties are parameters that portray the dispersion values during the measurement. In this way, the study of uncertainties aims to determine the capabilities of the measurement means. Uncertainty is determined by various components, particularly mounting patterns, instruments involved, and environmental factors.

### 2.3.6 Standards

Calibration is used to keep the measuring and control instruments within the specified limits, and for this, a standard is used to calibrate the equipment [83], [110]. It is necessary to have standards to be able to make a comparison between this and the equipment to be calibrated.

According to Wiemann *et al.* [106], "the reliability of quality assurance is closely connected with the traceability chain and the availability of measurement standards."
The units of physical quantities are reproducible with the help of reference standards and measuring devices, so these instruments play a vital role in the measurement unit. [92]
According to the VIM, standards are defined as a realization of the definition of a given quantity, with a determined value and associated measurement uncertainty, used as a reference.

Flores & Marques [95] define a measurement standard as an artefact, a measuring instrument, a reference material, or a measurement system intended to define, perform, conserve or reproduce a unit of one or more values of a quantity to serve as a reference. There is a hierarchical structure of reference standards (Figure 12) and their calibrations (Figure 13): International Standards; Primary Standards; Secondary Standards; and Work Standards. Through this structure and consecutive calibrations, it is possible to guarantee measurements from the primary standard to the working standard. The standards that rank higher in the hierarchy have greater quality and accuracy in measurements. That's why higher-level standards are used to calibrate lower-level ones.

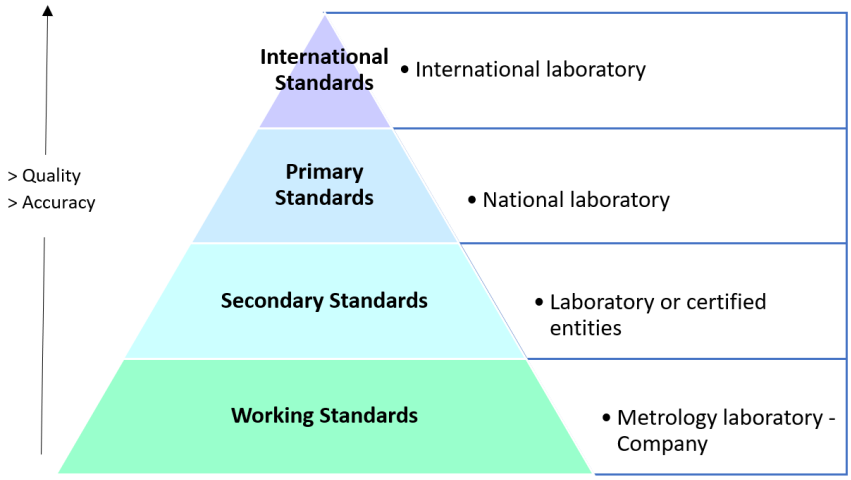



Figure 12 - Measurement Standards Hierarchy.

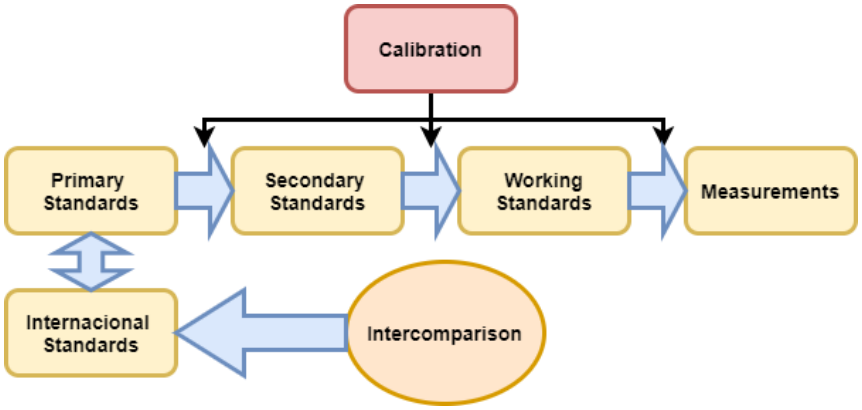



Figure 13 - Standard hierarchy and calibration.

Measurement standards traced back to the meter are used to establish a reliable traceability chain. Accredited calibration laboratories are responsible for offering calibration to working standards. This process involves a loss of precision at each step. To guarantee the production of the part within the required tolerances at the end of the

chain, the National Institutes of Metrology need to calibrate as accurately as possible [106].

### 2.3.7 International Standards

An international standard is a basis for recognizing a quantity that, later, sets standard values in the traceability chain. They are periodically evaluated and tested through absolute measurements of fundamental units. These standards are maintained at the *Bureau International des Poids et Measures (BIPM)* and are not available to the normal user for calibration [33], [40], [101], [111]. This Paris-based organization is responsible for:

- establish the standards of fundamental quantities and the scales of the main physical quantities, as well as to maintain international standards;
- compare national and international standards;
- ensure the coordination of the corresponding measurement techniques;
- carry out and coordinate the determinations regarding the physical constants involved in the activities (temperature conditions, humidity, etc.).

The BIPM ensures these powers through the authority of the Metro Convention, to which 55 nations belong, including Portugal.

According to Guedes [33], the international standard is recognized by an international agreement to serve as a basis, among several countries, for setting the values of other standards of magnitude to which it respects.

### 2.3.8 Primary Standards

A primary standard is a measurement standard with the highest accuracy in a country [92].

It has the highest metrological qualities, and the value is determined without reference to other standards of the same magnitude. Standard with the highest precision order calibrates a lower-level standard [40].

They possess the highest metrological qualities and whose value is accepted without reference by other standards of the same magnitude. In Portugal, the person responsible for these standards is the IPQ, which ensures the maintenance of primary standards and the accreditation of the Calibration Laboratories. These are standards for exclusive use in national laboratories, whose main function is to calibrate secondary standards. These

laboratories issue calibration certificates for the secondary standards, usually maintained by accredited calibration laboratories [84], [111].

According to Guedes [33], these standards have as their main function the traceability and calibration of secondary standards, not being made available for use outside the national laboratories.

The same author also explains that the quality of these standards is guaranteed by their traceability to the BIPM and by interlaboratory comparison programs with other reference laboratories.

### 2.3.9 Secondary Standards

A secondary standard is a measuring element of measurement that takes the value of one unit from the primary standard [40], [95].

These are considered reference standards used by accredited laboratories and are, usually, maintained by companies. These standards are periodically sent to national laboratories for calibration and comparison with primary standards [36], [84].

### 2.3.10 Working Standards

In addition to primary or national standards and secondary standards, the working standard is a reference standard generally used in companies' Metrology laboratories, mainly in tests and calibrations of other instruments used in the laboratory or another area of activity. These standards are periodically screened in accreditation laboratories for calibration and comparison with secondary standards [40], [101].

These working standards are commonly used in testing and calibrating other laboratory instruments or industrial applications.

It is a standard used routinely to calibrate or verify measurements of material, measuring equipment, or reference materials [33], [40], [95], [105], [112].

The calibrated values and related measurement uncertainties of high-level secondary standards that permit accredited calibration laboratories offer calibration services to working standards. These standards are like products and set the basis for the manufacturing industry to prove its measurement procedures and, ultimately, its quality control [33], [113].

A high-accuracy resistor manufacturer, for example, may use a standard resistor in the quality control department to check resistor testing equipment. In this case, the

manufacturer must verify that this industrial company processes according to the pre-established limits of accuracy [114].

### 2.3.11 Calibration strategies in a CBM methodology

According to Khosrow-Pour [16], "data veracity is about the certainty of data meaning. This feature expresses whether data reflect properly the reality or not. It depends on the way in which data are collected. It is strongly linked to the credibility of sources. For example, the veracity of the data collected from sensors depends on the calibration of sensors."

As Shah *et al.* [37] refer, "calibration can be considered the alkyl heel of prediction. For this reason, it is essential to approach this issue, demonstrating the importance of calibration in CBM and how it can be planned".

Davies [38] explains that there is a fundamental rule of monitoring; the instrumentation must be more reliable than the equipment to be monitored. This guarantee of the data provided is achieved through calibration. The same author also explains that there must be a calibration validity for long periods, or else the integration of calibration verification procedures against a reference standard that is an integral element of the process. Although the sensors are pre-calibrated, some manipulations can affect their measurements [39].

Through frequent calibration, it is possible to combat instrument error. Therefore, calibration should be done periodically related to the instrument type and its stability characteristics, using adequately maintained and traceable standards. Records of each instrument, tests used, traced test standards used, and any adjustments made must be kept. Due to the high price of traceable standards, it may be more economical to use specialized instrument calibration services. In this way, it is unnecessary to keep all the standards, except the most frequently used ones internally [38].

With the increase of sensors in the industrial process, there has also been a vast field of measurement of the equipment used. For this reason, it became necessary to acquire calibration management strategies. Shagluf *et al.* [41] explained that a complete calibration performed on large or complex machines could take several days. In this way, it is necessary to define a strategy for the calibration of the sensors to maximize the stoppings of the equipment in a programmed manner.

Essentially, there are two types of sensor calibration strategies: a preventive approach, based on time, where the sensors are forced to be calibrated periodically, which happens

practically only when there is a stop in the production; a strategy based on the condition of the reading equipment, where it is led to calibrate when necessary. Figure 14 shows, on a time basis, a type of preventive calibration done at regular intervals. These calibration cycles must not exceed the time required for the sensor to exceed its tolerance. This strategy is not adequate for large or complex machines due to the time needed for their calibration, as this is done more frequently and regularly.

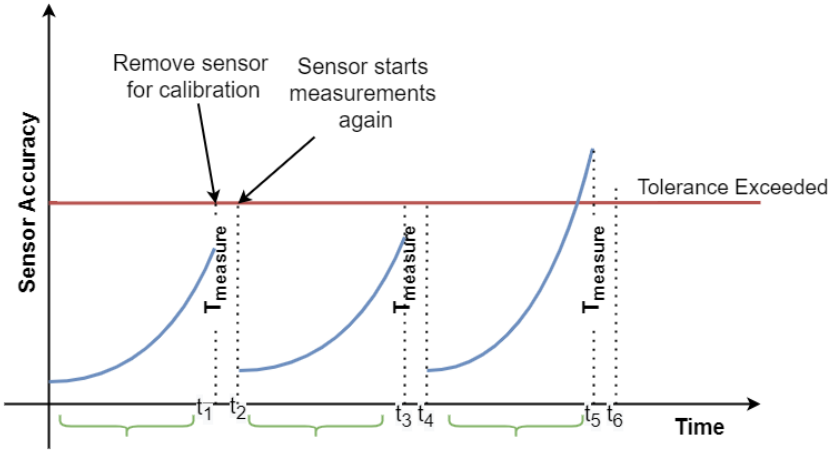



Figure 14 – Preventive Calibration by time.

Equipment that does not have specific calibration intervals must be examined at least every six months and calibrated at intervals of at least one year. Sensor calibrations should be done, where a minimum of 95% of equipment or standards of the same type must be within tolerance when subject to regularly scheduled recalibration. That is, if more than 5% of a specific kind of equipment is out of tolerance at the end of its range, the range should be reduced until less than 5% is defective.

The Predictive Calibration (PdC) strategy, which can be considered a subset of a correctly implemented Predictive Maintenance (PdM) strategy, can reduce unnecessary downtime for calibration and maintain sensor accuracy (Figure 15). In this type of approach, there are two action strategies: when periodic check is carried out with redundant instruments to determine the condition of the reading equipment (Figure 15); through online monitoring strategies, which is one of the focuses of this study (presented in more detail in the following sections). Periodic checks, together with the application of the necessary technical knowledge, management strategies, and decision-making skills, allowing to schedule a calibration through a more informed process. It is vital to establish a secondary Key Performance Indicators (KPIs) associated to the measurement method to indicate poor sensor accuracy performance. It is also necessary to define appropriate tolerances for the verifications carried out, thus stimulating more interventions [41].

According to Smith & Mobley [76] and Mobley [104], the check interval can be performed in terms of time (hourly, weekly, monthly), based on the amount of usage or each batch. It should be based on stability, purpose, and degree of use.

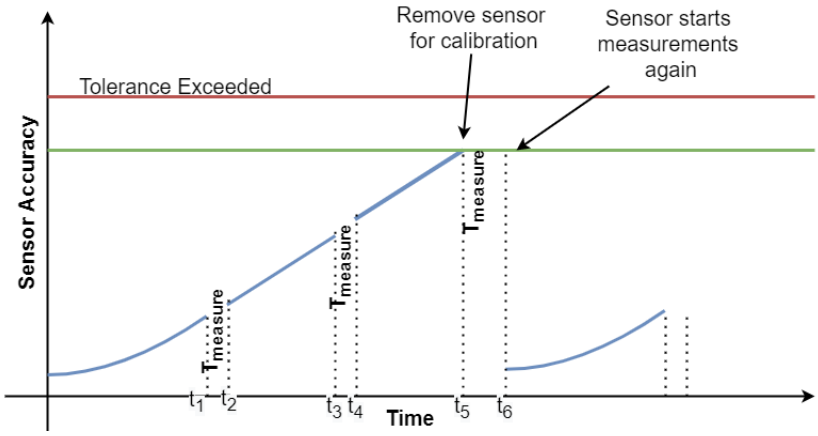



Figure 15 – Predictive calibration with periodic check.

The introduction of this strategy will bring new costs to the company. Therefore, it becomes essential that these are offset by optimizing operational efficiency and reducing the total cost of downtime related to the sensor's unexpected and unplanned quality problems. It will also allow for the reduction of the overall downtime of a manufacturing facility. It can eliminate the sensor as the root cause of any later failure in the production process [41].

Interaction of the instrument and its installation can generate false readings in several ways, so permanent instrument mounting should always allow for simple removal for calibration or built-in installations to permit calibration and reinstallation without the risk of disturbing any factors that may affect the readings [41].

## 2.4 Online monitoring to determinate the sensors calibration status

As we referred in previous sections, one of the predictive calibration strategies based on equipment conditions can be done through online monitoring. Since this is one of the main focuses of this thesis, this section will provide a general framework on the subject. In several industries, the strategy used for maintenance is traditionally based on the validation of sensors, involving their calibrations periodically. As we referred above, this causes the reading instrument to be turned off and removed to be calibrated, causing, in some cases, very high costs. To suppress these problems, some companies use faulty sensors, ignoring their data. This will cause wrong decisions to be made that can later

lead to more significant problems and, consequently, to unexpected and higher expenses than those previously avoided by calibrating or repairing the sensors.

Condition monitoring allows early detection of faults, as an essential issue in surveillance and diagnostics. This type of monitoring is called OnLine Monitoring (OLM) which, according to Hines *et al.* [42], can be implemented for instrument monitoring, equipment monitoring, or operation monitoring. However, the acronym OLM is commonly used to represent monitoring the calibration status of readout instruments.
Online monitoring systems, designed to monitor and diagnose sensors with online measurements during their measurements, were developed by researchers at Nuclear Power Plants (NPP). The Nuclear Regulatory Commission (NRC), from July 24, 2000, provided generic online monitoring approval to reduce process instrumentation calibration. In this way, it was verified that it is possible to extend the calibration periods of the sensors, what can improve the economic activity of the NPP's, with the value added about the failures that could occur during the calibration process, which can be avoided [42], [43], [115], [116].

According to Hoffmann [43], the OLM system offers an alternative approach to traditional time-driven calibration and opens the possibility of a calibration strategy based on sensor conditions, providing an assessment of instrument performance and a basis for determining when adjustments are needed. The elimination or reduction of unnecessary calibrations will commute to a decrease of associated labour costs, among others.

According to Hines [117], “On-line monitoring evaluates the deviation of an instrument with reference to its process parameter estimate as determined by one of the predictive algorithms”.

So, OLM, according to Boechat *et al.* [118], consists of estimating correct measurements that the sensors must have read and constantly monitoring the difference between the estimated values and the values read by the sensors. OLM is a non-invasive approach that provides a more frequent assessment of instrument calibration in the real operating environment. It has the potential to alleviate problems with current calibration practices, allowing the identification of sensors that have deviated from tolerance limits to direct calibration activities during interruptions [110]. It evaluates the instrument channel performance and, according to the experience of Hoffmann [43], in several factories, he demonstrates that this general approach is very effective in identifying instrument channels that have poor performance characteristics.

Schiff [113], Hines [117], and Hines & Seibert [119] explain that a typical OLM system collects data from the reading instruments and processes them on an off-line computer, which later makes a graphical representation of the individual deviation of the sensor channel from the estimation of the process channel as a function of time. The term online is used because data collection is done during the operation, not necessarily meaning that monitoring is carried out in real-time. OLM evaluates the deviation of a sensor concerning the estimation of process parameters, as determined by one of the predictive algorithms. Regardless of the algorithm used, the online monitoring technique will evaluate the sensor deviation about the estimated process parameters as determined by one of the monitoring systems. Then, to evaluate the sensor performance, the residual between the OLM model process estimates and the sensor output is used [119]. Most OLM software is automated, but this does not imply that human interaction is not required, as OLM predictions should never be trusted without a trained employee reviewing and interpreting the data [117]. OLM systems can also be used, not only to detect sensor deviations but also to detect anomalies or process failures, as the data collected has valuable information about factory operations and their dynamics [117]. So, the data collected in OLM systems can be used in reliability and production diagnostics, but this should not affect the acceptance of this system for sensor calibration extension.
To explain a basic monitoring system, we can use Figure 16 as an example. We have a vector of sensor measurements ($x$) inserted into a prediction model, calculating the best estimates of the sensors ($x'$). Then, these estimates are compared to the measured values, allowing the calculation of differences called residuals ($r$). Subsequently, a decision logic module determines whether the residuals are statistically different from zero, establishing the performance of each sensor. This module can also use predictive uncertainty values and deviation limits to assess the instrument channel condition [42], [120].

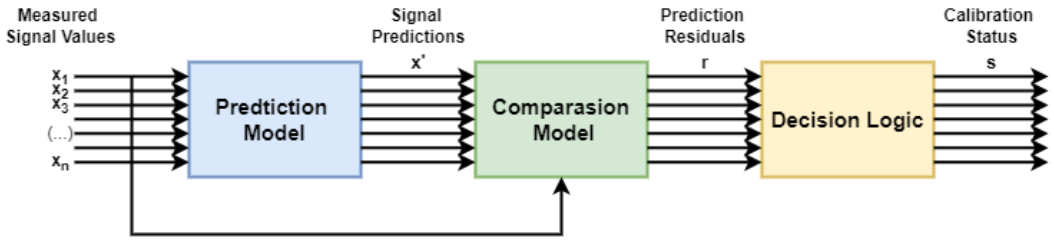



Figure 16 - Simple online monitoring system for OLM.

The difference between expected (modelled) and actual (measured) behaviour, called the residual, characterizes the deviations of the system from normal behaviour and can be used to determine whether the sensor or system is operating in an abnormal state [121].

Hines *et al.* [42] state that the OLM reduces the number of unnecessary calibrations and, consequently, reduces the interruption time. The OLM system can also provide faster problem discovery and allow for more timely and convenient corrective actions than traditional calibration. Periodic sensor calibration is expensive, time-consuming, and unnecessary maintenance actions can damage sensors. For example, when a sensor requires calibration, this can be skipped simply because the calibration interval has not passed, even if the sensor needs maintenance, which can lead to unexpected downtime and safety hazards. Another example, and a significant concern with periodic calibrations, is that performing maintenance on components that are working correctly can cause a failure when entering the system [42], [43]. On the other hand, OLM systems monitor the condition of the sensor channel, identifying those that have been degraded to the point of justifying their calibration. So, according to Hines *et al.* [42], online methods can help reduce maintenance costs, reduce the potential for calibration errors, increase instrument reliability and, consequently, reduce equipment downtime. To reinforce this idea, Hoffmann [43] refers that much of the conventional calibration effort is currently carried out in the verification of sensors that do not need maintenance and gives an example based on a case where it was verified that 646 sensors were calibrated and only three had deviated and required maintenance. Considering that an average calibration takes around 8-10 hours to perform, the potential for savings is clear, and a shorter refill interruption duration suggests that a different approach to instrument calibration is needed.

Dwivedi *et al.* [116] refer that "an essential component for quality control is a low cost recalibration system which can be applied on-line during the production cycles (…)."
According to Schiff [113], the advantages of online monitoring claimed by the Electric Power Research Institute in the topic report 104965 are as follows:

- Comparing to traditional calibration processes, online monitoring is non-intrusive, more frequent, and will result in a reduced number of field calibrations;
- Online monitoring can be ongoing and will identify calibration issues when they occur;
- Eliminating unnecessary field calibration, it will reduce associated labour costs, personnel radiation exposure, and the potential for calibration errors;
- Online monitoring allows instrument performance evaluation under normal operating operation, while routine calibration usually occurs during plant shutdown;

- By reducing the time spent on typical calibration, refuelling interruptions will be shorter;
- Long-term trends in instrument performance developed using online monitoring can be used for predictive maintenance planning.

An OLM system consists of several components, the most common being: an off-line computer, where there is a monitoring system; a communication hardware and software tools for collecting process data; a history of process data; an OLM software, responsible for performing the analysis and presenting the results of the sensor calibration performance [119]. Thus, the author explains that the first step in implementing online monitoring is installing, testing, and verifying the data acquisition system responsible for acquiring and storing the historical data files. The data acquisition system usually receives the instrument data in the form of a voltage output, which is converted to a digital signal. Data transfer between the data acquisition system and the OLM software can occur in batch mode or real-time. The term batch mode means that the data files are stored somewhere and accessed by online monitoring at discrete time intervals [119]. In Figure 17, we can verify the functioning of a generic OLM system.

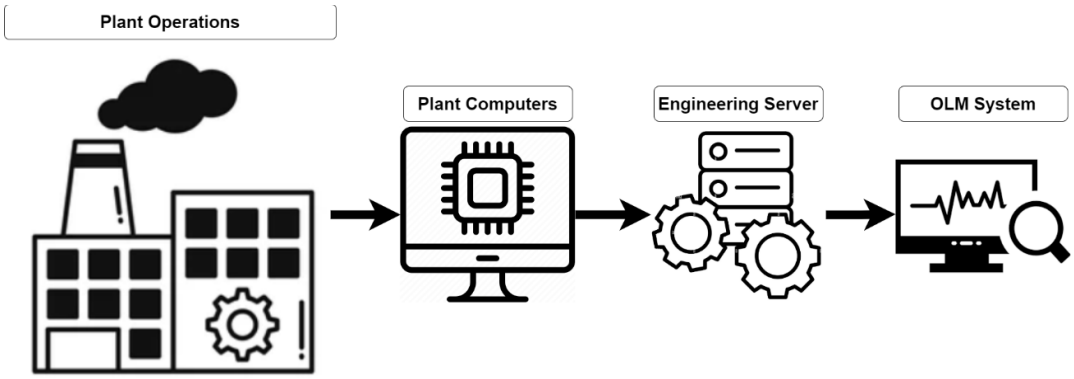



Figure 17 – OLM system setup.

According to Coble *et al.* [111] and Ramuhalli *et al.* [115], OLM typically involves two steps: modelling the expected detected values and evaluating the difference between the expected and actual behaviour for failures. In OLM, sensor calibration is assessed by comparing the measured data with the expected value of the sensor. The expected value can be measured using various models, including physics-based models, neural networks, non-parametric models, etc. [110]. Hines & Seibert [119] and Ramuhalli *et al.* [115] explain several empirical modelling techniques for online monitoring of instrument channel performance and say that these can be divided into two main categories: redundant and non-redundant. Hines & Seibert [119] refer that redundant modelling techniques only use measurements from a group of redundant instrument channels to

obtain parameter estimates, such as the simple average. On the other hand, non-redundant modelling techniques use a group of correlated but not truly redundant instrument channels to obtain the parameter estimation.

A comparison of the instrument readings is performed for redundant equipment to distinguish between process deviation and instrument deviation. According to Hines *et al.* [42], redundant modelling techniques are the ones that only use measurements from a group of redundant instrument channels to obtain the parameter estimation. The author explains that "redundant" describes instrument channels that measure the same process parameter in a similar operating range. Redundant techniques are generally considered more intuitive and easier to troubleshoot than non-redundant techniques. They cannot detect failures in common mode and can also be more affected by spillover, which is when a faulty sensor input occurs, what impairs the predictions of the other sensors in the group. Due to these problems and disadvantages of redundant models, non-redundant models have become the most used technique for OLM. Coble *et al.* [111] and Ramuhalli *et al.* [115] refer that redundant sensors can be paired with other related measurements, and the complete set can be monitored through non-redundant sensor modelling methods. Using redundant hardware to detect faulty sensors can be very expensive and not so useful when they oscillate in the same direction. Furthermore, Boechat *et al.* [118] give an example that these components in the oil industry can occupy valuable and limited space and consume precious energy. Empirical models, developed with historical data, are created by training the model. This, through the analysis of faultless training data, was obtained during normal operations.

For non-redundant instruments, we have the empirical modelling of the process. This estimate is updated frequently and compared with the output of the corresponding instruments to detect any deviation in the instrument's output. Coble *et al.* [111] and Ramuhalli *et al.* [115] also say that non-redundant sensors require sensors in a single model to contain related information (e.g., temperature and pressure of a gas), typically identified by linear correlations or a physical understanding of measurement relationships.

Several modelling methods are proposed to evaluate sensor performance of redundant and non-redundant sensor groups, which are presented by Coble *et al.* [111].

Empirical instrument channel calibration monitoring methods use historical data to build predictive models, which, after their completion, are put into monitoring mode to provide the best estimates of process variables for previously unseen data [119]. Hines

[117] tells that there must be a clear scheduling method and adequate documentation for an OLM system to succeed.

According to Hines & Seibert [119], for a modelling technique to be considered suitable for OLM, the model must:

- produce accurate results;
- produce repeatable and robust results;
- have a method for estimating the uncertainty of forecasts.

In this sequence, according to Hines & Seibert [119], a list of the basic steps for the development and implementation of the model is as follows:

1. Acquire "good" data
   - This first step must ensure that the data collected is carefully reviewed and that its quality is guaranteed.
2. Group sensors into ideal models
   - Here the sensors to be used in each model must be selected, and this choice is complicated, depending on whether sensors are used for redundant or non-redundant techniques.
3. Select training data
   - This step divides the data into training, verification, and validation datasets. The model uses the training datasets to learn the relationship among the sensors. The verification data optimize the model parameters to reduce predictive uncertainty. Validation data is used to quantify the model's performance measures.
4. Build and optimize predictive models
   - In this step, the models are optimized to minimize the predictive uncertainty, ensuring that they have a complexity corresponding to the complexity of the relationships to be modelled.
5. Evaluate the model
   - Here the model should be evaluated using criteria-based validation data. The following must be considered: accuracy, which is a measure of how well the model outputs correspond to the sensor data; robustness, which is a measure of how well a sensor prediction tracks the actual plant parameter when the sensor is drifting; overflow, which is a measure of how a drift sensor input affects the prediction of other sensor values; and predictive uncertainty of the model.

6. Uncertainty analysis
    - This step is performed after the model is developed and optimized, and thus its uncertainty needs to be quantified.
7. Transition to online mode
    - After the above six steps have been taken, the OLM system can be implemented in an online or batch monitoring mode. At this stage, it may be necessary to retrain the models with more up-to-date training data when changes in operational or environmental states are encountered.

A perfect model, as explained by Hines *et al.* [42], would be a model that would make sensor predictions, not be significantly affected by degraded inputs, and would be able to detect minor faults and anomalies in the sensor. But, as with any modelling paradigm, the predictions made have some level of uncertainty associated with them. Understanding and quantifying this uncertainty, it is an essential need in developing an OLM system for monitoring sensor performance [115].

The sources of uncertainty, according to Ramuhalli *et al.* [115] and Hines *et al.* [42], in OLM can be categorized as:

- process noise, which is the result of the normal fluctuation of physical process parameters (e.g., temperature, flow, pressure) over the true process value;
- measurement uncertainty, that occurs due to several factors, including sensor accuracy, calibration accuracy (e.g., calibration offset, error in converting sensor units to engineering units), environmental effects (due to temperature, vibration, pressure, etc.). These sources of uncertainty apply to the sensor;
- electronic noise, where the transmission of measurements through the instrumentation line can induce additional noise, along with the analogue to digital conversion at the computer input;
- modelling uncertainty, which arises from input uncertainty (related to process noise, measurement uncertainty, and electronic noise described above) and modelling error (resulting from model selection, model training, input selection, etc.).

The uncertainty inherent in model predictions affects the size of the flaw that can be reliably detected [42].

In short, we have the conclusions of the NRC on online monitoring systems [113]:

- The generic concept of an online monitoring technique is acceptable for online tracking of instrument performance;
- Online monitoring has several advantages, including timely detection of degraded instrumentation;
- Online monitoring can provide information about the direction in which instrument performance is going and, in this role, can be helpful in preventive maintenance activities;
- Although the proposed online monitoring technique, compared to the traditional calibration process, renders the results with less precision, it is considered acceptable that the precision provided by the estimation of the process parameters is sufficient to assess the operability of the instrument;
- Compared to the traditional calibration, the online monitoring technique offers greater assurance of instrument operability throughout the plant's operating cycle when used as a whole.

## 2.5 Hidden Markov Model - HMM

After being described the importance and necessity of using reliable data collected by sensors and how to guarantee it through OLM processes, next, it will be synthesised the Hidden Markov Model. This is the model used to classify the health state of equipment and to define the calibration status of the sensors used to collect data in CBM.

Markov Models are the result of the study carried on by the mathematician Andrei Andreyevich Markov. This scientist was born in Russia in 1856 and studied subjects associated with number theory, approximation theory, and continued fractions for a long time. At a certain stage of his life, Markov applied the method of continued fractions to probability theory, starting the study from which, he is known until today: Markov chains [122]. Hidden Markov Models (HMM) began to be developed in the late 60's and early 70's by Baum & Petrie [120], [123], [124]. Its application in word recognition began to be used in the mid-1970's by Baker [121], developed in his thesis to obtain a degree of Doctor of Philosophy in Speech and Computer Science. During last few years, HMM models have been used in several areas, ranging from speech recognition [125], language modelling, handwritten word recognition, online signature verification, human action learning, and detection of equipment failures. The study of Markov Models and their applications are very relevant for decision-making support systems. They start from a series of probabilities from which we can know which is the most likely choice or path that offers more significant guarantees when studying the projection of a given problem.

The Hidden Markov Model is a variant of the Markov chains. In addition to the probability distribution function associated with the states, there is a probability distribution function for the observations that can be perceivable in each state. It consists of a doubly stochastic process composed of a hidden (unobservable) process, which manifests itself through another stochastic process that produces the sequence of observed symbols. The two types of variables that a Hidden Markov Model is associated with are:

- Probabilities of emission of visible symbols;
- Hidden state transition probabilities.

These models have an excellent aptitude for solving problems based on the theory of stochastic processes, with numerous applications in various fields of scientific research, such as cell biology, bioinformatics, artificial intelligence, economics, meteorology, etc. There are three problems related to the Hidden Markov Models, which have solutions through the Forward, Backward, Viterbi, and Baum Welch algorithms (these algorithms will be detailed in the following section).

Explaining, in general, the processes in the real-world, it produces signals or sequences of observations, which can be modelled using deterministic or statistical techniques, where statistical models, such as the HMM, only seek to characterize the statistical properties of the signal. In a regular Markov model, the health status of a system is directly visible to the observer, and, consequently, the transition probabilities between each one are easily determined. In HMM models, the state is not now visible, but the variables influenced by the state are visible or measurable. For example, in production equipment working in a state of good functioning, the observable conditions measured will be different from the observable states measured for moments of its malfunction. Thus, each state presents a probability distribution on the possible observable outcomes produced. Consequently, once the sequence of results produced is known, the HMM can yield information regarding the sequence of hidden states. Since the state is not directly visible, it corresponds to an unknown variable. Then, this approach considers that each hidden state, represented by unobservable events, is correlated with observable events through a probability distribution of occurrence [122]. The challenge lies in determining the value of hidden variables from knowing the value of visible variables.

One of the objectives of this type of study is to characterize real-world signals through signal models collected by sensors, which can be modelled through predictive models that explain the respective condition of the system or equipment. Models that describe the behaviour of signals (emissions) can provide the basis for a theoretical description of

how the calibration states of a sensor can be found, allowing effective management decisions to be made about the calibration of these devices. HMM incorporate a dual stochastic process, with a stochastic process not visible, as it is not directly measurable (hence, the designation of "hidden"), but which can be correlated with indicators corresponding to another stochastic process, which produces a sequence of observations. The hidden processes consist of a group of states interconnected by transition probabilities between them. On the other hand, observable processes consist of a group of possible outputs or observations interconnected in a probabilistic way of the hidden states.

### 2.5.1 Elements of a Hidden Markov Model

An HMM translates to a more complex problem than the problem described by Markov chains. Each hidden state corresponds to an unobservable event; each hidden state is linked to a group of possible observable events, according to a probability distribution function. It becomes necessary to define a more general statistical model, the Bayesian model [122]:

- A set of random variables, $X = (W, Y)$, where the set $W$ corresponds to hidden variables and $Y$ represents a set of observed signals that describes events or attributes correlated with $W$;
- An HMM model is characterized by the parameters $\lambda = \{A, B, \pi\}$, corresponding to the stochastic model that best describes the class of symptoms of pathologies that are sought to be decoded, which are conditions that occur in the domains of the hidden variables $W$.

An HMM model is represented by the following notations [122], [124], [125], [126]:

- $N$ – Number of elements in the system's hidden state set. It translates the finite set of possible individual states identified by $Q = \{q_1; q_2; q_3; \dots; q_N\}$, which is supported in classes defined in the domains of variables $W$. The state at instant $t$ is specified by $S_t$, so, the sequence of states at the instants $t = 1, t = 2, t = 3, \dots, t = T$ will be $S = \{S_1; S_2; S_3; \dots; S_T\}$;
- $M$ - Number associated with the distinct observation classes for the different states. Individual classes symbols are specified by the set $V = \{V_1; V_2; V_3; \dots; V_M\}$ and are supported in classes defined in the domains of the variables $Y$. The sequence of symbols or observations at the instants $t = 1, t = 2, t = 3, \dots, t = T$ will be $O = \{O_1; O_2; O_3; \dots; O_T\}$;

- $A$ - The probability distribution of the state transition is represented by $A = \{a_{ij}\}$, where $a_{ij} = P[S_{t+1} = q_j | S_t = q_i], 1 \leq i, j \leq N$ is an array of non-negative real numbers, between $0$ and $1$, indexed by $N * N$. This comes from one or several periodic sequences of characterization of health states for one or several elements of the same series of an equipment. According to Simões [122], when a complete background of the same equipment or system is not known, covering a time interval of sufficient dimension to guarantee the monitoring of the condition of all the different phases of the life cycle, the option is the monitoring of different equipment at different stages of its life cycle. For new equipment, life cycle curves of equipment or systems with greater technical and functional similarities can be used. The parameters $A$ can be estimated through the following expression:

$$a_{ij} = P[S_{t+1} = q_j | S_t = q_i] = \frac{nº\ of\ transitions\ of\ q_i\ to\ q_j}{nº\ of\ time\ in\ state\ q_i} \qquad \text{Eq. (1)}$$

- $B$ - The probability distribution for the set of emission classes or combinations of emission observations, coded by $B = \{b_j(k)\}$, defines the probabilities with which the different emission classes occur for each of the hidden states $j$, with $j = 1,2,3, \dots, N$, where $b_j(k) = P[O_t = V_k | S_t = q_j], 1 \leq k \leq M$. It is a matrix of non-negative real numbers, between $0$ and $1$, indexed by $M * N$. This matrix comes from one or several periodic sequences to characterize the classes or combinations of emissions for one or several elements of the same series of an asset or equipment. Likewise, Simões [122] explains that when a complete background of the same equipment or system is not known, covering a time interval of sufficient dimension to guarantee the monitoring of all classes or combinations for the different phases of the life cycle, the option is to monitor different equipment at different stages of its life cycle. The calculation of the emission parameters is done through the following expression:

$$b_j(k) = \frac{nº\ of\ times\ in\ state\ q_j\ with\ the\ symble\ v_k}{nº\ of\ times\ in\ state\ q_j} \qquad \text{Eq. (2)}$$

- $\pi$ - The initial state distribution is represented by $\pi = \{\pi_i\}$ equivalent to $\pi = \{\pi_1, \pi_2, \pi_3, \dots, \pi_N\}$, where $\pi_i = P[S_1 = q_i], 1 \leq i \leq N$ is, by consequence, the vector of the initial probability distribution of the different states. It represents the occurrence probability of each state in the initial instant of a new chronological sequence that is intended to be known. It translates the estimated probability of the class of states $q_i$ before starting a new sequence. The expression determines the values of $\pi_i$:

$$\pi_i = Number\ of\ times\ in\ state\ i\ in\ instant\ t = 1 \qquad \text{Eq. (3)}$$

For convenience, the compact notation, $\lambda = (N, M, A, B, \pi)$ or, in a more reduced form, $\lambda = (A, B, \pi)$ is used to specify the complete set of model variables.

The characterization of the classes' hidden states and observable states, as well as the parameters of the matrices associated with them, can be intuitively analysed through Figure 18.

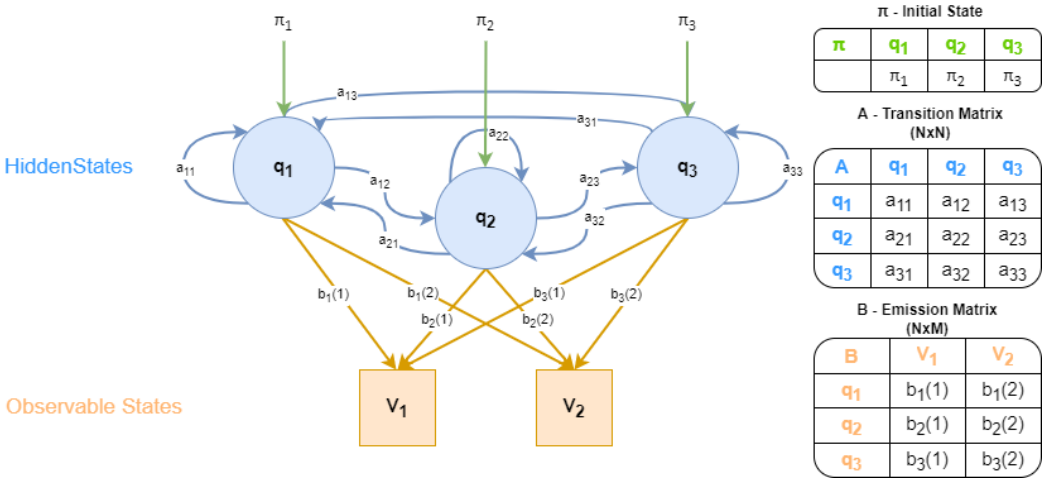
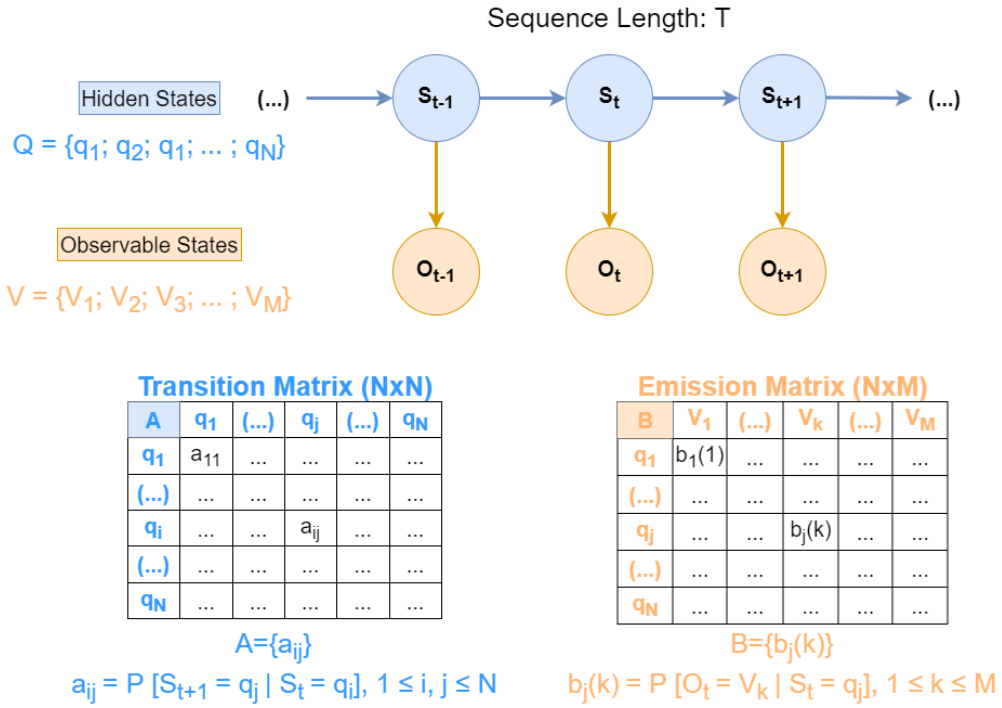



Figure 18 – Nomenclature of a HMM with 3 hidden states and two observables states.

The characterization of $N$ hidden states and $M$ observable states represented along a sequence of times $t = 1, t = 2, t = 3, \ldots, t = T$ can be represented by Figure 19.




| A | $q_1$ | (...) | $q_j$ | (...) | $q_N$ |
|---|---|---|---|---|---|
| $q_1$ | $a_{11}$ | ... | ... | ... | ... |
| (...) | ... | ... | ... | ... | ... |
| $q_i$ | ... | ... | $a_{ij}$ | ... | ... |
| (...) | ... | ... | ... | ... | ... |
| $q_N$ | ... | ... | ... | ... | ... |

| B | $V_1$ | (...) | $V_k$ | (...) | $V_M$ |
|---|---|---|---|---|---|
| $q_1$ | $b_1(1)$ | ... | ... | ... | ... |
| (...) | ... | ... | ... | ... | ... |
| $q_j$ | ... | ... | $b_j(k)$ | ... | ... |
| (...) | ... | ... | ... | ... | ... |
| $q_N$ | ... | ... | ... | ... | ... |

Figure 19 - N hidden states and M observable states represented along a sequence of times $t = 1, t = 2, t = 3, \dots, t = T$.

### 2.5.2 The three problems solved by HMM model

Three fundamental problems must be solved, so that an HMM can be used in real-world applications, as in the case under study:

- Problem 1 - It is an evaluation problem; given a sequence of observations $O = \{O_1; O_2; O_3; \dots; O_T\}$ and the model $\lambda = (A, B, \pi)$, how to efficiently calculate the associated probability to the emission sequence $P(O|\lambda)$?
- Problem 2 - It is a decoding problem, which consists of finding the most likely sequence of hidden states, given the sequence of observed emissions $O = \{O_1; O_2; O_3; \dots; O_T\}$ and the $\lambda$ model. That is, how to find a corresponding sequence of states $S = \{S_1; S_2; S_3; \dots; S_T\}$?
- Problem 3 – It is a model calibration problem, that seeks to maximize the likelihood, given that the states have already occurred. Given a sequence of emissions, or even a set of such sequences, how can the strongest correlation be found between the set of transition states and the measured observations of emissions? In other words, with a database of emission sequences known, how to adjust the model parameters $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$?

### 2.5.3 Problem 1

The evaluation problem referred to as problem 1 can be seen from how a given model corresponds and adjusts to a given sequence of observations of emissions. The most direct way to calculate the probability that a sequence of observations $O = \{O_1; O_2; O_3; \dots; O_T\}$ will fit a given model $\lambda = (A, B, \pi)$ is by enumerating all possible sequences of states of size T (the number of inspection instants), through the following expression:

$$P(O|\lambda) = \sum_{S_1, S_2, \dots, S_T} \pi_{S_1} b_{S_1}(O_1) a_{S_1 S_2} b_{S_2}(O_2) a_{S_2 S_3} b_{S_3}(O_3) \dots a_{S_{T-t} S_T} b_{S_T}(O_T) \quad \text{Eq. (4)}$$

The expression involves $2T * N^T - 1$ calculations, making its application unfeasible even with powerful computing resources. For $N = 5$ and $T = 43$, the calculations involve about 1032 multiplications. For this reason, the forward-backward procedure is used [124], [125]. The forward variable, $\alpha_t(i)$, corresponds to the probability of occurrence of a sequence of emissions $O_1; O_2; O_3; \dots; O_t$ (input signals to be recognized) associated with a symbol library, $V_1; V_2; V_3; \dots; V_M$ that, at time $t$, have state $i$, given the model $\lambda$, where:

$$\alpha_t(i) = P(O_1; O_2; O_3; \dots; O_t, S_t = q_i|\lambda) \quad \text{Eq. (5)}$$

Consequently, it results:

$$\alpha_T(i) = P(O_1; O_2; O_3; \dots; O_T, S_T = q_i|\lambda) \quad \text{Eq. (6)}$$

In this model, the emission observations correspond to symptoms that, in turn, are associated with the probable values of the status indicators. This correlation makes it possible to calibrate the model and generate the probabilistic function of states that estimates the symptom–state mapping $p(S_t = q_i|O_t)$ at the inspection instants.

The algorithm initialization is done randomly. After implementing the model, assuming conditional dependence assumptions, $p(S_t = q_i|O_1; O_2; O_3; \dots; O_T)$, it can be estimated without explicitly providing the values of the indicators $O_1; O_2; O_3; \dots; O_T$ as direct entries in the classification model, as they are embedded in the matrices.

#### 2.5.3.1 Algorithm forward

In general, it is possible to estimate and assess the probability of occurrence of the emission data observed up to instant $T + 1$, based on the knowledge of the HMM implementation parameters, previously calculated and assumed to be unchanged at

instant 1. Such parameters are designated generically by $\lambda$. Such probability can be calculated through the expression:

$$P(O|\lambda) = p[(O(T+1), O(T), O(T-\Delta t), \dots, O(1))|\lambda] \qquad \text{Eq. (7)}$$

For this calculation, it is convenient to work with an intermediate variable $\alpha$, called the Forward variable, where, $\alpha_{T+1}(j) = p(S_{T+1} = q_j, \lambda)$.

To find the corresponding *a posteriori* probability, given a given model, $\lambda$, it is sufficient to calculate the $\alpha_s$ for any instant t by applying Bayes' rule:

$$p(S_{t+1} = q_j|\lambda) = \frac{p(S_{t+1} = q_j, \lambda)}{p(\lambda)} = \frac{\alpha_{t+1}(j)}{p(\lambda)} = \frac{\alpha_{t+1}(j)}{\sum_{j=1}^{N} \alpha_{t+1}(j)} \qquad \text{Eq. (8)}$$

The estimate of $\alpha_{t+1}(j)$ is obtained recursively. The deduction has some constraints, since $P[AB] = P[A].P[B]$ only holds when A and B are independent.

$$\alpha_{t+1}(j) = P\big(O_1, \dots, O_t, O_{t+1}, S_{t+1} = q_j\big|\lambda\big) =$$

$$\sum_{i=1}^{N} P\big[O_1, \dots, O_t, O_{t+1}, S_{t+1} = q_j\big|S_t = q_i, \lambda\big].P[S_t = q_i|\lambda] =$$

$$\sum_{i=1}^{N} P[O_1, \dots, O_t|S_t = q_i, \lambda].P[S_t = q_i|\lambda].P[O_{t+1}, S_{t+1} = q_j|S_t = q_i, \lambda] =$$

$$\sum_{i=1}^{N} P[O_1, \dots, O_t, S_t = q_i|\lambda].P\big[O_{t+1}\big|S_{t+1} = q_j, S_t = q_i, \lambda\big].P[S_{t+1} = q_i|S_t = q_i, \lambda] =$$

$$b_j(O_{t+1})[\sum_{i=1}^{N} \alpha_t(i)a_{ij}]$$

There is another way to reach the same algorithm, as is presented below.

$$\alpha_{t+1}(j) = \sum_{i=1}^{N} p\big(S_{t+1} = q_j, \lambda, S_t = q_i\big)$$

$$= \sum_{i=1}^{N} p\big(S_{t+1} = q_j, O_{t+1}, \lambda, S_t = q_i\big)$$

$$= \sum_{i=1}^{N} p\big(S_{t+1} = q_j, O_{t+1}|(\lambda, S_t = q_i)\big)p(\lambda, S_t = q_i)$$

Considering the definition of $\alpha$:

$$\alpha_{t+1}(j) = \sum_{i=1}^{N} (p\left(S_{t+1} = q_j, O_{t+1} \middle| (\lambda, S_t = q_i)\right)) \alpha_t(i)$$

$$= \sum_{i=1}^{N} \left(p(O_{t+1}|S_{t+1} = q_j, S_t = q_i) p\left(S_{t+1} = q_j \middle| (\lambda, S_t = q_i)\right)\right) \alpha_t(i)$$

Assuming that $O_{t+1}$ is independent of past observations and previous states, but only of the present state, we have:

$$\alpha_{t+1}(j) = \sum_{i=1}^{N} p(O_{t+1}|S_{t+1} = q_j) p(S_{t+1} = q_j|(\lambda, S_t = q_i))) \alpha_i(t)$$

$$\alpha_{t+1}(j) = \sum_{i=1}^{N} p(O_{t+1}|S_{t+1} = q_j) p(S_{t+1} = q_j|S_t = q_i)) \alpha_i(t)$$

Assuming that $S_{t+1}$ is independent of past observations, but only of the previous state $S_t$, will come:

$$\alpha_{t+1}(j) = p(O_{t+1}|S_{t+1} = q_j) \sum_{i=1}^{N} a_{ij} \alpha_t(i) \qquad \text{Eq. (9)}$$

This equation establishes a basic recursive relationship to estimate the probabilities of a state occurring at any time $t + 1$. From this equation, the most practical recursive estimate is deduced as follows: First, the term $p(O_{t+1}|S_{t+1} = q_j)$, which, as mentioned before, is represented by $b_j(O_{t+1})$ is replaced by $p(S_{t+1} = q_j|O_{t+1})|p(S_t = q_i)$.

The summation term of Equation 9 is a linear combination of the $\alpha_s$ from the previous model. The terms $\alpha_t(i)$ are replaced by $p(S_t = q_i|\lambda)$.

These two substitutions are equivalent, dividing both sides of the equation by $p(\lambda)$, which leads to the following equivalent recursive relation:

$$p(S_t = q_i|\lambda) = \frac{p(S_t = q_i|O_{t+1}).}{p(S_{t+1} = q_j)} \sum_{i=1}^{N} a_{ij} p(S_t = q_i|\lambda) \qquad \text{Eq. (10)}$$

Thus, in the determination of $\alpha_t(i)$, the following inductive expressions must be used:

1. Initialization:

$$\alpha_1(i) = \pi_i b_i(O_1), 1 \leq i \leq N \qquad \text{Eq. (11)}$$

2. Induction: $$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_i(i)\alpha_{ij}\right] b_j(O_{t+1}), 1 \leq t \leq T-1 \; e \; 1 \leq j \leq N \quad \text{Eq. (12)}$$

3. Finishing: $$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \quad \text{Eq. (13)}$$

This procedure drastically reduces the number of calculations required. This method of calculating $P(O/\lambda)$ involves $N^2T$ multiplications. For the values $N = 5$ and $T = 25$, the calculations involve 625 multiplications.

### 2.5.3.2 Algorithm backward

Backward algorithm is defined by the variable $\beta_t(i)$ as the probability of occurrence of the sequence of classes of observations, from the instant $t + 1$ to the instant $T$, given the occurrence of the state $q_i$ at the instant $t$ given the model $\lambda$, where:

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, O_{t+3}, \dots, O_T | S_t = q_i, \lambda) \quad \text{Eq. (14)}$$

Again, a solution for $\beta_t(i)$ can be obtained using the following inductive (recursive) expressions:

1. Initialization: $$\beta_t(i) = 1, 1 \leq i \leq N \quad \text{Eq. (15)}$$

2. Induction: $$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), t = T-1, T-2, \dots 1 \; e \; 1 \leq i \leq N \quad \text{Eq. (16)}$$

3. Finalization: $$P(O|\lambda) = \sum_{i=1}^{N} \pi_i b_i(O_1) \beta_1(i) \quad \text{Eq. (17)}$$

Also, this method of calculating $P(O|\lambda)$ involves $N^2T$ multiplications.

### 2.5.4 Problem 2

In solving problem 2, the algorithm will determine the sequence of states $S = (S_1, S_2, S_3, \dots, S_T)$, so that the probability $P(S|O, \lambda)$ is maximized. It thus allows the estimation of the most probable sequence of states, i.e., the $máx\{p(S(T) = q_j, \dots, S(1) = q_i|\lambda)\}$. Problem 2 seeks a more expeditious way to discover the hidden part of the model, that is, to find the correct sequence of hidden states. This problem is normally solved using an optimal procedure. The Viterbi algorithm, based on dynamic programming, looks for the best sequence of states $S = (S_1, S_2, S_3, \dots, S_T)$ for a given sequence of observations, $O = (O_1, O_2, O_3, \dots, O_T)$, given the model $\lambda = \{A, B, \pi\}$.

### 2.5.4.1 Algorithm Viterbi

Corresponds to finding $S^* = \arg\max_S P(S|O,\lambda)$, since it turns out that P $P(S|O,\lambda) = \frac{P(S,O|\lambda)}{P(O|\lambda)}$ and since $P(O|\lambda)$ is constant concerning a sequence of states S, it is enough to calculate $S^* = \arg\max_S P(S|O,\lambda)$. This is because it is from the sequence of observations (fixed) that the sequence of states is intended to be calculated.

For this purpose, the auxiliary variable $\delta_t(i)$ is defined as:

$$\delta_t(i) = \max_{S_1,S_2,S_{3,\dots,}S_{t-1}} P(S_1, S_2, S_3, \dots, S_{t-1}, S_t = q_i, \quad O_1, O_2, O_3, \dots, O_t|\lambda) \qquad \text{Eq. (18)}$$

$\delta_t(i)$ corresponds to the maximum result, agreeing to the sequence of recorded observations, given that at the time $t$, at the end of the sequence, the state coincides with $q_i$.

By induction, we have:

$$\delta_{t+1}(j) = [\max_i \delta_t(i)a_{ij}]b_j(O_{t+1}) \qquad \text{Eq. (19)}$$

When predicting the sequence of states, it is necessary to extract the arguments that maximize the previous expression for each $t$ and $j$, creating the vector $\psi_t(j)$. In summary, the complete procedure to find the best sequence of states is as follows:

The Initialization algorithm corresponds to the starting state, $S_1$:

$$\delta_1 = P[S_1 = q_i, O_1|\lambda] = \pi_i b_i(O_1), 1 \le i \le N \qquad \text{Eq. (20)}$$
$$\psi_1(i) = 0$$

The index 1 corresponds to the observation at instant $t = 1$.

The Recursion to calculate the states corresponding to the sequence $O = \{O_1, O_2, O_3, \dots, O_T\}$, identified in moments following the evaluation of the present or initial state, corresponds to the instant $t = 1$:

$$\delta_t(j) = \max_{1\le i\le N}\left[\delta_{t-1}(i)a_{ij}\right]b_j(O_t), 2 \le t \le T \; e \; 1 \le j \le N \qquad \text{Eq. (21)}$$
$$\psi_t(j) = arg \max_{1\le i\le N}\left[\delta_{t-1}(i)a_{ij}\right], 2 \le t \le T \; e \; 1 \le j \le N \qquad \text{Eq. (22)}$$

The Finalization, at the end the value of $\max_{1\le i\le N} P[S, O|\lambda]$ can be calculated through the following expression:

$$P^* = \max_{1\le i\le N} P[S, O|\lambda] = \max_{1\le i\le N} \delta_T(q_i) = \max_{1\le i\le N} \delta_T(i) \qquad \text{Eq. (23)}$$

This results in knowledge of the most likely state at the end of the sequence (instant $T$).

$$S_T^* = \max_{1\le i\le N} \delta_T(q_i) \qquad \text{Eq. (24)}$$

The path or sequence of states, using an end to the beginning assessment (backtracking), taking the reverse-path or back.

$$S_t^* = \psi_{t+1}(S^*_{t+1}), t = T-1, T-2, \dots, 1 \quad \text{Eq. (25)}$$

Or,

$$S_t^* = arg \max_{1 \le i \le N} [\delta_t(i) a_{iS^*_{t+1}}] \quad \text{Eq. (26)}$$

Generate the sequence, $S_1, S_2, S_3, \dots, S_{t-1}, S_T$.

The Viterbi algorithm and the forward procedure have the same implementation except for the backtracking step. The only difference between them lies in the fact that the maximization in the Viterbi algorithm replaces the sum of the forward process.

### 2.5.5 Problem 3

Problem 3 is the most difficult to solve. The answer is to define a method capable of adjusting the model parameters, $\lambda = \{A, B, \pi\}$, to satisfy a specific criterion of optimization. The criterion used maximizes the probability $P = (O|\lambda)$. The calibration of a stochastic model is not, in general, a simple problem. The existence of an efficient algorithm for this problem is a fundamental condition for the applicability of this statistical model [43]. The sequence of observations used to adjust the model parameters is called the training sequence; therefore, it trains the HMM. There is no known analytical process for finding the set of model parameters that maximize the probability of the sequence of observations. However, we can choose the model $\lambda = \{A, B, \pi\}$ to calibrate it so, the probability $P = (O|\lambda)$ is locally maximized. For this purpose, an iterative procedure is used, such as the Baum-Welch method, a particular case of the EM (Expectation-Maximization) algorithm. The algorithm's $E$ (Expectation) component corresponds to calculating the expected value mathematically of the hidden variables, given $\lambda$. The $M$ component (Maximization) corresponds to calculating $\lambda^*$ so, $P(O|\lambda^*)$ is maximized, assuming the values of the hidden variables in step $E$. It translates into an iterative process that ends with a new value of parameters in the $\lambda$ model.

Therefore, the training of the HMM is done with the Baum-Welch algorithm, which gives the optimal parameters, $\lambda^*$. The model after calibration corresponds to $\lambda^* = arg \max_{\lambda} P(L|\lambda)$, where $L$ corresponds to the sequence of measured emissions, which is the model associated to a particular equipment health status.

#### 2.5.5.1 Algorithm Baum-Welch

It integrates several steps:

1. Step $E$ (Expectation) - Step $E$ of the EM algorithm, solved by an HMM, that corresponds to the calculation or prediction of the expected mathematical values and respective sequences for each model variable. It includes the following projections: number of times the model detects the different classes of states, number of times the model detects transitions from state $i$ to state $j$, number of times in state $i$ that the model associates to the emission symbol $v_k$, and number of times the model starts the algorithm from state $i$, that is, the number of times the model assigns state $i$ at instant $t = 1$ to each one of the time series.
2. Step $M$ (Maximization) - Calculates the new parameters $\hat{\lambda}$ so, $P(O|\hat{\lambda})$ is maximized, starting from the values of the hidden variables defined in step E, and compares them with the real values. When such hidden state values (obtained in step $E$) are known, calculating the model parameters $\lambda$ so, $P = (O|\lambda)$ is maximum, what becomes simple.
3. Iteration - If $P(O|\hat{\lambda}) - P(O|\lambda)$ is greater than a fixed value, then go to step $E$ (with the new parameters $\hat{\lambda}$); if not, terminate the execution and resume the new parameters $\hat{\lambda}$.

It can be proved that $P(O|\hat{\lambda}) \geq P(O|\lambda)$ and the sequence of models $\hat{\lambda}_i$ obtained with the EM algorithm converges to $\lambda^*$, a local maximum of variable $arg \max_{\lambda} P(O|\lambda)$. The Baum-Welch algorithm is fast and, generally gives a good local maximum. The constraints imposed on the model calibration lead to the imposition of several iterations while the difference between two successive iterations is greater than $10^5$. Alternatively, as the Baum-Welch algorithm only guarantees local maximization, it is usual practice to make five attempts to select the one that gives the highest probability value [122].

In summary, the algorithm aims to answer the following question:

- Which are the parameters of an HMM, $\lambda = (A, B, \pi)$, that maximize $P = (O|\lambda)$? That is, which $arg \max_{\lambda} P(O|\lambda)$?

It corresponds to a continuous optimization problem of $N^2 + NM + N$ parameters. This is the number of free parameters of the Markov model corresponding to a $N^2$ matrix A, plus $NM$ referring to matrix $B$ and $N$ corresponding to vector $\pi$. Therefore, if $N$ is too large, the determination with a high degree of precision and reliability of the matrix $A$

and $B$ parameters can become complicated when the number of training data is reduced [122].

To describe the procedure for re-estimating the HMM parameters, at the level of iterative updating with successive improvements in prediction efficiency, two auxiliary variables are defined:

- $\gamma_t(i)$ – Probability of the system to be in state $q_i$ at time $t$, when the entire sequence of observations and the model are known.

$$\gamma_t(i) = P(S_t = q_i|O, \lambda) \quad \text{Eq. (27)}$$

- $\xi_t(i,j)$ – Probability of the system to be in state $q_i$ at time $t$ and state $q_j$ at time $t+1$, given the model and the sequence of observed emissions.

$$\xi_t(i,j) = P(S_t = q_i, S_t = q_j|O, \lambda) \quad \text{Eq. (28)}$$

Note that $\sum_{i=1}^{T-1} \gamma_t(i)$ corresponds to the expected number of times in state $q_i$ or the expected number of transitions from state $q_i$ and $\sum_{i=1}^{T-1} \xi_t(i,j)$ represents the expected number of transitions from state $q_i$ to state $q_j$. Under these conditions, it becomes possible to relate $\gamma_t(i)$ and $\xi_t(i,j)$ through the sum of the values referring to the successive $j$, resulting in:

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j)\ t = 1 \dots T \quad \text{Eq. (29)}$$

Therefore, the method of calculating the HMM re-estimation parameters, referenced by the $\pi$, $A$, and $B$ matrices of the HMM, which are translated by the Baum-welch formulas, is made using the following expressions:

- Expected frequency or number of times spent in state $q_i$ at the instant $(t = 1)$ of the new sequence of application of the model:

$$\hat{\pi}_i = \gamma_t(i) \quad \text{Eq. (30)}$$

- Number of transitions from state $q_i$ to state $q_J$ divided by the expected number of transitions from state $q_i$:

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i,j)} \quad \text{Eq. (31)}$$

- Expected number of times that the system is in state $q_j$ and, at the same time, the symbol $v_k$ is divided by the expected number of times in which state $q_j$ is verified:

$$\hat{b}_{ij} = \frac{\sum_{t=1}^{T} \gamma_t(j) \; for \; O_t = v_k}{\sum_{t=1}^{T} \gamma_t(j)} \qquad \text{Eq. (32)}$$

The expression $\gamma_t(i)$ can be worked from the HMM parameters and from the forward and backward variables.

$$\gamma_t(i) = P(S_t = q_i|O, \lambda) = \frac{P[S_t = q_i, O|\lambda]}{P[O|\lambda]} \qquad \text{Eq. (33)}$$

Given the laws of probability:

$$P[S_t = q_i, O|\lambda] = P[O|S_t = q_i, \lambda].\, P[S_t = q_i|\lambda] =$$

$$P[O_1, O_2, \dots, O_t, O_{t+1}, \dots, O_T|S_t = q_i, \lambda].\, P[S_t = q_i|\lambda] =$$

$$P[O_1, O_2, \dots, O_t|S_t = q_i, \lambda].\, P[S_t = q_i, \lambda].\, P[S_t = O_{t+1}, \dots, O_T|S_t = q_i, \lambda] =$$

$$P[O_1, O_2, \dots, O_t, S_t = q_i, \lambda|\lambda].\, P[O_{t+1}, \dots, O_T|S_t = q_i, \lambda] = S_t = \alpha_t(i).\, S_t = \beta_t(i)$$

Since the sequence $O_1, O_2, \dots, O_t$ is independent of the sequence $O_{t+1}, \dots, O_T$.
In this way, the variable $\gamma_t(i)$ can be written from the HMM parameters and the forward and backward variables in the form:

$$\gamma_t(i) = P(S_t = q_i|O, \lambda) = \frac{P[S_t = q_i, O|\lambda]}{P[O|\lambda]} = \frac{\beta_t(i)\alpha_t(i)}{P[O|\lambda]} \qquad \text{Eq. (34)}$$

Also, $\xi_t(i, j)$ expression can be worked from the HMM parameters and the forward and backward variables.

$$\xi_t(i, j) = P\big(S_t = q_i, S_t = q_j\big|O, \lambda\big) = \frac{P[S_t = q_i, S_{t+1} = q_j, O|\lambda]}{P[O|\lambda]} = \qquad \text{Eq. (35)}$$

Complying with the laws of probability:

$$P\big[S_t = q_i, S_{t+1} = q_j, O\big|\lambda\big] = P\big[S_{t+1} = q_j, O\big|S_t = q_i, \lambda\big].\, P[S_t = q_i|\lambda] =$$

$$= P\big[O_1, O_2, \dots, O_t, O_{t+1}, \dots, O_T, S_{t+1} = q_j\big|S_t = q_i, \lambda\big].\, P[S_t = q_i|\lambda] =$$

$$= P[O_1, O_2, \dots, O_t|S_t = q_i, \lambda].\, P[S_t = q_i|\lambda].\, P\big[O_{t+1}, \dots, O_T, S_{t+1} = q_j\big|S_t = q_i, \lambda\big] =$$

$= \alpha_t(i).P\left[O_{t+2},\dots,O_T\middle|O_{t+1},S_{t+1}=q_j,S_t=q_i,\lambda\right].P[O_{t+1}|S_{t+1}=S_t=q_i,\lambda].P\left[S_{t+1}=q_j\middle|S_t=q_i,\lambda\right]$
$=$

$$= \alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)$$

Since $P[ABC|D] = P[A|BCD].P[B|CD].P[C|D]$, it follows that $O_{t+2},\dots,O_T$ is independent of $O_{t+1}$ and of $S_t$ and the even that $O_{t+1}$ is independent of $S_t$.
Consequently,

$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O|\lambda)} \qquad \text{Eq. (36)}$$

Finally:

$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{1=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)} \qquad \text{Eq. (37)}$$

Where the numerator term corresponds precisely to $P(S_t = q_i, S_{t+1} = q_j, O|\lambda)$ and division by $P(O|\lambda)$ allows obtaining the desired probability measure.

Note further that $\sum_{t=1}^{T-1}\gamma_t(i)$ encompasses the expected number of transitions from $q_i$ and $\sum_{t=1}^{T-1}\xi_t(i,j)$ fits the expected number of transitions from $q_i$ to $q_j$.

It is also possible to characterize this variable through:

$$\gamma_t(i) = \sum_{t=1}^{T-1}\xi_t(i,j) \qquad \text{Eq. (38)}$$

From the deductions presented, the final formulas of the Baum-welch algorithm result:

$$\hat{\pi}_i = \frac{\alpha_1(i).\beta_1(i)}{P[O|\lambda]} \qquad \text{Eq. (39)}$$

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1}\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{t=1}^{T-1}\alpha_t(i)\beta_t(i)} \qquad \text{Eq. (40)}$$

$$\hat{b}_j(k) = \frac{\sum_{\substack{t=1\\O_t=k}}^{T}\alpha_t(j)\beta_t(j)}{\sum_{t=1}^{T}\alpha_t(j)\beta_t(j)} \qquad \text{Eq. (41)}$$

Starting from an initial model, defined as $\lambda = (A, B, \pi)$, the respective reestimated model is defined as $\hat{\lambda} = \hat{A}, \hat{B}, \hat{\pi}$, calculated through the three expressions immediately above. As the Baum-Welch algorithm is a particular case of EM, one has $P[O|\hat{\lambda}] \geq P[O|\lambda]$ being the sequence of models $\hat{\lambda}_i$ obtained in this method converging to $\lambda^*$ in such a way $P[O|\lambda^*]$ is a local maximum. Baum proved that, if the initial model, $\lambda$, is at a point of zero gradients of the likelihood function, which was given the designation of the critical point, this corresponds to $\hat{\lambda} = \lambda$. Otherwise, the $\hat{\lambda}$ model is more promising than the $\lambda$ model, in the

sense that $P[O|\hat{\lambda}] > P[O|\lambda]$, that is, a new model $\hat{\lambda}$ was found, in which the sequence of observations are more likely to have occurred.

The Baum-Welch algorithm locally converges to a maximum, which is a fast process and, generally, the local maximum is typically a good maximum of the function. An essential aspect in the re-estimation process is related to the stochastic constraints of the HMM parameters, namely:

$$\sum_{i=1}^{N} \hat{\pi}_i = 1 \qquad \text{Eq. (42)}$$

$$\sum_{i=1}^{N} \hat{a}_{ij} = 1\,, 1 \leq i \leq N \qquad \text{Eq. (43)}$$

$$\sum_{i=1}^{N} \hat{b}_j(k) = 1\,, 1 \leq j \leq N \qquad \text{Eq. (44)}$$

The probabilistic nature of an HMM can be seen in Figure 18, above, where the correlation between three hidden states and two observable states is established.

Based on this procedure, $\hat{\lambda}$ will be used iteratively in place of $\lambda$. The reestimation calculation is repeated, thus improving the probability of the observations being reproduced by the model.

The method gains consistency with the integrated calculation of probabilities for each of the states of the current sampling interval referring to the posterior probability of the system being in each of the possible states, allowing for the assessment of eventual changes.

# Chapter 3

# Methodology

This section aims to provide knowledge on the literature review of the topics addressed in the methodologies presented in this thesis, namely, the data preparation, whose objective is to increase the model prediction quality. The feature generation is an essential process to select the type of information that wants to obtain from the data set. All the processes used for the "optimization" of the observable states, such as the dimensional reduction, are performed by the Principal Analysis Components (PCA) and the Clustering done by K-means. The HMM used to diagnose the health state of the equipment and, then a type of Deep Neural Network (DNN) named Gated recurrent unit (GRU) is used to perform the prognosis. Another DNN, Multi-Layer Perceptron (MLP), is also presented to perform value imputation on a small data set.

## 3.1 Data Preparation

### 3.1.1 Why is data pre-processed?

According to Yin *et al.* [127], it is since the early 1980s that more attention has been paid to Data Mining as a means of obtaining knowledge. Nowadays, with the abundance of raw data generated from various sources, Big Data has becoming a prominent approach in acquiring, processing, and analyzing large amounts of heterogeneous data to originate valuable information. The size, speed and formats in which the data is generated and processed affect the overall quality of the information. Therefore, Big Data Quality has become an important factor to ensure that data ownership is maintained at all stages of processing [128].

When we talk about predictive modelling, one of the main steps is data preparation. The raw data should be processed so that it can be analyzed as accurately as possible. Pre-processing aims to provide a structural, reliable, and integrated data source and is a very critical and complex step, which, if performed in the right way, it allows the success of the pattern discovery [129]. Kuhn & Johnson [130] explain that data preparation can impair the predictive ability of a model and, by consequence, well-executed transformations can improve modelling performance. Banhatti & Deka [131] refer that depending on the goal of the data analysis, pre-processing transformations should be

adapted to ensure a better quality of the analysis. To find meaningful information from Big Data, it is essential to perform data pre-processing [132]. Rinnan, Norgaard & Berg [133] say that this pre-processing step is extremely important to ensure reasonable results, whether the analysis is related to data exploration, classification or building a good and robust predictive model. To design an effective model, the input data must be provided in the appropriate quantity, structure, and format into the algorithm [134].

Kuhn & Johnson [130] refer that preprocessing techniques are responsible for adding, deleting, or transforming data to solve the problems that need to be fixed. Abdallah and Webb [135] add that to be able to do data analysis, it must go through a preparation process through manipulation and organization. This is done through an iterative process of manipulating raw data which, usually, appears poorly structured and confusing. Data preparation consists of several major activities including data characterization, data cleaning, eliminating duplication, compression, filtering, and format conversion [128], [135].

### 3.1.2 Importance of Data Pre-Processing

According to Banhatti & Deka [131], pre-processing techniques facilitate the stabilization of the mean and variance; these authors refer that data processing will eliminate irregularities in order to build effective computational models. Data pre-processing, according to these authors, refers to the analysis and transformation of input and output variables, with the aim of detecting trends, minimizing noise, highlighting important relationships, and leveling the distribution of variables. In this way, through analysis and transformation, it is possible to "help" the model learn relevant patterns.

Taleb, Dsouli & Serhani [128] also refer that to monitor the relevance of data, as well as the impact of its pre-processing, a concept of data quality is required. The authors define data quality in quality management as fitness for use or meeting user needs. They also state that the quality of the worked data always depends on the quality of the source data. For predictive models to have useful value, there must be high quality and volume of data. Data is often one of the biggest reasons why AI models fail [136]. Moreover, data quality is an aspect to be considered, as data in automated data collection processes may have some errors. This will cause a collection of inaccurate or incorrect data. Under this perspective, the reliability of the data transmitted by the measuring instruments will also be considered. This topic is very well explained in paper [137].

According to Zhang *et al.* [138], data cleaning and preparation consumes approximately 80% of the total data engineering effort. Nisbet *et al.* [134] also state this assumption,

emphasizing that 60% -90% of project time can be saved on data preparation activities. Abdallah & Webb [135] say that this is a process that may need many transformations and is repeated several times. In this way, data preparation can be considered the "bottleneck" in performing data analysis, especially in Big Data. Therefore, we can see that this phase of data pre-processing is a task of great importance to make a quality study, ensuring that the observable states given as input in the algorithms that do not contain missing or incorrect values. As Zhang *et al.* [138] state, this leads to:

i. distinguishing useful patterns that are hidden in the data;
ii. poor performance;
iii. poor quality results.

### 3.1.3 Data Quality Problems

According to Roadidea [139], the errors existing in the dataset are generated during the process. Since the purpose of the generated data is to describe reality, it is good to obtain data with a "one-to-one" correlation with that part of reality. When reality is converted into a signal through a sensor, transmission, and finally stored, it can be corrupted for various reasons. Common data quality problems (anomalies) include inconsistent data conventions between sources, data entry errors, inconsistent data formats, missing, incomplete, outdated or incorrect attribute values, data duplication, and irrelevant data [140]. Data anomalies are considered as a property of the data that misrepresent the actual "measurements". Therefore, anomaly can be characterized as duplication, inconsistency, missing values, outliers, noisy data, or any kind of distortion that can cause imperfections in the data [135]. According to Roadidea [139], some types of errors that can occur in data are:

- Measurement error

This is the first type of error that appears; always there will be a deviation between the actual value and the one obtained, due to imperfections in the measuring device (Figure 20).

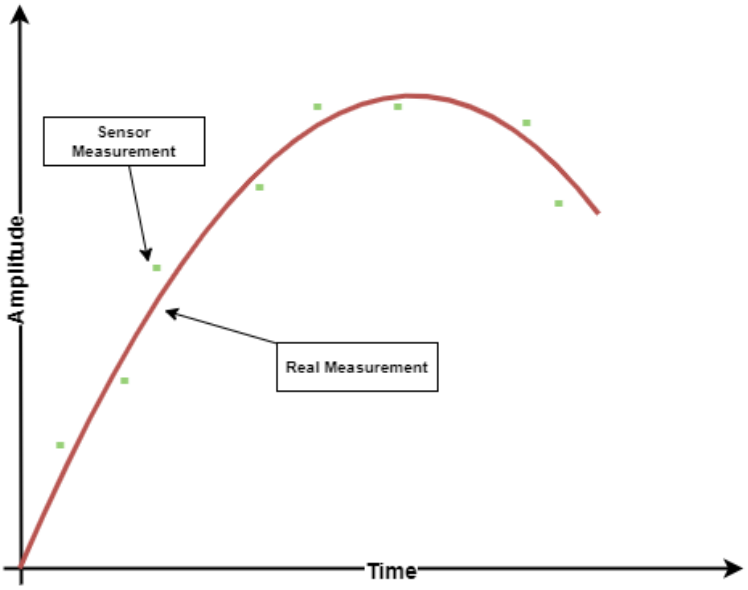



Figure 20 – Example of inconsistent values in a measurement.

- Inconsistent Data

Inconsistent data can be of various types; when an error is detected in the data acquisition chain, it is necessary to locate its source and to ensure that it never appears in future collection sessions.

- Duplicate data

When data is transmitted and stored, sometimes duplicate records appear for different reasons (Figure 21). The solution is to remove the cloned data, except for one of the records.

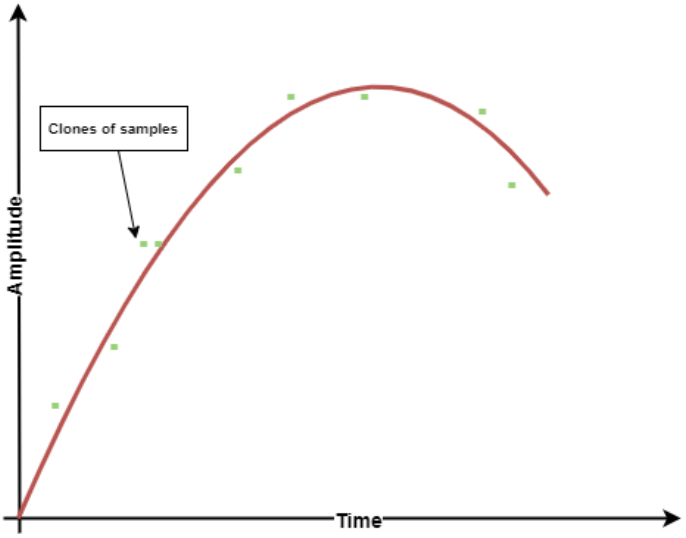



Figure 21 – Example of duplicate data in a measurement.

- Outliers

Outliers are values that do not follow the background pattern and are very rare or unlikely values (Figure 22) [139]. According to Abdallah & Webb [135], outliers are data that do not conform to the general data distribution. That is a type of data anomaly that also needs attention in the cleaning process.

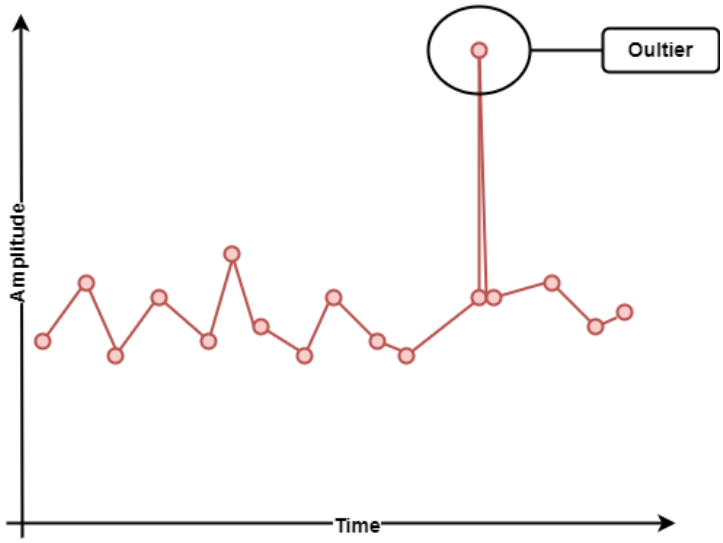



Figure 22 -Example of an outlier value in a measurement.

Kuhn & Johnson [130] define outliers as samples that are very distant from the main data stream. According to the authors, they can sometimes be difficult to define, even with a good knowledge of the data. As such, the first thing to do when suspecting outliers is to check that the value is scientifically valid and that there was no recording error. Outlier values may also represent a special part of the study population and are not considered abnormal. Sakr & Zomaya [141] say that outliers can be identified in visualizations and using statistical routines.

According to Nisbet *et al.* [134], Abdallah & Webb [135] and Roadidea [139], outliers can be classified into two different subgroups: natural outliers, of interest to the study, which contribute and improve the subsequent analysis or model building; outliers caused by faulty instruments, which contribute errors that make information extraction less accurate.

### 3.1.4 Data cleaning

Data cleaning is the first step performed in data preprocessing. Through a good data cleaning, it is possible to improve its performance and quality [129]. According to Ganti & Sarma [142], data cleaning is a term used to refer a variety of tasks aimed at improving

data quality. These tasks may need to be performed by joining several operations. According to Do *et al.* [143], it deals with the detection and removal of data errors and inconsistencies to improve data quality. The need for data cleaning increases when multiple data sources need to be integrated, for example, in data warehouses. As such, to obtain a quality database, it is necessary to consolidate different data representations and eliminate duplicate information. Noise must be eliminated from the data, since duplicate or missing values will produce incorrect or misleading statistics. According to Do *et al.* [143], data cleaning must satisfy several requirements, such as: detecting and removing all important errors and inconsistencies in individual data sources and when integrating multiple sources. This approach should be supported by tools to limit manual inspection and programming effort; it should not be done in isolation, but in conjunction with schema-related data transformations based on comprehensive metadata. According to Van Den Broeck, Cunningham & Herbst [144], data cleaning should be based on knowledge of technical errors and expected ranges of normal values.

According to Do *et al.* [143], data cleaning involves the following phases:

- Data analysis - to detect which types of errors and inconsistencies should be removed;
- Definition of transformation workflow and mapping rules - a database with different sources will increase data heterogeneity and data noise so, there should be a high number of transformation and cleaning steps;
- Verification - within the transformation workflow, testing on its accuracy and effectiveness should be done at this stage;
- Transformation - execution of the transformation steps by implementing the workflow to load and update a data warehouse.

According to Chen & Zhang [145], data cleansing consists of four actions:

i. Detect and eliminate data anomalies;
ii. Detect and delete duplicate records;
iii. Integrate data issues;
iv. Specify areas of data cleaning where particular application of clean procedure preparation is required.

### 3.1.5 Dealing with missing values

Many variables may contain values or blanks that damage the dataset, which must be excluded. This should be done as their use would damage the modelling and decrease its predictive ability. Missing values are one of the problems in data collection, then, it is important to detect them and ensure that appropriate measures are taken to enable learning systems modelling. This phase is essential for the data cleaning and conversion process [135]. Missing values are errors that will impair the quality of the data, and this occurs when the value is not stored for a variable. Roy *et al.* [146] say that missing values can occur due to several technical reasons, such as:

- Malfunction or physical limitations of the measurement equipment;
- Some missing values may not have seemed important during the data collection phase, and corrective action may not have been taken at a later stage to determine them;
- Some values may have been removed due to inconsistency with other recorded data.

Missing values can be quickly identified by the programmer and then deleted or replaced. According to Sakr & Zomaya [141], users can remove these data, or perform interpolation or imputation, depending on the analysis required. Similarly, outliers can also be removed or replaced with an imputed quantity, depending on the objectives of the analysis. The process of replacing missing data values is called imputation values. We should use the technique very carefully because, although imputed values are good for aggregate analysis, no individual imputed value can be reliable (because it is an estimate) [147]. According to Nisbet *et al.* [134], when data for a variable are missing, they should be replaced by some intuitive data if possible. This action consists of replacing the missing data with estimated values, using the information about the measured values. Although it may not provide the best prediction for the missing values, it can increase the integrity of the dataset [146].

## 3.2 Features Generation

The principle of Feature Generation (FG) prediction is to maximize the exploitation of information generated exclusively from time and process data [148]. According to Silva & Leong [149], Machine Learning (ML) algorithms can be seen as techniques to derive one or more hypotheses from a set of observations so, one of the ways to improve the input (observations) and output (hypothesis) is to select features that maximize the performance of an ML architecture [149].

FG is of paramount importance in any pattern recognition task, since given a set of measurements, the goal is to discover compact and informative representations [150]. FG is also known as construction, feature extraction, or feature engineering. There are different interpretations of the terms FG, construction, extraction, and feature engineering [151]. It enriches the observation language with additional constructed and derived features [149]. A good FG framework should have the inherent ability to scan this large feature space to extract the best results.

Feature Selection (FS) is important in ML tasks as it can significantly improve performance by eliminating redundant and irrelevant features, while speeding up the learning task [149], [151], [152]. Thus, FS helps to improve both dimensionality and the learning task. In a typical ML problem, there is usually an optimal number of features that provide the least error (highest accuracy) [149]. Feature irrelevance is the problem that some features are simply not correlated with the dependent feature. These features can even have a negative effect on the performance of a model [151]. This means that most of the classification-related information is "compressed" into a relatively small number of features, leading to a reduction in the required size of the feature space. We sometimes refer to these processing tasks as dimensionality reduction techniques [150], as will be discussed further below.

FG is of paramount importance in any pattern recognition task. Given a set of measurements, the goal is to discover compact and informative representations of the obtained data [150]; these authors also say that, if the transformation is chosen correctly, the features in the transformation domain can exhibit high information packing properties compared to the original input samples. The chosen transformation can exploit and remove information redundancies, which usually exist in the set of samples obtained by the measurement devices. With compact and useful representations, FG is used to maximize the exploitation of the information created from the data [153].

Then, FG is necessary for feature creation of the dataset to gain more information about the data collected from the equipment under study, since the continuous and time-varying features can provide a prediction of possible failures [154]. FG aims to acquire more and better information that improves the accuracy of the performance evaluation by comparing with the original signal [155]. To make a feature generation and, consequently, a dimensional reduction of the data, a time window processing method is used [154]. Here, time windows are created with different intervals depending on the study to be performed. Each time window will be represented by a certain number of samples that will be used to extract several features. In this thesis, the methods of feature

generation are performed in the time domain and frequency domain, where several features are created in each of the time windows created. The statistical features chosen for the characterization of the equipment behaviour, according to the papers [153], [156], were chosen, whose show that they relate well to the detection of equipment performance deviations.

To better understand the degradation process of a component using vibration signals, the features should be generated from the raw vibration signal [157], which can build multiple features in both time and frequency domains [158]. Frequency analysis has become a fundamental tool for vibration signal processing [159]. To perform FG in the frequency domain, the Fast Fourier Transform (FFT) is usually used, which, as explained in paper [157], it is one of the most widely used techniques in signal frequency domain analysis being responsible for converting the monitoring signal into a frequency spectrum.

## 3.3 Standardization

Data formats can be transformed so, they can meet the assumptions of a statistical inference procedure or to improve interoperability [146]. Sakr & Zomaya [141] say that data should undergo corrective transformations or be standardized as necessary in order to ensure data quality. Data transformations are necessary to support any changes in the structure, representation, or content of the data. These transformations become necessary in many situations; for example, to deal with schema evolution, migration from an old system to a new information system, or when multiple data sources to be integrated [143].

Abdallah & Webb [135] state that it is sometimes necessary to transform data from one representation to another, and explain that many of the reasons for this may be:

- To generate symmetric distributions instead of the original skewed distributions;
- To improve the visualisation of data that may be tightly clustered with respect to some outliers;
- To achieve better interoperability;
- To improve the compatibility of the data with the assumptions underlying a modelling process.

Data transformation can be done by standardizing the data, where a transformation function is used to change the domains of the data. Roy *et al.* [146] say that the term standardization means applying a transformation so, the transformed data are normally

distributed. Abdallah & Webb [135] and Roy *et al.* [146] show different types of transformation, explaining the purpose of each technique, which will depend on the nature of the data. Each technique used has its own purpose and depends on the nature of the data.

Data standardization is the process of transforming raw data values into another format with properties more suitable for modelling and analysis. The standardization process focuses on scaling the data in terms of range and distribution, for a variable $X$ represented by a vector $\{x1, x2, x3, \ldots, xn\}$.

In this thesis, the approach used to perform standardization is the *Z-Score*. It is a standardization method that transforms, not only the magnitude of the data, but also its dispersion. It will convert a range of variables with some mathematical heuristics, allowing all variables to have the same range. This standardization of values is not essential in machine learning algorithms but, it can make patterns in the data more visible [134]. Approaches, such as Linear Discriminant Analysis (LDA), PCA, and Kernel Principal Component Analysis require features to be on the same scale to find directions that maximize variance (under the constraints that in the case of PCA, the principal components are orthogonal). *Z-score* standardization overcomes the problem of variables with different units as it transforms the variables, so they are centred at *0* with a standard deviation of 1 [135]. This method transforms the values of the variables based on their mean and standard deviation. In other words, through standardization, the amplitude of the initial continuous variables also contributes to the analysis. The features with larger amplitudes do not overlap over the features with smaller amplitudes, not leading to one-sided results [160]. So, the data are first transformed to comparable scales. This is done using Z-score standardization, where the mean is subtracted and divided by the standard deviation for each value of each feature, aiming that all features be standardized to 0 mean and 1 standard deviation (Eq. 45).

$$Z_{Score} = \frac{x_i - \bar{X}}{std_{dev}(X)} \qquad \text{Eq. (45)}$$

Where:

$\boldsymbol{x_i}$ is the Z-score value of $x$;

$\boldsymbol{\bar{x}}$ is the line mean of $x$;

$\boldsymbol{std}$ is the standard deviation.

## 3.4 Dimensional Reduction

Dimensional reduction techniques are classified into two major categories: those based on attribute selection "Feature selection", which retain only the attributes considered useful, according to some functional evaluation; and those based on attribute extraction "Features Generation" (discussed in the previous section), which creates new attributes, either by combinations or transformations from the original attributes [152].

There are some ML systems that have difficulty to be processed and are slow in processing when deal with high dimensional data so, it is advisable to design the data into a lower dimensional space. According to Merola & Abraham [161], dimension reduction methods were initially used in data analysis as a purely descriptive (exploratory) tool, useful for isolating and, possibly, visualizing selected features from a set of variables in fewer dimensions. The same authors explain that the representation of multivariate data in a two or three dimensional graph makes it easier to "see" trends, clusters, and relationships in data, otherwise hidden or scattered by high dimensionality. To overcome the poor performance of ML algorithms with high dimensional datasets, dimension reduction is vital. Thus, Dimension Reduction is a technique for taking a high dimensional dataset (data objects with many features), replacing it with a much smaller dimensional dataset, preserving the similarities between the data objects, being useful for reducing the memory requirements for storing a dataset, as well as for speeding up algorithms [141]. Data reduction techniques are another class of predictor transformation. These methods reduce data by generating a smaller set of predictors that seek to capture most of the information in the original variables [130].

Dimension Reduction, according to Roy *et al.* [146], is simply the removal of unnecessary or insignificant attributes. The presence of irrelevant attributes in the dataset can make the problem intractable and mislead the analysis, which is often called the Dimensionality Curse. The same author also says that dimensionality reduction should be carried out in such a way that, as far as possible, the information content of the original data remains unchanged and, using the reduced dataset, does not change the result. Dimension reduction can be achieved by selecting only the relevant attributes and ignoring the remaining.

Then, according to Anowar *et al.* [162], dimensionality reduction algorithms aim to solve the curse of dimensionality, with the goal of improving data quality by reducing the data complexity.; according the same authors and Kacprzyk [152], dimensionality reduction methods bring several benefits, including:

- Identifying the relevant factors;
- Reducing the storage space required;
- Improving ML performance through less misleading and redundant features;
- Avoiding overfitting through fewer resources;
- Reducing computation time;
- Easily visualise and interpret the data.

There are several techniques to determine which variables are unnecessary for the model before training it. In the case of this thesis, we will only use PCA, which is the approach used in the methodologies presented below.

### 3.4.1 Principal Components Analysis (PCA)

Principal components were first found by Pearson [163] and were considered by the statistical community when Hotelling [164] proposed them as the estimates of linear combinations of a set of random variables that retained the highest possible variance. It is an unsupervised learning method of feature extraction and dimensionality reduction (moving p-dimensional data to a lower dimensional linear subspace of m-dimensionality) in ML, which retains the original features of the data and selects the main properties of the dataset [154], [162], [165]–[167]. It analyses a data table in which observations are described by several inter-correlated quantitative dependent variables and is widely used due to its ability to extract interpretable information by efficiently removing redundancies [160], [168]. It is used to project feature vectors in orthogonal space for robustness and computing efficiency [169]. That is, it is responsible for doing a dimensional reduction to replace a high dimensional dataset (with large features), with a much smaller dimensional dataset with only a few Principal Components (PCs) [153].

PCA is a statistical unsupervised linear transformation procedure, commonly used to do dimensional reduction of large sets of time series observations [170], which uses an orthogonal transformation to map a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called Principal Components [146], [162], [171]. In this way, they preserve the highest percentage of information [168], [172], [173], and the original data are mapped into a much smaller number of dimensions. In other words, it is a method of projecting large dimensional measurements towards a space of minimum dimension and preserving maximum variance [174], compressing sensory data according to their spatial and temporal correlations [175].

PCA is a classical multivariate statistical technique used to reduce data, in which the first few PCs account for most of the variability in the original data [171]. When this is done, the original variables are replaced by the first several principal components. The higher the degree of correlation between the original variables, the fewer the number of new variables needed [176]. Thus, PCA is a method that seeks to find linear combinations of predictors, the PCs, that capture the greatest possible variance [130]. The extracted PCs refer to variables that can explain the maximum variance (variance) in all the data. Linear combinations of the original variables are produced to generate new axes, the PCs. The number of PCs is equal to the number of original variables, but the number of significant PCs becomes smaller. That is, the transformation is defined such that the first principal component has as high a variance as possible, possessing the greatest variability of the data, and each following component, in turn, has the greatest possible variance under the constraint that it is not correlated with the previous components [160], [172]. Then, the first Principal Component ($PC_1$) is removed to have the most significant variability of all possible linear combination data and is of the highest importance. The second Principal Component ($PC_2$) is extracted in an orthogonal link to the existing PC to reflect data variations that $PC_1$ was not able to fully explain and so on [177]. According to the author, the minimum number of CPs is based on 85% or more of the cumulative contribution rate. Mathematically, the $j^{th}$ PC can be written as [130] (Eq. 46):

$$PC_j = (a_{j1} * Predictor1) + (a_{j2} * Predictor2) + \cdots + (a_{jp} * PredictorP) \qquad \text{Eq. (46)}$$

Where:

$P$ is the number of predictors;

The coefficients $a_{j1}, a_{j2}, \ldots, a_{jp}$ are called the component weights, helping to understand which predictors are most important for each PC.

PCA draws the high-dimensional dataset into a new subspace where the orthogonal axes, or PCs, are considered as the directions of maximum variation of the data [162].

PCA is one of the most widely used techniques for reducing dimensionality because it is the oldest and, therefore the most studied method. It is relatively easy to calculate and, above all, it is the solution to a number of different problems involving the dimensional reduction of a set of variables [146], [161], [162], [178]. The main advantage of PCA, and the reason why it has maintained its popularity as a data reduction method, is that it creates uncorrelated components, which improves the numerical stability of the model [130].

Briefly, PCA is mathematically defined as a linear transformation that converts the data into a new coordinate system so, the first PC is, in the coordinate system, the one that has the highest variance per projection of the data; the second PC is in the coordinate with the second highest variance, and so on [160]. The PCA algorithm is implemented in five steps, as shown in Table 4, [153], [162], [177].

TABLE 4 - STEPS OF THE ALGORITHM PCA

| |
|---|
| Input: $\boldsymbol{X \epsilon R^{n*d}}$ |
| 1 – Calculate the correlation of data sets ($\boldsymbol{X}.\boldsymbol{X^T}$); |
| 2 - Calculate eigenvectors and eigenvalues applying Linear Eigen decomposition $\boldsymbol{X}.(\boldsymbol{X^T})$; |
| 3 – Sort Eigen Values in decreasing order to sort Eigen Vectors; |
| 4 – Build a Matrix $\boldsymbol{W}(\boldsymbol{D} * \boldsymbol{k})$ with KTop Eigen Vectors; |
| 5 – Transform $\boldsymbol{X}$ using $\boldsymbol{W}$ to obtain the new subspace $\boldsymbol{Y} = \boldsymbol{X}.\boldsymbol{W}$; |
| Output: $\boldsymbol{X \epsilon R^{n*k}}$ |

## 3.5 Clustering

As the volume of data has increased, the application of Clustering algorithms has begun to expand [179]. Clustering is a fundamental Data Mining technique that is used in a wide variety of areas for data analysis [180]. It is a widely used unsupervised classification application in data mining. Clustering analysis represents a collection of multivariate statistical methods whose goal is to identify groupings of objects within data, in which each given object is assigned to a cluster based on similarity or distance measures [181]; the authors add that, since the assignment of group membership, it is not known in advance, and the algorithm cannot learn the characteristics of the clusters, making these learning algorithms unsupervised: "Clustering, like unsupervised learning, provides for labels of unknown data by exploring the latent structure of the data" [182].

Clustering is an unsupervised data mining technique used in a wide range of data analysis areas where multivariate statistical methods allow clustering analysis. It is a technique that divides data into meaningful groups and the main goal is to reduce the amount of input data by grouping similar data items [176]. It is especially suitable for datasets where many features are highly similar to each other because these features are grouped and reduced [183]. Having several observations, clustering aims to design a classification scheme to group objects into several classes so, the instances within a class are similar in some aspects but distinct from those in other classes. Thus, it is possible to identify a finite set of categories to describe a given data set, maximizing the homogeneity within each grouping and the heterogeneity between different groupings; that is, a method to explore data in order to find groups of objects that are similar but different

from objects in other groups [131], [153], [164], [178], [180]–[182], [184]. It aims to minimize intra-class similarity while maximizing inter-class dissimilarity [185]. These goals of homogeneity within groups and heterogeneity among groups are usually achieved according to a traditional approach based on some distance measures (dissimilarity), or by assessing the underlying density of the data [186]. In short, clustering is a procedure that aims to explore relationships and patterns in data by grouping objects according to their similarity; that is, similar objects will belong to the same cluster [187], [188].

Clustering methods can be categorised according to the need for prior knowledge of the data [176]. There are many clustering methods available, and each one can provide different groups of datasets [185]. There are two groups of Clustering: partition clustering; and hierarchical clustering. In this case, we will talk about a type of clustering belonging to the partition clustering group, called *K-means*, which is the model used in this thesis.

### 3.5.1 K-means

The *k-means* algorithm was first proposed by Stuart Lloyd, in 1957, as a modulation technique, with a more efficient version being proposed and published in Fortran by Hartigan & Wong in 1975/79, and Subbalakshmi *et al.* [189], [185]. *K-means* is a popular method in partition clustering techniques, widely used due to its simplicity and efficiency. It is a Clustering, non-hierarchical, unsupervised ML method from the branch of multivariate statistical analysis, where the number of clusters k is determined and the observations closest to the centre of the cluster are included in it [185], [189]–[194]. K-means is used to perform classification of unlabeled data in which the specific response variables are unknown [195]. To achieve optimal clusters, the variance in the distance difference between each cluster and the observations within that cluster should be minimized [196]. In this process, the similarity between observations in the same cluster increases and the similarity with data from other clusters decreases [196], [197].

The algorithm divides the $n$ observations into $k$ clusters in which the observation belongs to the cluster with the closest mean [185]. It is an algorithm used to group data into k clusters [197], usually comparing the data using Euclidean distance (which is the one also used in this thesis) [198], [199]. In other words, it can perform clustering based on the spatial position of each object to be grouped [191].

This k-means analysis manages each observation in the data with a location in space and finding a partition in which the objects within each cluster are located as close as possible

to each other and as far as possible from objects in other clusters [189], [200]. These authors also explains that each cluster in the partition is defined by its member objects and its centroid; the centroid being the point at which the sum of the distances of all objects in that cluster is minimized [189]. Thus, the clustering process consists of the initial definition of centroids and the assignment of events that are inserted as input into the cluster with the nearest centroid [187]. As objects are inserted and assigned to the clusters, the centroids are updated until they have little or no change, reaching the convergence condition [187], [184]. Then, K-means is an iterative algorithm that minimizes the sum of the distances from each object to its cluster centroid, across all clusters, moving objects between clusters, until the sum can be further reduced; the result is a set of clusters as compact and well-separated as possible. The process of estimating data belonging to a given cluster is executed iteratively until the convergence condition is satisfied [179]. That is, given an initial set of centres (usually started randomly), the K-means algorithm alternates the two steps: (1) for each centre it identifies the subset of points that is closer to it than any other centre; (2) the averages of each feature for the data points in each cluster are computed, and this average vector becomes the new centre for that cluster. These two steps are iterated until convergence to stabilize the data assignments to clusters [195], [201]. K-means has a sufficiently fast convergence speed and does not require large computational resources; it is an easy algorithm to implement and apply even on large dataset. It is an appropriate algorithm for continuous data due to its simplicity and efficiency [189].

The main goal of Clustering is to group similar data points and discover the underlying pattern [202]. For this purpose, k-means requires a fixed number of clusters ($k$). Each of the clusters is assigned a centroid, which is a location of the centre of the cluster. After $k$ is chosen, each data point is allocated to the nearest cluster by summing the squared distances of the Euclidean distances between the items and the centroid (Eq. 47), minimizing intracluster variation [202].

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \qquad \text{Eq. (47)}$$

Here, $x_i$ is the $i^{th}$ data point of cluster $k(C_k)$ and $\mu_k$ is the mean value of the points in cluster $k$. The total variation within the cluster is defined as Equation 48. The total Sum of Square Error (SSE) within the cluster measures the quality of the clustering, which increases as the sum of squares measures decrease [203].

$$Total\ within\ cluster\ variation = \sum_{k=1}^{k} W(C_k) \qquad \text{Eq. (48)}$$

Then, as Borlea *et al.* [204] and Peng *et al.* [205] explain, K-means is used to process a dataset $D = \{x_1, x_2, \dots, x_n\} \in \Re^d$, where $x$ is a dataset record defined as $X_i = [x_{i1}, x_{i2}, \dots, x_{id}]^T \in \Re^d, i = 1 \dots n$, $d$ is the dimension of a dataset record and $T$ stands for matrix transpose. The algorithm divides the dataset $D$ into a set of $k$ predefined numbers of clusters $C_{j,}\ j = 1 \dots k$. Each cluster $C_j$ is composed of a centre of mass called centroid and defined as $C_j = [C_{j1}, C_{j2}, \dots, C_{jd}]^T \in \Re^d, j = 1 \dots k$. The total number of points assigned to each cluster, with the notation is $n$, with the cluster expression $C_j = (c_j, n_{C_j})$. The centroid array is defined as $c = [c_1^T, c_2^T, \dots, c_k^T]^T \in \Re^{dk}$, which represents the centroids of all existing clusters. The main objective of the algorithm is to minimize the intracluster variance (Eq. 49):

$$c^* = arg \min_{c \in \Re^{dk}} V(c), \qquad V(c) = \sum_{j=1}^{k} \sum_{\substack{i=1 \\ x_i \in C_j}}^{n_{C_j}} \|x_i - c_j\| \qquad \text{Eq. (49)}$$

where $c_j$ is the centroid of the cluster $C_{j,}\ j = 1 \dots k$, $V$ is the objective function or the criterion, and $c^*$ is the optimal arrangement of centroids.

The K-means clustering process has the following steps [185], [202]–[204], [206]–[208]:

- Step 1 - Create $k$ initial clusters, choosing $k$ as the number of random data points;
- Step 2 - Calculate the arithmetic mean of each cluster formed, randomly selecting $k$ data points as initial centroids;
- Step 3 - Each record is assigned to the closest cluster, finding the similarity between the points using the formula, i.e., assigns the data set $x_i$ to the nearest centroid $c_j$, using the Euclidean distance (Eq. 50):

$$d_{x_i, c_j} = \|x_i - c_j\| = \sqrt{(x_{i1} - c_{j1})^2 + (x_{i2} - c_{j2})^2 + \cdots + (x_{id} - c_{jd})^2} \qquad \text{Eq. (50)}$$

- Step 4 - Reassign each record to the most similar cluster and recalculate the arithmetic mean of all clusters in the data set;
- Step 5 - The process continues from step 3 until no data points are reassigned and the k-means procedure is completed.

Since the initial cluster centre, and the similarity measure clustering criterion, function easily converges to the local minimum, selecting different initial clustering centres that will lead to different clustering results [209].

K-means requires the number of k clusters to be set in advance; according to Tang & Miyamoto [210], the clustering results based on this algorithm are severely affected by the setting of the initial parameters. The optimal number of clusters is a fundamental problem in cluster partitioning in the K-means method, being necessary to know it in advance to obtain a good result [197]. The inappropriate value of $k$ will affect the clustering effect, the number of iterations of the algorithm, and the accuracy and complexity of the algorithm [190], [191], [196].

The task of estimating the correct number of natural clusters for a dataset is, in general, difficult. One of the reasons for this limitation is that in most cases these are samples from a population, and it is necessary to estimate the number of actual conglomerates in the population [180]. There are several types of methods that can help us to decide about the optimal number of conglomerates [202]. Subbalakshmi *et al.* [189] present some of them: the general rule; the cross-validation; the elbow method; the information criterion approach; the kernel matrix; the silhouette analysis. The authors also tell us that most of the methods need to perform the clustering process to decide the right number of clusters. In this thesis, we will only talk about the silhouette analysis method and the elbow method which are the ones used in the created methodologies.

Many of the methods used to define the number of clusters are based on some measure of distance between objects and clusters. According to Menardi [186], an exploratory tool that relies on this idea is the Silhouette analysis. The author also says that the diagnostics used to evaluate the quality of a partition should be consistent with the clustering method adopted to produce that partition. In particular, the diagnostic-based distance is inadequate to evaluate the identified clusters using a technique based on density estimation. As such, silhouette analysis is a method that matches well with the k-means Clustering algorithm. K-means is designed to minimise the sum of the distances between instances and their nearest centroids [211]. As with the Silhouette criterion, good partitions are also obtained when this minimisation is performed correctly, as well as when the clusters are well separated, and there is a symbiosis between k-means and the Silhouette criterion. Silhouette information evaluates the quality of the partition detected by a clustering technique, since it is based on a measure of distance between the clustered observations; its standard formulation is not suitable when a density-based clustering technique is used [186]. Thus, as stated by Nidheesh [180], the Silhouette

index has been a widely used technique to determine the number of natural clusters for the k-means algorithm.

Silhouette analysis method for the validation of cluster analysis can be consulted for an in-depth explanation, as can be seen in [180], [186], [212]. In summary, the Silhouette Index (SI) is calculated for each feature and measures through the ratio between $a(k)$ and $b(k)$, the intra-cluster disparity, respectively, as shown in Equation 51, generating a value between -1 and 1 [183]. This measure ranges from +1, indicating points that are far away from neighbouring clusters, to 0, indicating points that are not distinctly in one cluster or another, to - 1, indicating points that are likely to be assigned to the wrong cluster [200].

$$S(k) = \frac{b(k) - a(k)}{\max\{a(k), b(k)\}} \quad \text{Eq. (51)}$$

The Sum of Squared Error (SSE), or elbow method as it is known from the graph it forms, is one of the most popular cluster evaluation criteria [203], where the cost function is calculated when a different number of $k$ clusters are formed during Clustering [192]. We use different numbers of $k$ and calculate the total of them within the sum of squares for each value of $k$ and plot, where k is represented in the graph at a kink (elbow) location - we consider this the optimal number $k$. Then, for better optimisation, the number of clusters can be identified using Equation 52 of the elbow method [213]:

$$E_k = \sum_{r=1}^{k} \frac{1}{nr} D_r \quad \text{Eq. (52)}$$

Where:

$k$ - denotes the size of the cluster;

$nr$ - represents the number of data points in the cluster;

$D_r$ - is the sum of the distances among all points within the cluster.

The elbow method is comprised of four steps [214]:

- To run a centroid-based clustering variance of each clustering result, e.g., sum-of-squares-errors algorithm, e.g., k-means, for each $k \in N$;
- To calculate the (SSE) for k-means;
- To plot the results on a graph;
- To select the elbow curve in the graph.

$k \in N$ is plotted on the $x$-axis and the corresponding graph shows it visually, after which makes the reduction of the SSE, becoming negligible when increasing the value of $k$.

## 3.6 Deep Neural Network (DNN)

### 3.6.1 Multi-Layer Perceptron (MLP)

The Multi-Layer Perceptron (MLP) is a supervised learning method that works with nonlinear problems thanks to its multilayer construction. It is one of the most useful feedforward networks [215], [216]. MLP is divided into three layers: input layer; hidden layer; and output layer, having the same structure as a single-layer neural network but with more hidden layers [217]. Each neuron is fully connected, and each neuron has weights that are used to calculate whether the input data has the information we need by adjusting the weights between the elements [215], [218], [219]. Thus, the MLP has an input layer of neurons that act as receivers, one or more hidden layers of neurons that calculate the data and undergo iterations, and the output layer that predicts the output [220]. Each hidden layer is composed of several artificial neurons, and the output of an artificial neuron is obtained using Equation 57 [177], [221], [222], [223]:

$$f\left(\sum_{i=1}^{p}(W_i X_i + b)\right) \qquad \text{Eq. (53)}$$

Where:

- $f$ is the activation function (e.g., sigmoidal function, exponential linear unit, hyperbolic tangent function, or rectified linear unit);
- $p$ is the number of inputs;
- $W$ is the input weight;
- $X$ it's the input;
- $b$ represents the polarization value.

During the training phase, weights and biases are calibrated using accessible input-output datasets. During training, two operations occur: activation propagation (feedforward), sending the information forward, and error propagation (backpropagation) [224]. During the first step, each neuron generates an output signal based on Equation 57 that is compared with the target, and errors are propagated back through the network during the second step. The weights and variances are adjusted based on an optimization algorithm selected to minimize the predefined loss function [224]. The same author also explains that several iterations are required during the

training process until a convergence criterion is satisfied [224]. Hyperparameters regarding the model structure, such as the number of inputs, layers, number of neurons per layer, activation function, etc., should be carefully selected and changed since they have a significant impact on the training efficiency and performance. "(...) a good choice of hyperparameters is usually necessary to make them work well on real-world problems, and tricks are often used to make the most efficient use of these methods and to extend their capabilities" [225]. In Figure 24, we can see a scheme of an MLP model.

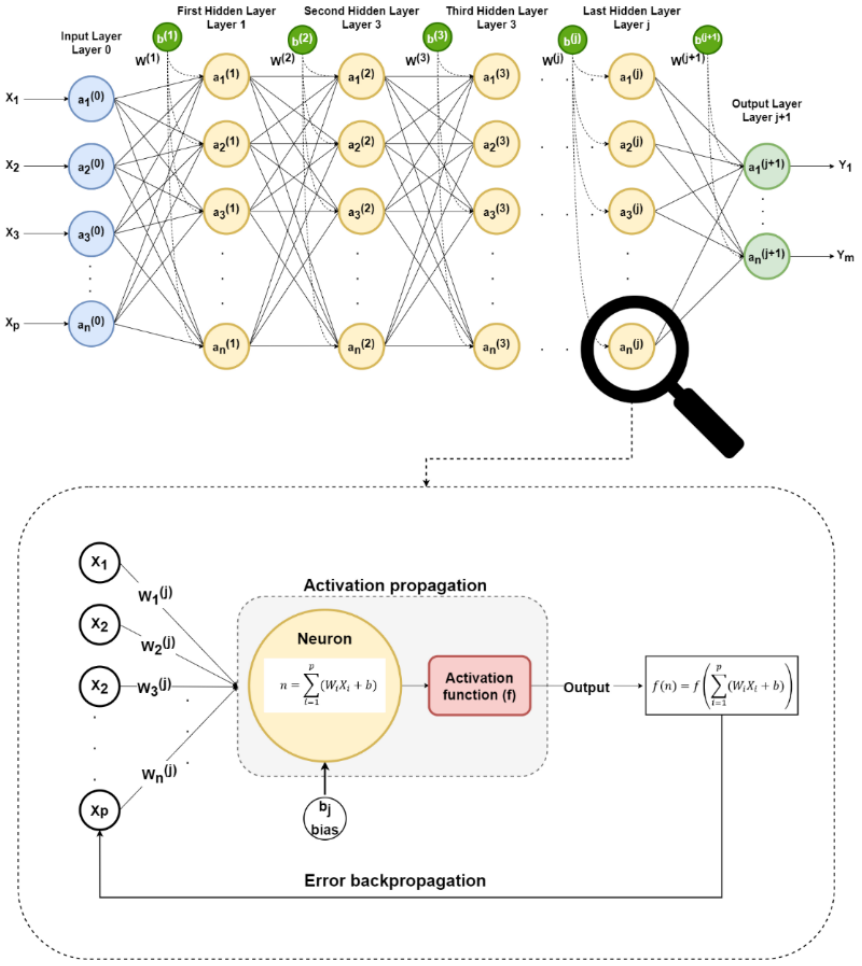



Figure 23 - Schematic of the operation of a MPL neural network.

### 3.6.2 Gated Recurrent Unit (GRU)

There are some improvements in the Recurrent Neural Network (RNN) model that can address the shortcomings of the application of the long-term sequence of this model [226]. Long Short-Term Memory (LSTM) can overcome the deficiencies of RNN that cannot handle well the long-distance dependency, while GRU can make the model structure simpler and more efficient while maintaining the effect of LSTM [226]. Both

LSTM and GRU have the uniform goal of tracking long-term dependencies, while alleviating the vanishing gradient problems that often occur in the training phase of model RNNs [227]. LSTM and GRU are the two improved versions of RNN, which are the most widely used and found in the literature [228]. As the author explains, LSTM and the GRU are called gated RNNs, since both rely on different types of gates to regulate the flow of data within their units. LSTM was initially introduced by [229] in 1997 and its unit consists of two gates: the input gate is used to control the input of information (the degree of input of new information); the output gate is used to control the degree of the current state of the unit is filtered (information filtering) [226]. In turn, the GRU, which was introduced by [230], subsequently proposed based on the LSTM, with a relatively simple structure that requires fewer parameters for training, having only two ports (Figure 23) [226]–[228], [231]–[234]: reset port, that can help to encrypt the model to determine how much past information can be forgotten; update port, which combines the input port and forgets the LSTM to determine the degree of the previous hidden state that will be used to update the current state. Both reset port and update port are used to solve the leakage gradient problem, which are two vectors that can manipulate the information in networks/layers that flow to the desired output [231]. As the author explains, the two gates can be trained to maintain long-term memories without removing relevant information that is meaningful for future predictions.

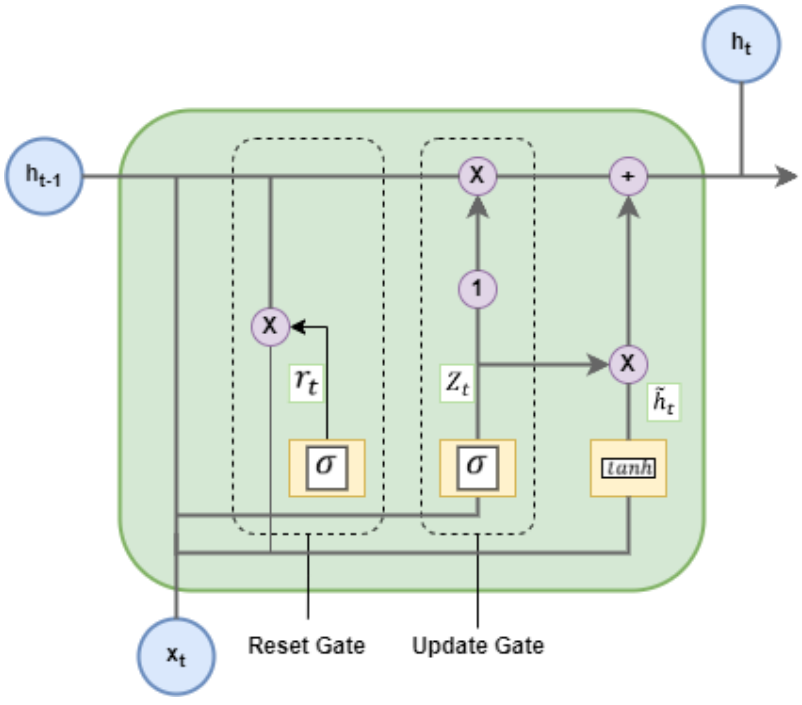



Figure 24 – Example of Gated Recurrent Unit (GRU).

GRU has an update port $z$ and a reset port $r$ to simplify the memory block structure of the original LSTM network (Figure 23). The input of GRU network being $x_t$, the formula to calculate the next output and state value in GRU is [228], [226], [231], [232], [235], [236]:

1- The algorithm starts with the calculation of the update port $z_t$ for time step t (Eq. 53);

$$z_t = \sigma(W_z.[h_{t-1}, x_t]) \quad \text{Eq. (54)}$$

When $x_t$ is connected to the network unit, it is multiplied by its own weight $W_z$. The same applies to $h_{t-1}$ which holds the information of the previous $t-1$ units. Both results are summed, and a sigmoid activation function is applied standardizing the result between 0 and 1.

2- The reset port uses the model to decide how much of the passed information is forgotten (Eq. 54);

$$r_r = \sigma(W_r.[h_{t-1}, x_t]) \quad \text{Eq. (55)}$$

3- Here, a new memory content is introduced, which will use the reset port to store the most important information of the past. After getting the switch signal, it uses the switch reset function, get the reset data, then it combines it with the $tanh$ activation function to get $\tilde{h}_t$ (Eq. 55).

$$\tilde{h}_t = tanh(W_h.[r_t * h_{t-1}, x_t]) \quad \text{Eq. (56)}$$

$tanh$ can control the range of values between - 1 and 1. It can be seen that the input data is included, and the hidden information is controlled.

4- Finally, it is necessary to calculate the vector $h_t$ that holds the information of the current unit and transmits it to the network, which determines what to collect from the previous steps $h_{t-1}$ (Eq. 56):

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad \text{Eq. (57)}$$

It can be extracted from the forward training that the state of the reset port $r_t$ controls the combination of the input $x_t$ and the previous time state $h_{t-1}$. The update port $z_t$ determines the use of the information from the current phase to the previous phase.

# Chapter 4

# Results of the Methodology

This chapter presents all the studies carried out and the implementation of the methodology developed. Each study presented here gave rise a paper that was submitted to international journals Scopus of Q1 or Q2.

## 4.1 Methodology and results to diagnose the behaviour patterns in equipment

This methodology gave origin to a paper submitted to the Applied Science journal, indexed Scopus, evaluation Q2, with DOI: 10.3390/app11167685 (Appendix B). It focuses on the determination of the fault diagnosis status of equipment, more specifically, of a pulp drying press used in a Portuguese paper industry. The objective is to use the data collected by the sensors attached to the equipment, to diagnose its state of health, to determine whether it is in "Good Functioning", "Alert State", or "Fault State". The specificity of this methodology adds value to the evaluation of the equipment's diagnosis, in particular, because it does not require prior knowledge of the fault conditions. The study was carried out using *MatLab* software and its features.

### 4.1.1 Methodology

Real-world Processes usually create observable outputs that can be classified as signals [125]. In the case of this methodology, the observations collected by the sensors will be sent to the HMM model, which will provide the hidden states that will correspond to the status of the equipment. A series of approaches are used to perform the "optimization" of the observable states, to improve these inputs into the HMM model (Figure 25).

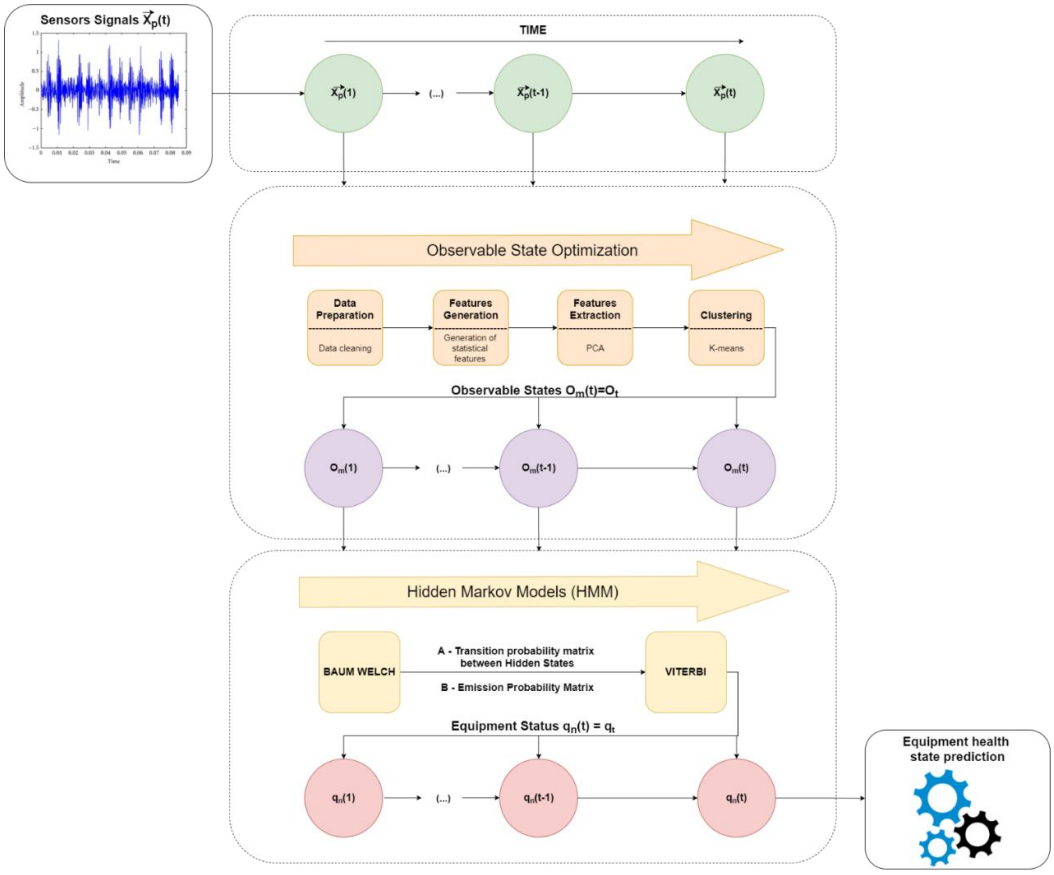



Figure 25 - Methodology for optimising observations for input to the HMM health state classifier.

Figure 25 shows how the sensor data will be changed to improve the efficiency of the algorithm in terms of computation time and overall prediction performance.
The sensors are first used to collect data over time, which are then processed aiming to improve the observable states. The data preparation is done during the pre-processing stage, which follows the data collection, with the objective of raising its quality.

After the data preparation stage, a data grouping into time periods, to create appropriate statistical values for each time window, is done. Thus, using this feature generation, more information can be extracted from the original data, since statistical features based on the time domain perform well in identifying patterns and changes in the data [237].

After generating time domain features, the data are standardization and incorporated into a dimensional reduction model, Principal Components Analysis (PCA). An important step in pre-processing the features is to standardize them to a conventional normal distribution of mean 0 and standard deviation 1, using Z-Score standardization for this purpose. Since we are interested in the PCs that maximise the variance of the data, it is crucial to perform the PCA approach using Z-Score standardization. If one

feature varies less than another due to their respective scales, the PCA method defines the direction of the maximum variance that is closest to the axis, which varies more if these features are not standardized. To make the algorithm more sensitive to changes that may take place in various metrics, it is crucial to standardize the data. standardization is necessary in this methodology because this model needs features on the same scale to determine the directions that will maximise the variation in the data. Only the principal components (PCs) needed for the investigation and to reduce the dimensionality of the dataset will be obtained during the dimensional reduction phase. The axes are redefined using the PCs instead of the original variables, and the observations are now described in a new orthogonal space. In order to decrease the computational burden, the new Features, the PCs, characterised by PCA will be used. The PCA method was used for this research since it is a relatively simplified and user-friendly technique. It is an unsupervised method that finds stronger features with a higher percentage of information using a more meaningful coordinate system of the dataset. Moreover, it is fast, effectively eliminates overfitting, and can be applied as a method for noise removal and data compression [146], [161], [162], [178]. It eliminates a significant part of the experimental noise, because the noise results in random errors that are not related to the information in the data matrix, which enhances the numerical stability of the model [130].

After the PCA, the K-means clustering is performed, where the data will be differentiated through clusters. In other words, clustering will group observations as heterogeneously as possible between clusters and as homogeneously as possible within each cluster. Well-functioning observations will be gathered in one cluster, while observations of failing equipment will be grouped in a different cluster. Each cluster created will be used as an observable state and will be fed into the HMM. K-means is a clustering method that uses the Euclidean Distance metric and is simple to implement and use, even on large datasets [162], [164]. It is a model that uses the unsupervised learning method, has a sufficiently fast convergence speed, and does not need many CPU resources. It is a method that can be applied after reorienting the source data of the central metric structure on the cloud, in a process done by PCA. That is, placing the data centred with respect to an origin, approximately equidistant from the centre and positioned in the different quadrants of the Rp space (p being the Principal Components of the PCA), allows the k-means algorithm to work well.

Finally, after obtaining the observable states from the k-means method, the HMM is applied revealing the diagnosis of the equipment through its hidden states.

### 4.1.2 Results

#### 4.1.2.1 Data Preparation

For the application of the developed methodology, six sensors have been used that measure respectively: current intensity; hydraulic unit level; torque; VAT pressure; rotation speed; hydraulic unit temperature. These sensors collect an observation every minute, being the data collected between August 2017 and October 2020. Consequently, the data were collected over a period of three years and three months (Table 5).

TABLE 5 - QUANTITY OF OBSERVATIONS COLLECTED BY EACH OF THE SENSORS FOR THE DRYING PRESS UNDER STUDY

| *Equipment* | *Sensors* | *Metric* | *Nº of data by day (each sensor)* | *Nº of data by year (each sensor)* | *Total of data (each sensor)* |
|---|---|---|---|---|---|
| *Drying Press* | Sensor 1 | Hydraulic unit temperature | 1440 | $1440 * 365 = 525600$ | $(1440 * 153) + (525600 * 3) = 1710720$ |
| | Sensor 2 | Hydraulic unit level | | | |
| | Sensor 3 | Rotation Speed Press | | | |
| | Sensor 4 | Torque Press | | | |
| | Sensor 5 | Current intensity | | | |
| | Sensor 6 | VAT Pressure | | | |

As can be seen from Table 5, there is a dataset of 10,264,320 samples. However, as already explained, there are occasional errors in sensor data collection so, not all the information contained in this dataset is reliable and has quality. Data cleaning should be based on an understanding of technical errors and expected ranges of normal values [144]. Thus, for this dataset, the quality of the data obtained may be determined by applying a confidence index to the data collection. This confidence index protects the data collection system by indicating the quality of data collection using only two numbers, 0 or 100: where 100 represents well-collected data that was valid for collecting information; 0 represents bad data collection and the data is classified as "rubbish". Thus, it was decided to eliminate from the data set all the values that had been collected with a confidence index equal to 0. To eliminate the equipment stoppages throughout the study time, it was decided to replace by the average all the values that were in the time period whose values of intensity, torque, VAT pressure, and rotation speed were below a certain threshold. This ensures that the downtime of the equipment will not affect the HMM prediction.

According to the most frequently cited sources in the literature, such as references [134], [147], [150], the imputation of values by their mean, it is the most generally accepted procedure when the data of a variable are missing. Based on this methodology, the

imputation of the values was exercised respecting some criteria: when a slice of one of the signals was replaced, the others were all replaced by their corresponding averages. We can visually examine the original data set with the data after its preparation (Figure 26).

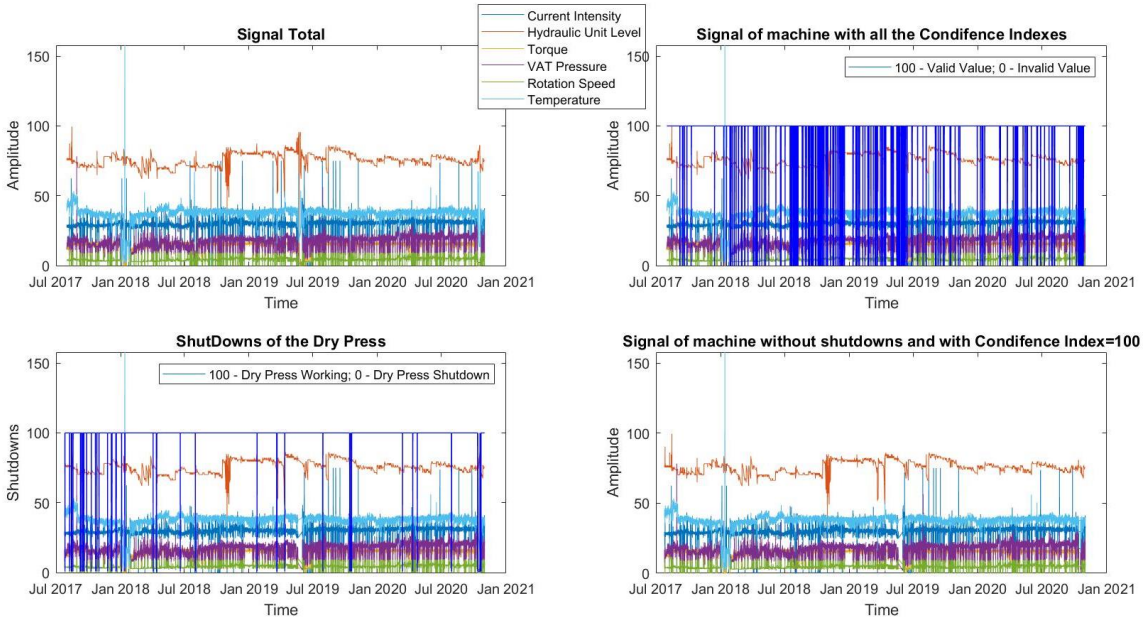



Figure 26 - Observations collected by the sensors before and after data pre-preparation.

A set of 1,640,880 data was left after 69,840 samples were removed using the confidence index. Regarding the equipment disconnections, they were replaced by the average of the signal, thus maintaining the value of 1,640,880 data. It is common knowledge that outliers can be divided into two groups: those that are faulty, and those that are caused by faulty equipment [134], [135], [139]. Thus, no additional filtering was done since outliers may not be considered outliers but observations that add value to the prediction. Furthermore, the following data processing models complement the noise "filtering" of the data set so, no additional data cleaning is required.

#### 4.1.2.2 Features Generation

To Generate Features, in this study it was decided to create temporal windows of 6 hours, whose objective is to cover 4 shifts of the working day. To this end, 4,558 windows were formed, with 360 samples in each one of them. Once the temporal windows had been created, statistical characteristics were generated in the time domain with the objective of obtaining a greater amount of information in the detection of patterns of the signals under study. Then, 10 statistical features were chosen to better characterize the signals: 1. mean; 2. standard deviation; 3. variance; 4. kurtosis; 5. skewness; 6. coefficient of

variation; 7. maximum; 8. minimum; 9. mode; 10. median. These 10 features are probabilistic representations capable of explaining the frequency of occurrence of events, thus making it possible to absorb more information about the equipment's behaviour and, consequently make its diagnosis. As there are no vibration signals, no analysis is performed on the frequency spectrum. Furthermore, statistical features based on the time domain provide high performance to characterize trends and changes in the signals [237]. All the features used can be found in the papers [237], [238]. To get a better understanding of the information that each feature provides, we can see on Figure 27 its behaviour for all signals over the study time.

After generating the 10 features for each sensor in each time window, we have a matrix with a dimension of 4.558*60. For better understanding, in the first-time window, we get the average of all sensors, the standard deviation of all sensors and so on.

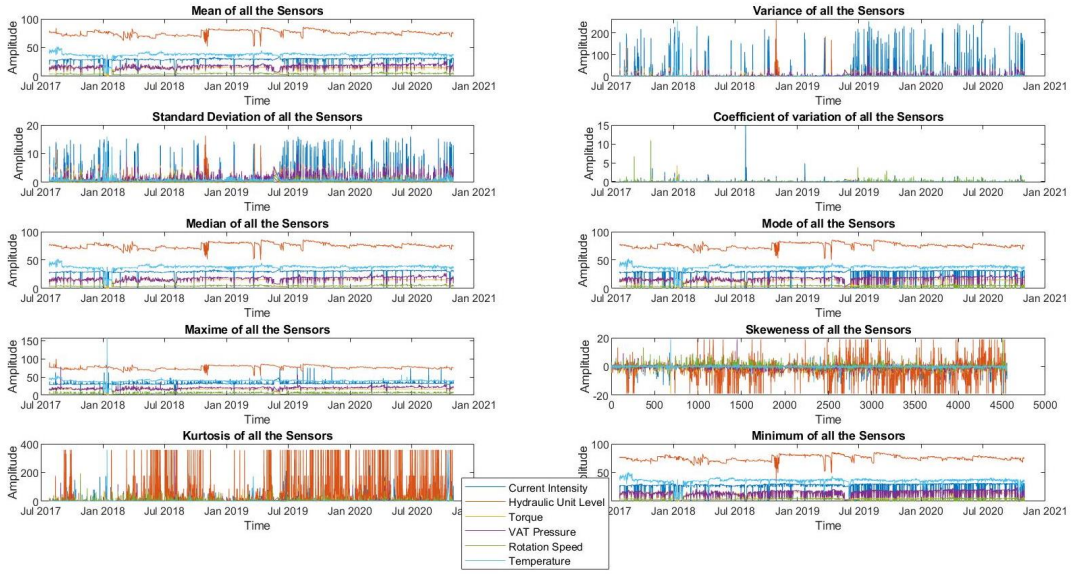



Figure 27 - representation of each of the features for each of the sensors over time.

### 4.1.2.3 PCA

Through the PCA a feature extraction is done, whose objective is to create new features with greater amount of information and, consequently, to make a dimensional reduction of the matrix created in the previous section. Applying the PCA, the matrix of 4.558*60 now has only 10 features with a dimension of 4.558*10. As it was described in the theoretical framework, the first PCs are the ones that have bigger data variance and, by consequence, the biggest percentage of information. With that, through the Table 6, we can verify the percentage variability that each one of the PCs has.

TABLE 6 - PERCENTAGE OF VARIABILITY EXPLAINED BY EACH MAIN PC.

| PRINCIPAL COMPONENTS | % OF PRESERVED DATA |
|---|---|
| **1** | 26,8299362662924 |
| **2** | 11,3444018431582 |
| **3** | 10,5209646496898 |
| **4** | 7,43400606597570 |
| **5** | 6,75048105033166 |
| **(...)** | (...) |

As can be seen, the $1^{st}$ PC has the most data variability, the $2^{nd}$ PC has the second most and so on. With this, adding up the amount of percentage of preserved data, we find that 10 PCs have ≈82% of the data. Through the Pareto diagram (Figure 28) it is possible to verify this.

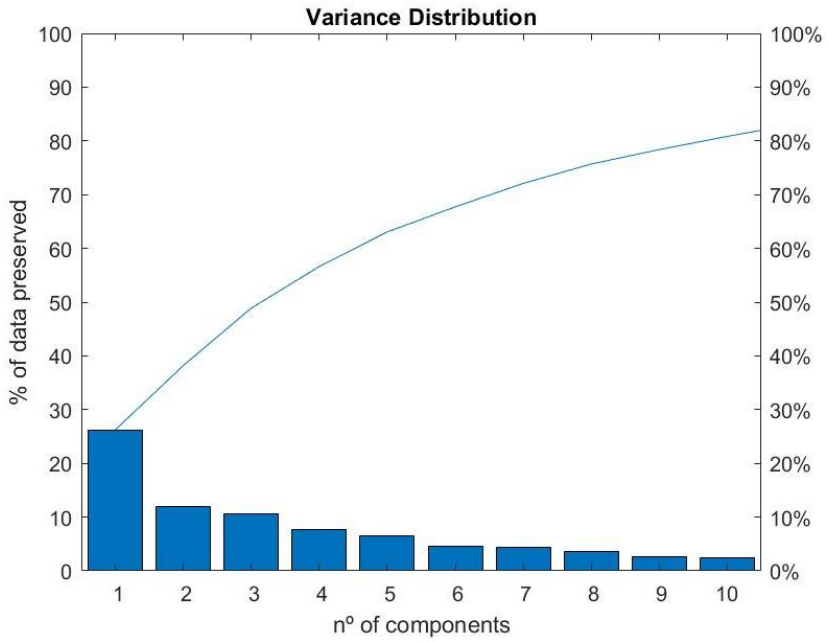



Figure 28 - Pareto diagram to explain the percentage of information preserved in each of the PCs.

In order to complement the study of this stage, the features that contribute the most to each of the PCs were identified. To this end, a heatmap matrix was made (Figure 29) which, through the coefficients of the principal components, shows which variables more contribute to the study.

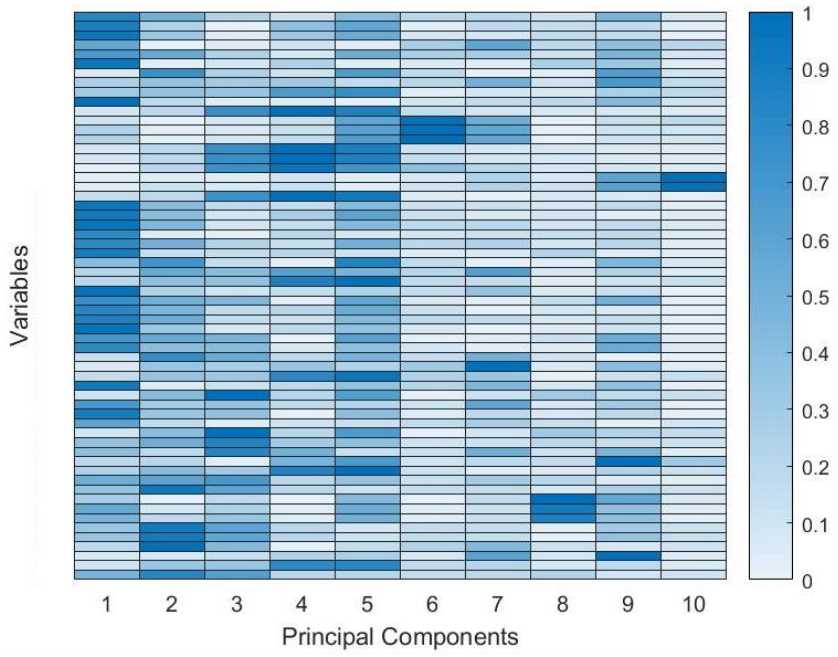



Figure 29 - Contribution of each of the features in each of the PC.

In Figure 29, the columns represent the PCs that are arranged in decreasing order of component variance, the rows represent the 60 features, generated in the second step, where the first 10 rows represent the features generated for the first sensor and so on, respecting the order of the sensors and features described above. Through this complementary study, it is possible to verify, through the darkest blue colour, which variables have greater contribution in each one of the PCs. Being the $1^{st}$ PC the one that has more data variance, it can be inferred that the features that have more influence are the ones that contribute more to this principal component.

#### 4.1.2.4 Clustering

Once the PCA is done, now the Clustering is performed, which aims to group the data in the most homogeneous way possible inside each cluster and heterogeneous between clusters. Each cluster will be identified with an observable state that has a set of points that identify each other and, consequently, a condition of the production equipment. In other words, data on good equipment functioning will be within the same cluster while the data referring to bad functioning will remain in a different cluster. Each cluster will be an observable state and each observation represents a good working state or a failed state; this will be decoded later by the HMM.

Then K-means will be used to do the clustering of the data coming from the PCA. For this, it is necessary to first determine which number of clusters best suits this data set. For that we used the silhouette analysis where, through the silhouette criterion it is

possible to evaluate which the $k$ clusters quantity is ideal. Based on Figure 30, it was determined for this set of data that $k = 3$ is the ideal value.

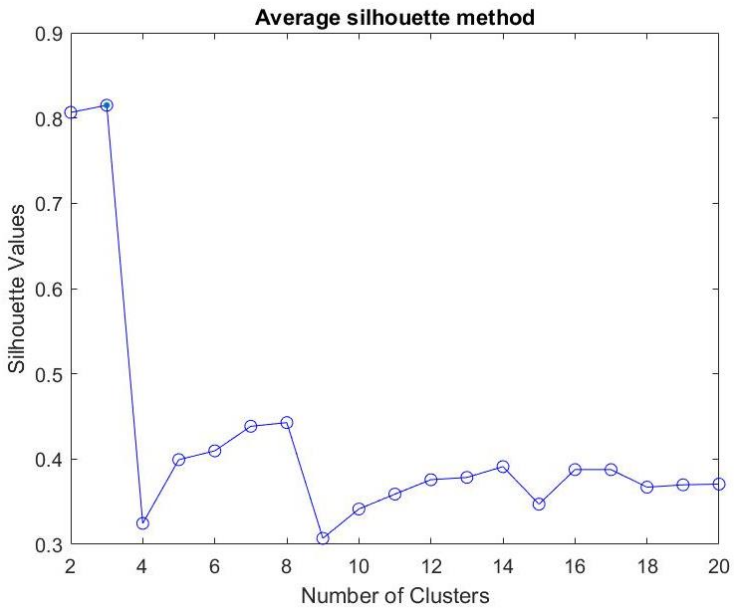



Figure 30 - Silhouette index to determine ideal number of clusters.

As each cluster is considered an observable state and they were sorted in descending order. Thus, it is possible to ensure that the $1^{st}$ Cluster is the one with the most data, the $2^{nd}$ Cluster is the second with most data and so on. In this case, the $3^{rd}$ cluster is the one that has less amount of data and, consequently, the one that is observed less during the study. In this case, the third cluster is the one that has the least amount of data and, therefore the one that is observed the fewest times during the study. This will be proven further, in the classification stage of the diagnosis of the equipment made by the HMM.

Once the number of clusters is chosen, the Clustering can be done, and each cluster represented over time. In Figure 31 it is possible to verify the development of clusters (observable states) over time.

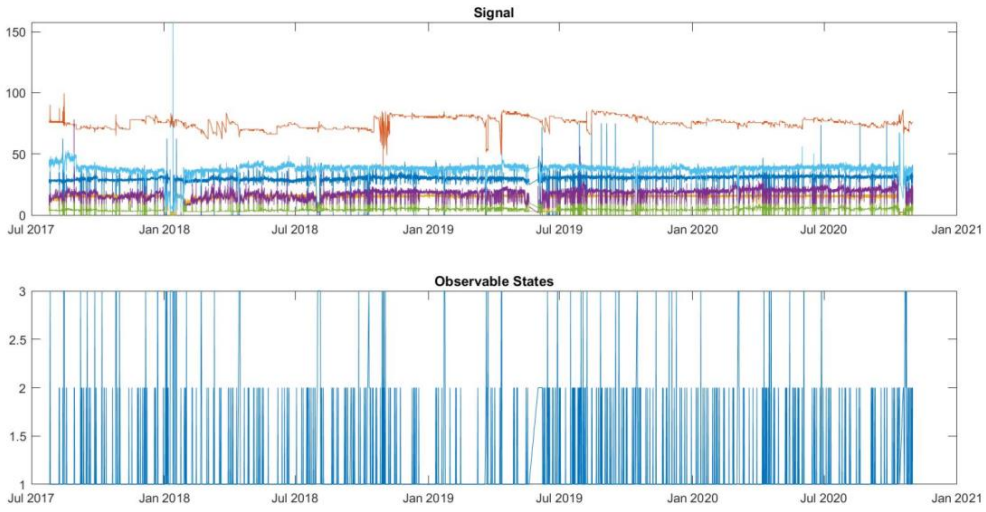



Figure 31 - Sequence of observable states over time.

To understand if Clustering performed by k-means makes sense, a check can be made through a comparison with the original signal. As can be seen in Figure 31, cluster 3, considered the rarest observable state, appears at the same time as the original observations run away from the common patterns of operation. This means that K-means did a good clustering of the data.

#### 4.1.2.5 HMM - Equipment Diagnosis

In this section, we will use the previously defined clusters as observable states to give input into the HMM. With this, it will be possible to determine the hidden states that will be the condition of the health state of the equipment. For this purpose, 3 hidden states were chosen, namely: $1^{st}$ hidden state - represents the equipment's good operation; $2^{nd}$ hidden state - represents the alert state; $3^{rd}$ hidden state - represents the equipment's bad operation.

To determine the accuracy of the HMM model, the amount of correctly validated estimates for the types of state classes of a system is checked, divided by the total monitoring of the condition of a system [239]. Then, the clusters coming from the K-means are divided into training data, with 70%, and test data with 30%. This aims to validate the HMM model. The percentages chosen to represent these values in ML models are the common practice. Since the data are time series, the division of the clusters into training and test data is also done, respecting the temporal epoch. The first 70% of the data will be used for training and 30% of the remaining data will be used for testing.

Once this is done, through the training data it is possible to do the Baum-Welch algorithm to determine the parameters, $\lambda$, of the HMM model. The transition and emission matrices obtained with training are shown below:

Transition matrix:

$$\begin{bmatrix} Stage & State\ 1 & State\ 2 & State\ 3 \\ State\ 1 & 0.9113 & 0.0857 & 0.0031 \\ State\ 2 & 0.5080 & 0.4627 & 0.0294 \\ State\ 3 & 0.0615 & 0.3291 & 0.6094 \end{bmatrix}$$

Emission Matrix:

$$\begin{bmatrix} Stage & Obs\ 1 & Obs\ 2 & Obs\ 3 \\ State\ 1 & 0.9892 & 0.0108 & 0.0000 \\ State\ 2 & 0.3133 & 0.6431 & 0.0436 \\ State\ 3 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$$

The data was split in a temporal manner, where 70% of the training data runs from August 2017 to mid-October 2019, and the remainder for the 30% of the test data. A metric was done to understand if this way of data splitting was correct. A comparison was made between the matrices trained with just the training data and the matrices trained with the full data. The metric used was the Root Mean Square Error (RMSE), represented in Equation 58.

$$RMSE = \sqrt{\frac{1}{n}\sum_{ij}^{n}(a_{Data\_Total}(i,j) - a_{\text{Data\_Train70\%}}(i,j))^2} \qquad \text{Eq. (58)}$$

The results obtained were:

$$RMSE_{\text{Transition matrix}} = 0.0709$$

$$RMSE_{\text{Emission matrix}} = 0.0289$$

Using the RMSE, we verify that the temporal split performed meets its objective and the model does not need all the data to be trained.

Having obtained the model parameters through training, it is now possible to generate observable states. This will generate, through the training matrices, a sequence of observable states with the same amount of 30% of the test data. Thus it is possible to make a comparison between the real test data, coming out of the clustering, and the data generated by the HMM, making it possible to calculate the probability, $y$, of such sequences overlapping [122]. It is possible to determine whether the observable states

generated by the training matrices correspond to the real data and to infer the prediction quality of the HMM model. This may be done using Equation 59.

$$y = (sum(HMM\_Observations == Data\_Test)/n) * 100 \quad \text{Eq. (59)}$$

Since the HMM is a statistical probability model, the set of observable states generated can vary. As such, 10,000 sequences of observable states were generated, and accuracy was performed with the test data to determine an average. With this, the accuracy of the HMM model was equal to 78.05%.

After creating the model and validating its predictive ability, it is now possible to make a graphical scheme of it, as shown in Figure 32. In the scheme, we can verify the observable states and the hidden states as well as the transition probabilities between hidden states and the emission probabilities of the hidden states for the observable states.

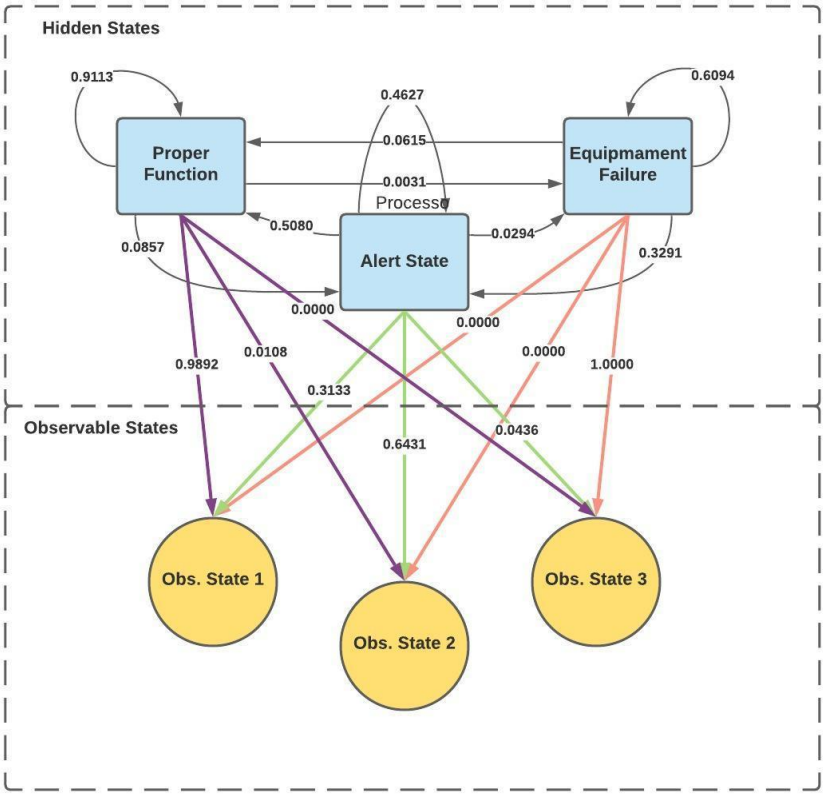



Figure 32 - Schematic of the HMM model with observable and hidden states and respective emission and transition probabilities.

Having trained the model and validated its predictive ability, we can now run the Viterbi algorithm to determine the most probable sequence of hidden states over the study time and thus determine the diagnosis of the health state of the equipment.

Figure 33 shows the evolution of the equipment's states over time.

The observable states extrapolated from the set and the "raw" signal of each sensor is also represented in Figure 33 to help understand the evolution of the hidden states.

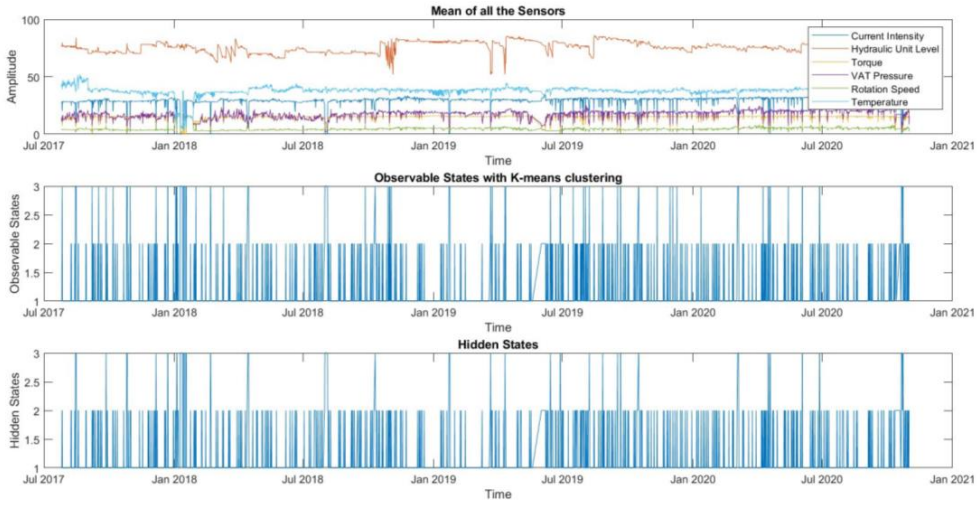



Figure 33 - Hidden states over time compared with optimised observable states and the original signal.

## 4.2 Methodology and results to diagnose behaviour patterns through vibration data and data imputation through MLP

This methodology gave rise to a paper that is currently under review in the journal Applied Soft Computing, indexed Scopus, evaluation Q1. It is already in pre-print in a journal indexed at Applied Soft Computing, with DOI:10.2139/ssrn.4194601 (Appendix C). The objective of this section is to demonstrate that, through observations collected by vibration sensors, it is possible to characterise the operating state of a component of production equipment, namely a compression roller of a drying press. It also demonstrates how it is possible to perform data imputation using a Deep Neural Network, Multi-Layer Perceptron (MLP), for a dataset with few samples. It is based on determining the health status of a pressure roller of a drying press, using values collected by vibration sensors over a period of about one year (Start date: 2020-11-05 and End date: 2021-08-21). Based on this, the objective will be to classify, during the study period, the state of the equipment through a "traffic light" classification, having three states:

1. Good working condition, shown by green colour;
2. Warning states, shown by yellow colour;
3. Malfunctioning states, shown by red colour.

### 4.2.1 Methodology

For this case, the methodology developed for the diagnosis of the production equipment component status goes through three main stages (Figure 34). Each stage is composed by different approaches that complement each other in obtaining the desired information in the context of Condition Based Maintenance (CBM).

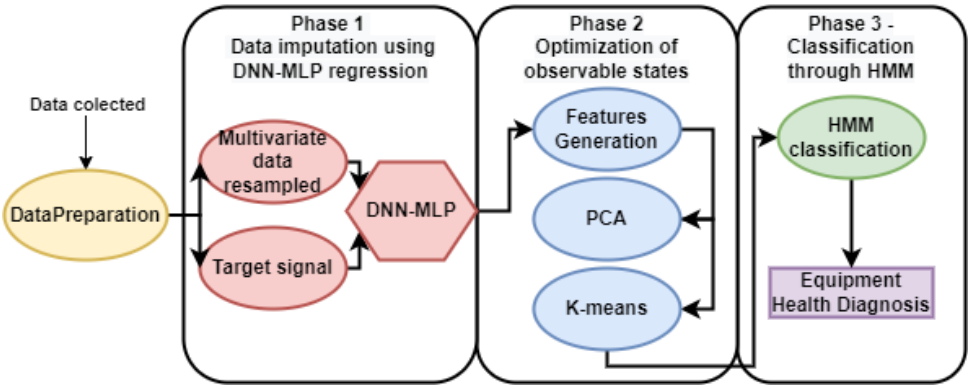



Figure 34 - Schematic of the general methodology for equipment diagnosis through a small dataset.

*Phase 1*

This first phase is responsible for imputing values in a dataset collected by a set of sensors. That is, the goal will be to impute vibration observations into a collection made over time. This is necessary since the observations were collected in widely spaced periods creating a relatively small dataset for the CBM study. Since in the same equipment there are other sensors that make data collection in a shorter period of time, it is possible, through multivariate MLP analysis (Figure 35), to perform a data imputation in the vibration dataset. Thus, it is possible to fill the missing temporal spaces, converting the vibration dataset into a vector with the same number of samples as the other sensors.

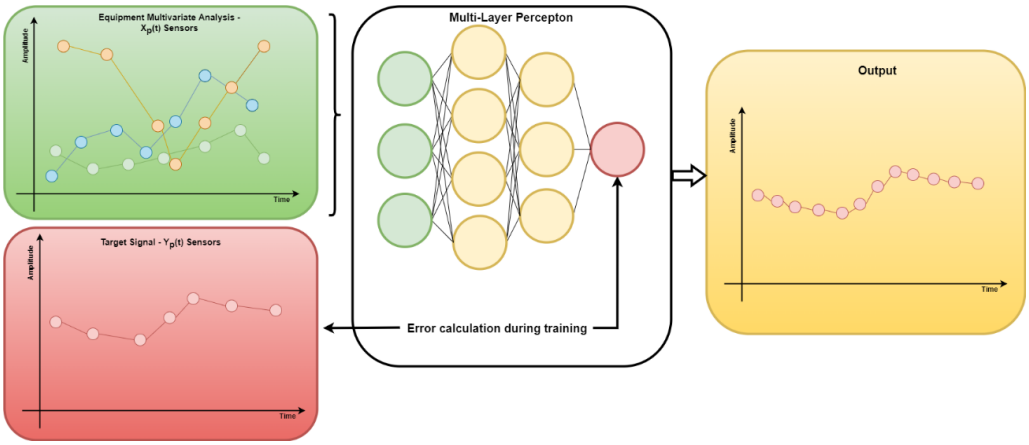



Figure 35 - Schematic diagram of how the MLP method works.

So, to use MLP, we start by resampling the multivariate sensors data to put them with the same sample size as the vibration sensors. This is done with the goal of being able to train the MLP network and get through the multivariate sensors, the vibration vectors.

As Shaik *et al.* refer [217], MLP is one of the most widely used networks for most tasks among the numerous categories of Neural Networks (NN). It has several advantages over other more sophisticated models, such as ease of implementation, short training period and creates high-quality models [215].

In all NNs, one of the main problems is to determine which Hyperparameters to use and which network architecture to use [240]. For this, in this methodology, a library in Python called KerasTuner is used that will help to choose the best Hyperparameters and the best network architecture. KerasTuner is a scalable and user-friendly Hyperparameter optimization Python library [241]. It has three search optimization algorithms, Bayesian Optimization, Hyperband, and Random Search, which help to find the best Hyperparameter values for the models. Thus, in this methodology, the Hyperparameters of the network are chosen through the library, thus obtaining more confidence in the MLP model. After obtaining an architecture with low errors, the model is used to generate a vibration vector that has the same set of samples as the dataset of the multivariate signals before resampling. Based on this, it is possible to assign new vibration observations and significantly increase the set of samples in the dataset.

*Phase 2*

Subsequently, using the new vibration observations, it is possible to initialise phase 2 of the methodology. This phase consists of optimizing the observations from the vibration sensors. It is composed of a sequence of methods (Figure 36), whose objective is to reduce the number of observations that will enter phase 3. Based on this, the computational capacity and prediction quality will be increased since new features with a greater amount of information will be used. Briefly, phase 1 is responsible for increasing the number of temporal observations on the x-axis (time) and phase 2 is responsible for decreasing the number of observations on the y-axis (amplitude), which is where the information about the health status of the component will be collected. So, through these two phases, we will have more information for more periods of time.

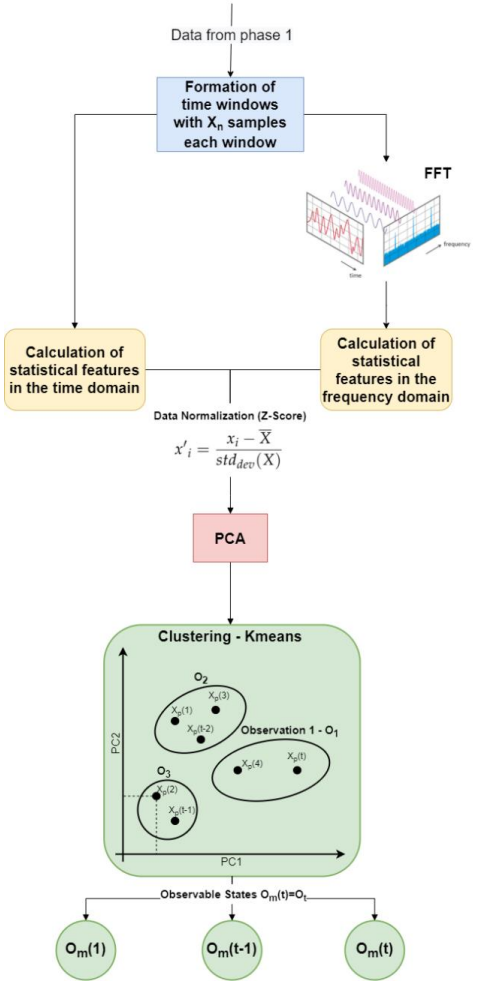



Figure 36 - Optimization scheme of observable states using FG in time and frequency domain.

Phase 1 is essential to increase the integrity of phase 2 since it starts by creating temporal windows, where each window will have $x_n$ for each block, where $n = 1, 2, \ldots, N$, and $N$ is the number of signal samples. That is, for a vector with more samples, a one-hour time window will have more samples than for a vector with few samples. Thus, by increasing the number of samples through the MLP network, we will increase the number of samples per window and, consequently obtain more reliable information. This happens because in each temporal block we will generate statistical features in the time domain and in the frequency domain. If the time window has more samples, the statics generated will be more reliable. So, it is possible to better characterize the signal over time, obtaining more critical information. A vibration signal in the time and frequency domain is usually used by the analyst to detect faults in a system [238]. As Cheng *et al.* [157] explained, the time domain degradation characteristic is one of the simplest and most effective methods of vibration signal analysis, using calculations by statistical analysis of the vibration signal. The vibration signal frequency analysis is also widely used to

diagnose equipment faults [67]. For rotational components, as is the case of the component under study, the vibration signal in the frequency domain is obtained using Fast Fourier Transform (FFT) analysis. This is the most frequently used approach that converts the temporal waveform data into frequency components [60].

Then, to generate static parameters in the time domain and in the frequency domain were used the features represented in Tables 7 and 8. Table 7 shows the mathematical equations corresponding to the statistical features generated in the time domain, where $x_n$ is the series of signals of the time domain occupied in each time block. Table 8 shows the features generated in the frequency domain, where $y_m$ is the Fourier transform for $m = 1,2,.,M$, where $M$ is the number of lines in the spectrum and $f_m$ is the frequency value of the $m^{th}$ line of the spectrum.

These features were chosen based on equipment fault diagnostics [169], [237], [238], [242].

TABLE 7 - FEATURES USED FOR TIME DOMAIN

| Parameter | Mathematical Equation | Parameter | Mathematical Equation |
|---|---|---|---|
| Mean | $T_1 = \frac{\sum_{n=1}^{N} x(n)}{N}$ | A Factor | $T_{12} = \frac{T_5}{T_2 . T_3}$ |
| Standard Deviation | $T_2 = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - T_1)^2}{N-1}}$ | B Factor | $T_{13} = \frac{T_7 . T_8}{T_2}$ |
| Variance | $T_3 = \left(\frac{\sum_{n=1}^{N}\sqrt{\lvert x(n)\rvert}}{N}\right)^2$ | SRM | $T_{14} = \left(\frac{\sum_{n=1}^{N}\sqrt{x(n)}}{N}\right)^2$ |
| RMS | $T_4 = \sqrt{\frac{\sum_{n=1}^{N}(x(n))^2}{N}}$ | SRM Shape Factor | $T_{15} = \frac{T_{14}}{T_1}$ |
| Absolute Maximum | $T_5 = \max \lvert x(n)\rvert$ | Latitude Factor | $T_{16} = \frac{T_5}{T_{14}}$ |
| Coefficient of Skewness | $T_6 = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - T_1)^3}{(N-1)T_2^3}}$ | Fifth Moment | $T_{17} = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - T_1)^5}{(N-1)T_2^5}}$ |
| Kurtosis | $T_7 = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - T_1)^4}{(N-1)T_2^4}}$ | Sixth Moment | $T_{18} = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - T_1)^6}{(N-1)T_2^6}}$ |
| Crest Factor | $T_8 = \frac{T_5}{T_4}$ | Median | $T_{19} = meadian\ x(n)$ |
| Margin Factor | $T_9 = \frac{T_5}{T_3}$ | Mode | $T_{20} = mode\ x(n)$ |
| RMS Shape Factor | $T_{10} = \frac{T_4}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$ | Minimum | $T_{21} = \min x(n)$ |
| Impulse Factor | $T_{11} = \frac{T_5}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$ | | |

TABLE 8 - FEATURES USED FOR FREQUENCY DOMAIN

| Parameter | Mathematical Equation | Parameter | Mathematical Equation |
|---|---|---|---|
| Mean | $F_1 = \frac{\sum_{m=1}^{M} y(m)}{M}$ | D Factor | $F_8 = \sqrt{\frac{\sum_{m=1}^{M} f_m^4 y(m)}{\sum_{m=1}^{M} f_m^2 y(m)}}$ |
| Variance | $F_2 = \frac{\sum_{m=1}^{M} (y(m) - F_1)^2}{(M-1)}$ | E Factor | $F_9 = \frac{\sum_{m=1}^{M} f_m^2 y(m)}{\sqrt{\sum_{m=1}^{M} y(m) \sum_{m=1}^{M} f_m^4 y(m)}}$ |
| Third Moment | $F_3 = \frac{\sum_{m=1}^{M} (y(m) - F_1)^3}{M(\sqrt{F_2})^3}$ | G Factor | $F_{10} = \frac{F_6}{F_5}$ |
| Forth Moment | $F_4 = \frac{\sum_{m=1}^{M} (y(m) - F_1)^4}{M(F_2)^2}$ | Third Moment 1 | $F_{11} = \frac{\sum_{m=1}^{M} (f_m - F_5)^3 y(m)}{M F_6^3}$ |
| Grand Mean | $F_5 = \frac{\sum_{m=1}^{M} f_m y(m)}{\sum_{m=1}^{M} y(m)}$ | Forth Moment 1 | $F_{12} = \frac{\sum_{m=1}^{M} (f_m - F_5)^4 y(m)}{M F_6^4}$ |
| Standard Deviation 1 | $F_6 = \sqrt{\frac{\sum_{m=1}^{M} (f_m - F_5)^2 y(m)}{M}}$ | J Factor r | $F_{13} = (F_7 + F_8)/F_1$ |
| C Factor | $F_7 = \sqrt{\frac{\sum_{m=1}^{M} f_m^2 y(m)}{\sum_{m=1}^{M} y(m)}}$ | | |

Having done feature generation in the time domain and frequency domain, it is now necessary to perform a standardization of the data. Since data clustering is based on the difference between value points, it is essential to have all features on the same scale; otherwise high values will be clustered regardless of their defaults [198]. This is achieved by using the Z-Score method which will remove the volume difference in the data, this being a key step of data pre-processing. It allows scaling the features through normal distribution with a mean of 0 and a standard deviation of 1. Also, because we will use the PCA method, which gives us the components that maximize the variance, standardization is essential. We assume that one feature varies less than the other due to scale differences. In that case, PCA defines the direction of the maximum variance that is closest to the axis, which varies the most if these features are not scaled [153]. Thus, standardization is essential for the PCA method to be more responsive to changes in different measurements. PCA "imposes" that the features are at the same scale, to better detect the directions that maximize the variation of the data.

Then, the dimensional reduction, performed by PCA, will be responsible for extracting new features, called Principal Components (PCs). The data starts to be distributed in a new space, where the new axes are represented by the PCs instead of the features generated in the previous step. PCA is responsible for identifying the coordinate system that best detects the variance of the data to find new features with the most information. It is a simple to use, very simple and fast method that minimizes overfitting efficiency. It can also be used to do dimensional and noise reduction in the dataset [153], [162], [178].

With the completion of the PCA method, the various features were reduced in a few PCs, which allowed to minimize a large part of the experimental noise, which could create random errors unrelated to the information contained in the data matrix, thus increasing the numerical stability of the model [153].

Now, the new features extracted through the PCs will go through a clustering phase. This is a most important unsupervised learning method, being the beginning to obtain knowledge that deals with the partitioning of databases under unexpected conditions [187]. It aims to group similar data in the same cluster that will be distinct from data in other clusters. The observations within each cluster will be homogeneous and heterogeneous across clusters, which will cause well-functioning data to be in the same cluster while poorly functioning data to be in a distinct cluster. Thus, each cluster can be seen as an observable state, where the cluster with the highest amount of data will be the observation corresponding to the good functioning of the equipment. In this way, the clusters are ordered in descending order so, the first cluster is the one with the most data and the last one contains the least amount of data. This will enable the last cluster to represent the observation of equipment malfunction since it is the one that occurs least frequently. So, this methodology will be used for clustering the K-means method that uses the Euclidean metric. This is an algorithm well regarded by the scientific community, easy to implement, unsupervised, with fast convergence speed and that does not require large computational resources [187]. For everything else, it is an algorithm that works well in conjunction with PCA, since it can be applied after reorienting the data source from the metric frame of the point cloud core. In other words, k-means works better since the data are centred with respect to an origin, being relatively equidistant from the centre and placed in the different quadrants of the space $Rp$; where $p$ represents the principal components of PCA [153].

*Phase 3*

After phase two, which aimed to optimize the observations that enter the HMM, we now begin phase three. This phase is the classification of the health status of the production equipment component. For this, the HMM method will be used which, through the observable states that come out from the clustering will classify them into hidden states that represent the operating status of the component. We can verify in Figure 37 how the HMM works in this phase.

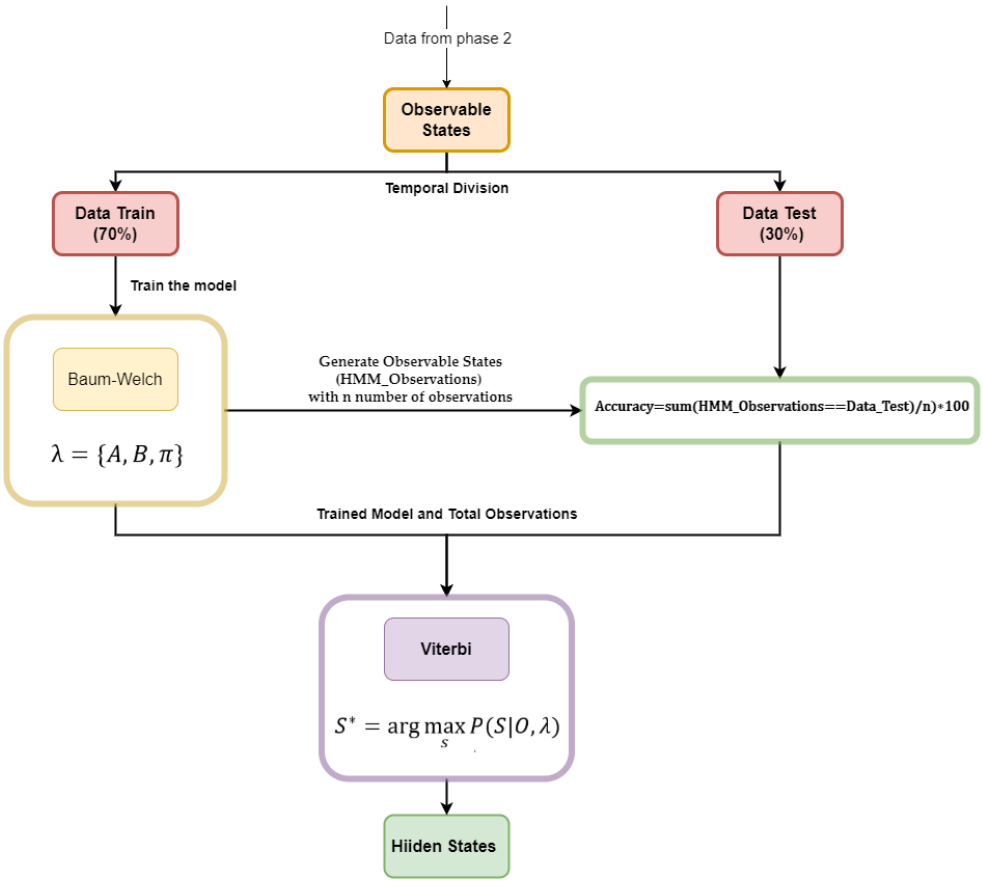



Figure 37 - Operation of the HMM algorithm in the diagnostic classification phase.

The observable states from k-means will be divided into training data (70%) and test data (30%). With the training observations, the model will be trained using the Baum-Welch algorithm to obtain the HMM model parameters, $\lambda = \{A, B, \pi\}$. After obtaining the transition and emission probability matrices, it is now possible to generate a sequence of observable states and compare it with the sequence of observable states from the clustering test data. Being able to obtain the Accuracy through the equation shown in Figure 37, we will now run the *Viterbi* algorithm which aims to determine, through the observable states and the model parameters, which is the best path to the hidden states over time. These hidden states represent the state of health of the equipment and, for this study three states were defined: State 1- Good Operation; State 2 - Alert; State 3- Failure.

### 4.2.2 Results

#### 4.2.2.1 Data Preparation

For this study, 4 different types of vibration sensors were used: *vibra-m acc*; *vibra-m op acc*; *vibra-facc*; *vibra-f op acc*. The sensors collect data at each 4 hours (Figure 38),

which corresponded to 6 observations per day. To create one-day time windows and generate features, 6 observations correspond to a low amount of data. To solve this problem, observations of other sensors were used, also attached to the same equipment, which collect samples every 5 min. These complementary sensors were composed by 6 different variables (Figure 39): Electric current; Hydraulic unit level; Torque pressure; VAT pressure; Press rotation speed; Hydraulic unit temperature.

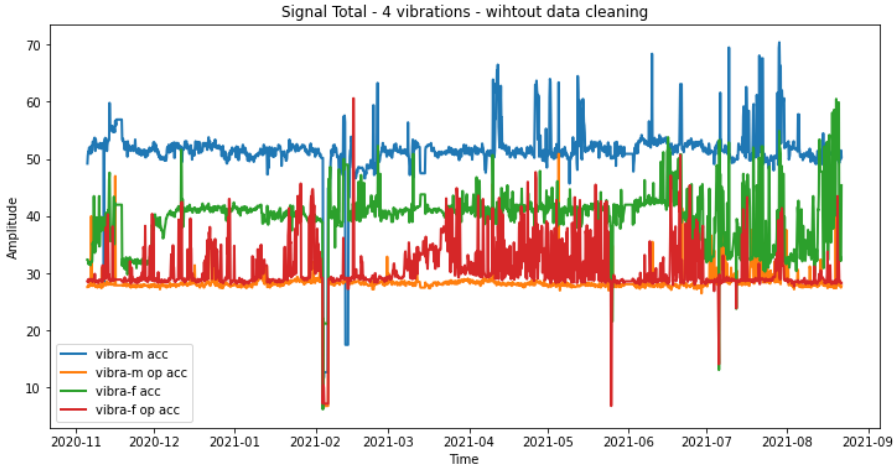



Figure 38 - Four vibration sensors with data collected every four hours.

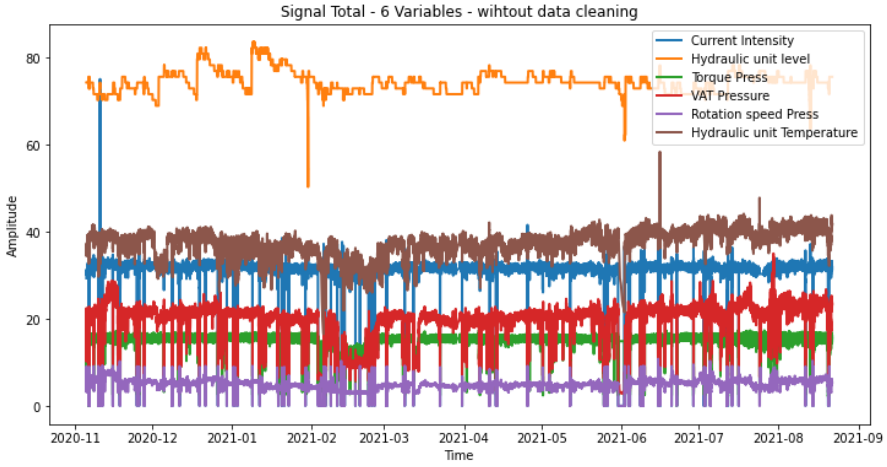



Figure 39 - Six complementary sensors with data collected every five minutes.

To improve the integrity of the dataset, both vibration and complementary sensors, some preparations were made on the data, namely, to replace the incoherent values $(+inf; -inf; -inf; Nan)$ with the average of the respective signal. Equipment shutdowns were also detected. It was considered that whenever Current Intensity, Torque Pressure, VAT Pressure, and Rotation Speed were below a certain value all at the same time this would be considered as a shutdown of the equipment (Figure 40). These shutdown periods were also replaced by the respective average of each signal, for all

signals. In this way, it is prevented the equipment shutdowns affect the prediction of the adopted methodology.

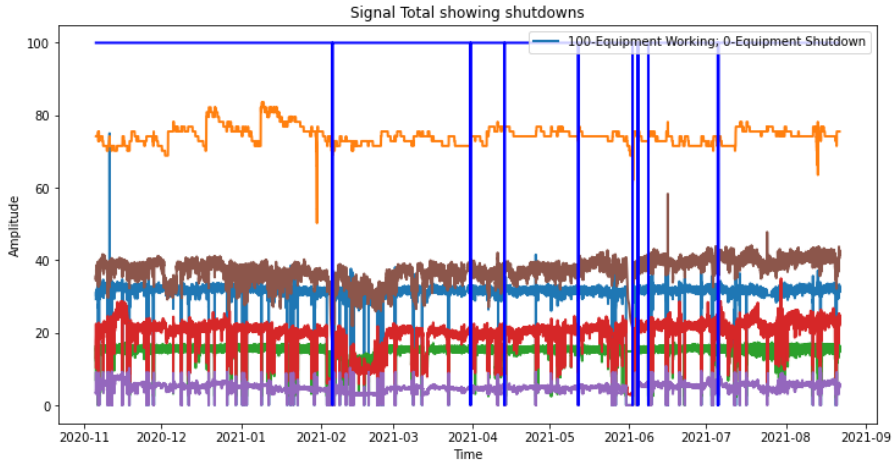



Figure 40 – Representation of equipment shutdowns.

No additional treatment was added to the data since here we consider the unusual values as something that could add value to the prediction and not outliers. Furthermore, the methodology used in phase 2 makes all the filtering and reduces the noise in the data set.

#### 4.2.2.2 Data imputation using MLP

After increasing the integrity of the data in the dataset preparation, the values of the complementary sensors were resized to have a vector of the same size as the vibration sensors (Figure 41). Thus, we now have a dataset with values collected only every 4 hours that can be used to create an MLP model that can predict the observations of the vibration sensors. After creating the MLP model, it is possible, using the original data from the complementary sensors (collected every 5 minutes), to generate vibration signal data with observations every five minutes.

The six complementary sensors were used as input variables for each of the vibration signals used as the output of the MLP model. After the rescaling of the complementary sensors, the data were divided into training data (70%) and test data (30%). Using the KerasTuner library and random search, the best architectures with the best Hyperparameters were chosen for each of the MLP models (for each of the vibration signals), which obtained the lowest Mean Absolute Error (MAE). Using KerasTuner the following were determined: Number of layers; Number of neurons per layer; Activation function; Learning rate.

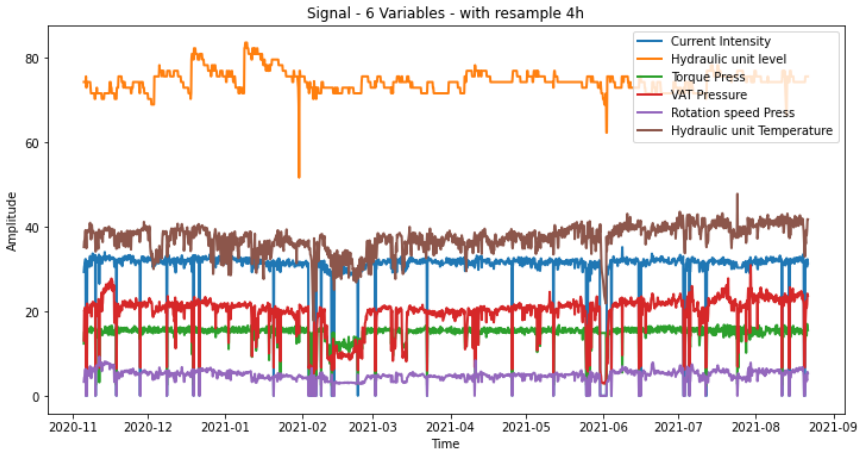



Figure 41 - Six complementary variables with resizing to values each four hours.

Through Table 9 it is possible to see which structures were determined by KerasTuner for each of the MLP models created to obtain each of the vibration signals.

TABLE 9 – MLP NETWORK STRUCTURE FOR EACH VIBRATION SIGNAL

| **Hyperparameters** | ***vibra-m acc*** | ***vibra-m op acc*** | ***vibra-f acc*** | ***vibra-f op acc*** |
|---|---|---|---|---|
| **Number of Layers:** | 8 | 9 | 3 | 4 |
| **1st Layers** | 288 Neurons | 128 | 352 Neurons | 160 Neurons |
| **Activation function** | tanh | tanh | tanh | relu |
| **2nd Layer** | 160 Neurons | 160 | 160 Neurons | 160 Neurons |
| **Activation function** | tanh | tanh | relu | tanh |
| **3rd Layer** | 320 Neurons | 192 | 384 Neurons | 192 Neurons |
| **Activation function** | tanh | tanh | tanh | relu |
| **4th Layer** | 416 Neurons | 64 | | 352 Neurons |
| **Activation function** | relu | tanh | | relu |
| **5th Layer** | 128 Neurons | 384 | | |
| **Activation function** | tanh | relu | | |
| **6th Layer** | 448 Neurons | 256 | | |
| **Activation function** | relu | relu | | |
| **7th Layer** | 224 Neurons | 384 | | |
| **Activation function** | relu | tanh | | |
| **8th Layer** | 352 Neurons | 512 | | |
| **Activation function** | | relu | | |
| **9th Layer** | | 32 | | |
| **Activation function** | | relu | | |
| **Learning rate** | 0.0001 | 0.001 | 0.001 | 0.001 |
| **Score - MAE** | 1.8357 | 0.9170 | 2.9033 | 2.5250 |

After choosing the best architecture for each one of the MLP models, the predictions were made for each of the vibration signals. Through Figures 42-46, it is possible to verify the

comparison between the original vibration data and the data predicted by the MLP models.

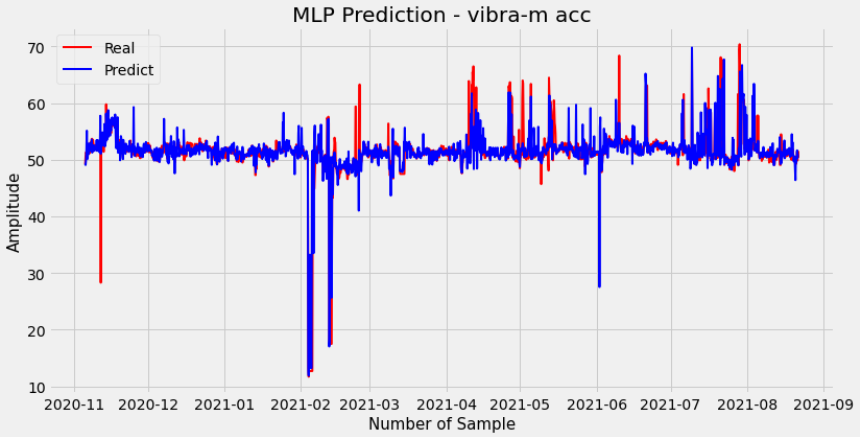



Figure 42 - MLP prediction *versus* Actual values for the *Vibra-m acc* signal.

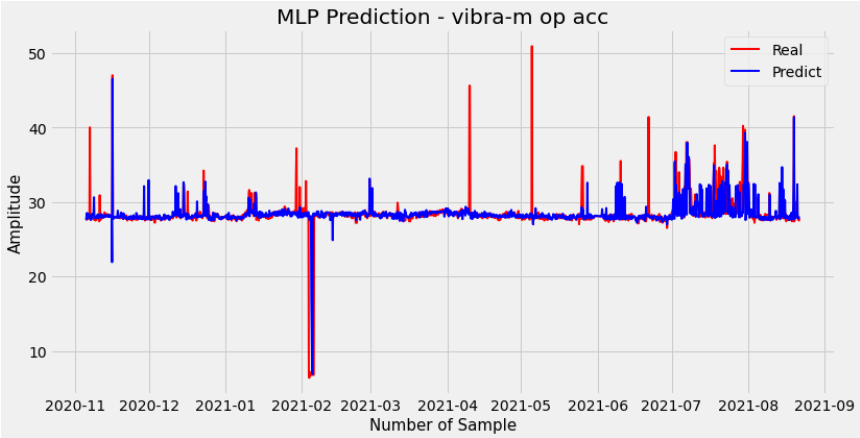



Figure 43 - MLP prediction *versus* Actual values for the *Vibra-m op acc* signal.

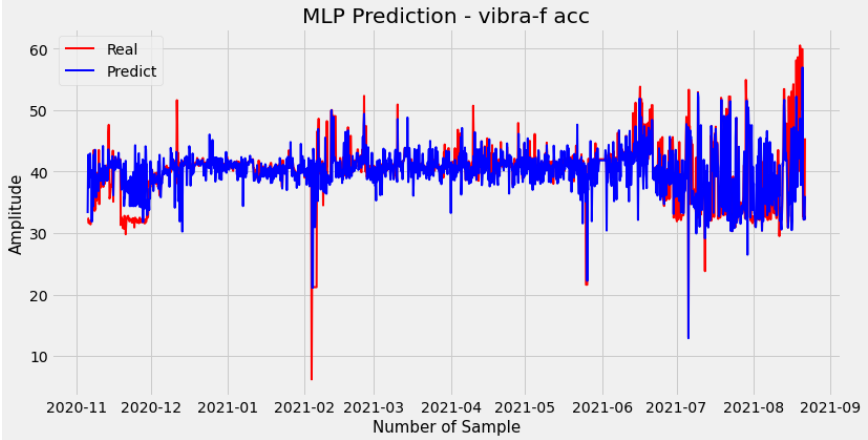



Figure 44 - MLP prediction *versus* Actual values for the *Vibra-f acc* signal.

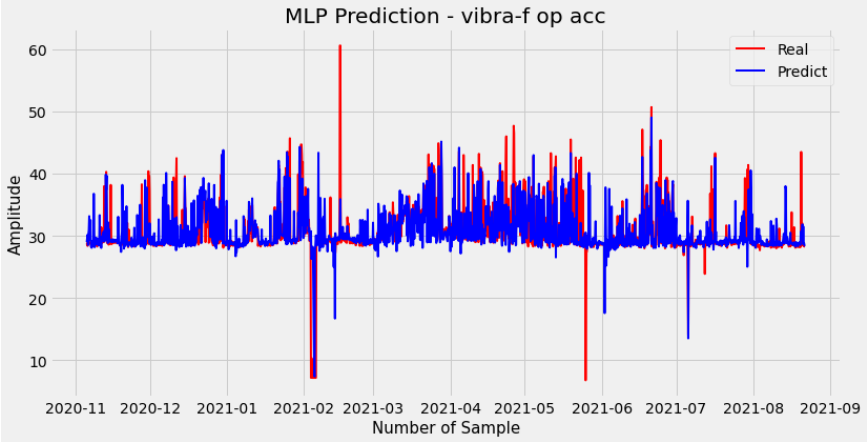



Figure 45 - MLP prediction *versus* Actual values for the *Vibra-f op acc* signal.

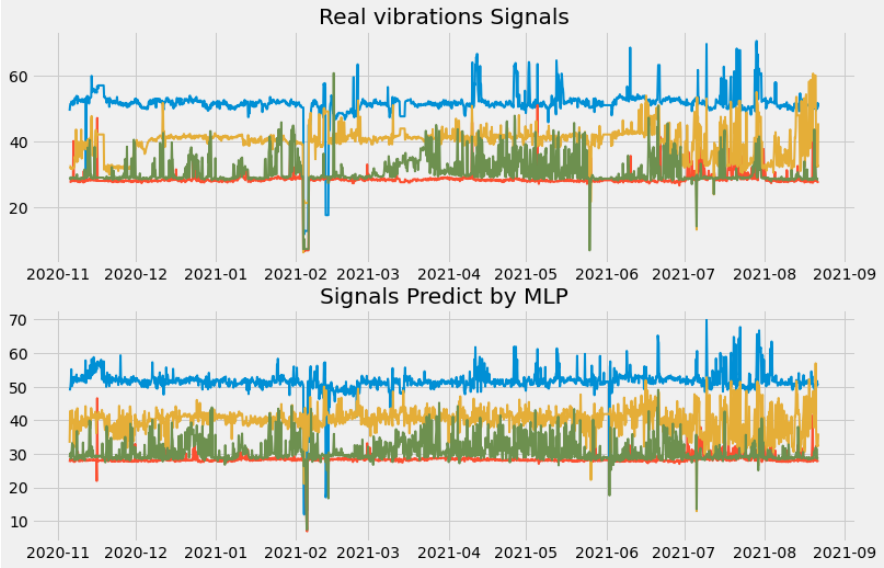



Figure 46 - All Actual Vibration Data *versus* MLP Predicted Vibration Data.

As we can see in Figures 42-46, the MLP models created were able to make good predictions of the vibration sensors. Therefore, we can use these models to create vibration signals with observations every 5 minutes. For this, the created models and the original complementary sensors were used with collections made every 5 minutes. Based on this, the vibration signals were created with observations every 5 minutes (Figure 47) which will be the data used for the continuation of the study of the health diagnosis of the pressure roller of the drying press.

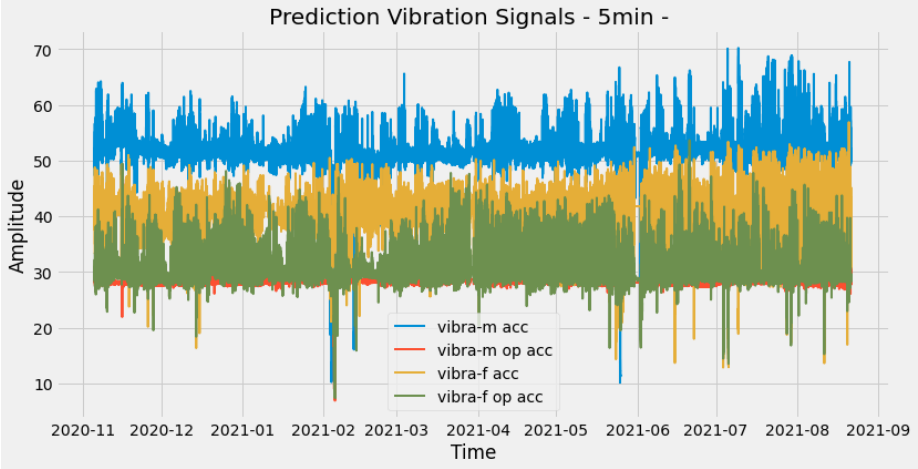



Figure 47 - Five-minute vibration data predicted by MLP.

#### 4.2.2.3 Features Generation

Having a dataset with a large set of observations, it is possible to create one-day time windows. With observations every 5 minutes the dataset will have 83,329 samples. Thus, each one-day time window will have a set of 288 observations per chunk with a total of 290 chunks. Through phase 1, we no longer have 6 samples per time window, but 288 samples. In this way it is possible to create features for a more considerable number of observations. Statistical features will be created in the time domain and frequency domain, described in Tables 7 and 8. In this way, for each sensor, 21 statistical features were created in the time domain and 13 in the frequency domain, thus obtaining a $290 * 140$ matrix. After generated the matrix of the features, it is necessary to return to standardization to remove the difference in volume of the data. Because of this the Z-Score method was used, where the data assume a conventional normal distribution, having a mean of 0 and a standard deviation of 1. We can verify, through Figure 48, the behaviour of the features generated, standardized along the study time.

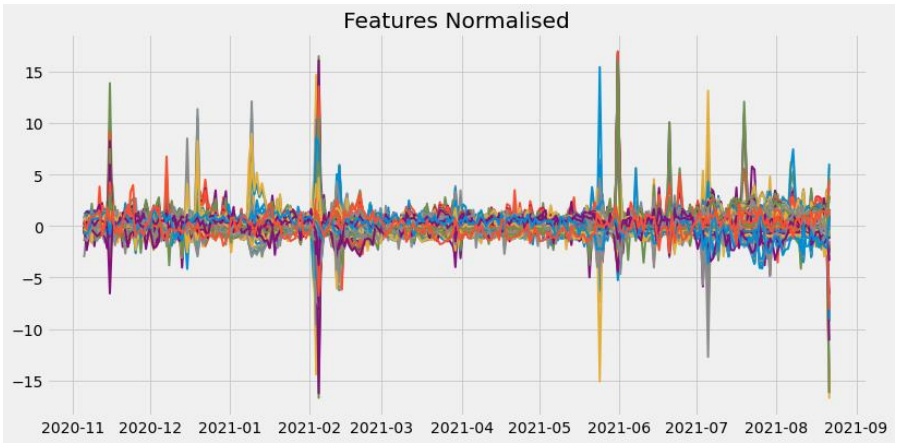



Figure 48 – Standardized generated features values over time.

#### 4.2.2.4 Principal Components Analysis (PCA)

Now the PCA method is used with the objective of making a dimensional reduction of the matrix coming from the generation of characteristics and the extraction of new variables with a greater amount of information. So, through this step, we go from a $290 * 140$ matrix to a $290 * 10$ matrix. Thus, we no longer have the 140 characteristics generated, but only 10 PCs. This was the number of PCs chosen since it preserves about $\cong 85\%$ of the variability of the data. This can be seen through a Pareto analysis in Figure 49. Through this method, a considerable dimensional reduction was made and most of the percentage of information stored in the data was kept.

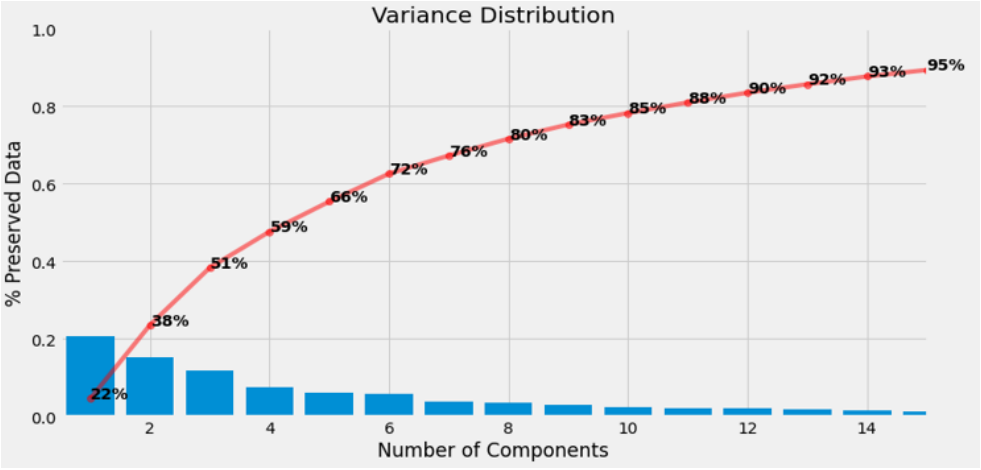



Figure 49 - Pareto diagram with the percentage of data variability for each PC.

PCA can also be used to get a better visualization of the data. As we saw above in the theoretical framework, the first PCs are the ones with higher data variation. As such, in Figure 50 we can see in 3D how the data from the first 3 PCs are distributed and what is the variability percentage of each one of them.

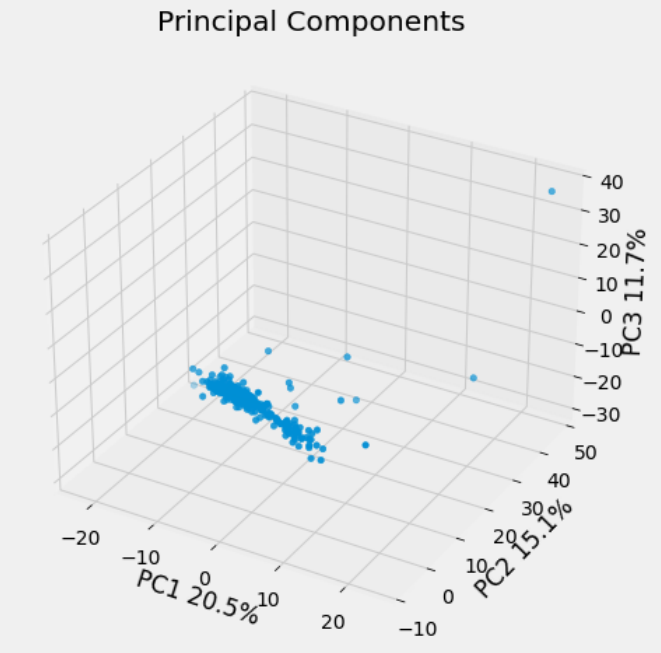



Figure 50 - Distribution of data in the first three PCs.

#### 4.2.2.5 K-means

After orthogonal redistribution and the creation of new variables, the PCs, the data will be clustered in the matrix resulting from the previous step. Here, data will be grouped into different clusters, where each cluster will correspond to an observable state that will be entered later in phase 3. As the method defined to be used here was the K-means, it is necessary to determine, *a priori*, the number of clusters k. K-means is the most popular clustering method for minimizing the Sum of Squared Error (SSE) [243]. Therefore, the elbow method was the method for determining the k clusters chosen for this study. This is one of the most widely used methods, which serves to measure the quality of a clustering structure and uses a graph where the SSE *versus* number of clusters is plotted (Figure 51) [244].

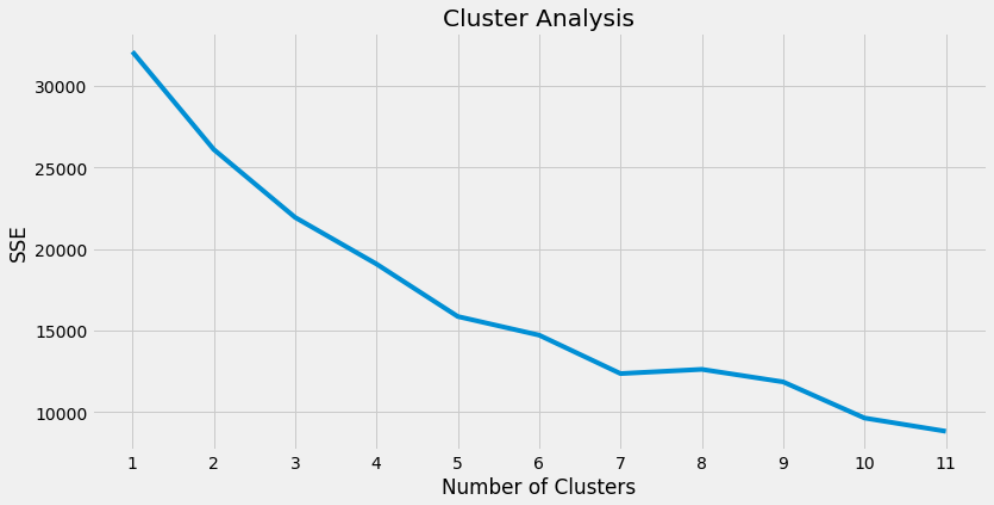



Figure 51 - SSE vs $k$ Clusters – Elbow method.

As we can see from the graph, the elbow is formed with $k = 7$. From that number of clusters, the SSE value starts to decrease more slowly. This leads us to believe that from that point, the clusters formed are smaller clusters formed from the ones chosen previously and do not add improvements to the SSE [245]. Thus, the number of clusters chosen for this study is 7. Based on Figure 52 it is demonstrated how the clusters are divided in the first three variables coming from the PCA.

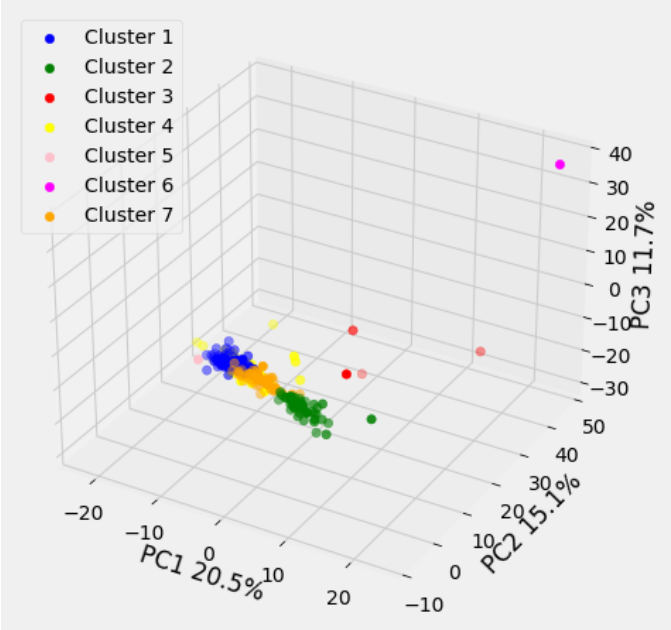



Figure 52 – Each Clusters represented in the first 3 Principal Components.

Then, we have 7 observable states that are ordered in descending order, where the first cluster has a larger amount of data and the last cluster has the least amount. This aims to ensure that observable state 7 is the rarest to happen and observation 1 the one that

happens more often. This could mean that the first observable states are related to the good functioning of the component while the last states could represent the bad functioning. This will be subsequently classified by the HMM method. Now, the clusters are defined, being represented in a temporal form, as can be seen in Figure 53.

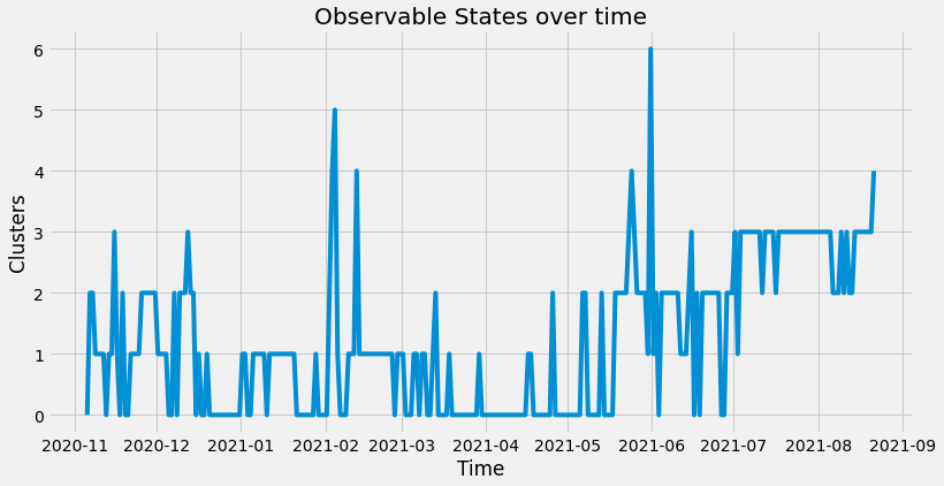



Figure 53 - Time series of clusters over study time.

#### 4.2.2.6 Hidden Markov Models (HMM)

In this last phase, the observable states are classified and converted into hidden states through the HMM method. For this purpose, three hidden states were chosen, corresponding to the diagnosis of the component under study and representing respectively: $1^{st}$ State of Good Operation; $2^{nd}$ State of Warning; $3^{rd}$ the State of Bad Operation.

First, the HMM model must be trained, using the Baum-Welch algorithm, to determine their $\lambda$ parameters. After training, it is possible to determine which hidden state sequence best represents the observable states over time. For this, the *Viterbi* algorithm is used.
Starting with training, as we saw in the description of the methodology, it is necessary to divide the observable states into training data and test data. As is common practice in ML, for the training data 70% of the data is used and for the test 30%. Then, after performing the training of the HMM model with the training observations it is possible to generate, through the λ parameters, a sequence of observations with the same size as the sequence of observations of the test data. Thus, as explained in the methodology, it is possible to determine the accuracy of the model by dividing the number of successfully validated estimates for all types of events or state classes of a system by the entire condition monitoring of a system [126], [153]. To determine the sequence, it is necessary to overlap of the observable states created by the HMM model with the actual 30%

observable states from clustering and determine the accuracy of the model, the following equation is used (Eq. 60):

$$Accuracy = \left(\sum \frac{(HMM_{Observation} = Data_{Test})}{n_{samples}}\right) * 100 \qquad \text{Eq. (60)}$$

As the HMM parameters, $\lambda$, are based on probabilities, each sequence of observations can be different. As such, 10,000 sequences of observations were created, and the respective accuracy was calculated for each of them. Then, the model's accuracy was averaged. For this model, with seven observable states and three hidden states, the model only obtained an accuracy of 32.26%.

To understand the low evaluation of the model, some tests were made with other HMM models, where the number of hidden states was varied and the order of the signal of observable states was modified. That is, the signal was divided into 40%, 30% and 30% temporal blocks. After that, the 30% temporal blocks were changed, leaving the signal altered as is shown in Figure 54.

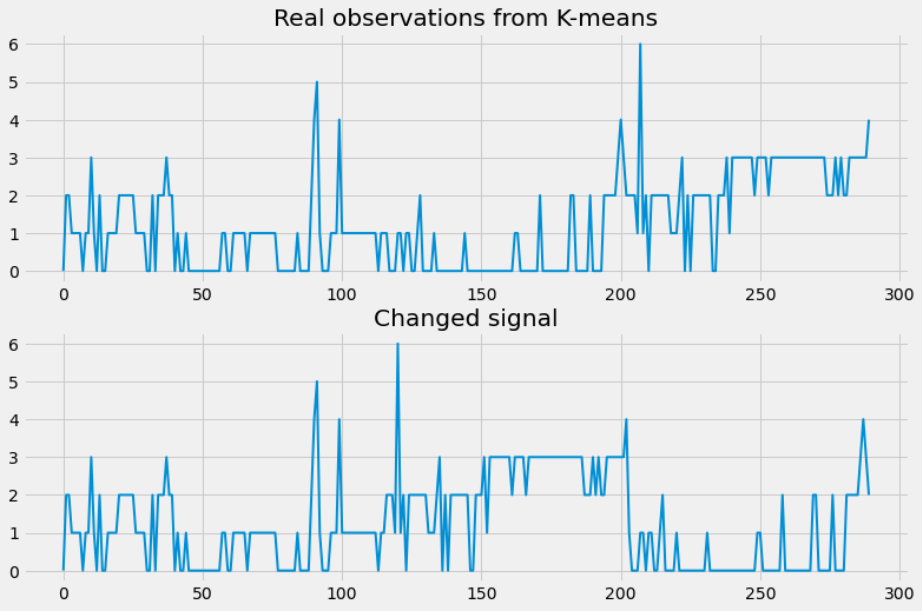



Figure 54 - Original Observations and Modified Observations.

Once the signal order is switched, all observations are now in the training data. So, here the goal is to understand if the signal size has an impact on the accuracy of the model. The HMM is a model that needs that all observations to have occurred in the training phase for it to perform well, whose results are presented in Table 10.

TABLE 10 – ACCURACY AND LOG-LIKELIHOOD FOR EACH SEQUENCE OF OBSERVATIONS AND DIFFERENT NUMBER OF HIDDEN STATES

| | **Real Observation from K-means** | **Changed Signal** |
|---|---|---|
| **Nº. of Hidden States** | **3** | |
| Accuracy | 0.32264367816091954 | 0.48747126436781607 |
| Log-Likelihood | -248.86347140364788 | -285.5781288565171 |
| **Nº. of Hidden States** | **4** | |
| Accuracy | 0.2695402298850575 | 0.6603448275862069 |
| Log-Likelihood | -245.69196936616808 | -282.9993414180359 |
| **Nº. of Hidden States** | **5** | |
| Accuracy | 0.27206896551724136 | 0.7096551724137931 |
| Log-Likelihood | -244.29181521910746 | -278.9731023761346 |
| **Nº. of Hidden States** | **6** | |
| Accuracy | 0.37091954022988505 | 0.7295402298850575 |
| Log-Likelihood | -242.7765380922869 | -278.67744793543966 |
| **Nº. of Hidden States** | **7** | |
| Accuracy | 0.3671264367816092 | 0.7045977011494253 |
| Log-Likelihood | -242.77622604230996 | -278.7458634574436 |

As can be seen in the results obtained in Table 10, there is an improvement in the model evaluation when we change the sign of observable states and put all observations in the training data. Based on this, it can be inferred that model has a more complete training when the Baum-Welch algorithm uses all observations coming from clustering. Together with Accuracy, it was also calculated the logarithmic probability, which is responsible for characterizing the ability of the model to adapt to the observations and, consequently, to use the Viterbi algorithm. The log-likelihood is calculated to verify that each data sequence fits the model [246]. We can verify from Table 10 that the log-likelihood values are all in the same order of magnitude. This means that although the model does not have good Accuracy, because the sequence of observations is short and there are no cycles of repetitions of observations, it is able to adapt to the observations and make a good classification of the hidden states. The model would have better accuracy if the sequence of observable states was longer, collected over a longer period, where there was already, at least, one cycle of observations represented in the training data.

Once we have verified that the algorithm performs well in the classification of hidden states, we can now apply the Viterbi algorithm to the original set of observable states with the training parameters defined by the Baum-Welch algorithm. Based on this, we obtain the sequence of hidden states demonstrated in Figure 55. The traffic light method was applied where the green colour is represented for state 1, yellow represents state 2 and red colour state 3.

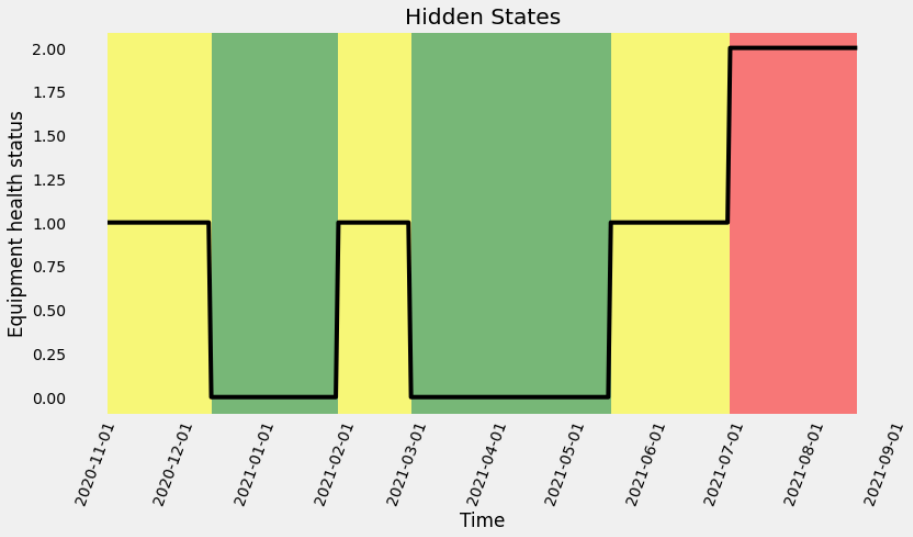



Figure 55 - Health status of the equipment over time with traffic light signalling for each of the states.

## 4.3 Methodology and results to prognostic 7 days ahead in the equipment patterns

This methodology gave origin to a paper submitted to the Energies journal, indexed Scopus, evaluation Q1, with DOI: 10.3390/en16062651 (Appendix D). Aims to demonstrate a methodology for diagnosis and prognosis of production equipment, namely a drying press used in the paper industry, using several unsupervised ML and DNN tools. Through the generation of features in time windows and the use of several ML tools, such as dimensional reduction and clustering, it is possible, through HMM, to classify the health state of the equipment. To perform the prognosis and predict in a period of 7 days ahead, the state of health of the equipment used the GRU tool, that directly manages the Inputs and Outputs of the HMM. With this, 3 hidden states are classified, resulting from the HMM, representing the status of the equipment: State 1 - Good Operation; State 2 - Alert; State 3 - Bad Operation.

### 4.3.1 Methodology

To better explain the methodology used in this study, we can return to the diagram represented in Figure 56, where the whole process is schematised.

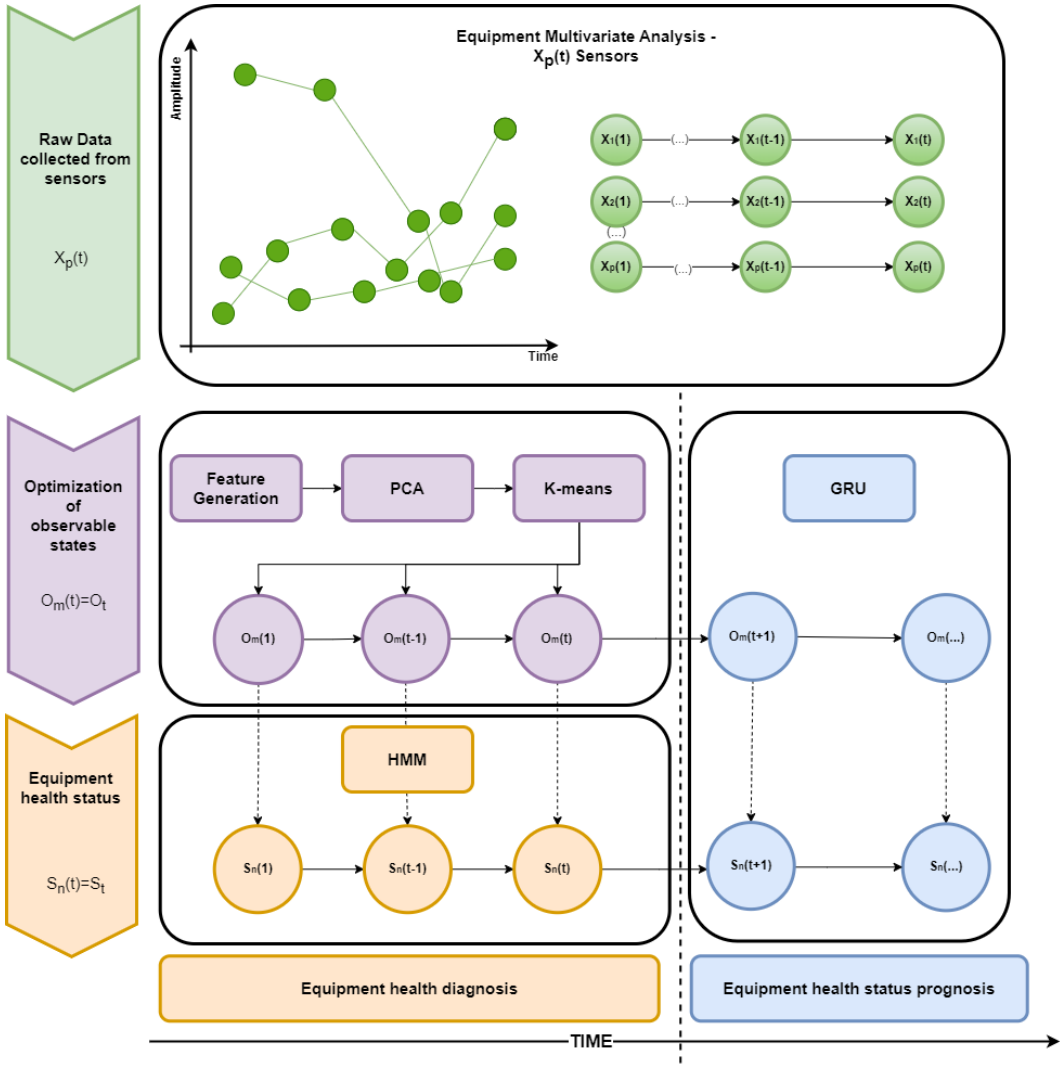



Figure 56 - Methodology used to perform the diagnosis and prognosis of the state of health of a production equipment.

The methodology starts by collecting data over time, through several different sensors, $X_p(t)$, attached to the production equipment. After collecting the data, it goes through a cleaning stage. Here, the goal is to eliminate or replace incoherent data that reduce the integrity of the dataset and, consequently the predictive capacity of the methodology created. Next, a phase of optimization of the observations collected by the sensors is performed, called the optimization phase of observable states. The goal is to improve the observations collected by the sensors to give input to the HMM algorithm, which will then map into hidden states. Therefore, the optimization of observable states will start by a Features Generation (FG) step in the time domain. In this way, it will be possible to obtain more information about the data, since continuous and time-varying features can predict possible failures [154]. This step allows, through the created features to acquire more and better information, enhancing the prediction performance compared to the

original signals [155]. To create the features, it is first necessary to use a temporal window processing method [154]. Thus, temporal windows can be created with variable time intervals, depending on the study that is being performed. The method is used in the time domain, where several features are generated that can obtain more information about the health status of the equipment. For this purpose they were used characteristics collected from the following papers: [153], [156], [237], [238].

By creating features in the time domain, the number of variables has increased. It is necessary to make a feature selection and, consequently a dimensional reduction. For that, it is used the PCA method. This method is responsible for creating new features, the PCs, which retain more data variance and, therefore more information. In this way, it is possible to work with more valuable features and, at the same time, to solve the curse of dimensionality and increase computational capacity. PCA is among the most widely used dimensional reduction and data compression methods, demonstrating superior performance in terms of reconstruction error [175]. The decision-making process is accelerated from the perspective of equipment health diagnosis due to the reduction of class representation [174]. Furthermore, PCA is applied to remove the highly correlated features based on the correlation matrix. Since highly correlated features lead to overfitting, this step will increase the prediction accuracy of the model [247]. The data is transformed into uncorrelated features that help in converting data from a high-dimensional space to a low-dimensional one, retaining as much information as possible [248]. It also considers the potential correlations between the response variables without increasing the computational complexity [249]. To improve the input of the variables in the PCA method, since they have large differences in amplitudes, it is necessary to use standardization of the data. In this way, the amplitude of the initial continuous variables will contribute equally to the analysis. We thus guarantee that variables of higher amplitudes will not overlap with features of lower amplitudes, and not cause one-sided results [160]. Therefore, to transform the data into comparable scales, the Z-Score standardization (Eq. 61) is used, being each feature value subtracted from the mean and divided by the standard deviation. This ensures that all features are standardised to a mean of 0 and standard deviation of 1.

$$Z_{score} = \frac{x_i - \bar{X}}{std_{dev}(X)} \qquad \text{Eq. (61)}$$

After passing through a standardization, the data enter into the PCA algorithm which will be responsible for reorienting the orthogonal axes moving them in the direction of the biggest variability of the data. These new axes will be called Principal Components and contribute to improving data visualization and evaluation so, the differences

between observations are more visible. The original data are mapped to a new vector, transforming the original response variables into uncorrelated principal components [249]. PCA is also a method that matches very well when clustering is needed later, as is the case with this methodology [156]. PCA uses techniques of projecting high-dimensional spaces to lower-dimensional subspaces, trying to retain all relevant information [250].

After creating the new variables, the PCs will feed into the clustering process. This is a multivariate analysis technique used to understand the degree of similarity between data and their classification into different clusters [190], [201]. Clustering is performed in a temporal manner to understand which clusters are forming over time. They are ordered in a decreasing order relative to the amount of data they have. In this way, cluster 1 is the one with the most data and the last cluster has the least amount of data. Cluster one will be the one that appears most frequently over time and the last cluster, since it has less data, is the one that appears most rarely. The clustering in this methodology is performed by the K-means algorithm. This is an unsupervised learning algorithm, used to highlight the intrinsic properties and laws of the data [199]. This algorithm was chosen because it is [199], [251]: easy to use; easy to implement; has fast convergence; has strong interpretation ability; has ability to handle large amounts of data effectively. K-means is among the most fruitful methods for data mining, summarization probability density estimation and many other essential tasks [252]. It is among the most widely used algorithms in data analysis due to the conciseness of its results and its high scalability [253]. It is suitable for large-scale original failure scenario reduction [254]. To start the algorithm, it is first necessary to choose the number of clusters, k. It aims to minimize the Sum of Squared Error (SSE) of the distances between the data points and the centroid of the assigned cluster. The distance between the data points and the centroid of the cluster is calculated allowing the data to be assigned to the closest cluster. These calculations are repeated several times until convergence [201]. Once the clusters are created, they can now be managed as observable states where each cluster represents an observation - these will be the new inputs to the HMM. Through the observable states, created by k-means, the HMM will be able to train and, subsequently map the observations into hidden states, which will represent the diagnosis of the equipment. Since the HMM has a hidden layer and an observable layer, it fits well with the detection procedure whose spectrum states are unknown, but the receiver can be obtained [255]. In this case, the collected and processed observations come from the sensors and the hidden states that represent the health state of the equipment. Then, the HMM aims to determine the best sequence of hidden states applied to the sequence of observable states over time. For this methodology, three hidden states are defined, corresponding to the

health state of the equipment. Namely: State 1 represents the "Well-Functioning State"; state 2 represents the "Warning State"; state 3 represents the "Failure State". Then, the first step of the HMM is to determine, through the observable states, which are the best λ parameters for the model [256]. For this, the Baum-Welch algorithm is used, which will determine which model parameters can best justify the sequence of observable states. Mathematically explained, $\lambda$ is determined, such that the probability, $P(O|\lambda)$, of the observation sequence O, given the model $\lambda$, is maximized [257]. Baum-Welch determines the maximum probability of the model parameters based on the observed state sequence [258]. Having obtained the model parameters, it is possible to perform the diagnosis of the equipment through the hidden states. To determine which hidden states best apply over time, the Viterbi algorithm is used. This algorithm uses dynamic programming to find the maximum likelihood path and, finally to perform a prediction of the HMM [258]. The most likely sequence of hidden states, called the Viterbi path, results in a sequence of observed events [259]. HMM can then be used to perform diagnosis classification of equipment since it is a suitable model for continuous dynamic signals under processing [260]. The function of HMM is to determine the most probable sequence of hidden states (equipment diagnosis), using for this the sequence of observable states (clusters coming from the K-means). In statistical learning theory, the HMM is most efficient in pattern recognition processing [260], and has often been used for: recognising changing behaviours of dynamic features of a system [256]; modelling time series-based phenomena, due to their computational efficiency and, because they can be used to build data-driven models that provide characteristic indicators [257]; modelling non-stationary and complex random physical processes of machine condition deterioration, being able to perform both monitoring and diagnosis [261].

After performing the diagnosis with the HMM method, the equipment's prognosis can follow. That is, to predict the health state of the equipment in the future. For this, a DNN method is used for time series prediction. DNN is one of the most active and promising areas of research, being the most used algorithms [236]: the Recurrent Neural Network (RNN), the Long Short Term Memory (LSTM), the Convolution Neural Network (CNN), and the Gated Recurrent Unit (GRU). For the methodology developed in this thesis, the GRU Network is going to be used since it is a simplification of the LSTM architecture, able to avoid gradient problems and to train faster due to fewer tuning parameters [262]. According to recent studies, recurrent units tend to become simpler, requiring the use of RNNs with fewer memory requirements and less demanding training algorithms. GRU can solve the escape gradient problem of a standard RNN because it uses the so-called update and reset gate, which are the two vectors that decide what information should be

passed to the output [236]. The GRU network, as the author explains, has as special feature the fact that it can be trained to keep long-term information without forgetting information or removing information that is irrelevant to the prediction. Therefore, GRU is the neural network used in this methodology to make the prediction of a few days ahead of both the observable and hidden states. This is done to figure out which of the cases will get better results in the GRU model. This GRU network is also a good option for this type of analysis, since both in the observable and hidden states, the amount of data is smaller. GRU shows better results on smaller and less frequent datasets [263], [264]. This is one of the advantages compared to the LSTM architecture. GRU has as powerful cells as the cells of LSTM, even for smaller datasets [264].

### 4.3.2 Results

#### 4.3.2.1 Data preparation

The data used for this study were collected in a paper industry company, taken from sensors placed in a pulp drying press. The purpose of this equipment is to remove the moisture from the pulp, being an important process regarding the production flow to the finished product. This makes that this equipment needs a PdM maintenance. To perform this type of maintenance, 6 sensors were attached to the equipment, which collects data every 5 minutes. To perform multivariate analysis, each sensor collects data of different magnitudes, namely: Current; Hydraulic Level; Torque; Pressure; Rotation Speed; Temperature. Through multivariate analysis, it is possible to obtain a better picture of the state of health of the equipment. The data supplied by the company has a history of approximately one year, containing 83,329 data collected by each of the sensors (Figure 57).

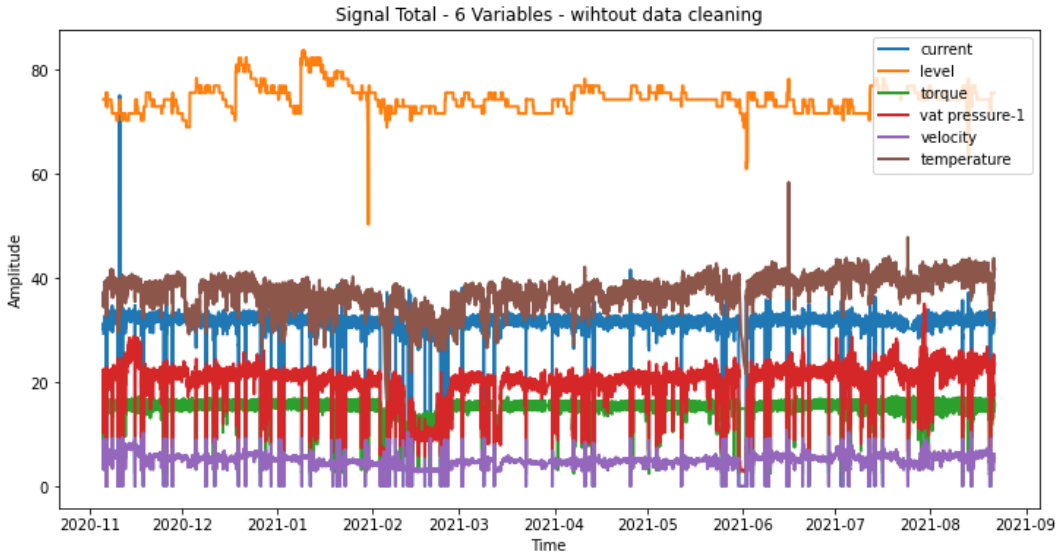



Figure 57 - Data collected by the sensors.

Since the process of data collection is done in an automated way there may be some weaknesses, then it is necessary to perform a data preparation. The aim here is to improve the quality and integrity of the data set, because inaccurate or incorrect values decrease the quality of diagnosis and prognosis of the equipment. To find meaningful information from big data, it is essential to perform data pre-processing. This step is of utmost importance to ensure reasonable results, aiming the analysis with exploratory data mining, classification or building a good and robust prediction model [132]. In this way, a data cleaning was performed to remove all inconsistencies and increase its quality. All duplicate data, non-existent data and zeros were replaced by the mean of the respective signal. No filter was added because it could remove values that would not be outliers, but values that could represent equipment malfunction. So, these values were kept to increase and add value to the prediction. The equipment downtimes were also replaced by the respective average of each signal. This was done because the stops could be confused with equipment malfunction and, consequently reduce the model prediction. To detect equipment downtimes an algorithm was created which, when the current, torque, pressure and speed were below a certain threshold at the same time, it is seen as an equipment downtime (Figure 58). After all this preparation, the data used for the study are shown in Figure 59.

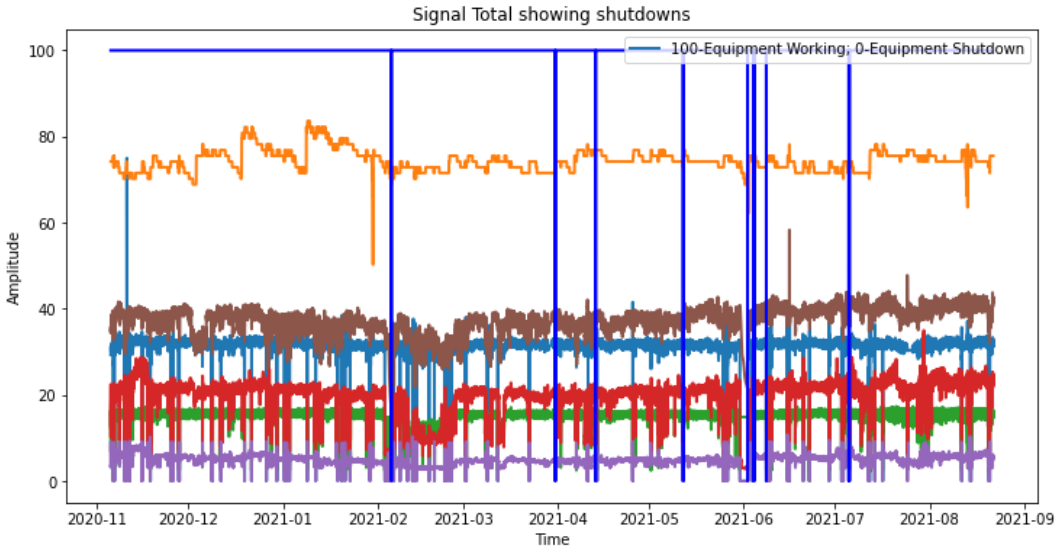



Figure 58 - equipment stoppage over time.

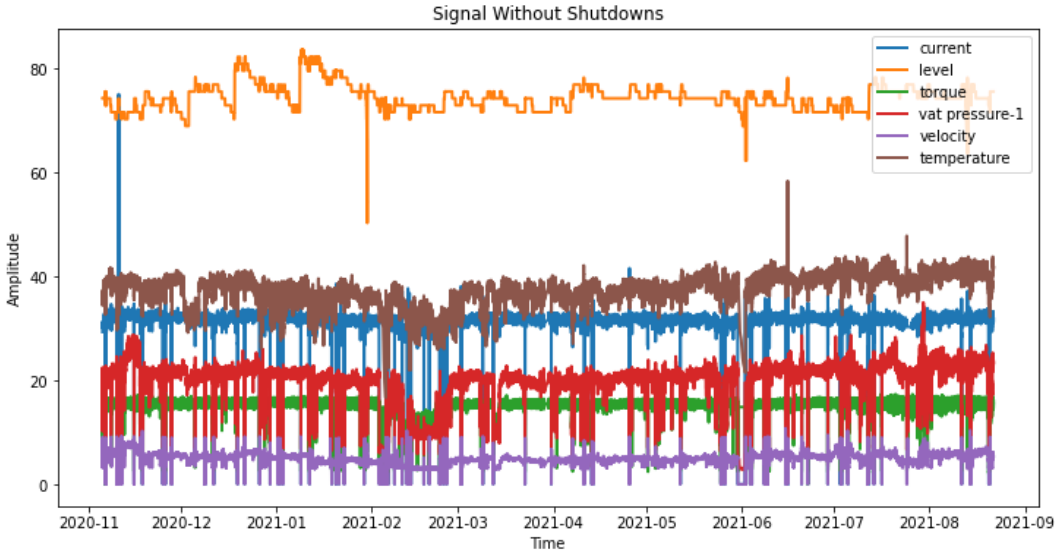



Figure 59 - data after cleaning and preparation.

### 4.3.2.2 Feature Generation

In this phase, the data prepared in the previous phase are used and divided into temporal windows. The objective is to make a feature generation for each one of the temporal windows. For that, the features are represented in Table 11.

TABLE 11 - TIME DOMAIN FEATURES

| Parameter | Mathematical Equation | Parameter | Mathematical Equation |
|---|---|---|---|
| Mean | $T_1 = \frac{\sum_{n=1}^{N} x(n)}{N}$ | A Factor | $T_{12} = \frac{T_5}{T_2 . T_3}$ |
| Standard Deviation | $T_2 = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - T_1)^2}{N-1}}$ | B Factor | $T_{13} = \frac{T_7 . T_8}{T_2}$ |
| Variance | $T_3 = \left(\frac{\sum_{n=1}^{N}\sqrt{\lvert x(n)\rvert}}{N}\right)^2$ | SRM | $T_{14} = \left(\frac{\sum_{n=1}^{N}\sqrt{x(n)}}{N}\right)^2$ |
| RMS | $T_4 = \sqrt{\frac{\sum_{n=1}^{N}(x(n))^2}{N}}$ | SRM Shape Factor | $T_{15} = \frac{T_{14}}{T_1}$ |
| Absolute Maximum | $T_5 = \max \lvert x(n)\rvert$ | Latitude Factor | $T_{16} = \frac{T_5}{T_{14}}$ |
| Coefficient of Skewness | $T_6 = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - T_1)^3}{(N-1)T_2^3}}$ | Fifth Moment | $T_{17} = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - T_1)^5}{(N-1)T_2^5}}$ |
| Kurtosis | $T_7 = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - T_1)^4}{(N-1)T_2^4}}$ | Sixth Moment | $T_{18} = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - T_1)^6}{(N-1)T_2^6}}$ |
| Crest Factor | $T_8 = \frac{T_5}{T_4}$ | Median | $T_{19} = meadian\ x(n)$ |
| Margin Factor | $T_9 = \frac{T_5}{T_3}$ | Mode | $T_{20} = mode\ x(n)$ |
| RMS Shape Factor | $T_{10} = \frac{T_4}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$ | Minimum | $T_{21} = \min x(n)$ |
| Impulse Factor | $T_{11} = \frac{T_5}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$ | | |

Time windows of 6h were created to cover 4 daily operation shifts. For each time window there is a set of 72 data for a total of 1158 windows. Since 21 features are generated for each one of the 6 sensors, a matrix of 1158*126 is created.

#### 4.3.2.3 PCA

Now, the 1158*126 matrix from the previous step, it will go through the PCA method. Here, a dimensional reduction will be done where new characteristics will be created, the PCs. In this way it will be obtained a new matrix with a lower dimension that, later will

be input to the K-means. The goal of this step is to increase computational speed and use more important features for prediction. The PCs created will be oriented in the direction of greater data variability so, the first PCs contain most of the information of the entire data set.

Since the data coming from the FG are very scattered, possessing different amplitudes, first it is necessary to restart a Z-Score standardization (Figure 60).

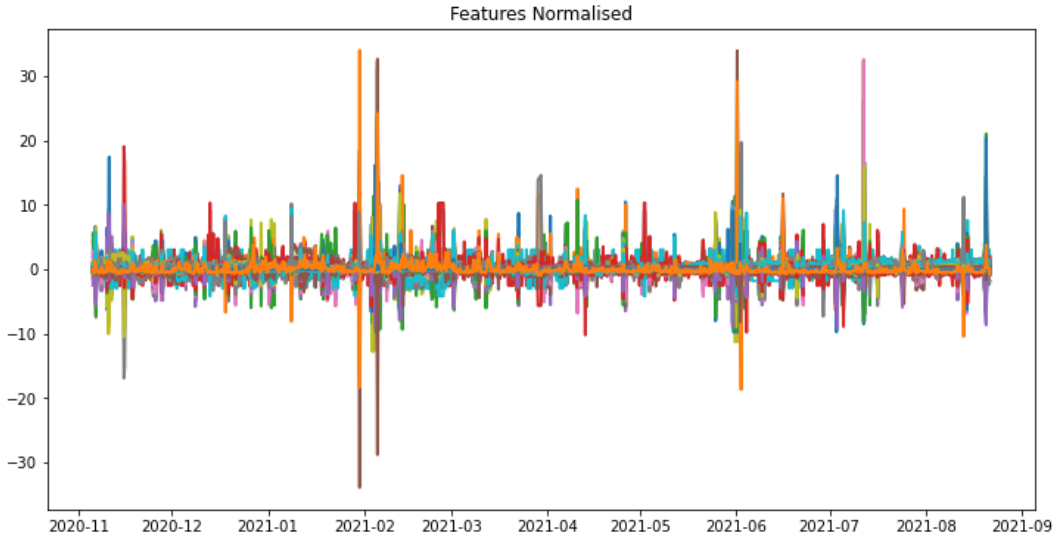



Figure 60 - Features standardized over study time.

After the standardization, the features enter the PCA. Through the study of eigenvectors and eigenvalues, we can verify that 10 PCs preserve about 85% of the data variance. This can be verified, in Figure 61, through the Pareto chart created.

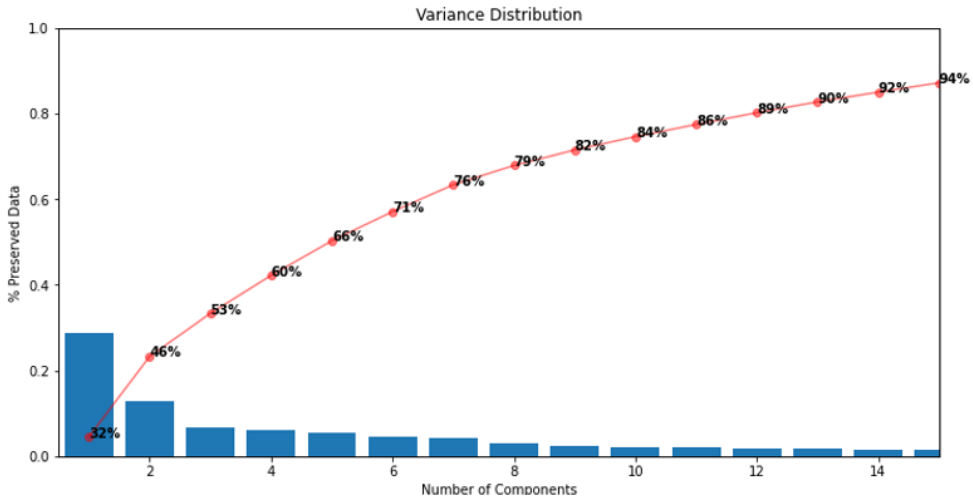



Figure 61 - Percentage distribution of data variability by each PC.

In this way, the initial matrix of 1158*126 becomes an inferior representation with dimension 1158*10.

#### 4.3.2.4 K-means

The PCs now go through a clustering process whose goal is to group the data by similarity gathering in the same cluster the data are most identified. The k-means is the algorithm responsible for clustering the data, with the following procedures [190], [213], [251], [253], [265]:

- Choose the initial number of clusters $k$.
- Randomly select the k initial centroids $c_{j,}\ j = 1,2, \dots, k$ in the observations.
- Calculate the distance between observation and the initial centroid and assign the observed object to the cluster closest to the result (Eq. 62).

$$dist(x_i, c_j) = \sqrt{\sum_{j=1}^{d} (x_{i,d} - c_{j,d})^2} \qquad \text{Eq. (62)}$$

- Define a new centroid based on the average of the group variables (Eq. 63).

$$cj = \frac{1}{N_j} \sum_{x_i \in S_j} x_i \qquad \text{Eq. (63)}$$

- Repeat Step 3 using the new centroid until the observed objects are not relocated to another group.

This process is iterated, moving the centroids to minimise the total variance within the cluster [195]. There are two conditions to terminate the iteration [192], [199], [251]: the specified number of iterations is reached; there are no further changes to the centroid of the cluster. For this study, the second form is used.

To start the K-means algorithm it begins choosing the number of clusters that best fit this dataset. The Elbow method is used to support the decision of choosing the number of clusters. This is one of the most used methods that uses Sum of Squares Error (SSE) *versus k* Clusters (Figure 62).

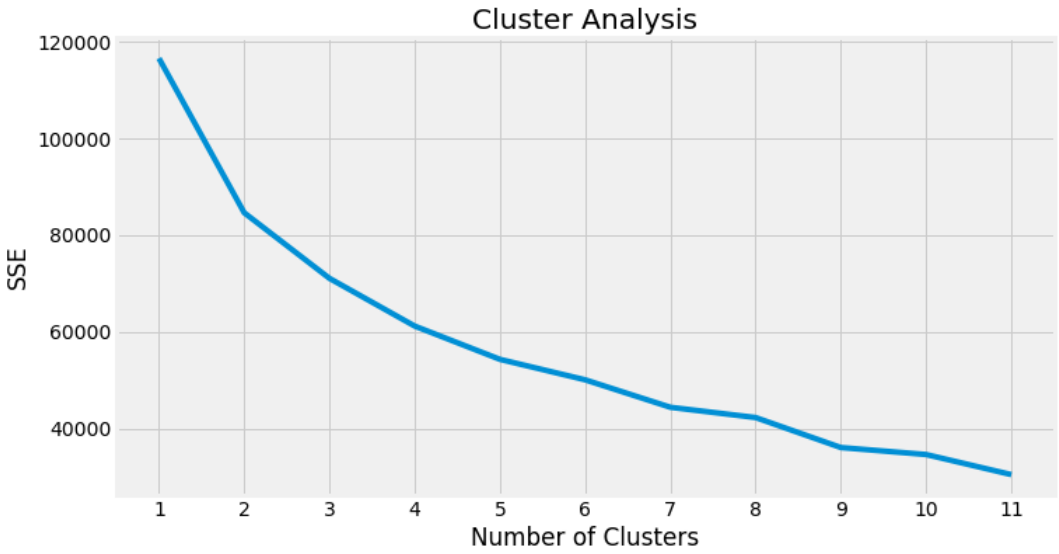



Figure 62 - Elbow method for choosing the ideal cluster number.

As can be seen from the graph created, the elbow is created between the values of $k = 4$ until $k = 6$. It is from these values that SSE decreases more slowly, which means that from $k = 6$ no more worthwhile clusters are created. To help deciding which k value to choose between 4 and 6, a silhouette study was also done (Figure 63) to help deciding the optimal k choice. Through the silhouette graph (Figure 63) we can verify that, among $k = 4$, $k = 5$ or $k = 6$, $k = 4$ is the one with the highest silhouette index. Thus, K-means will group the data for a set of 4 clusters.


Figure 63 - Silhouette method to define the optimal number of clusters.

The silhouette index is also a widely used method for determining the number of clusters in the K-means algorithm. It takes into consideration measures of coherence and separation of events in a cluster [187]. According to the author, the silhouette function calculates the average silhouette coefficient of all samples based on the average Intra cluster distance and the average distance of the nearest cluster for each sample, and its index varies between [−1,1] [187]. Higher index values represent good solutions for clustering.

Once cluster number is chosen and clustering is done, where the clusters are sorted in descending order, we can see their representation in temporal form (Figure 64). Cluster 1 is the one with the most points, then it is the one that appears most often. On the contrary, cluster 4 is the one that happens more rarely.

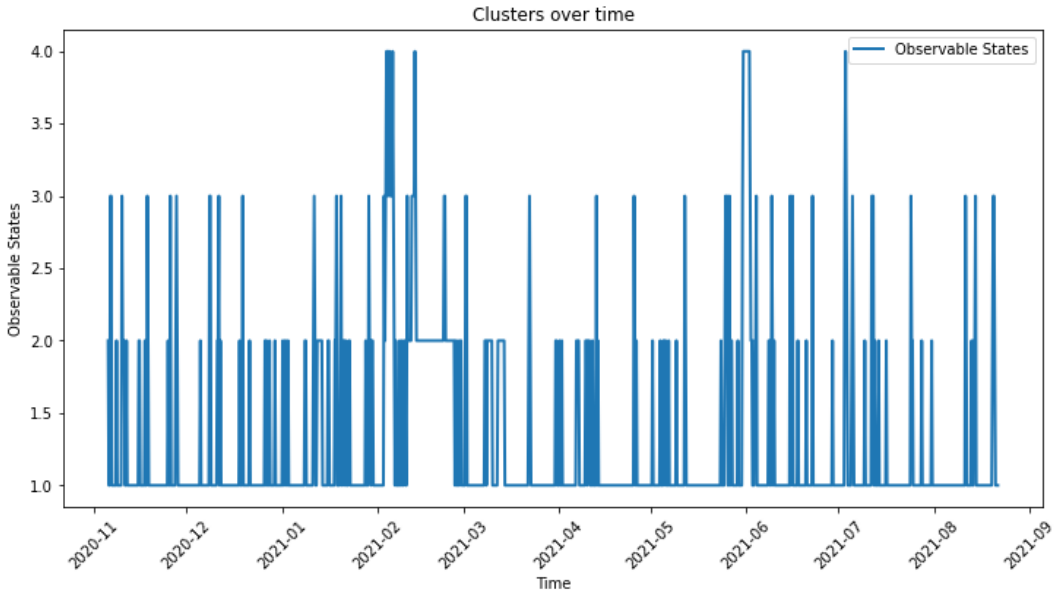



Figure 64 - Clusters movement over time.

### 4.3.2.5 HMM

We now begin the classification phase which aims to make the diagnosis of the health state (condition) of the equipment. For this purpose, three hidden states are defined for the HMM model: Hidden State 1 represents the "good working state; State 2 represents the "alert state"; State 3 is the "bad working state". The HMM is a doubly stochastic model where there are probabilities of emission from each observable state (Clusters coming from the HMM) to each hidden state (equipment diagnosis) and probabilities of transition between the hidden states.

First, the observable states are used to train the HMM model. For this purpose, the Baum-Welch algorithm is used to maximize the expectation to find local maxima of P(O|λ) [257]. Thus, we obtain the best possible parameters for the selected sequence of observations which, in this case, is a sequence of clusters over time. The training algorithm is then responsible for determining: the transition probability between hidden states $A = a_{ij}$; the probability of emission of observable states into hidden states $B = b_{jk}$; the probability of the initially hidden state $\pi_i$.

After obtaining the parameters of the trained model, the model is evaluated by calculating Accuracy. To do this, initially the sequence of observable states is temporally divided into training data (70%) and test data (30%). After training the model with 70% of the training data and obtaining the model parameters, it is possible to create a sequence of observations with the same number of samples as the test data. Once this is done you can determine the Accuracy by determining the number of times the observable states generated coincide with the real ones (Eq. 64).

$$Accuracy = \frac{\sum(HMM_{generated\ observations} = Data_{Test})}{n_{Samples}} * 100 \qquad \text{Eq. (64)}$$

Since the HMM parameters represent probity matrices and the generated sequences are not always equal, 10000 sequences were generated, and the Accuracy was calculated for each one of them. After that, it was made the average of all Accuracy, having been obtained the value of approximately 72%. This represents a good Accuracy value showing that the model was well-trained and that it can now proceed to the determination of the sequence of hidden states. Using the observable states and the trained model parameters, the Viterbi algorithm can be used to determine the most likely sequence of hidden states. Figure 65 shows the evolution of the hidden states and the respective diagnosis of the equipment throughout the study period.

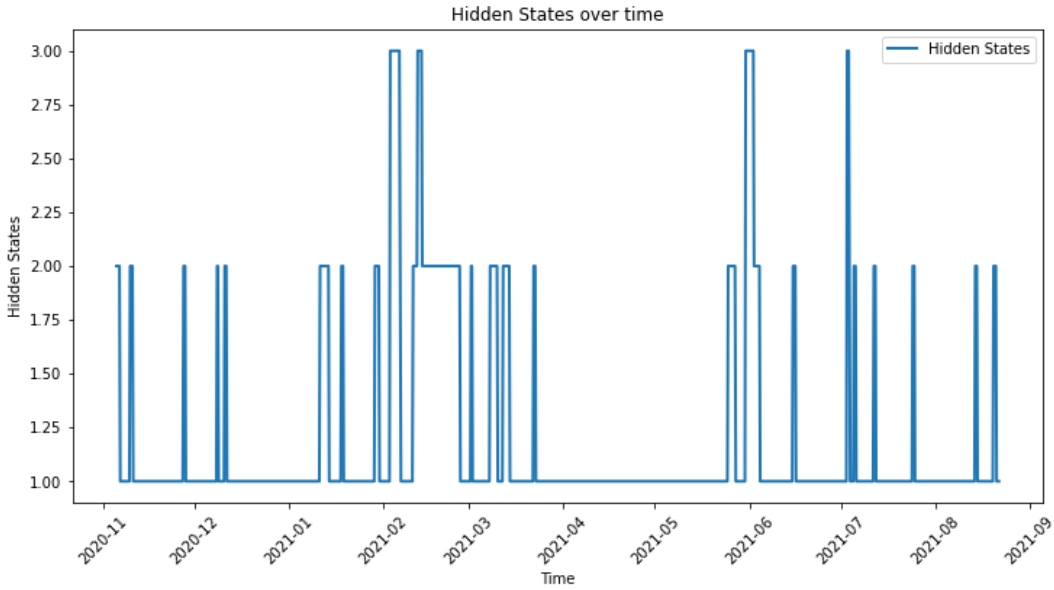



Figure 65 - Hidden states movement over time.

#### 4.3.2.6 Prognostic with GRU model

Once the diagnosis has been completed, we can now proceed to a prognosis about the equipment health state. To do so, the GRU tool will be used. The network is prepared for both the observable states and the hidden states of the HMM to make a prediction of a 7 day ahead. Using the recurrent neural networks with the encoder and decoder structure, a prediction of the observable states in a period of 7 days forward, which corresponds to one week, was performed. A GRU recurrent neural network of 5 units was created with a delay window of 3 days, which corresponds a prediction of 7 days for the observable states, as is shown in Figure 66. The net structure presents a Relu activation function in both the first and second layers. After the prediction is made by the GRU net, it is necessary to use the Equation 65 to scale the predicted data to the scale of the observable states. That is, the goal is to scale the predicted data, $x_n$, to the same values of the observable states, as is shown in Figure 67.

$$ScaleValues = round\left([x_n - \min(x_n)] * \left[\frac{n_{States} - 1}{\max\left(x_n - \min\left(x_n\right)\right)}\right]\right) + 1 \qquad \text{Eq. (65)}$$

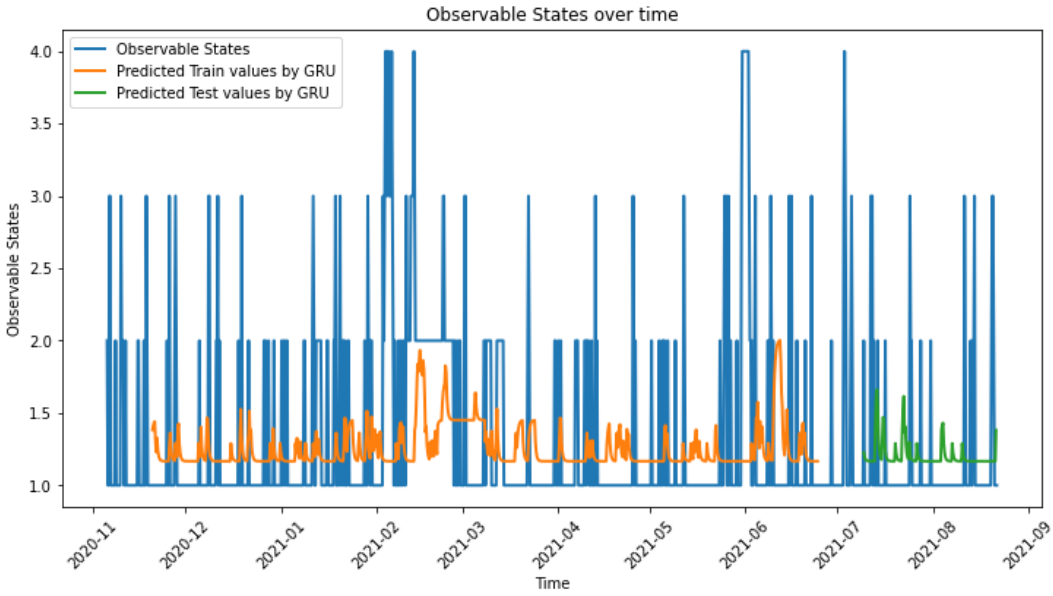



Figure 66 - Prediction made by the GRU network for the sequence of observable states.

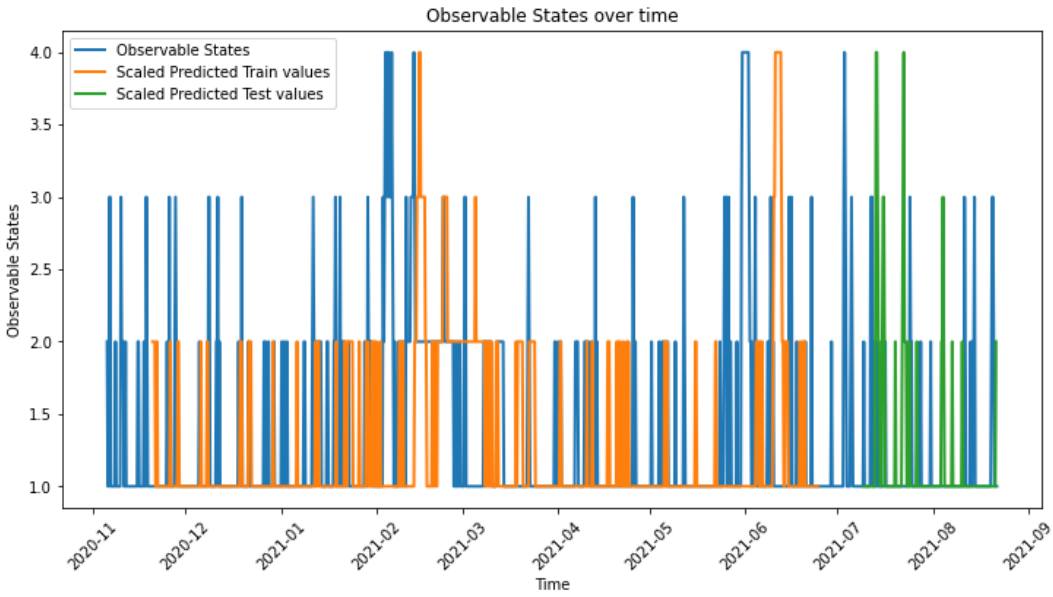



Figure 67 - Prediction performed by the GRU network with predicted values scaled to the sequence of observable states.

To validate the model, the following evaluation methods were used: Mean Absolute Percentage Error ($MAPE$) (Eq. 66); Root Mean Square Error (R$MSE$) (Eq.67); Mean Absolute Error ($MAE$) (Eq.68); $R^2$ (Eq.69).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - y_i|}{x_i} \qquad \text{Eq. (66)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}} \quad \text{Eq. (67)}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \quad \text{Eq. (68)}$$

$$R^2 = 1 - \sum_{i=1}^{n}\frac{(x_i - y_i)^2}{(x_i - \bar{y}_i)^2} \quad \text{Eq. (69)}$$

Where:

$x_i$ – is the actual value;

$y_i$ – is the value predicted by the model;

$n$ – is the total number of observations.

For each preceding evaluation method, the following values were obtained, respectively: 5.49, 0.33, 0.086 and 0.68. Through the observation of these values, it is possible to conclude that the model had a good evaluation.

The same network with the same architecture was used to make the 7 day ahead prediction in the hidden states, i.e., directly in the designation of the equipment health state. Only the activation function of the last layer was changed from Relu to sigmoid. The 7-day prediction was performed resulting as is shown in Figure 68. The same scaling process is performed for the predicted values to be at the measure of the hidden states (Figure 69). Also, applying the same evaluation method through the error calculations of $MAPE$, $RMSE$, $MAE$ and $R^2$, the following values were obtained, respectively: 2.79, 0.22, 0.035 and 0.71.

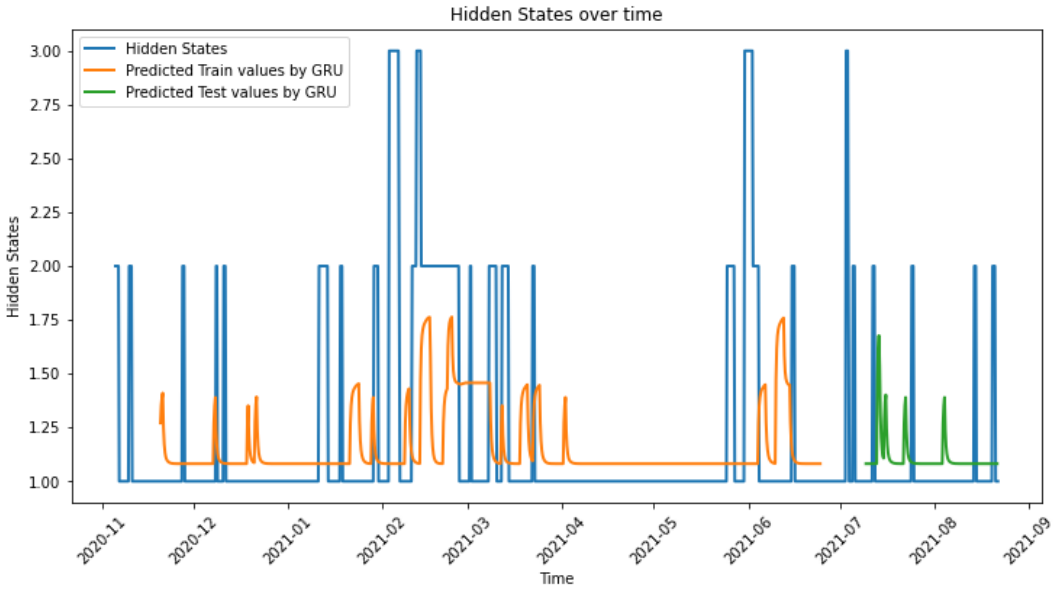



Figure 68 - Prediction made by the GRU network for the sequence of hidden states.

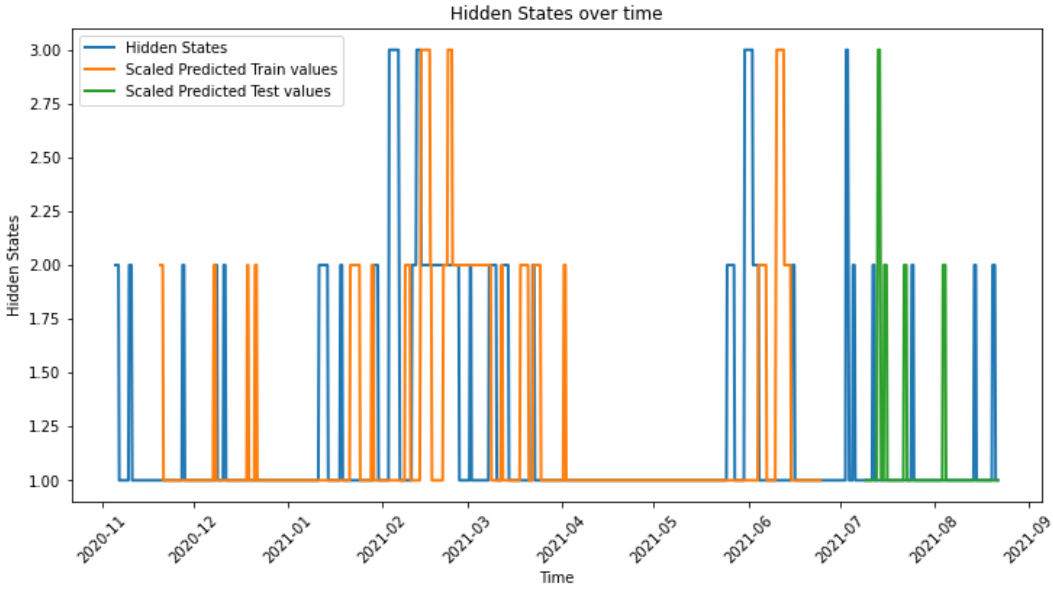



Figure 69 - Prediction performed by the GRU network with predicted values scaled to the sequence of hidden states.

## 4.4 Methodology to diagnose the equipment condition and reading devices status

This methodology gave origin to a paper submitted to the Sensors journal, indexed Scopus, evaluation Q1, with DOI: 10.3390/s23052402 (Appendix E). Aims to demonstrate how the health status of production equipment and reading devices can be detected. In other words, through the same set of data, a methodology that can extract

two different pieces of information is explained. Through feature generation, it is possible to obtain distinct information from the same data set. First, through temporal correlations is determined the health state of the production equipment. After that, an HMM filter is made where all the values of the original signal that occur during periods of malfunction of the production equipment are eliminated. Thus, only remain the errors due to deviations in the sensors. To demonstrate this methodology, it was simulated a signal with equipment error and a deviation in one of the sensors.

### 4.4.1 Methodology and simulated signal results

#### 4.4.1.1 Signal Simulation

To explain the methodology developed, a multivariate signal was created with 4 non-redundant sensors (Figure 70). Each sensor is responsible for measuring a different physical phenomenon. Two different types of failure were simulated to the signal, one being a failure assigned to the production equipment and another to the reading equipment. The first failure simulates a fault of the production equipment where all sensors react to changes in the equipment's behaviour. After performing maintenance, the sensors returned to the normal operational functioning in the production equipment. The second fault simulates the error of one of the sensors, who is the only one that had a deviation in the data collection. Since no other sensor had a measurement deviation reaction in that period of time, this leads that it is an error of the reading equipment and not of the production equipment.

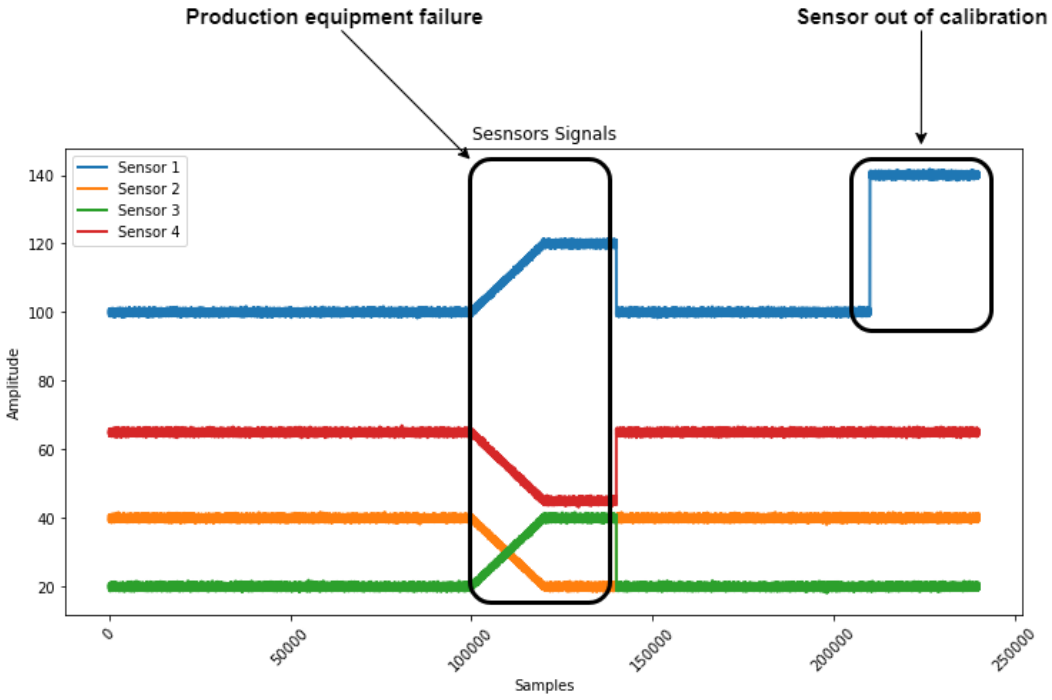



Figure 70 - Simulated signal with production equipment error and deviation in one of the sensors.

For the operation of this methodology, it is necessary to group the sensors based on their correlation in the occurrence of equipment failure. That is, it is necessary to choose a group of sensors that react to the failure of the production equipment since the model performance is based on the correlation among the sensors [266]. Modelling methods developed for non-redundant sensor groups require that the sensors in a single model contain related information (e.g., temperature and pressure of a gas), usually identified by linear correlations or a physical understanding of the measurement relationships [267]. In multivariate analyses with a heterogeneous sensor set, where there are two or more sensors, there is always the possibility that two or more variables are correlated [268]. Methodologies developed for non-redundant sensor groups will, usually, work well on a redundant sensor set. Heterogeneous sensor data tends to exhibit a robust correlation in space and time, improving the anomaly detection problem [269]. Therefore, the use of multiple sensors can improve the performance of any individual sensor [270].

When analysing the correlations of the simulated signal sensors (Figure 71), the correlation is only high when there is a change in the behaviour (fault) of the production equipment. During the rest of the time, the correlation among sensors is practically zero, even when there is a deviation in one of the readings devices.

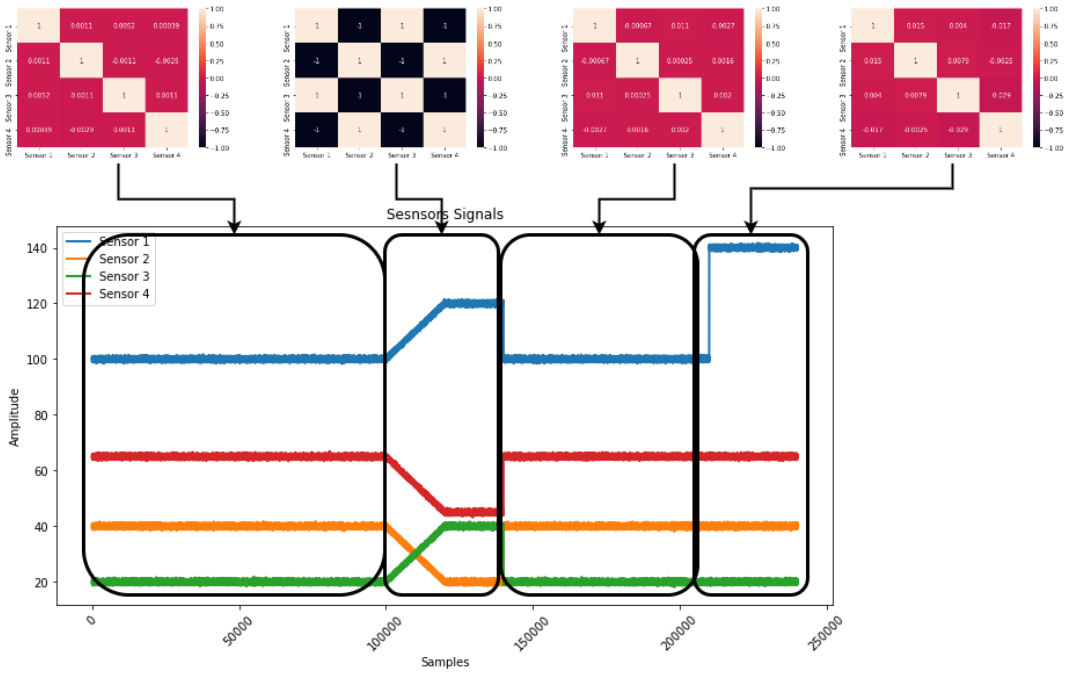



Figure 71 - Correlations presented in each one of the phases of the simulated signal.

### 4.4.1.2 Generation of Correlation Features

Feature generation and selection, depending on the problem at hand, are among the first steps to be performed in any Machine Learning (ML) modelling [271]. In this case, the goal is to create correlation values over time to determine nonsense correlations and detect production equipment failures. To do this, it is necessary to join a dataset in blocks [271]. Time windows are created, where depending on the length of each window it will have a certain set of samples. After creating the temporal windows, it is possible to determine the correlations for the defined time period. To increase the relationship between all the sensors, first, the subtraction and ratio between all of them were done. The methodology was tested without this step without obtaining such good results. Then, after increasing the ratio between the variables, the temporal windows were created, with $X = \{x_1, x_2, \dots, x_n\}$, where $n$ represents the number of samples per window. In this simulation, temporal windows with 288 samples were chosen. Based on this, we obtained 833 temporal windows since the simulated data set had a size of 24000 observations (Table 12). If the data were collected every 5 min, this would mean that the time windows were one day.

TABLE 12 - TIME WINDOWS WITH DATA SUBTRACTED AND RATIOS BETWEEN EACH SENSOR PAIR

| | **Col 1** | **Col 2** | **(…)** | **Col 24** |
|---|---|---|---|---|
| | **Sensor1-Sensor 2** | **Sensor 1 - Sensor 3** | | **Sensor 4 / Sensor 3** |
| | 60.633244195683154 | 79.78937757904451 | (…) | 3.0538616746523806 |
| **1$^{st}$ Chunk** | 59.64240130685382 | 79.86135304565789 | | 3.062736422782496 |
| | (…) | (…) | | (…) |
| | 59.262459386540584 | 79.87646380056816 | | 3.4567110848650304 |
| | 59.82199978785137 | 78.56797868615018 | (…) | 3.0538616746523806 |
| **2$^{nd}$ Chunk** | 60.19302454746067 | 78.74172104561583 (…) | | 3.062736422782496 (…) |
| | (…) | 80.59289809317407 | | 3.4567110848650304 |
| | 60.122575793590606 | | | |
| **(…)** | (…) | (…) | (…) | (…) |
| | 100.12391557061736 | 119.97019459713032 | (…) | 3.210611636252731 |
| **834$^{th}$ Chunk** | 99.36665172104591 | 119.88307859672575 | | 3.3163521246018797 |
| | (…) | (…) | | (…) |
| | 100.67188973084929 | 120.40164877024282 | | 3.2317746352961993 |

After creating the temporal windows, we can now generate the correlation features between each of the columns (Table 13). In the case of multiple sensors, performance can be improved by performing a cross-correlation between all paired sensor combinations [270]. Correlation is a mathematical operation that allows portraying how two or more variables fluctuate together [272]. A Spatial-temporal correlation exists to define several physical phenomena. Spatial correlation is responsible for mapping the correlation among two or more variables in time $t$ [273], [274]. There are several correlation measures that can support expressing the fluctuation between variables. As the preceding authors explain, correlation can be defined as a degree of linearity between the variables $X$ and $Y$. The level of correlation between the two positions is demonstrated by the degree of correlation, which also represents the association of structural response data between the two positions. Through the intervals ranging from -1 to 1, it is possible to determine how correlated the two variables are. The closer the correlation value is to 1, the more correlated the two variables are. If the correlation value approaches zero, it means that more weakly the two variables are correlated [275], [276].

Three types of correlation were used in this methodology, namely:

*Pearson* Correlation

This technique is a covariance measure of the degree of correlation between two input set estimates [276], and has been widely used in some research related to trait selection [277]. Pearson correlation between two variables $X$ and $Y$ is defined as (Eq.70) [268], [275], [273]:

$$Coo[X,Y] = \frac{Cov[X,Y]}{\sqrt{var[X], var[Y]}} \quad \text{Eq. (70)}$$

Where:

$cov[X,Y]$ - is the covariance of $X$ and $Y$;

$var[X]$ - is the variance of the random variable $X$.

*Spearman* Correlation

Unlike Pearson's correlation, which evaluates linear relationships, Spearman's correlation evaluates monotone relationships, whether linear or non-linear. It is a variant correlation measure of Pearson correlation, where observations are corresponding $rg(X_i)$ and $rg(Y_i)$ ranks instead of the actual random transport values [268], [278].

It can be represented by the Equation 71:

$$r_s = \frac{\text{cov}(rg(X), rg(Y))}{\sigma_{rg(X)}\sigma_{rg(Y)}} \quad \text{Eq. (71)}$$

Where:

$\text{cov}(rg(X), rg(Y))$ – is the covariance of the variables in ranks;

$\sigma_{rg(X)}$ $e$ $\sigma_{rg(Y)}$ – are the standard deviations of the variables in ranks.

Another popular formula for representing spearman correlation is (Eq. 72):

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_1^2}{n(n^2-1)} \quad \text{Eq. (72)}$$

Where:

$d_i = rg(X_i) - rg(Y_i)$;

$n$ – is the number of observations.

*Kendall* Correlation

Any pair of observations $(x_i, y_i)$ and $(x_j, y_j)$, where $i \neq j$, are concordant if the ratings of both elements agree with each other, i.e., if $x_i > x_j$ and $y_i > y_j$. They are discordant if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant nor discordant. Kendall's coefficient τ is defined as (Eq. 73) [268]:

$$\tau = \frac{n_1 - n_2}{n(n-1)/2} \qquad \text{Eq. (73)}$$

After generating all correlations among all columns of the Table 13, we can see how the features are represented over time (Figure 72). We can also see how the correlations are represented in tabular form (Table 13).

TABLE 13 - CORRELATIONS BETWEEN EACH PAIR OF SENSORS IN EACH OF THE TIME WINDOWS

| | **Col 1** | **Col 2** | **(...)** | **Col 1656** |
|---|---|---|---|---|
| | **Corr. Pearson Col1-Col2** | **Corr. Pearson Col1-Col3** | | **Corr. Kendall Col24-Col23** |
| **1$^{st}$ Chunk** | 0.5382026016598894 | 0.5620274177793831 | (...) | 0.08560782036391794 |
| **2$^{nd}$ Chunk** | 0.495471883817057 | 0.4213862868717626 | (...) | 0.1498741773132017 |
| **(...)** | (...) | (...) | (...) | (...) |
| **834$^{th}$ Chunk** | 0.43642132554147606 | 0.5029634629836655 | (...) | 0.0557017543859649 |

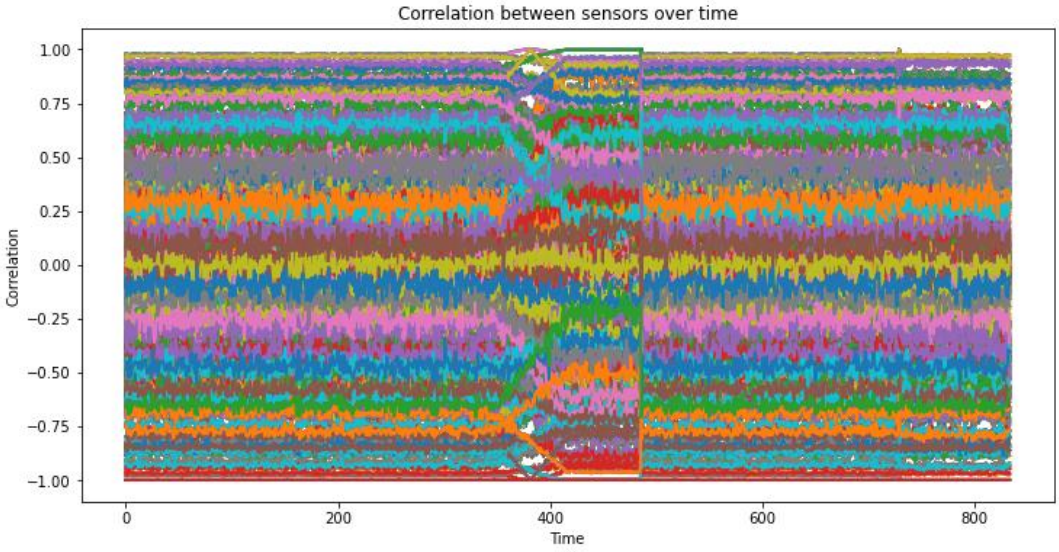



Figure 72 - Correlations between each sensor pair over signal time.

#### 4.4.1.3 Standardization

As can be seen in Figure 72, it is necessary to perform a standardization of the data to ensure better prediction results [279]. It is necessary to convert the magnitude and time scale more uniformly [269]. Standardization is a method that is usually required in ML models to obtain better and more consistent results [280]. For this methodology, the standardization process using the Z-Score method was used to obtain a mean of 0 and a standard deviation of 1, i.e., $Z \sim N(0,1)$. This method not only converts the amplitude of the data but also transforms its dispersion [153]. This author also explains that this method converts a range of variables with some mathematical heuristics, allowing all variables to have the same range, and can make the patterns in the data more visible.

By applying standardization to the data, all initial continuous variables will contribute equally to the study. Features with larger amplitudes will not overlap with the others, creating inequitable results [280]. The authors also explain that it is recommended to scale features to the same range of values as this will also increase the speed of training. Standardization is one of several data transformation techniques, which has the effect of reducing the parameters to a common range, providing a measure that allows the relative importance of any factor or interaction to be more clearly identified [113].

After performing standardization, it is possible to better understand the behaviour of the data as is verified in Figure 73. In addition to all this, standardization will also allow the k-means clustering algorithm [281], which will be used later in the methodology, to work better.

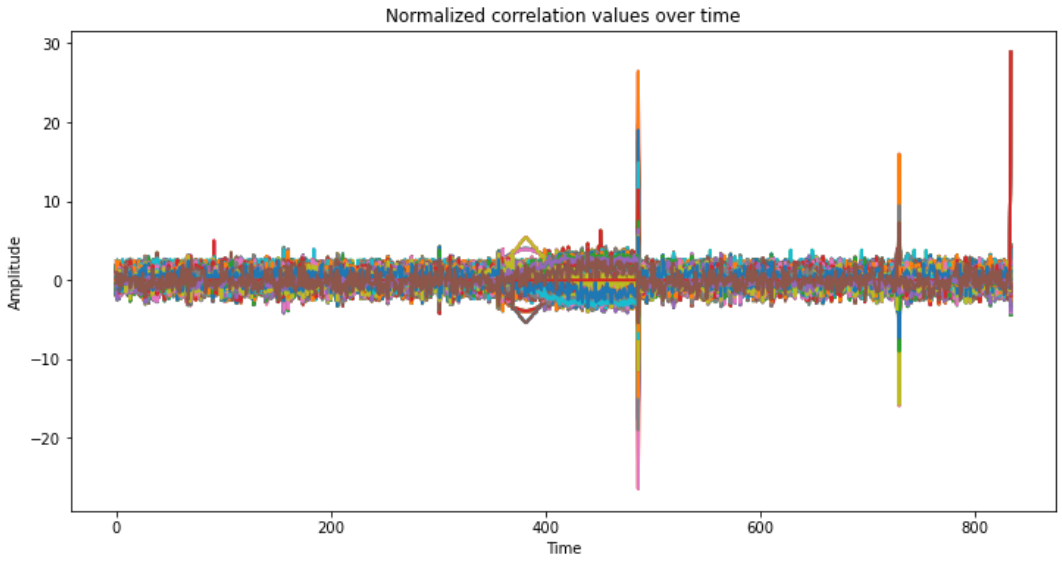



Figure 73 - Correlation features standardized by Z-Score over study time.

#### 4.4.1.4 Dimensional Reduction through PCA

Now, a dimensional reduction technique will be used, namely Principal Components Analysis (PCA). This is a fault detection method based on statistical analysis [282]. It is one of the most widely used methods to do feature extraction and dimensional reduction in ML [154], [277], [283]. It also ends up doing noise removal from the data through dimensional reduction [205], which will also help to improve the quality and integrity of the data that will be used in prediction. Feature extraction and dimensional reduction are done by rotating the orthogonal axes of the coordinate system. The axes are oriented in the direction of the greatest variability of the data. Since new features, called Principal Components (PCs), will be extracted, containing more information, and the processing speed increases. So, we will also improve the prediction ability. PCA is a statistical method, responsible for converting several potentially correlated variables into a new group of uncorrelated variables, the PCs. The number of PCs is equal to the number of original variables, with the first PCs having most of the information [284]. This allows one to choose only a few of the initial PCs to be able to extract all the desired information from the total data set. PCA finds a new coordinate system of the dataset centred on the cloud mean, with the axes being perpendicular and having maximum variance in descending order [285]. This means that $PC_1$ is the one with the highest variance of the data, $PC_2$ is the second one with the highest possible variance, under the constraint that it is orthogonal to the previous PC and succinctly [283]. The process of the PCA algorithm can be described as follows [286]: In the scaled feature space of dimension $N$, a direction that maximizes the variance of the data can be found. After that, this direction is used as the first principal direction and projects the dataset into a space of dimension $N - 1$ by removing the first principal direction. The process is repeated $M$ times, where $M \leq N$, and we obtain a transformed dataset in the principal dimensions.

The new features extracted by PCA, the PCs, are generated as columns in the matrix $X'$. This new matrix $X'$ becomes $m \times n'$ in dimension. To keep the information of the original features, the default number of the desired dimension of the lower dimensional space $n' = n - 1$, where $n$ represents the dimensions of the original dimensional space.

The variances of each PC are accumulated sequentially to assess the proportion of the original information contained in the features in different dimensions. As shown in Figure 74, after the cumulative variance calculation, the first 10 PCs contain about 95% of the variance.

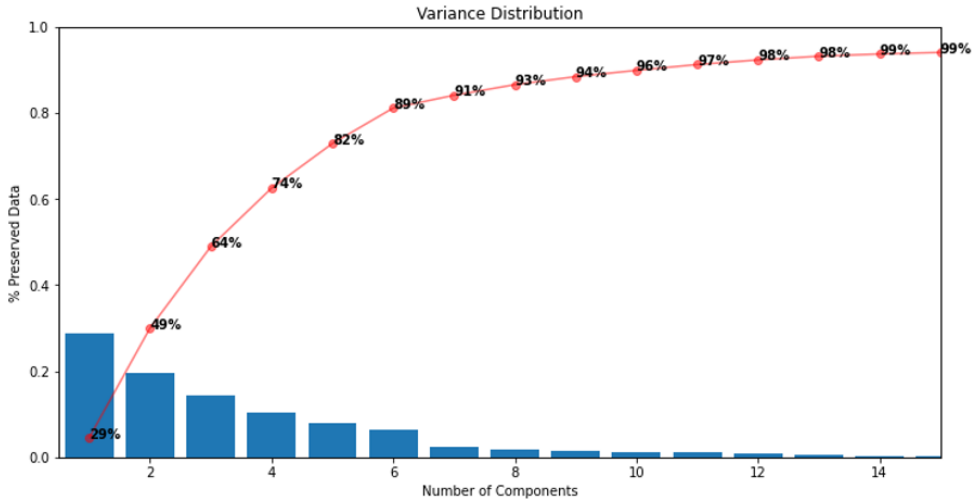



Figure 74 - Pareto chart with the percentage of information for each PC.

Thus, using the first 10 PCs, it is possible to preserve about 95% of the information of the original data. Thus, it is possible to decrease the matrix from $X = [x_{ij}]_{834*1656}$ to the matrix $X' = [x'_{ij}]_{834*10}$.

In Figure 75 it can also be observed the movements of the new data set over the study time. It is already possible to identify more clearly the areas that deviate from the normal operating patterns of the equipment.


Figure 75 - Data movement of the first 10 PCs processed by PCA.

#### 4.4.1.5 Clustering through K-means

For a high feature set, i.e., a high dimensional data set, usually, a preliminary PCA is performed and then, a clustering for the first few PCs [287]. Once the PCA is performed, then the clustering of the data is done. The goal of clustering is to [288]:

- investigate the underlying structure of the data;
- do classification, to determine the degree of similarity between data points;
- do reduction, to organize and summarize data into comprehensible groups.

For this study, the clustering algorithm used is K-means. In this algorithm, data are clustered based on similarity characteristics [207], [205]. Thus, well-functioning data of the equipment will be in the same cluster and malfunctioning data will remain in a different cluster. This is one of the first algorithms to be proposed and assumes that each data has a connection to only a single cluster and is assigned to the closest cluster [289]. It is an unsupervised algorithm and is popular for identifying similarities between points based on distance vectors [288]. One of the most used clustering methods has the following main advantages [203], [206], [214], [288]: brevity; simplicity; efficiency; speed; less computational power. Its main goal is to group similar data to detect underlying patterns [202]. It is necessary to initially determine a fixed number of clusters ($k$). For each $k$ cluster, there is a centroid, which is located at the centre of the cluster. After defining the number of clusters, each piece of data in the dataset is assigned to the nearest cluster. This is done by summing the squared distances of the Euclidean distances between the items and the centroid (Eq. 74), minimizing the Intra cluster variance [202].

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \qquad \text{Eq. (74)}$$

Where:

$x_i$ is the $i^{th}$ data point of cluster $k(C_k)$;

$\mu_k$ is the mean value of the points in cluster $k$.

The Total Within Cluster Variation is defined by Equation 75, with the total Sum of Square Error (SSE) within the cluster measuring the quality of the clustering, which increases as the sum of squares measures decreases [203].

$$Total\ within\ cluster\ variation = \sum_{k=1}^{k} W(C_k) \qquad \text{Eq. (75)}$$

Therefore, according to Borlea *et al.* [204] and Peng *et al.* [205], K-means processes a dataset $D = \{x_1, x_2, \dots, x_n\} \in \Re^d$, where $x$ is a dataset record defined as $X_i =$

$[x_{i1}, x_{i2}, \dots, x_{id}]^T \in \Re^d, i = 1 \dots n$. $d$ is the dimension of a dataset record and $T$ stands for matrix transpose. The algorithm divides the dataset $D$ into a set of $k$ predefined numbers of clusters $C_j, j = 1 \dots k$. Each cluster $C_j$ is composed by a centre of mass called centroid and defined as $C_j = [C_{j1}, C_{j2}, \dots, C_{jd}]^T \in \Re^d, j = 1 \dots k$. The total number of points assigned to each cluster with the notation is $n$, with the cluster expression $C_j = (c_j, n_{C_j})$. The centroid array defined as $c = [c_1^T, c_2^T, \dots, c_k^T]^T \in \Re^{dk}$, which represents the centroids of all existing clusters. The main objective of the algorithm is to minimize the intracluster variance (Eq. 76):

$$c^* = arg \min_{c \in \Re^{dk}} V(c), \qquad V(c) = \sum_{j=1}^{k} \sum_{\substack{i=1 \\ x_i \in C_j}}^{n_{C_j}} \|x_i - c_j\| \qquad \text{Eq. (76)}$$

Where:

$c_j$ is the centroid of the cluster $C_j, j = 1 \dots k$;

$V$ is the objective function or the criterion;

$c^*$ is the optimal arrangement of centroids.

The K-means Clustering sequence has the following steps [202]-[204], [206]-[208]:

- Specify the number of clusters (k);
- Randomly select $k$ data points as initial centroids;
- Assign the data set $x_i$ to the nearest centroid $c_j$, using the Euclidean distance (Eq. 77);

$$d_{x_i, c_j} = \|x_i - c_j\| = \sqrt{(x_{i1} - c_{j1})^2 + (x_{i2} - c_{j2})^2 + \cdots + (x_{id} - c_{jd})^2} \qquad \text{Eq. (77)}$$

- Then, all data points are redistributed according to the previous method to obtain the second clusters. Continuing in the same way, the K-means clustering algorithm stops until the points in each cluster do not change.

Since the initial cluster centre, the similarity measure clustering criterion function easily converges to the local minimum; selecting different initial clustering centres it will lead to different Clustering results [209]. Thus, as this author explains, the correct selection of the initial clustering centre in the K-means algorithm has a great influence on the quality of the clustering results. To choose the optimal number of k clusters, there are different methods, as explained by Wickramasinghe *et al.* [202]. For this methodology, the SSE method, or elbow method, as it is better known, will be used. This is one of the most widely used methods for choosing cluster numbers [203]. It is necessary to use several $k$ clusters and calculate the SSE for each of them. This is done by plotting the

graph, as represented in Figure 76, where an elbow-shaped kink is detected, that is where the optimal k is considered. So, this method can be performed in 4 stages [214]:

- To run a centroid-based clustering variance of each clustering result, e.g., sum-of-squares-errors algorithm, e.g., k-Means, for each k ∈ R;
- To calculate the (SSE) for k -Means;
- To plot the results on a graph;
- To select the elbow curve in the graph.

From the formed graphic (Figure 76), the attempt to minimize SSE stops having so much expression from $k = 4$, and the clusters from this value will reduce the sum of squares in small values. In this way the number of clusters chosen was $k = 4$.

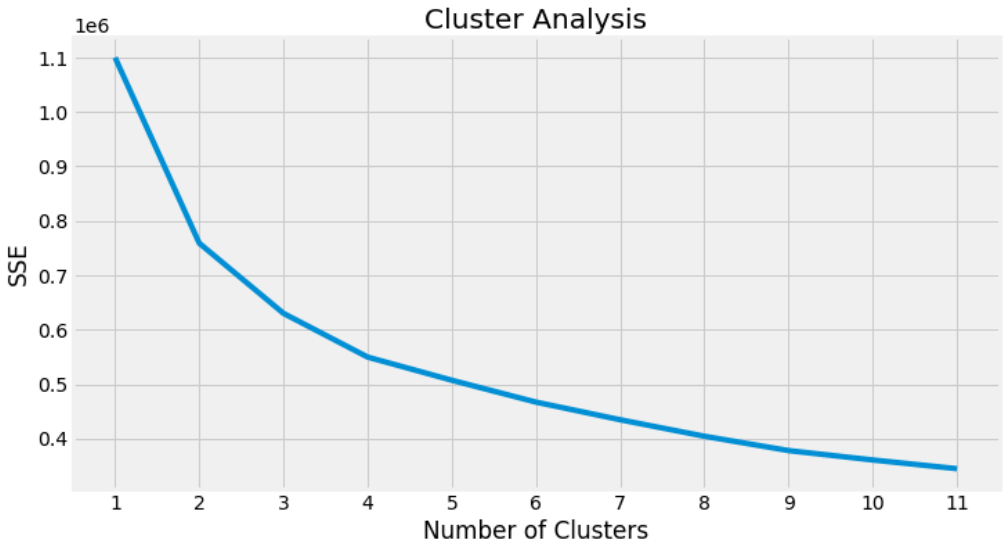



Figure 76 - SSE Analysis to determine the number of Clusters.

Once the number of clusters is chosen, it is possible to proceed with the k-means algorithm. This algorithm will run following all the steps explained above. Once the data is grouped in its respective clusters, each cluster, $C_j$, will be represented over time, as shown in Figure 77. The clusters $C_1, C_2, C_3, C_4$ will be considered in the new observable states, which were optimized from the original data through this ML tools methodology. The clusters formed are also sorted in descending order by the amount of data, $n_{C_1} > n_{C_2} > n_{C_3} > n_{C_4}$. In this way, it is possible to ensure that $C_1$ appears most often over time and $C_4$ the rarest to occur.

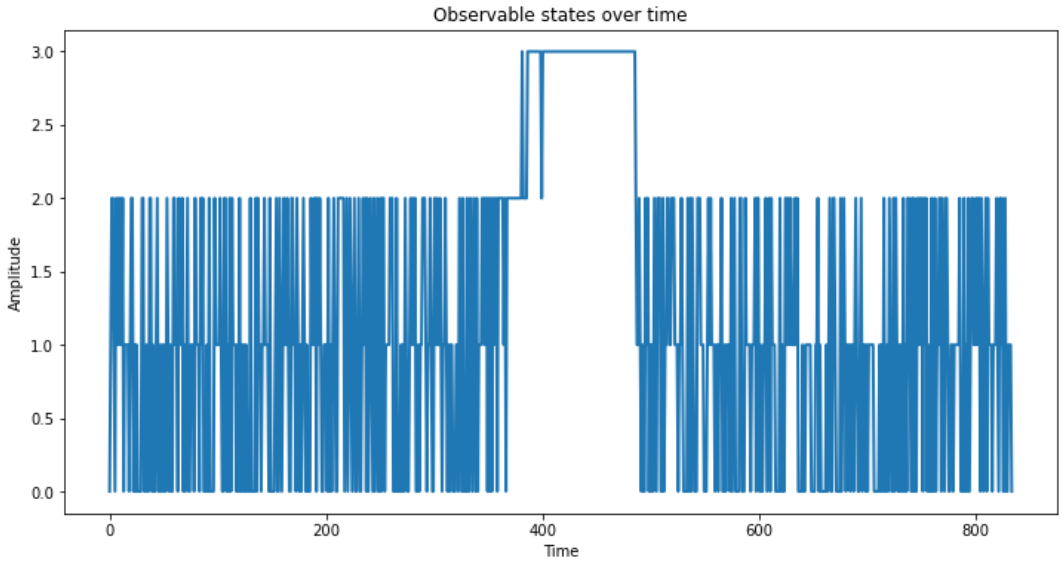



Figure 77 - Cluster (Optimal Observable States) over study time.

#### 4.4.1.6 Classification of production equipment behaviour through the HMM

To classify the optimised observable states into equipment health states, the Hidden Markov Model (HMM) algorithm is used. This is a doubly stochastic process that has both observable and hidden states [255]. We can now use the clustering formed by k-means, which integrate multivariate data from sensors attached to the equipment, as observable states to give input to the HMM model. The observable states represented at time instant $t$ are represented by $O_t \in C, = \{C_1, C_2, C_3, C_4\}$. The hidden states, which represent the health state of the production equipment, are represented at time instant $t$ by $q_t \in S, S = \{S_1, S_2, \dots, S_N\}$. As stated, HMM is a doubly stochastic process that comes from Markov Chains where there are only transition probabilities between states. The hidden states of HMM are a Markov chain, which is obtained through the observable states by the emission probability matrix. That is, through the observable states, it is possible to determine which hidden states best represent each of the observations that occurs at a certain time $t$ and thus deduce the diagnosis of the equipment. For this purpose, three hidden states are defined in this methodology, $S = \{S_1, S_2, S_3\}$, to represent the state of health of the production equipment. Also, the hidden states will be ordered by several points, $n_{S_1} > n_{S_2} > n_{S_3}$. In this way, it is guaranteed that $S_1$ is the one that happens less often and $S_3$ the rarest to happen. This means that $S_1$ represents the good working state of the equipment and $S_2$ and $S_3$ as they happen less, often represent the alert and failure states of the equipment.

To use the HMM and do the intended classification, the model has three basic problems that need to be determined [126], [153], [255]:

- Evaluation - via the forward-backwards algorithm, calculates the probability of the observed fusion outcome sequence $O = \{O_1, O_2, \dots, O_T\}$, given the model $\lambda = (\pi, A, B)$;
- Training – is done using the Baum-Welch algorithm, adjusts the model parameters, $\lambda = (\pi, A, B)$, to maximise the probability of the observed sequence. That is, given a chain of observable states, which model λ best fits, $P(O|\lambda)$;
- Prediction - solved by Viterbi's algorithm, it calculates the most likely hidden state sequence according to the observable sequence and the model parameters. Using the model, $\lambda$, and the observation sequence, $O$, it is possible to detect the best-hidden state sequence, $S$.

So, firstly the HMM model is trained, through the Baum-Welch algorithm, using the observable states coming from the clusters formed by K-means. In this way, it is possible to obtain the parameters of the $\lambda$ HMM. Using the trained parameters and the observable states it is possible to determine the best sequence of hidden states by using the Viterbi algorithm. Figure 78 illustrates the behaviour of the hidden states over time.

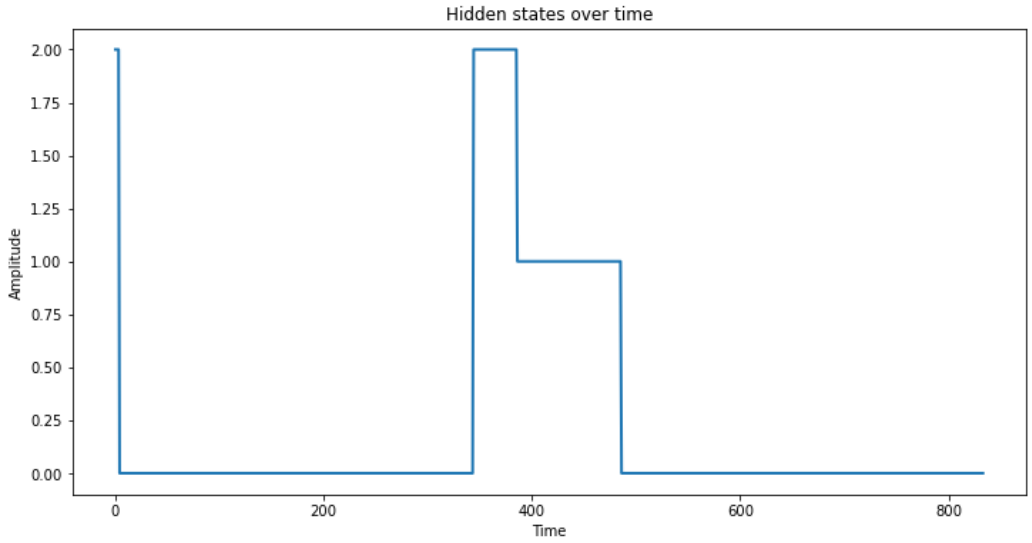



Figure 78 - Hidden States (health states of the production equipment) over the study time.

### 4.4.1.7 HMM Filter

It is possible to verify that the developed methodology works well. The HMM was able to detect non-zero hidden states at the same time that the production equipment malfunction was simulated, having ignored the sensor malfunction. This proves that, by

using a correct sensor group, it is possible to detect only the behaviour of the production equipment. It starts with hidden state 2 due to problems with the initialization of the Viterbi algorithm and these initial values can be ignored.

After determining the periods of the malfunctioning production equipment, it is possible to eliminate them in the original signal. For this, an "HMM Filter" is used, which uses the values of the hidden states over time. Then, the periods where the hidden states are $S_1$ or $S_2$ are saved to, later in the original data, eliminate the data that are in the same epoch. To illustrate the "HMM Filter", the hidden states of the HMM were placed in the original data (Figure 79). In this figure is possible to verify that the HMM really detects only the behaviours of the production equipment. After eliminating the errors of the production equipment in the original signal, we left the data as is represented in Figure 80.

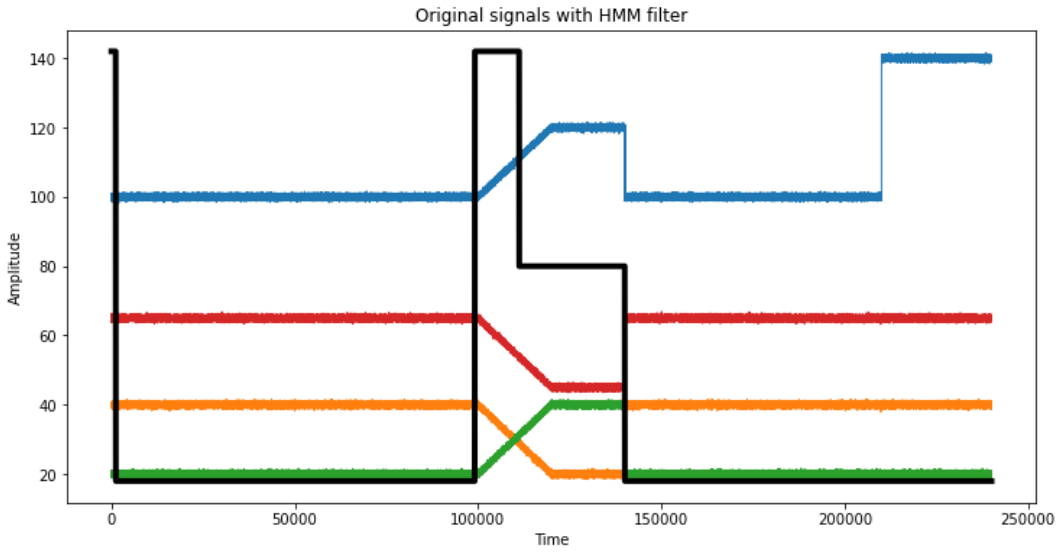



Figure 79 - Original signal with the HMM states overlaid.

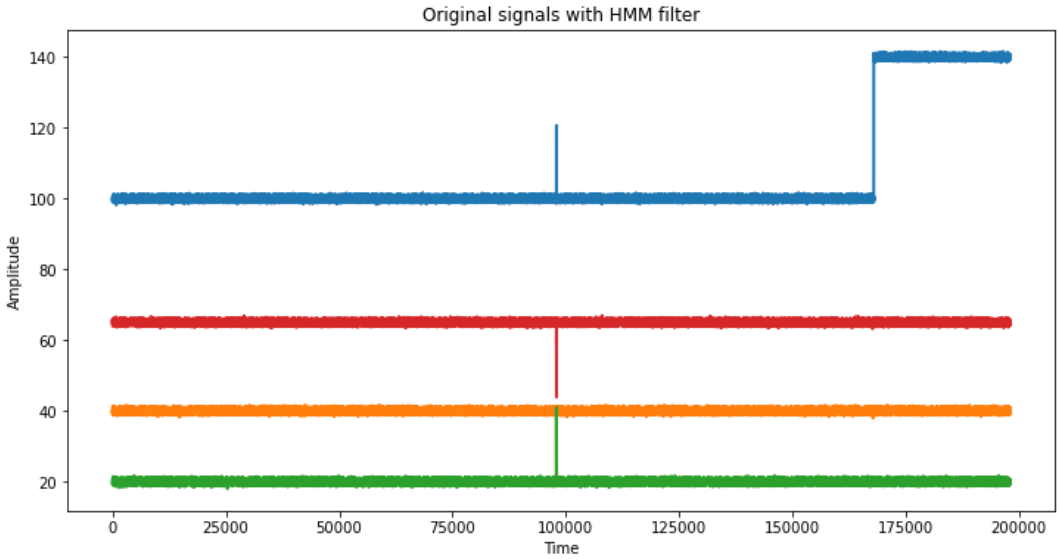



Figure 80 - Original signal after HMM filter.

#### 4.4.1.8 Classification of sensor behaviour

The sensor deviations are not correlated with each other, despite the correlation among the process variables [290]. This is verified by the methodology developed. Now, only the values that can contain the sensor errors are used for the study. Therefore, it is necessary to study individually each of the sensors to understand if they present deviations and, consequently need to be calibrated. The same methodology that was used above, and will be used with a difference in the generation of features. Now the features generated will be statistical features of the time domain and the methodology will be used from sensor to sensor. The statistical features based on the time domain provide a good basis for performance to characterize patterns and changes in behaviour in equipment [237]. The features used are represented in Table 14 and were taken from another paper that made detection of patterns of equipment behaviour over time [153], [156], [238]. In this way, we will understand the behaviour of each sensor individually since we no longer need their merged information.

TABLE 14 - TIME DOMAIN STATISTICAL FEATURES USED TO DETECT SENSOR BEHAVIOUR

| Parameter | Mathematical Equation | Parameter | Mathematical Equation |
|---|---|---|---|
| Mean | $T_1 = \frac{\sum_{n=1}^{N} x(n)}{N}$ | A Factor | $T_{12} = \frac{T_5}{T_2 . T_3}$ |
| Standard Deviation | $T_2 = \sqrt{\frac{\sum_{n=1}^{N} (x(n) - T_1)^2}{N-1}}$ | B Factor | $T_{13} = \frac{T_7 . T_8}{T_2}$ |
| Variance | $T_3 = \left( \frac{\sum_{n=1}^{N} \sqrt{\lvert x(n) \rvert}}{N} \right)^2$ | SRM | $T_{14} = \left( \frac{\sum_{n=1}^{N} \sqrt{x(n)}}{N} \right)^2$ |

| | | | |
|---|---|---|---|
| RMS | $T_4 = \sqrt{\frac{\sum_{n=1}^{N}(x(n))^2}{N}}$ | SRM Shape Factor | $T_{15} = \frac{T_{14}}{T_1}$ |
| Absolute Maximum | $T_5 = \max \lvert x(n) \rvert$ | Latitude Factor | $T_{16} = \frac{T_5}{T_{14}}$ |
| Coefficient of Skewness | $T_6 = \sqrt{\frac{\sum_{n=1}^{N}(x(n)-T_1)^3}{(N-1)T_2^3}}$ | Fifth Moment | $T_{17} = \sqrt{\frac{\sum_{n=1}^{N}(x(n)-T_1)^5}{(N-1)T_2^5}}$ |
| Kurtosis | $T_7 = \sqrt{\frac{\sum_{n=1}^{N}(x(n)-T_1)^4}{(N-1)T_2^4}}$ | Sixth Moment | $T_{18} = \sqrt{\frac{\sum_{n=1}^{N}(x(n)-T_1)^6}{(N-1)T_2^6}}$ |
| Crest Factor | $T_8 = \frac{T_5}{T_4}$ | Median | $T_{19} = meadian\ x(n)$ |
| Margin Factor | $T_9 = \frac{T_5}{T_3}$ | Mode | $T_{20} = mode\ x(n)$ |
| RMS Shape Factor | $T_{10} = \frac{T_4}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n) \rvert}$ | Minimum | $T_{21} = \min x(n)$ |
| Impulse Factor | $T_{11} = \frac{T_5}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n) \rvert}$ | | |

Applying the whole methodology individually and with the features in the time domain, it was possible to obtain the result represented in Figure 81. Through the hidden states of the HMM, we verified that only the first sensor presents significant errors, which leads to the need for calibration.

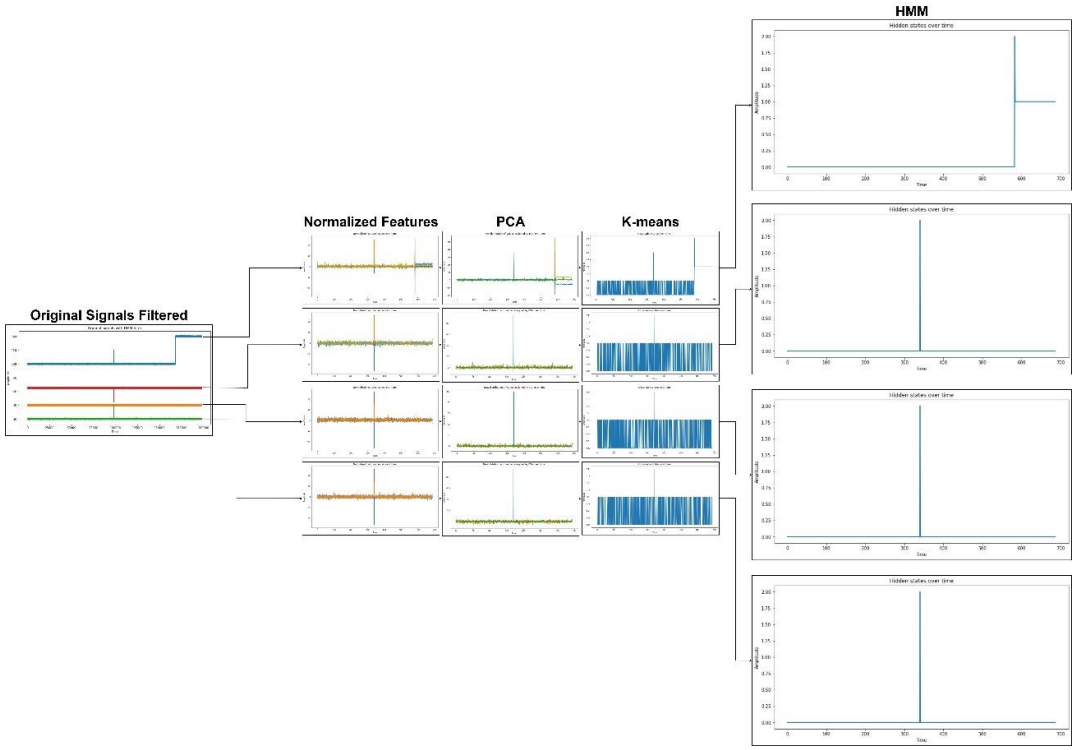



Figure 81 - Methodology used to detect sensor errors individually.

# Chapter 5

# Discussion

## 5.1 Main results and methodology development

The tests developed were aimed at demonstrating methodologies to determine the behaviour of equipment patterns. It was a gradual evolution with the objective of creating a unique methodology using the HMM as a classifier model of the health status of production equipment and reading equipment using for that the same data set.

In the first study, a method of creating optimised observable states was developed to give input to the HMM. In this methodology, only 10 features were used in the time domain and, through the HMM, the health state of a drying press of the paper industry was determined. The results obtained show that the developed methodology is capable of detecting the health status of the equipment throughout the study time. In addition, it is also possible to perceive that this methodology can work online mode, on any type of equipment with any type of sensors. In addition, to being a generic, unsupervised methodology, it also detects equipment operating patterns without prior information or technical knowledge about the equipment. Regarding the results of this study, the equipment had several alert states that return to a good functioning condition, in principle through (preventive) maintenance. Equipment failures happened several times over the three years, showing the need for improvement of a maintenance strategy, mainly CBM. The HMM obtained good evaluation results. Through a graphical analysis, it can be verified that the periods where the hidden states 2 and 3 of the HMM are found, coinciding with the abnormal values of the original signals coming from the sensors (Figure 33).

In the second study, the methodology was improved to be able to work with smaller datasets. For this, a DNN component was implemented, the MLP, which imputes data through complementary sensors to increase the dataset and, consequently improve the features generation component, which is the phase of generating more information about the data. Also, to improve this stage, new features, capable of better detecting the equipment behaviour, were added. Not only in the time domain but also in the frequency domain where, through FFT additional features can be generated in the study with vibration signals. Good results in fault detection can also be observed in this study. The

MLP demonstrated good model evaluation results and, as can be seen from the graphs obtained, the predicted values follow almost identically the behaviour of the original data (Figure 55). It can also be demonstrated the need for a dataset with all cycles of behaviours to train the HMM. But, it was shown through loglikelihood that the model was able to adapt the optimized observable states succeeding in making a good prediction of the hidden states (Table 10). Not only by the mathematical evaluations but also by a final graphical analysis, it can be seen that the malfunctioning hidden states determined by the HMM coincided with the data that came out outside the behaviour pattern of the original signals.

The third study aimed to create an HMM-GRU method, in order to add the prognosis to the methodology since it was already capable of performing the diagnosis of the equipment's health status. For the prognosis, the DNN tool GRU network was used. As in the literature, there is already a prediction for the future direction of the data collected by the sensors; in this methodology, the prognosis was made directly in the optimised observable states coming out from the Clustering and also in the hidden states of the HMM that already represent the equipment health state. It obtained a good evaluation, both in the Accuracy of the HMM to perform the diagnosis and the evaluation of the GRU model to make the prognosis. The results show that the prognosis can be made for a period of 7 days ahead, both on the inputs of the HMM, and also directly on the health status of the equipment. Thus, the methodology is capable of, within a period of 7 days ahead, directly describing the equipment health state, without having to make the prediction on the sensor data and then having to run the whole methodology from the beginning of the HMM.

Finally, the last study demonstrates that the methodology can also be used to classify the calibration status of sensors since the data is collected by this reading equipment. Through manipulation in the generation of features and an HMM filter, the methodology demonstrated good results, having the ability to remove two distinct pieces of information from the same set of data. Through the results and graphical analysis, it is possible to see that the developed methodology works with good results. First, it was able to detect only the malfunction data of the production equipment ignoring the errors that come from the sensors' behaviour (Figure 78). After that, it also proves to be able to detect the errors from the sensors that presented a deviation (Figure 81).

## 5.2 Advantages and limitations of the proposed methodology

As for the contribution of the developed methodology, it allows a diagnosis and prognosis using data from sensors attached to equipment, making it possible to detect failures of both production and reading devices. It allows for improving a CBM and PdM maintenance policies by increasing equipment availability as well as increasing profits. Furthermore, it allows the use of OLM methodology, increasing the extension or elimination of periodic sensor calibration intervals, improving the safety and reliability of observations through a higher precision and reliability of the sensors used, thus improving the reliability of data, what also improves CBM and PdM maintenance.

Furthermore, it is a methodology that does not require prior information on the equipment's status. That is, it classifies the health status of equipment without having prior information or technical knowledge about it. Through several unsupervised ML tools, it is capable of manipulating sensor data to detect patterns outside the regular operation of the equipment through the HMM, also being able, through DNN, to predict its anomalies in the future. It is also able, through DNN, to manage time series data to better diagnose the equipment state.

It is a generic methodology, easy to adapt the number of features depending on the equipment to be detected. It can be used in online mode to provide real-time insights data and support a predictive maintenance strategy of the condition of production equipment as well as an evaluation of the calibration status of reading equipment. Furthermore, the model can be generalised to any number of sensors and any equipment. The major limitation in the development of this methodology lies in the fact that there is no information on the equipment health status over time to be able to confirm the results obtained. Besides, another disadvantage is due to the diagnostic mode of the production equipment, which indicates the existence of a fault in the equipment, but does not specify which type of fault or which component. For this, it will always be necessary to support of a specialised technician.

# Chapter 6

# Conclusions and Future work

## 6.1 Problem Summary

A maintenance strategy based on equipment conditions requires the use of values collected by sensors attached to the production equipment, but it is necessary to ensure the quality and reliability of the data collected by those sensors. Only calibrated sensors guarantee the quality of the measured data. So, we can conclude that industrial metrology is the basis of good CBM and PdM maintenance.

A time-based type of maintenance strategy can lead to excessive maintenance or, even maintenance performs after the required time. Like maintenance, a sensor calibration strategy done at regular periods can lead to over-calibration or performing calibrations long after it is needed. This is why Online Calibration Monitoring (OLM) strategy is essential. This will make sure that sensors are only calibrated when needed and save immense labour in periodic checks.

So, this thesis aims to develop a real-time support methodology regarding CBM maintenance strategies as well as OLM calibration strategies using the Hidden Markov Model (HMM) for this purpose.

## 6.2 Contributions of the work to the state of art

The results of the studies carried out demonstrate that the methodology developed can detect failures in production equipment as well as in reading equipment. The main value added to the state of the art is the fact that a single developed methodology can integrate various functionalities autonomously. The methodology can detect equipment behaviour patterns without human intervention, prior information or technical knowledge. Furthermore, it is capable of imputing data in a small dataset also autonomously, since it uses optimisation tools to create the architecture of DNNs.

As for the prognosis, it is performed on classification states instead of predicting the original sensor data. This allows faster fault detection since it predicts the future direction of the equipment health state.

The health state of equipment is defined by a classification method rather than by linear regression methods. Although, the classification is already used to diagnose the state of production equipment, the case of reading devices is something innovative, even using the HMM method.

Also, something that stands out in the complement of the state of the art is the creation of a methodology that allows to easily manipulate the generation of features. This makes it possible to extract different information from the same data set for different types of equipment. That is, through the same set of data, this methodology can classify the state of health of production equipment as for reading equipment.

The HMM filter developed also allows new studies to be integrated to use linear regression methods already developed to detect deviations in reading equipment. These already developed methods worked, assuming that the historical data used to train the models did not contain errors from the production equipment. Thus, through the HMM filter, we can use data with equipment errors that will be eliminated without knowing in advance the periods of the equipment failures.

## 6.3 Future works

In future work, it is intended to further develop the methodology to create new capabilities and further support the decision method regarding CBM and OLM.

In the future, to retrain the HMM model with the new observable states that may happen, a supervised classification model, such as Support Vector Machines (SVM), will be built. The aim is to direct the new observable data read by the sensors directly to optimised observable states without going through the data optimisation phase. A test will also be performed using this method to classify the new data collected by the sensors directly to the hidden states of the HMM. In this way, it is possible to determine the health state of the equipment right away without going through the whole methodology. In this case, the methodology has to be trained to later teach the supervised classification method. For this, an adaptive methodology will be created, where, through a metric of the distance of entry of new points an alert will be given, if a set of new points appears very distant from the rest. If this happens, it means that there is a new behaviour in the equipment that had never appeared in the historical data cycle. Therefore, the whole methodology needs to be retrained.

Another aspect that can be improved is the fine-tuning of the methodology with other equivalent methods in each of the stages. Depending on the type of data, there may be

other tools more suitable for the study. As such, it will be necessary to compare different algorithms. In the first stage, a comparison of classes of Artificial Neural Networks can be made, such as: GRU, MLP; LSTM; CNN; etc. This is for the case of value imputation to perform the prognosis of the health status of the equipment. In the second phase, for dimensional reduction, the following methods can be compared: PCA; Linear Discriminant Analysis (LDA); Independent Component Analysis (ICA); etc. For Clustering, the following tools would be used: K-means; Density-based Spatial Clustering of Applications with Noise (DBSCAN); Gaussian Mixture Models (GMM); etc. Through this study and the combination of different models, it is possible to determine the best segment of the methodology for a given equipment.

Also, in feature generation, new features will be tested, and optimisation algorithms will be applied to determine the best features for using in each type of equipment.

# References


[1] J. D. Campbell and A. K. S. Jardine, *Maintenance Excellence: Optimizing Equipment Life-Cycle Decisions*, 1st Editio. New York: Marcel Dekker, Inc., 2001.

[2] H. Li and A. K. Parlikad, "Study of dynamic workload assignment strategies on production performance," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 13710–13715, 2017, doi: 10.1016/j.ifacol.2017.08.2547.

[3] J. M. T. Farinha, *Asset Maintenance Engineering Methodologies*, 1st Editio. Taylor & Francis Group, LCC, 2018.

[4] L. Da Xu, E. L. Xu, and L. Li, "Industry 4.0: state of the art and future trends," *Int. J. Prod. Res.*, vol. 56, no. 8, pp. 2941–2962, Apr. 2018, doi: 10.1080/00207543.2018.1444806.

[5] E. Oztemel and S. Gursev, "Literature review of Industry 4.0 and related technologies," *J. Intell. Manuf.*, vol. 31, no. 1, pp. 127–182, Jan. 2020, doi: 10.1007/s10845-018-1433-8.

[6] B. P. Santos, A. Alberto, T. D. F. M. Lima, and F. M. B. Charrua-Santos, "Indústria 4.0: desafios e oportunidades," *Rev. Produção e Desenvolv.*, vol. 4, no. 1, pp. 111–124, 2018, [Online]. Available: https://www.researchgate.net/publication/325060590_INDUSTRIA_40_DESAFIOS_E_OPORTUNIDADES.

[7] S. M. Lee, D. Lee, and Y. S. Kim, "The quality management ecosystem for predictive maintenance in the Industry 4.0 era," *Int. J. Qual. Innov.*, vol. 5, no. 4, pp. 1–11, 2019, doi: 10.1186/s40887-019-0029-5.

[8] J. Smit, S. Kreutzer, C. Moeller, and M. Carlberg, "Industry 4.0," *Publ. Off. Eur. Union*, pp. 1–90, 2016, doi: 10.2861/665669.

[9] J. M. G. Faustino, "Data Science - The State of the Art," Instituto Universitário de Lisboa, 2017.

[10] A. Wodecki, *Artificial Intelligence in Value Creation: Improving Competitive Advantage*. Cham: Springer International Publishing, 2019.

[11] J.-L. Monino and S. Sedkaoui, *Big Data, Open Data and Data Development*, vol. 3. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2016.

[12] L. Zhang, "Big Data Analytics for Fault Detection and its Application in

Maintenance," Luleå University of Technology, 2014.

[13] C. Zhu, "Big Data as a Governance Mechanism," *SSRN Electron. J.*, pp. 1–57, 2018, doi: 10.2139/ssrn.3164624.

[14] M. M. V. Inês, "Framework BigData," Faculdade de Engenharia da Universidade do Porto, 2018.

[15] C. R. da S. Teixeira, "Implementation of a data virtualization layer applied to insurance data," Faculdade de Engenharia da Universidade do Porto, 2016.

[16] M. Khosrow-Pour, *Big data: Concepts, Methodologies, Tools, and Applications*. Igi-global, 2016.

[17] V. Grover, R. H. L. Chiang, T.-P. Liang, and D. Zhang, "Creating Strategic Business Value from Big Data Analytics: A Research Framework," *J. Manag. Inf. Syst.*, vol. 35, no. 2, pp. 388–423, 2018, doi: 10.1080/07421222.2018.1451951.

[18] J. F. C. Nereu, "Open Source Private Cloud Platforms for Big Data," Instituto Superior de Engenharia do Porto, 2017.

[19] A. Oussous, F.-Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, "Big Data technologies: A survey," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 30, no. 4, pp. 431–448, Oct. 2018, doi: 10.1016/j.jksuci.2017.06.001.

[20] M. I. Baig, L. Shuib, and E. Yadegaridehkordi, "Big data adoption: State of the art and research challenges," *Inf. Process. Manag.*, vol. 56, no. 6, pp. 1–18, 2019, doi: 10.1016/j.ipm.2019.102095.

[21] Q. Qi and F. Tao, "Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison," *IEEE Access*, vol. 6, pp. 3585–3593, 2018, doi: 10.1109/ACCESS.2018.2793265.

[22] M. O. Gabriel, "Artificial Intelligence Opportunities and an End-To-End Data-Driven Solution for Predicting Hardware Failures," Massachusetts Institute of Technology, 2016.

[23] C. S. Krishnamoorthy and S. Rajeev, *ARTIFICIAL INTELLIGENCE and EXPERT SYSTEMS for ENGINEERS*, 1st Editio. Taylor & Francis Group, 1996.

[24] J. Kaplan, *Artificial Intelligence: What Everyone Needs to Know*. Oxford University Press, 2016.

[25] P. F. da S. P. Dinis, "Análise e Monitorização de Condição em Motores de Indução Trifásicos," Faculdade de Engenharia da Universidade do Porto, 2017.

[26] J. Lee, H. Davari, J. Singh, and V. Pandhare, "Industrial Artificial Intelligence for

industry 4.0-based manufacturing systems," *Manuf. Lett.*, vol. 18, pp. 20–23, 2018, doi: 10.1016/j.mfglet.2018.09.002.

[27] Altice, "Artificial intelligence," *Altice Labs*. pp. 1–56, 2017, [Online]. Available: https://www.alticelabs.com/content/WP-Artificial-Intelligence-MLearning.pdf.

[28] R. Pinto and T. Cerquitelli, "Robot fault detection and remaining life estimation for predictive maintenance," *2nd Int. Conf. Emerg. Data Ind. 4.0*, vol. 151, pp. 709–716, 2019, doi: 10.1016/j.procs.2019.04.094.

[29] T. Borangiu, D. Trentesaux, P. Leitão, A. G. Boggino, and V. Botti, *Service Oriented, Holonic and Multi-agent Manufacturing Systems for Industry of the Future*. Springer Nature Switzerland AG 2020, 2019.

[30] A. Mosavi, S. Mohsen, S. F. Ardabili, -timon Rabczuk, S. Shamshirband, and A. R. Varkonyi-Koczy, "State of the art of machine learning models in energy systems, a systematic review," *Energies*, vol. 12, no. 7, pp. 1–42, 2019, doi: 10.3390/en12071301.

[31] E. Sezer, D. Romero, F. Guedea, M. Macchi, and C. Emmanouilidis, "An Industry 4.0-Enabled Low Cost Predictive Maintenance Approach for SMEs," in *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 2018, pp. 1–8, doi: 10.1109/ICE.2018.8436307.

[32] B. X. Yong and A. Brintrup, "Multi Agent System for Machine Learning Under Uncertainty in Cyber Physical Manufacturing System Bang," in *Service Oriented, Holonic and Multi-agent Manufacturing Systems for Industry of the Future: Proceedings of SOHOMA 2019*, 2019, pp. 244–257, doi: 10.1007/978-3-030-27477-1_19.

[33] P. Guedes, *Metrologia Industrial*, 1st Editio. Lisboa: ETEP - Edições Técnicas e Profissionais, 2011.

[34] H. Eren, "Calibrations in Process Control," in *Instrument Engineers' Handbook*, 4th ed., vol. 3, no. January, B. G. Liptak and H. Eren, Eds. CRC Press, 2012, pp. 141–149.

[35] F. Tse and I. Morse, *Measurement And Instrumentation in Engineering: Principles and Basic Laboratory Experiments*. CRC Press, 2009.

[36] J. Juran and A. Godfrey, *JURAN'S QUALITY HANDBOOK*, 5th ed. McGraw-Hill Companies, Inc, 1998.

[37] N. D. Shah, E. W. Steyerberg, and D. M. Kent, "Big Data and Predictive Analytics," *JAMA*, vol. 320, no. 1, p. 27, 2018, doi: 10.1001/jama.2018.5602.

[38] A. Davies, *Handbook of Condition Monitoring: Techniques and Methodology*. Springer Netherlands, 1998.

[39] H. M. A. Fahmy, "WSNs Applications," in *Wireless Sensor Networks: Concepts, Applications, Experimentation and Analysis*, Springer, Singapore, 2016, pp. 69–213.

[40] A. Grous, *Applied Metrology for Manufacturing Engineering*, 1st Editio. Hoboken, NJ, USA: John Wiley & Sons, Inc, 2011.

[41] A. Shagluf, A. P. Longstaff, and S. Fletcher, "Maintenance Strategies to Reduce Downtime Due to Machine Positional Errors," *Maint. Perform. Meas. Manag.*, 2014, doi: 10.14195/978-972-8954-42-0_25.

[42] J. W. Hines, R. Seibert, and A. Usynin, *Technical Review of On-Line Monitoring Techniques for Performance Assessment. Volume 2: Theoretical Issues*, vol. 2. United States Nuclear Regulatory Commission, 2008.

[43] M. Hoffmann, "PEANO — a tool for on-line calibration monitoring," *OECD Halden React. Proj. - Inst. Energy Technol.*, no. January, 2004, [Online]. Available: https://www.researchgate.net/publication/228577605_PEANO-a_tool_for_on-line_calibration_monitoring.

[44] U. Leturiondo, "Hybrid Modelling in Condition Monitoring," Lulea University of Technology, 2016.

[45] C. V. Pinto, *Organização e Gestão da Manutenção*. Monitor, 2002.

[46] L. A. Ferreira, *Uma Introdução á Manutenç*, First Edit. Publindústria, Edições Técnicas, 1998.

[47] NP EN 13306, *Maintenance - Maintenance terminology*. CEN-CENELEC, EUROPEAN COMMITTEE FOR STANDARDIZATION, 2017.

[48] A. M. V. Conceição, "Manutenção Baseada na Condição na Tabaqueira, E.I.T," Faculdade de Engenharia da Universidade do PortoFaculdade de Engenharia da Universidade do Porto, 2013.

[49] P. Barbosa, "Elaboração do Plano de Manutenção e respetivos documentos de suporte para uma PME de carroçarias de veículos automóveis," Faculdade de Ciências e Tecnologia da Universidade de Coimbra, 2017.

[50] N. Faria, "Elaboração e implementação de um plano geral de manutenção preditiva, preventiva e curativa na Lipor – Serviço Intermunicipalizado de Gestão de Resíduos do Grande Porto," Faculdade de Engenharia da Universidade do Porto, 2013.

[51] P. Calvão, “Desenvolvimento do Plano de Manutenção para uma Enchedora Assética,” Faculdade de Ciências e Tecnologia da Universidade de Coimbra, 2018.

[52] J. Levitt, *Complete Guide to Preventive and Predictive Maintenance Second Edition*, Second Edi. Industrial Press Inc., 2011.

[53] A. P. Sponseller, “New Approaches to the Integration of Organizational Maintenance and Failure Data During Submarine Maintenance Planning,” Massachusetts Institute of Technology, 2019.

[54] J. P. S. Cabral, *Organização e Gestão da Manutenção: dos conceitos à prática…*, Sixth Edit. Lidel, 2006.

[55] H. Raposo, “Modelos de optimização do tempo de substituição de autocarros de transporte urbano com indexação à frota de reserva,” Universidade de Coimbra, 2017.

[56] J. Machado, “Controlo operacional da Manutenção,” Faculdade de Engenharia da Universidade do Porto, 2016.

[57] L. R. Higgins and R. K. Mobley, *Maintenance Engineering Handbook*, 6th Editio. McGraw-Hill Professional, 2002.

[58] S. Soares, “Introdução da Manutenção Preventiva Sistemática nos SASUC,” Universidade de Coimbra, 2015.

[59] M. S. Forsthoffer, “Predictive and Preventive Maintenance,” in *Forsthoffer’s More Best Practices for Rotating Equipment*, M. S. Forsthoffer, Ed. Butterworth-Heinemann, 2017, pp. 501–546.

[60] B. Guo, S. Song, A. Ghalambor, and T. R. Lin, “An Introduction to Condition-Based Maintenance,” in *Offshore Pipelines*, Second Edi., B. Guo, S. Song, A. Ghalambor, and T. R. Lin, Eds. 2014, pp. 257–297.

[61] R. Ahmad and S. Kamaruddin, “An overview of time-based and condition-based maintenance in industrial application,” *Comput. Ind. Eng.*, vol. 63, no. 1, pp. 135–149, 2012, doi: 10.1016/j.cie.2012.02.002.

[62] A. K. S. J. Ã, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006, doi: 10.1016/j.ymssp.2005.09.012.

[63] D. C. M. de P. Campos, “Implementação de técnicas e ferramentas para manutenção preditiva na Colep,” Faculdade de Engenharia da Universidade do Porto, 2011.

[64] G. Allenby, “Condition based maintenance,” in *Condition Monitoring and Diagnostic Engineering Management*, R. B. K. . Rao, J. Au, and B. Griffiths, Eds. Dordrecht: Springer, 1990, pp. 155–161.

[65] W. Wang, “Condition-based Maintenance Modelling,” in *Complex System Maintenance Handbook*, K. A. H. Kobbacy and D. N. P. Murthy, Eds. London: Springer, 2008, pp. 111–131.

[66] D. Campos, “Implementação de técnicas e ferramentas para manutenção preditiva na Colep,” Faculdade de Engenharia da Universidade do Porto, 2011.

[67] D. Goyal and B. S. Pabla, “Condition based maintenance of machine tools — A review,” *CIRP J. Manuf. Sci. Technol.*, vol. 10, pp. 24–35, 2015, doi: 10.1016/j.cirpj.2015.05.004.

[68] E. Quatrini, F. Costantino, G. Di Gravio, and R. Patriarca, “Condition-Based Maintenance — An Extensive Literature Review,” *Machines*, vol. 8, no. 2, pp. 1–28, 2020, doi: 10.3390/machines8020031.

[69] A. Prajapati, J. Bechtel, and S. Ganesan, “Condition based maintenance : a survey,” *J. Qual. Maint. Eng.*, vol. 27, no. 4, pp. 384–400, 2012, doi: 10.1108/13552511211281552.

[70] E. Florian, F. Sgarbossa, and I. Zennaro, “Machine learning-based predictive maintenance : A cost-oriented model for implementation,” *Int. J. Prod. Econ.*, vol. 236, no. February, p. 108114, 2021, doi: 10.1016/j.ijpe.2021.108114.

[71] P. Wild, *Industrial Sensors and Applications for Condition Monitoring*. London: Mechanical Engineering Publications, MEP, 1994.

[72] C. Emmanouilidis and P. Pistofidis, “Wireless Condition Monitoring and Embedded Novelty Detection,” in *Engineering Asset Management Review*, vol. 1, M. J. Amadi-Echendu J., Brown K., Willett R., Ed. Springer, London, 2010, pp. 195–238.

[73] R. K. Mobley, “Condition based maintenance,” in *Handboock of Condition Monitoring*, First edit., A. Davies, Ed. Springer, 1998, pp. 35–53.

[74] S. Sadiki, M. Faccio, M. Ramadany, D. Amgouz, and S. Boutahar, “Impact of intelligent wireless sensor network on predictive maintenance cost,” in *2018 4th International Conference on Optimization and Applications (ICOA)*, Apr. 2018, pp. 1–6, doi: 10.1109/ICOA.2018.8370573.

[75] A. B. Martins, J. Rodrigues, and J. Farinha, “Laboratório de Metrologia no apoio à Manutenção de Condição,” *15.º Congr. Nac. manutenção*, 2019, [Online].

Available: https://www.academia.edu/42144637/Laboratório_de_Metrologia_no_apoio_à_Manutenção_de_Condição.

[76] R. Smith and R. K. Mobley, *Industrial Machinery Repair: Best Maintenance Pratices Pocket Guide*, 1st Editio. Elsevier Science (USA), 2003.

[77] K. K. Iniewski, *Smart Sensors for Industrial Applications*. CRC Press, 2017.

[78] M. J. McGrath and C. N. Scanaill, *Sensor Technologies: Healthcare, Wellness and Enviromental Applications*. Berkeley, CA: Apress, 2013.

[79] A. Martins, "Gestão de Laboratório de Calibração de Equipamentos de Análises de Vibrações," Instituto Superior de Engenharia de Coimbra, 2019.

[80] J. Wilson, *Sensor Technology HandBook*. Elsiver Inc., 2005.

[81] D. Scott, *Industrial Process Sensors*, 1st Editio. CRC Press, 2008.

[82] M. J. McGrath and C. N. Scanaill, "Sensing and Sensor Fundamentals," in *Sensor Technologies*, no. July 2016, Berkeley, CA: Apress, 2013, pp. 15–50.

[83] KeysightTechnologies, "Fundamentals of Sensor Measurements Using a Digital Multimeter," 2016. https://www.keysight.com/zz/en/assets/7018-06443/white-papers/5992-3555.pdf (accessed Mar. 05, 2020).

[84] M. M. Zanjireh and H. Larijani, "A Survey on Centralised and Distributed Clustering Routing Algorithms for WSNs," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, May 2015, vol. 2015, no. April, pp. 1–6, doi: 10.1109/VTCSpring.2015.7145650.

[85] M. Jaatinen, "Evaluation and Improving the Performance of the Arrowhead Framework," FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO, 2019.

[86] J. P. Talampas, "Electricity Sensors for Resource Efficiency and Supply Chain Visibility in Factories," Massachusetts Institute of Technology, 2019.

[87] W. Bolton, *Programmable Logic Controllers*, Fourth Edi. Elsevier Newnes, 2006.

[88] R. Nicolescu, M. Huth, P. Radanliev, and D. De Roure, "State of The Art in IoT - Beyond Economic Value," Iot Uk, 2018.

[89] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep Learning for IoT Big Data and Streaming Analytics: A Survey," *IEEE Commun. Surv. Tutorials*, vol. 20, no. 4, pp. 2923–2960, 2018, doi:

10.1109/COMST.2018.2844341.

[90] S. Addepalli, Y. Zhao, and L. Tinsley, *Advances in Through-life Engineering Services*. Cham: Springer International Publishing, 2017.

[91] C. T. A. Lai, W. Jiang, and P. R. Jackson, “Internet of Things enabling condition-based maintenance in elevators service,” *J. Qual. Maint. Eng.*, vol. 25, no. 4, pp. 563–588, Oct. 2019, doi: 10.1108/JQME-06-2018-0049.

[92] S. G. Rabinovich, *Measurements and Uncertainties: Theory and Practice*, 3rd Editio. Springer, 2005.

[93] J. F. F. Braz, “Conceção e Validação de uma Mesa de Vibração para a Calibração de Acelerómetros,” Universidade Nova de Lisboa, 2015.

[94] A. R. F. Pires, “Organização de um Processo de Gestão de Recursos de Monitorização e Medição,” Universidade de Coimbra, 2018.

[95] P. Flores and F. Marques, “Sobre a metrologia,” Universidade do Minho, 2018.

[96] A. I. A. Castro, “Implementação de um Laboratório de Implementação de um Laboratório de Metrologia Elétrica,” Instituto Superior de Engenharia do Porto, 2015.

[97] J. F. F. Braz, “Conceção e Validação de uma Mesa de Vibração para a Calibração de Acelerómetros,” Universidade Nova de Lisboa, 2015.

[98] C. Sousa, “Metrologia Notas Históricas,” 2010.

[99] R. Krishnakumar, “Entangling Atoms for Quantum Metrology,” Stanford University, 2017.

[100] T. Lopes, “Implementação de procedimentos de metrologia na empresa Milon,” Faculdade Engenharia da Universidade do Porto, 2015.

[101] J. C. da S. Neto, *Metrologia e Controle Dimensional*. Elsivier, 2012.

[102] P. Cabral, C. Sousa, E. Filipe, P. Couto, and P. Girão, “A Metrologia como ferramenta transversal da Engenharia,” Lisboa, 2015.

[103] P. flávio da S. Santos, “A Metrologia na Indústria Aeronáutica,” Instituto Superior de Engenharia do Porto, 2015.

[104] R. K. Mobley, *Maintenance Fundamentals*, 2nd Editio. Elsevier Inc., 2004.

[105] S. Lee and J. Chai, “An enhanced prediction model for the on-line monitoring of the sensors using the Gaussian process regression,” *J. Mech. Sci. Technol.*, vol. 33, no. 5, pp. 2249–2257, May 2019, doi: 10.1007/s12206-019-0426-7.

[106] A.-K. Wiemann, M. Stein, and K. Kniel, “Traceable metrology for large involute

gears," *Precis. Eng.*, vol. 55, no. September 2018, pp. 330–338, Jan. 2019, doi: 10.1016/j.precisioneng.2018.10.001.

[107] V. V. Skliarov, P. Neyezhmakov, and A. Prokopov, "Metrological assurance and traceability for Industry 4.0 and additive manufacturing in Ukraine," in *Smart Structures and NDE for Industry 4.0*, Mar. 2018, no. March, p. 26, doi: 10.1117/12.2292821.

[108] K. Singhal, Q. Feng, R. Ganeshan, N. R. Sanders, and J. G. Shanthikumar, "Introduction to the Special Issue on Perspectives on Big Data," *Prod. Oper. Manag.*, vol. 27, no. 9, pp. 1639–1641, Sep. 2018, doi: 10.1111/poms.12939.

[109] I. Zakharov and P. Neyezhmakov, "Peculiarity of measurement instruments verification by results of their calibrations," in *2017 11th International Conference on Measurement*, May 2017, no. Imi, pp. 19–22, doi: 10.23919/MEASUREMENT.2017.7983526.

[110] J. Coble, P. Ramuhalli, R. Meyer, H. Hashemian, B. Shumaker, and D. Cummins, "Calibration Monitoring for Sensor Calibration Interval Extension Identifying Technical Gaps," 2012, doi: 10.1109/FIIW.2012.6378348.

[111] J. B. Coble *et al.*, "A Review of Sensor Calibration Monitoring for Calibration Interval Extension in Nuclear Power Plants," no. August. Pacific Northwest National Laboratory, pp. 1–52, 2012, [Online]. Available: https://www.pnnl.gov/main/publications/external/technical_reports/pnnl-21687.pdf.

[112] F. Alves, "ABC da Metrologia Industrial," Instituto Superior de Engenharia Do Porto, 2003.

[113] A. J. Schiff, *Power Plant Surveillance and Diagnostics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002.

[114] M. F. Alves, "ABC Da Metrologia Industrial," Instituto Superior de Engenharia do Porto, 2003.

[115] P. Ramuhalli *et al.*, *Uncertainty Quantification Techniques for Sensor Calibration Monitoring in Nuclear Power Plants*, no. April. Washington: Pacific Northwest National Laboratory, 2014.

[116] S. Dwivedi, Al. Verma, and J. Sneckenberger, "CAD/CAM Robotics and Factories of the Future '90," in *International Conference on CAD/CAM, Robotics, and Factories of the Future*, 1989, doi: 10.1007/978-3-662-39962-0.

[117] J. W. Hines, "On-Line Monitoring for Calibration Extension : an Overview and

Introduction," U.S. Nuclear Regulatory Commission, 2009.

[118] A. A. Boechat, U. F. Moreno, and D. Haramura, "On-line calibration monitoring system based on data-driven model for oil well sensors," *IFAC Proc. Vol.*, vol. 45, no. 8, pp. 269–274, 2012, doi: 10.3182/20120531-2-NO-4020.00037.

[119] J. W. Hines and R. Seibert, *Technical Review of On-Line Monitoring Techniques for Performance Assessment. Volume 1. State-of-the-Art*, vol. 1. U.S. Nuclear Regulatory Commission, 2006.

[120] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," in *The Annals of Mathematical Statistics*, 1970, pp. 164–171, [Online]. Available: https://www.jstor.org/stable/2239727.

[121] J. k. Baker, "STOCHASTIC MODELING AS A MEANS OF AUTOMATIC SPEECH RECOGNITION," Carriegie-Mellon University.

[122] A. Simões, "Manutenção condicionada às emissões poluentes em autocarros urbanos," Instituto Superior Técnico, 2011.

[123] L. Baum and T. Petrie, "Statistical Inference For Probabilistic Functions Of Finite State Markov Chains," *Ann. Math. Stat.*, vol. 37, no. April, pp. 1554–1563, 1966, [Online]. Available: https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-37/issue-6/Statistical-Inference-for-Probabilistic-Functions-of-Finite-State-Markov-Chains/10.1214/aoms/1177699147.full.

[124] L. R. Rabiner and B. H. Juang, "An Introduction to Hidden Markov Models," *IEEE ASSP Mag.*, vol. 3, no. 1, pp. 4–16, 1986.

[125] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989, doi: 10.1109/5.18626.

[126] A. Simões, J. M. Viegas, J. T. Farinha, and I. Fonseca, "The State of the Art of Hidden Markov Models for Predictive Maintenance of Diesel Engines," *Qual. Reliab. Eng. Int.*, vol. 33, no. 8, pp. 2765–2779, 2017, doi: 10.1002/qre.2130.

[127] X. Yin, Z. Zhou, N. Li, and S. Chen, "An Approach for Data Filtering Based on Rough Set Theory," *Lect. Notes Comput. Sci.*, vol. 2118, pp. 367–374, 2001, doi: 10.1007/3-540-47714-4_33.

[128] I. Taleb, R. Dsouli, and M. Serhani, "Big Data Pre-Processing : A Quality Framework," *IEEE Int. Congr. Big Data*, 2015, doi:

10.1109/BigDataCongress.2015.35.

[129] G. S. Prasad, N. V. S. Reddy, and U. D. Acharya, “Knowledge Discovery from Web Usage Data : A Survey of Web Usage Pre-processing Techniques,” in *Information Processing and Management. BAIP 2010. Communications in Computer and Information Science*, 2010, vol. 10, pp. 505–507, doi: 10.1007/978-3-642-12214-9_88.

[130] M. Kuhn and K. Johnson, “Data Pre-processing,” in *Applied Predictive Modeling*, New York: Springer Science+Business Media, 2013.

[131] A. G. Banhatti and P. C. Deka, *Effects of Data Pre-processing on the Prediction Accuracy of Arti fi cial Neural Network Model in Hydrological Time Series*. Springer International Publishing, 2016.

[132] W. Cho and E. Choi, “Big data pre-processing methods with vehicle driving data using MapReduce techniques,” *J. Supercomput.*, 2017, doi: 10.1007/s11227-017-2014-x.

[133] A. Rinnan, L. Norgaard, F. Berg, J. Thygesen, R. Bro, and B. Engelsen, “Data Pre-processing,” in *Infrared Spectroscopy for Food Quality Analysis and Control*, Elsivier Inc., 2009.

[134] R. Nisbet, G. Miner, and K. Yale, “Data Understanding and Preparation,” in *Handbook of Statistical Analysis and Data Mining Applications*, Elsevier Inc., 2018, pp. 55–82.

[135] Z. S. Abdallah and G. I. Webb, “Data Preparation,” in *Encyclopedia of Machine Learning and Data Mining*, 2017, no. January, doi: 10.1007/978-1-4899-7687-1_62.

[136] T. Taulli, *Implementing AI Systems: Transform your Business in 6 Steps*. Monrovia: Apress, 2021.

[137] A. B. Martins, J. T. Farinha, and A. M. Cardoso, “Calibration and Certification of Industrial Sensors – a Global Review,” *WSEAS Trans. Syst. Control*, vol. 15, pp. 394–416, Sep. 2020, doi: 10.37394/23203.2020.15.41.

[138] S. Zhang, C. Zhang, and Q. Yang, “Data preparation for data mining,” *Appl. Artif. Intell.*, no. Nov, pp. 375–381, 2003, doi: 10.1080/713827180.

[139] ROADIDEA, “Data filtering methods,” *Eur. Com. under 7th Community Framew. Program. Res. Technol. Dev.*, 2008.

[140] R. R. Deshmukh and V. Wangikar, “Data Cleaning: Current Approaches and Issues,” 2015.

[141] S. Sakr and A. Zomaya, *Encyclopedia of Big Data Technologies*. Switzerland: Springer, 2019.

[142] V. Ganti and A. Sarma, *Data Cleaning Cleaning: A Pratical Perspective*. Morgan &Claypool, 2013.

[143] H. H. Do, A. Broder, J. Dean, M. R. Henzinger, K. Deshmukh, and S. Sarawagi, "Bulletin of Technical Committee on Data Engineering," in *Special Issue on Data Cleaning*, 2000, vol. 23, no. 4.

[144] J. Van Den Broeck, S. Cunningham, and A. Herbst, "Data Cleaning : Detecting , Diagnosing , and Editing Data Abnormalities," *PLoS Med.*, vol. 2, no. October, 2005, doi: 10.1371/journal.pmed.0020267.

[145] X. Chen and X. Zhang, "Extract-Transform-Load of Data Cleaning Method in Electric Company," *Int. Conf. Artif. Intell. Comput. Intell.*, pp. 345–349, 2010, doi: 10.1109/AICI.2010.310.

[146] S. Roy, P. Sharma, K. Nath, D. Bhattacharyya, and J. Kalita, "Pre-Processing : A Data Preparation Step," *Encycl. Bioinforma. Comput. Biol.*, pp. 1–9, 2018, doi: 10.1016/B978-0-12-809633-8.20457-3.

[147] T. Dasu and T. Johnson, "Data Quality : Techniques and Algorithms," in *Exploratory Data Mining and Data Cleaning*, Wiley-Interscience, 2003, p. 224.

[148] Z. Han, Q. He, and M. Von Davier, "Predictive Feature Generation and Selection Using Process Data From PISA Interactive Problem-Solving Items : An Application of Random Forests," vol. 10, no. November, pp. 1–15, 2019, doi: 10.3389/fpsyg.2019.02461.

[149] A. M. De Silva and P. H. W. Leong, *Grammar- Based Feature Generation for Time-Series Prediction*. London: Springer, 2015.

[150] S. Theodoridis and K. Koutroumbas, "Feature Generation I: Data Transformation and Dimensionality Reduction," in *Pattern Recognition*, 2009, pp. 323–409.

[151] S. Van Den Bosch, "Automatic feature generation and selection in predictive analytics solutions," Radboud Universiteit, 2017.

[152] J. Kacprzyk, "Studies in Computational Intelligence," in *Software Engineering Research, Management and Applications*, 2012, doi: 10.1007/978-3-642-30460-6.

[153] A. Martins, I. Fonseca, J. T. Farinha, J. Reis, and A. M. Cardoso, "Maintenance prediction through sensing using hidden markov models—a case study," *Appl.*

*Sci.*, vol. 11, no. 16, pp. 1–20, 2021, doi: 10.3390/app11167685.

[154] H. Chen, J. Hsu, J. Y. Hsieh, H. Hsu, C. Chang, and Y.-J. Lin, "Predictive maintenance of abnormal wind turbine events by using machine learning based on condition monitoring for anomaly detection," *J. Mech. Sci. Technol.*, vol. 35, no. 12, pp. 5323–5333, 2021, doi: 10.1007/s12206-021-1105-z.

[155] M. Zhang *et al.*, "Predictive Maintenance for Remanufacturing Based on Hybrid-Driven Remaining Useful Life Prediction," *Appl. Sci.*, vol. 12, no. 3218, pp. 1–15, 2022, doi: 10.3390/app12073218.

[156] A. Martins, I. Fonseca, F. J. Torres, J. Reis, and A. J. M. Cardoso, "Prediction Maintenance based on Vibration Analysis and Deep Learning - A case study of a drying press supported on Hidden Markov Model," *SSRN*, 2022, doi: 10.2139/ssrn.4194601.

[157] Y. Cheng, K. Hu, J. Wu, H. Zhu, and C. K. M. Lee, "A deep learning-based two-stage prognostic approach for remaining useful life of rolling bearing," *Appl. Intell.*, Aug. 2021, doi: 10.1007/s10489-021-02733-4.

[158] H. Pranesh, K. Suresh, S. S. Manian, R. Jegadeeshwaran, G. Sakthivel, and T. M. Alamelu Manghi, "Vibration-based brake health prediction using statistical features – A machine learning framework," *Mater. Today Proc.*, vol. 46, pp. 1167–1173, 2021, doi: 10.1016/j.matpr.2021.02.060.

[159] L. S. Dhamande and M. B. Chaudhari, "Detection of Combined Gear-Bearing Fault in Single Stage Spur Gear Box Using Artificial Neural Network," *Procedia Eng.*, vol. 144, pp. 759–766, 2016, doi: 10.1016/j.proeng.2016.05.082.

[160] N. K. Booker, P. Knights, J. D. Gates, and R. E. Clegg, "Applying principal component analysis ( PCA ) to the selection of forensic analysis methodologies," *Eng. Fail. Anal.*, vol. 132, p. 105937, 2022, doi: 10.1016/j.engfailanal.2021.105937.

[161] G. Merola and B. Abraham, "Dimension Reduction Methods Used in Industry," *Handb. Stat.*, vol. 22, 2003, doi: 10.1016/S0169-7161(03)22029-0.

[162] F. Anowar, S. Sadaoui, and B. Selim, "Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)," *Comput. Sci. Rev.*, vol. 40, no. 100378, pp. 1–13, 2021, doi: 10.1016/j.cosrev.2021.100378.

[163] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Mag. J. Sci. LIII*, 1901, doi: 10.1080/14786440109462720.

[164] H. Hotelling, “Relations Between Two Sets of Variates,” *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[165] S. Tipler, G. D. Alessio, Q. Van Haute, A. Parente, F. Contino, and A. Coussement, “Predicting octane numbers relying on principal component analysis and artificial neural network,” *Comput. Chem. Eng.*, vol. 161, p. 107784, 2022, doi: 10.1016/j.compchemeng.2022.107784.

[166] K. S. Kiangala and Z. Wang, “An Effective Predictive Maintenance Framework for Conveyor Motors Using Dual Time-Series Imaging and Convolutional Neural Network in an Industry 4 . 0 Environment,” *IEEE Access*, vol. 8, pp. 121033–121049, 2020, doi: 10.1109/ACCESS.2020.3006788.

[167] Y. Gu, X. Zhou, D. Yu, and Y. Shen, “Fault diagnosis method of rolling bearing using principal component analysis and support vector machine,” *J. Mech. Sci. Technol.*, vol. 32, no. 11, pp. 5079–5088, 2018, doi: 10.1007/s12206-018-1004-0.

[168] A. Kamari and C. Schultz, “A combined principal component analysis and clustering approach for exploring enormous renovation design spaces,” *J. Build. Eng.*, vol. 48, p. 103971, 2022, doi: 10.1016/j.jobe.2021.103971.

[169] T. Qiu, M. Liu, G. Zhou, L. Wang, and K. Gao, “An Unsupervised Classification Method for Flame Image of Pulverized Coal Combustion Based on Convolutional Auto-Encoder and Hidden Markov Model,” *Energies*, vol. 12, no. 13, p. 2585, Jul. 2019, doi: 10.3390/en12132585.

[170] Y. Lim, J. Kwon, and H. Oh, “Principal component analysis in the wavelet domain,” *Pattern Recognit.*, vol. 119, p. 108096, 2021, doi: 10.1016/j.patcog.2021.108096.

[171] J. Warmenhoven *et al.*, “PCA of waveforms and functional PCA : A primer for biomechanics,” *J. Biomech.*, vol. 116, p. 110106, 2021, doi: 10.1016/j.jbiomech.2020.110106.

[172] Z. Zhang, Y. Wang, and K. Wang, “Intelligent fault diagnosis and prognosis approach for rotating machinery integrating wavelet transform , principal component analysis , and artificial neural networks,” *Int J Adv Manuf Technol*, vol. 68, pp. 763–773, 2013, doi: 10.1007/s00170-013-4797-0.

[173] T. D. Popescu, D. Aiordachioaie, and A. Culea-Florescu, “Basic tools for vibration analysis with applications to predictive maintenance of rotating machines : an overview,” *Int. J. Adv. Manuf. Technol.*, vol. 118, pp. 2883–2899, 2022, doi: 10.1007/s00170-021-07703.

[174] M. K. Babouri, A. Djebala, N. Ouelaa, B. Oudjani, and R. Younes, “Rolling bearing faults severity classification using a combined approach based on multi-scales principal component analysis and fuzzy technique,” *Int. J. Adv. Manuf. Technol.*, vol. 107, pp. 4301–4316, 2020, doi: 10.1007/s00170-020-05342-6.

[175] T. Zhu, X. Cheng, W. Cheng, Z. Tian, and Y. Li, “Principal component analysis based data collection for sustainable internet of things enabled Cyber – Physical Systems,” *Microprocess. Microsyst.*, vol. 88, p. 104032, 2022, doi: 10.1016/j.micpro.2021.104032.

[176] X. Wang, *Data Mining and Knowledge Discovery for Process Monitoring and Control: Advances in Industrial Control*. London: Springer-Verlag, 1999.

[177] D. Han and S. Kwon, “Application of Machine Learning Method of Data-Driven Deep Learning Model to Predict Well Production Rate in the Shale Gas Reservoirs,” *Energies*, vol. 14, no. 12, p. 3629, Jun. 2021, doi: 10.3390/en14123629.

[178] F. W. Townes, S. C. Hicks, M. J. Aryee, and R. A. Irizarry, “Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model,” *Genome Biol.*, vol. 20, no. 295, pp. 1–16, 2019, doi: 10.1186/s13059-019-1861-6.

[179] R. M. Alguliyev, R. M. Aliguliyev, and L. V Sukhostat, “Computers & Industrial Engineering Parallel batch k-means for Big data clustering,” *Comput. Ind. Eng.*, vol. 152, no. December 2020, p. 107023, 2021, doi: 10.1016/j.cie.2020.107023.

[180] N. Nidheesh, “A Hierarchical Clustering algorithm based on Silhouette Index for cancer subtype discovery from genomic data,” *Neural Comput. Appl.*, vol. 32, pp. 11459–11476, 2019, doi: 10.1007/s00521-019-04636-5.

[181] T. Wendler and S. Gröttrup, “Cluster Analysis,” in *Data Mining with SPSS Modeler: Theory, Exercises and Solutions*, Switzerland: Springer International, 2016, pp. 587–712.

[182] L. Haoran, H. Fazhi, and C. Yilin, “Learning dynamic simultaneous clustering and classification via automatic differential evolution and firework algorithm,” *Appl. Soft Comput. J.*, vol. 96, p. 106593, 2020, doi: 10.1016/j.asoc.2020.106593.

[183] D. Peralta and Y. Saeys, “Robust unsupervised dimensionality reduction based on feature clustering for single-cell imaging data,” *Appl. Soft Comput. J.*, vol. 93, p. 106421, 2020, doi: 10.1016/j.asoc.2020.106421.

[184] R. M. Alguliyev, R. M. Aliguliyev, and L. V Sukhostat, “Parallel batch k-means for Big data clustering,” *Comput. Ind. Eng.*, vol. 152, no. December, p. 107023, 2021,

doi: 10.1016/j.cie.2020.107023.

[185] S. Mary and T. Selvi, “A Study of K-Means and Cure Clustering Algorithms,” *Int. J. Eng. Res. Technol.*, vol. 3, no. 2, pp. 1985–1987, 2014.

[186] G. Menardi, “Density-based Silhouette diagnostics for clustering methods,” *Stat. Comput.*, vol. 21, no. January, pp. 295–308, 2011, doi: 10.1007/s11222-010-9169-0.

[187] V. Ferreira, A. Pinho, D. Souza, and B. Rodrigues, “A New Clustering Approach for Automatic Oscillographic Records Segmentation,” *Energies*, vol. 14, no. 6778, pp. 1–18, 2021, doi: 10.3390/en14206778.

[188] Z. Zhang, A. Cheshmehzangi, and S. P. Ardakani, “A Data-Driven Clustering Analysis for the Impact of COVID-19 on the Electricity Consumption Pattern of Zhejiang Province, China,” *Energies*, vol. 14, no. 8187, pp. 1–22, 2021, doi: 10.3390/ en14238187.

[189] C. Subbalakshmi, G. R. Krishna, S. K. Mohan, and P. V. Rao, “A Method to Find Optimum Number of Clusters Based on Fuzzy Silhouette on Dynamic Data Set,” *Procedia - Procedia Comput. Sci.*, vol. 46, pp. 346–353, 2015, doi: 10.1016/j.procs.2015.02.030.

[190] J. H. Park and Y. J. Kang, “EVALUATION INDEX FOR SPORTY ENGINE SOUND REFLECTING EVALUATORS ’ TASTES , DEVELOPED USING K-MEANS CLUSTER ANALYSIS,” *Int. J. Automot. Technol.*, vol. 21, no. 6, pp. 1379–1389, 2020, doi: 10.1007/s12239-020-0130-8.

[191] Y. Yang, Q. Liao, J. Wang, and Y. Wang, “Application of multi-objective particle swarm optimization based on short-term memory and K-means clustering in multi-modal multi-objective optimization,” *Eng. Appl. Artif. Intell.*, vol. 112, p. 104866, 2022, doi: 10.1016/j.engappai.2022.104866.

[192] L. I. Voronova, V. Voronov, and N. Mohammad, “Modeling the Clustering of Wireless Sensor Networks Using the K-means Method,” *Int. Conf. Qual. Manag. Transp. Inf. Secur. Inf. Technol. (IT QM IS)*, pp. 740–745, 2021, doi: 10.1109/ITQMIS53292.2021.9642747.

[193] K. P. Sinaga and M. Yang, “Unsupervised K-Means Clustering Algorithm,” *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[194] K. P. Sinaga, I. Hussain, and M. Yang, “Entropy K-Means Clustering With Feature Reduction Under Unknown Number of Clusters,” *IEEE Access*, vol. 9, pp. 67736–67751, 2021, doi: 10.1109/ACCESS.2021.3077622.

[195] Q. Liao, L. Xue, G. Lei, X. Liu, S. Sun, and S. Patil, “Statistical prediction of water fl ooding performance by K-means clustering and empirical modeling,” *Pet. Sci.*, pp. 1–14, 2022, doi: 10.1016/j.petsci.2021.12.032.

[196] J. Yoo, Y. Park, and S. Han, “Predictive Maintenance System for Wafer Transport Robot Using K-Means Algorithm and Neural Network Model,” *Electronics*, vol. 11, no. 1324, pp. 1–14, 2022, doi: 10.3390/electronics11091324.

[197] F. Aksan *et al.*, “Clustering Methods for Power Quality Measurements in Virtual Power Plant,” *Energies*, vol. 14, no. 5902, pp. 1–20, 2021, doi: 10.3390/en14185902.

[198] H. Talei, D. Benhaddou, C. Gamarra, H. Benbrahim, and M. Essaaidi, “Smart Building Energy Inefficiencies Detection through Time Series Analysis and Unsupervised Machine Learning,” *Energies*, vol. 14, no. 6042, pp. 1–21, 2021, doi: 10.3390/ en14196042.

[199] P. Han, W. Wang, Q. Shi, and J. Yue, “A combined online-learning model with K-means clustering and GRU neural networks for trajectory prediction,” *Ad Hoc Networks*, vol. 117, p. 102476, 2021, doi: 10.1016/j.adhoc.2021.102476.

[200] R. Llet, M. C. Ortiz, L. A. Sarabia, and M. S. Sánchez, “Selecting variables for k - means cluster analysis by using a genetic algorithm that optimises the silhouettes,” *Anal. Chim. Acta*, vol. 515, pp. 87–100, 2004, doi: 10.1016/j.aca.2003.12.020.

[201] L. Chen, W. Shan, and P. Liu, “Identification of concrete aggregates using K-means clustering and level set method,” *Structures*, vol. 34, pp. 2069–2076, 2021, doi: 10.1016/j.istruc.2021.08.048.

[202] A. Wickramasinghe, S. Muthukumarana, D. Loewen, and M. Schaubroeck, “Temperature clusters in commercial buildings using k - means and time series clustering,” *Energy Informatics*, vol. 5, no. 1, pp. 1–14, 2022, doi: 10.1186/s42162-022-00186-8.

[203] A. E. Ezugwu, A. M. Ikotun, O. O. Oyelade, and L. Abualigah, “A comprehensive survey ofclustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects,” *Eng. Appl. Artif. Intell.*, vol. 110, no. February, p. 104743, 2022, doi: 10.1016/j.engappai.2022.104743.

[204] I. Borlea, R. Precup, and A. Borlea, “Improvement of K-means Cluster Quality by Post Processing Resulted Clusters,” *8th Int. Conf. Inf. Technol. Quant. Manag.*, vol. 199, pp. 63–70, 2022, doi: 10.1016/j.procs.2022.01.009.

[205] K. Peng, J. Tan, and G. Zhang, “A Method of Curve Reconstruction Based on Point Cloud Clustering and PCA,” *Symmetry (Basel).*, vol. 14, no. 726, pp. 1–14, 2022, doi: 10.3390/ sym14040726.

[206] J. Wang, K. Wang, R. Jia, and X. Chen, “Research on Load Clustering Based on Singular Value Decomposition and K-means Clustering Algorithm,” in *2020 Asia Energy and Electrical Engineering Symposium (AEEES)*, 2020, pp. 831–835, doi: 10.1109/AEEES48850.2020.9121555.

[207] M. Ali *et al.*, “A K-Sensor correlation-based evolutionary optimization algorithm to cluster contamination events and place sensors in water distribution systems,” *J. Clean. Prod.*, vol. 319, no. April 2020, pp. 1–13, 2021, doi: 10.1016/j.jclepro.2021.128763.

[208] I. Separation *et al.*, “Meta-Heuristic Optimization-Based Two-Stage Considering Intra-Cluster Compactness and Inter-Cluster Separation,” *IEEE Trans. Ind. Appl.*, vol. 56, no. 4, pp. 3375–3384, 2020, doi: 10.1109/TIA.2020.2984410.

[209] P. Zong, J. Jiang, and J. Qin, “Study of High-Dimensional Data Analysis based on Clustering Algorithm,” in *2020 15th International Conference on Computer Science Education (ICCSE)*, 2020, pp. 638–641, doi: 10.1109/ICCSE49874.2020.9201656.

[210] H. Tang and S. Miyamoto, “Sequential Extraction of Clusters for Imbalanced Data,” in *IEEE International Conference on Granular Computing (GrC)*, 2013, pp. 281–285, doi: 10.1109/GrC.2013.6740422.

[211] R. J. G. B. Campello and E. R. Hruschka, “A fuzzy extension of the silhouette width criterion for cluster analysis,” *Fuzzy Sets Syst.*, vol. 157, pp. 2858–2875, 2006, doi: 10.1016/j.fss.2006.07.006.

[212] P. J. Rousseeuw, “Silhouettes : a graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[213] S. Visalaxi and T. Sudalaimuthu, “Endometrium Phase prediction using K-means Clustering through the link of Diagnosis and procedure,” *8th Int. Conf. Signal Process. Integr. Networks*, pp. 1178–1181, 2021, doi: 10.1109/SPIN52536.2021.9566041.

[214] M. Fritz, M. Behringer, D. Tschechlov, and H. Schwarz, “Efficient exploratory clustering analyses in large-scale exploration processes,” *VLDB J.*, 2021, doi: 10.1007/s00778-021-00716-y.

[215] F. Haghighat, “Predicting the trend of indicators related to Covid-19 using the combined MLP-MC model,” *Chaos, Solitons and Fractals*, vol. 152, 2021, doi: 10.1016/j.chaos.2021.111399.

[216] A. Z. Shirazi, M. Hatami, M. Yaghoobi, S. Javad, and S. Chabok, “An Intelligent Approach to Predict Vibration Rate in a Real Gas Turbine,” *Intell. Ind. Syst.*, vol. 2, no. 3, pp. 253–267, 2016, doi: 10.1007/s40903-016-0057-6.

[217] M. E. Shaik, M. M. Islam, and Q. S. Hossain, “A review on neural network techniques for the prediction of road traffic accident severity,” *Asian Transp. Stud.*, vol. 7, no. March, p. 100040, 2021, doi: 10.1016/j.eastsj.2021.100040.

[218] D. Magare, S. Labde, M. Gofane, and V. Vyawahare, “Imputation of missing data in time series by different computation methods in various data set applications,” *ITM Web Conf.*, vol. 32, p. 03010, Jul. 2020, doi: 10.1051/itmconf/20203203010.

[219] S.-W. Tan, S.-W. Huang, Y.-Z. Hsieh, and S.-S. Lin, “The Estimation Life Cycle of Lithium-Ion Battery Based on Deep Learning Network and Genetic Algorithm,” *Energies*, vol. 14, no. 15, p. 4423, Jul. 2021, doi: 10.3390/en14154423.

[220] M. Desai and M. Shah, “An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN),” *Clin. eHealth*, vol. 4, pp. 1–11, 2021, doi: 10.1016/j.ceh.2020.11.002.

[221] S. G. Kumbhar, R. G. Desavale, and N. V. Dharwadkar, “Fault size diagnosis of rolling element bearing using artificial neural network and dimension theory,” *Neural Comput. Appl.*, vol. 33, no. 23, pp. 16079–16093, Dec. 2021, doi: 10.1007/s00521-021-06228-8.

[222] J. Rynkiewicz, “General bound of overfitting for MLP regression models,” *Neurocomputing*, vol. 90, pp. 106–110, Aug. 2012, doi: 10.1016/j.neucom.2011.11.028.

[223] P. M. R. Bento, J. A. N. Pombo, M. R. A. Calado, and S. J. P. S. Mariano, “Stacking Ensemble Methodology Using Deep Learning and ARIMA Models for Short-Term Load Forecasting,” *Energies*, vol. 14, no. 21, p. 7378, Nov. 2021, doi: 10.3390/en14217378.

[224] J. Chen *et al.*, “Powernet: SOI Lateral Power Device Breakdown Prediction With Deep Neural Networks,” *IEEE Access*, vol. 8, pp. 25372–25382, 2020, doi: 10.1109/ACCESS.2020.2970966.

[225] G. Montavon, W. Samek, and K.-R. Müller, “Methods for interpreting and

understanding deep neural networks," *Digit. Signal Process.*, vol. 73, pp. 1–15, Feb. 2018, doi: 10.1016/j.dsp.2017.10.011.

[226] K. Ke, S. Hongbin, Z. Chengkang, and C. Brown, "Short-term electrical load forecasting method based on stacked auto- encoding and GRU neural network," *Evol. Intell.*, vol. 12, pp. 385–394, 2019, doi: 10.1007/s12065-018-00196-0.

[227] T. Lin, M. Wang, M. Yang, and X. Yang, "A Hidden Markov Ensemble Algorithm Design for Time Series Analysis," *Sensors*, vol. 22, no. 2950, pp. 1–13, 2022, doi: 10.3390/s22082950.

[228] M. Benhaddi and J. Ouarzazi, "Multivariate Time Series Forecasting with Dilated Residual Convolutional Neural Networks for Urban Air Quality Prediction," *Arab. J. Sci. Eng.*, vol. 46, pp. 3423–3442, 2021, doi: 10.1007/s13369-020-05109-x.

[229] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput. 9,* vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[230] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation.," *Proc. 2014 Conf. Empir. Methods Nat. Lang. Process.*, pp. 1724–1734, 2014, doi: 10.48550/arXiv.1406.1078.

[231] X. Liu, Z. Lin, and Z. Feng, "Short-term offshore wind speed forecast by seasonal ARIMA - A comparison against GRU and LSTM," *Energy*, vol. 227, p. 120492, 2021, doi: 10.1016/j.energy.2021.120492.

[232] V. Gugnani and R. Kumar, "Analysis of deep learning approaches for air pollution prediction," *Multimed. Tools Appl.*, vol. 81, pp. 6031–6049, 2022, doi: 10.1007/s11042-021-11734-x.

[233] V. Veeramsetty, K. R. Reddy, M. S. Arjun, and M. Gaurav, "Short-term electric power load forecasting using random forest and gated recurrent unit," *Electr. Eng.*, vol. 104, pp. 307–329, 2022, doi: 10.1007/s00202-021-01376-5.

[234] Y. Wang, R. Zou, F. Liu, L. Zhang, and Q. Liu, "A review of wind speed and wind power forecasting with deep neural networks," *Appl. Energy*, vol. 304, no. April, p. 117766, 2021, doi: 10.1016/j.apenergy.2021.117766.

[235] S. Wang, J. Chen, H. Wang, and D. Zhang, "Degradation evaluation of slewing bearing using HMM and improved GRU," *Measurement*, no. June, 2019, doi: 10.1016/j.measurement.2019.06.038.

[236] J. Becerra-rico, M. A. Aceves-fernández, K. Esquivel-escalante, and J. C. Pedraza-ortega, "Airborne particle pollution predictive model using Gated

Recurrent Unit ( GRU ) deep neural networks," *Earth Sci. Informatics*, vol. 13, pp. 821–834, 2020, doi: 10.1007/s12145-020-00462-9.

[237] J. J. Saucedo-Dorantes, A. Y. Jaen-cuellar, M. Delgado-prieto, R. D. J. Romero-troncoso, and R. A. Osornio-rios, "Condition monitoring strategy based on an optimized selection of high-dimensional set of hybrid features to diagnose and detect multiple and combined faults in an induction motor," *Measurement*, vol. 178, no. March, p. 109404, 2021, doi: 10.1016/j.measurement.2021.109404.

[238] L. S. Dhamande and M. B. Chaudhari, "Compound gear-bearing fault feature extraction using statistical features based on time-frequency method," *Measurement*, vol. 125, no. January, pp. 63–77, 2018, doi: 10.1016/j.measurement.2018.04.059.

[239] A. Simões, M. Viegas, T. Farinha, and I. Fonseca, "The State of the Art of Hidden Markov Models," *Qual. Reliab. Eng.*, vol. 33, no. 8, 2017, doi: 10.1002/qre.2130.

[240] A. Kara, "A data-driven approach based on deep neural networks for lithium-ion battery prognostics," *Neural Comput. Appl.*, vol. 33, no. 20, pp. 13525–13538, Oct. 2021, doi: 10.1007/s00521-021-05976-x.

[241] Keras, "Keras Tuner." https://keras.io/keras_tuner/ (accessed Jan. 17, 2022).

[242] J. Yu, "Adaptive hidden Markov model-based online learning framework for bearing faulty detection and performance degradation monitoring," *Mech. Syst. Signal Process.*, vol. 83, pp. 149–162, Jan. 2017, doi: 10.1016/j.ymssp.2016.06.004.

[243] W. Kwedlo, "A clustering method combining differential evolution with the K -means algorithm," *Pattern Recognit. Lett.*, vol. 32, no. 12, pp. 1613–1621, 2011, doi: 10.1016/j.patrec.2011.05.010.

[244] R. J. Kuo, Y. R. Zheng, T. Phuong, and Q. Nguyen, "Metaheuristic-based possibilistic fuzzy k -modes algorithms for categorical data clustering," *Inf. Sci. (Ny).*, vol. 557, pp. 1–15, 2021, doi: 10.1016/j.ins.2020.12.051.

[245] V. D. T. Thuy, T. Do, and P. N. Nguyen, "A Data Mining-Based Approach for Exploiting the Characteristics of University Lecturers," in *Industrial Conference on Data Mining*, 2016, no. July, doi: 10.1007/978-3-319-41561-1_4.

[246] P. Arpaia *et al.*, "Fault detection on fluid machinery using Hidden Markov Models," *Measurement*, vol. 151, p. 107126, Feb. 2020, doi: 10.1016/j.measurement.2019.107126.

[247] N. Gokilavani and B. Bharathi, "Test case prioritization to examine software for

fault detection using PCA extraction and K-means clustering with ranking," *Soft Comput.*, vol. 25, pp. 5163–5172, 2021, doi: 10.1007/s00500-020-05517-z.

[248] A. Gang and W. U. Bajwa, "A linearly convergent algorithm for distributed principal component analysis," *Signal Processing*, vol. 193, p. 108408, 2022, doi: 10.1016/j.sigpro.2021.108408.

[249] N. Fard, H. Xu, and Y. Fang, "A unique solution for principal component analysis-based multi-response optimization problems," *Int J Adv Manuf Technol*, vol. 82, pp. 697–709, 2016, doi: 10.1007/s00170-015-7358-x.

[250] A. Sancho, J. C. Ribeiro, M. S. Reis, and F. G. Martins, "Cluster analysis of crude oils with k-means based on their physicochemical properties," *Comput. Chem. Eng.*, vol. 157, p. 107633, 2022, doi: 10.1016/j.compchemeng.2021.107633.

[251] K. Lakshmi, N. K. Visalakshi, and S. Shanthi, "Data clustering using K -Means based on Crow Search Algorithm," *Sādhanā*, vol. 43, no. 190, pp. 1–12, 2018, doi: 10.1007/s12046-018-0962-3.

[252] S. Reddy, S. Nethra, E. A. Sherer, and A. Amritphale, "Prediction of the number of COVID-19 confirmed cases based on K-means-LSTM," *Array*, vol. 11, p. 100085, 2021, doi: 10.1016/j.array.2021.100085.

[253] Y. Li, X. Chu, D. Tian, J. Feng, and W. Mu, "Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm," *Appl. Soft Comput.*, vol. 113, p. 107924, 2021, doi: 10.1016/j.asoc.2021.107924.

[254] Y. Ni *et al.*, "Faulty feeder detection of single phase-to-ground fault for distribution networks based on improved K-means power angle clustering analysis," *Int. J. Electr. Power Energy Syst.*, vol. 142, p. 108252, 2022, doi: 10.1016/j.ijepes.2022.108252.

[255] Y. Feng, W. Xu, Z. Zhang, and F. Wang, "Continuous Hidden Markov Model Based Spectrum Sensing with Estimated SNR for Cognitive UAV Networks," *Sensors*, vol. 22, no. 2620, pp. 1–23, 2022, doi: 10.3390/s22072620.

[256] S. Park, W. Lim, and M. Sunwoo, "ROBUST LANE-CHANGE RECOGNITION BASED ON AN ADAPTIVE HIDDEN MARKOV MODEL USING MEASUREMENT UNCERTAINTY," *Int. J. Automot. Technol.*, vol. 20, no. 2, pp. 255–263, 2019, doi: 10.1007/s12239-019-0025-8.

[257] C. López, Á. Naranjo, S. Lu, and K. Moore, "Hidden Markov Model based Stochastic Resonance and its Application to Bearing Fault Diagnosis," *J. Sound Vib.*, vol. 528, p. 116890, 2022, doi: 10.1016/j.jsv.2022.116890.

[258] Z. Shang, Y. Zhang, X. Zhang, Y. Zhao, Z. Cao, and X. Wang, “Time Series Anomaly Detection for KPIs Based on Correlation Analysis and HMM,” *Appl. Sci.*, vol. 11, no. 11353, pp. 1–23, 2021, doi: 10.3390/app112311353.

[259] A. Jandera and T. Skovranek, “Customer Behaviour Hidden Markov Model,” *Mathematics*, vol. 10, no. 1230, pp. 1–10, 2022, doi: 10.3390/math10081230.

[260] L. Sun, Y. Li, H. Du, P. Liang, and F. Nian, “Fault Diagnosis Method of Low Noise Amplifier Based on Support Vector Machine and Hidden Markov Model,” *J. Electron. Test.*, vol. 37, pp. 215–223, 2021, doi: 10.1007/s10836-021-05938-0.

[261] W. Yang and L. Chen, “Machine condition recognition via hidden semi-Markov model,” *Comput. Ind. Eng.*, vol. 158, p. 107430, 2021, doi: 10.1016/j.cie.2021.107430.

[262] Z. Lee, W. Ean, A. Najah, A. Marlinda, and A. Malek, “A systematic literature review of deep learning neural network for time series air quality forecasting,” *Environ. Sci. Pollut. Res.*, vol. 29, pp. 4958–4990, 2022, doi: 10.1007/s11356-021-17442-1.

[263] Y. Chung, J., Gulcehre, C., Cho, K., & Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” *NIPS 2014 Deep Learn. Represent. Learn. Work.*, 2014, doi: 10.48550/arXiv.1412.3555.

[264] N. Gruber and A. Jockisch, “Are GRU Cells More Specific and LSTM Cells More Sensitive in Motive Classification of Text?,” *Front. Artif. Intell.*, vol. 3, no. June, pp. 1–6, 2020, doi: 10.3389/frai.2020.00040.

[265] J. Guo, L. Wang, I. Fukuda, and K. Ikago, “Data-driven modeling of general damping systems by K-means clustering and two-stage regression,” *Mech. Syst. Signal Process.*, vol. 167, p. 108572, 2022, doi: 10.1016/j.ymssp.2021.108572.

[266] M. Hoffmann, “PEANO — a tool for on-line calibration monitoring PEANO - A Tool for On-Line Calibration Monitoring,” *OECD Halden React. Proj. - Inst. Energy Technol.*, no. January 2004, 2014, [Online]. Available: https://www.researchgate.net/publication/228577605_PEANO-a_tool_for_on-line_calibration_monitoring.

[267] J. Coble, P. Ramuhalli, R. Meyer, H. Hashemian, B. Shumaker, and D. Cummins, “Calibration Monitoring for Sensor Calibration Interval Extension: Identifying technical gaps,” in *Future of Instrumentation International Workshop (FIIW) Proceedings*, 2012, pp. 1–4, doi: 10.1109/FIIW.2012.6378348.

[268] G. Rajesh and A. Chaturvedi, “Correlation analysis and statistical characterization of heterogeneous sensor data in environmental sensor

networks," *Comput. Networks*, vol. 164, no. 106902, pp. 1–16, 2019, doi: 10.1016/j.comnet.2019.106902.

[269] N. Berjab, H. H. Le, and H. Yokota, "A Spatiotemporal and Multivariate Attribute Correlation Extraction Scheme for Detecting Abnormal Nodes in WSN," *IEEE Access*, vol. 9, pp. 135266–135284, 2021, doi: 10.1109/ACCESS.2021.3115819.

[270] N. Kleber, C. R. Dietlein, and J. D. Chisum, "Cooperative Cross-Correlation Algorithm to Optimize Linearity of Fused RF Sensors," *IEEE Sens. J.*, vol. 20, no. 7, pp. 3766–3776, 2020, doi: 10.1109/JSEN.2019.2959255.

[271] G. C. Karmakar, R. Das, and J. Kamruzzaman, "IoT Sensor Numerical Data Trust Model Using Temporal Correlation," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2573–2581, 2020, doi: 10.1109/JIOT.2019.2957201.

[272] A. Alqahtani, M. Ali, and X. Xie, "Deep Time-Series Clustering : A Review," *Electronics*, vol. 10, no. 3001, pp. 1–29, 2021, doi: 10.3390/electronics10233001.

[273] P. Biswas and T. Samanta, "A Method for Fault Detection in Wireless Sensor Network Based on Pearson ' s Correlation Coefficient and Support," *Wirel. Pers. Commun.*, vol. 123, no. 3, pp. 2649–2664, 2022, doi: 10.1007/s11277-021-09257-7.

[274] Y. Lai *et al.*, "Identifying malicious nodes in wireless sensor networks based on correlation detection," *Comput. Secur.*, vol. 113, pp. 1–12, 2022, doi: 10.1016/j.cose.2021.102540.

[275] W. Lu, R. Wen, J. Teng, X. Li, and C. Li, "Data correlation analysis for optimal sensor placement using a bond energy algorithm," *Measurement*, vol. 91, pp. 509–518, 2016, doi: 10.1016/j.measurement.2016.05.089.

[276] L. Dash *et al.*, "A Data Aggregation Approach Exploiting Spatial and Temporal Correlation among Sensor Data in Wireless Sensor Networks," *Electronics*, vol. 11, no. 989, pp. 1–17, 2022, doi: 10.3390/ electronics11070989.

[277] X. Zhang, S. Wang, W. Li, and X. Lu, "Heterogeneous sensors-based feature optimisation and deep learning for tool wear prediction," *Int. J. Adv. Manuf. Technol.*, vol. 114, pp. 2651–2675, 2021, doi: 10.1007/s00170-021-07021-6.

[278] B. Kendall, "A NEW MEASURE OF RANK CORRELATION," *Biometrika*, vol. 30, no. 1–2, pp. 81–93, 1938, doi: 10.1093/biomet/30.1-2.81.

[279] S. Awasthi, C. Travieso-González, G. Sanyal, and D. Singh, *Artificial Intelligence for a Sustainable Industry 4.0*. Switzerland: Springer, 2021.

[280] K. Azar, Z. Hajiakhondi-meybodi, and F. Naderkhani, “Semi-supervised clustering-based method for fault diagnosis and prognosis : A case study,” *Reliab. Eng. Syst. Saf.*, vol. 222, no. February, pp. 1–16, 2022, doi: 10.1016/j.ress.2022.108405.

[281] H. Rafiee, M. Aminizadeh, E. M. Hosseini, H. Aghasafari, and A. Mohammadi, “A Cluster Analysis on the Energy Use Indicators and Carbon Footprint of Irrigated Wheat Cropping Systems,” *Sustainability*, vol. 14, no. 4014, pp. 1–19, 2022, doi: 10.3390/su14074014.

[282] L. Li, G. Liu, L. Zhang, and Q. Li, “Sensor fault detection with generalized likelihood ratio and correlation coef fi cient for bridge SHM,” *J. Sound Vib.*, vol. 442, pp. 445–458, 2019, doi: 10.1016/j.jsv.2018.10.062.

[283] X. Shen, X. Lin, and K. Zhang, *Encyclopedia of Wireless Networks*. Springer, 2020.

[284] S. Stapathy, V. Bhateja, M. Favorskaya, and T. Adilakshmi, “Smart Innovation, Systems and Technologies,” in *Proceedings of the Fourth International Conference on Smart Computing and Informatics*, 2021, vol. 1, doi: 10.1007/978-981-16-0878-0.

[285] A. Yousefi-darani, M. Babor, O. Paquet-durand, and B. Hitzmann, “Model-based calibration of a gas sensor array for on-line monitoring of ethanol concentration in Saccharomyces cerevisiae batch cultivation,” *Biosyst. Eng.*, vol. 198, pp. 198–209, 2020, doi: 10.1016/j.biosystemseng.2020.08.004.

[286] Y. Zhang, L. O. H. Wijeratne, S. Talebi, and D. Lary, “Machine Learning for Light Sensor Calibration,” *Sensors*, vol. 21, no. 6259, pp. 1–17, 2021, doi: 10.3390/s21186259.

[287] M. C. Thrun and A. Ultsch, “Using Projection-Based Clustering to Find Distance- and Density-Based Clusters in High-Dimensional Data,” *J. Classif.*, vol. 38, pp. 280–312, 2021, doi: 10.1007/s00357-020-09373-2.

[288] D. Abdullah, S. Susilo, A. Ahmar, R. Rusli, and R. Hidayat, “The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data,” *Qual. Quant.*, pp. 1–9, 2021, doi: 10.1007/s11135-021-01176-w.

[289] X. Zhao, F. Nie, R. Wang, and X. Li, “Neurocomputing Improving projected fuzzy K-means clustering via robust learning,” *Neurocomputing*, vol. 491, pp. 34–43, 2022, doi: 10.1016/j.neucom.2022.03.043.

[290] A. A. Boechat, U. F. Moreno, and D. Haramura, “On-line calibration monitoring system based on data-driven model for oil well sensors,” in *Proceedings of the 2012 IFAC Workshop on Automatic Control in Offshore Oil and Gas Production*, 2012, pp. 1–6, doi: 10.3182/20120531-2-NO-4020.00037.

# Appendices

# Appendix A

# Calibration and certification of industrial sensors – a global review

[1,2]Alexandre Batista Martins; [3,4]José Torres Farinha; [2]António Marques Cardoso
[1]Industrial Eng. and Management, Univ. Lusofona. Lisbon, Portugal
[2]CISE – Electromechatronic Systems Research Centre, Univ. Beira Interior, Portugal
[3]CEMMPRE - Centre for Mechanical Engineering, Materials and Processes, Univ. of Coimbra, Portugal
[4]ISEC/IPC - Polytechnic Institute of Coimbra, Portugal

*Abstract:* - The main objective of this paper is to describe the state of the art of the calibration and certification of industrial sensors, giving an approach on how companies can manage the sensors' calibration, namely the ones that are incorporated in the equipment.
For any industry it is essential to make products that satisfy the customers in terms of quality and, simultaneously, to be competitive in the market. The data obtained from the sensors placed in the equipment is one of the sources that helps to increase the performance, adding value for the competitiveness of the company in the market.
Sensors are responsible, in many sectors, to guarantee equipment availability and product quality. In this way, the certification of the company and of the equipment, not only guarantee quality and prevent unwanted and unforeseen costs, but also gives company credibility, namely from the customer's point of view.
This paper focus on the state of art of industrial metrology, namely sensors, standards and measuring tools, calibrations and the certification of calibration. It also includes a theoretical section about sensors, types of sensors, their operation and characteristics. Next, it is presented the theme of metrology and the measurement science, responsible for ensuring the quality and veracity of the data sent by the reading equipment. It also addresses the importance of metrological traceability and certified management systems, which are required to certify the sensors calibration, in order to guarantee the effectiveness of the measurements and, consequently, the rigor of operation of the associated quality system.

*Key-Words:* Metrology; Sensors; Traceability; Calibration; Certification; Condition Monitoring

Received: February 8, 2020. Revised: August 25, 2020. Accepted: September 12, 2020. Published: September 17, 2020.

## 1 Introduction

When we talk about a company's quality image, we know that it is influenced by many factors. Its aptitude and competitiveness in the market are strongly indexed to the quality and reliability of the data used. We can also talk about the importance of this data in the maintenance policy implemented, and in the service and quality of maintenance performance. Through high-quality input elements it is possible to establish a reliable production process, ensuring high availability and stable product quality. [1]

When we talk about the use of sensors, we are talking about the reliability of the data obtained through them. This is where metrology, the science of measurement, comes in, and it is through it and the existing traceability pyramid that the company can guarantee the effectiveness of the sensors. Through an interrupted set of calibrations, from the pyramid top patterns of the industrial sensor, it is then possible to guarantee the reliability of the equipment.

A calibration corresponds to an issue of a calibration certificate and the placement of a label. Based on this information, a user can decide if the instrument is suitable for the application in question.

It is inferred that, without metrology and sensor calibration, the measurements performed are not reliable, and may lead to unwanted and harmful results in many situations. This means that without the support of metrology based on sensing, there would be measurement errors and unreliable data, which could result in uncontrollable risks and costs and, consequently, the loss of market confidence.

So, with the increase of the competitiveness among companies, namely based on the increasing of robotic and automation, and the increasing of customer demand, the certification of management systems has become an essential tool for their survival in the market. A certified company guarantees a set of established and recognized procedures, through which the conformity of products or systems is determined. Also, the calibration certification of the sensors is extremely important, because it guarantees the quality of the data. According to [2], the EN ISO 9001 standard involves the assessment of the company in several areas, one of which is the measurement instruments and its quality management, namely about inspection and testing.

Some requirements of this standard are as follows:

- selection of the appropriate equipment for the measurements to be made;
- equipment calibration at regular intervals, using certified standards;
- use of documented procedures;
- guarantee that the equipment has the required accuracy;
- certainty that the equipment indicates the calibration status, and calibration certificates are maintained;
- environmental conditions, storage, handling and safety are appropriate in order to maintain the validity of the calibrations.

This paper is structured in three sections: the first section is the introduction, where the topic under study is described; section two describes the state of the art and covers all topics related to the study, starting with the "industrial sensors", where an approach to this theme is made explaining their importance in the industry and, mainly, in the industrial maintenance area. In this way, it is described: the sensor's operation; its functions; the types of existing sensors; their performance characteristics. Then, in order to be able to explain how to maintain the reliability of the sensors for their perfect functioning, it is referred the importance of the "metrology". Here, the relation between this subject with the data reliability provided by the sensors is emphasized, which can, later, be used in the equipment conditioning maintenance. The metrology concept is described, as well as its areas of activity, wherein this research is focused on and the technical area where it is related. Additionally, it must be taken into account the traceability chain, which goes from the primary standard to the working standard, supported by consecutive calibrations. Therefore, it they are referred the concepts of: metrological traceability; standards; and calibration. Then, an approach to the type of calibration strategy that can be addressed by companies is made. They are referred some relevant concepts as follows: Preventive Calibration; and Predictive Calibration and On-line Monitoring (OLM), being this discussed in big detail. Finally, the calibration certificates is discussed: only through certified measuring instruments is possible to guarantee the reliability of the measurement results and to find the metrological traceability. Therefore, this chapter addresses this issue explaining the certification procedure, as well as the entities responsible for issuing the certificates (Calibration laboratories accredited with 17025: 2017). Finally, in section three, a conclusion of the study is made explaining the illation drawn from this paper.

# 2 State of the Art

## 2.1 Industrial Sensors

Sensors are one of the largest data sources used in Industry. Probably one of the largest sources in condition monitoring maintenance, since this type of maintenance is based on the equipment condition. It is necessary to make constant measurements to evaluate the health equipment. The sensors are responsible for continuously interpret the equipment condition.

According to [3], sensory technologies have grown very quickly in several areas such as: science; product design; electronics; photonics; mechanics; chemistry and biology. Sensors are used in the individual's daily life, as well as in companies, and are responsible for detecting audible, movement, optical or magnetic signals. The ability to have many small devices that transmit in real-time data physically distributed close to the detected objects brings new opportunities to observe and act in the world, which could bring significant benefits to humanity. [3]

So, sensors are responsible for making direct connection between a physical phenomenon and the data acquisition system, converting signals of physical quantities into electrical signals. They can convert the data received from a physical phenomenon, transforming it in order to be able to read it (Figure 1).

Then, leaving away the various nuances of domains and applications, a sensor simply measures something of interest and provides an output that can be useful. [4]

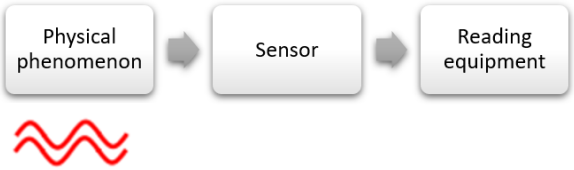



Fig. 1 - Basic diagram of the operation of a transducer

With the advancement of technology, namely computing, widespread communications, connectivity to the Web, smart mobile devices and integration in the cloud, the number of sensors increased. This evolution is illustrated in Figure 2.

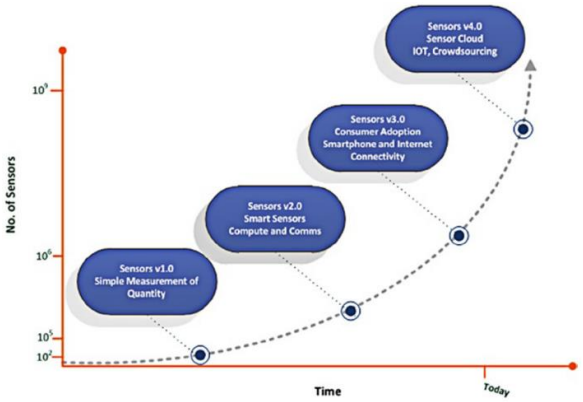



Fig. 2 - Evolution of sensors [Source: [4]]

The sensors convert a physical phenomenon into a measurable electrical signal. Some sensors do not respond naturally to changes in physical phenomena and it is necessary to recourse to a signal condition. Before digitizing the sensor output, the signal may need other components and circuits to produce a signal that can take advantage of all features of the measurement hardware and reduce the effects of noise from external interferences [5]. [6] adds that the sensors do not operate alone. They need a larger system consisting of signal conditioners and several analogue or digital signal processing circuits. This system can be a measurement system, data acquisition system or process control system, for example. The same author also says that sensors and their associated circuits are used to measure various physical properties, such as: temperature; force; pressure; flow; position; light intensity; etc. These are responsible for stimulating the sensor and the output of this is conditioned and processed in order to provide the corresponding measurement of the physical property.

[7] says that sensors can be used alone or together in order to monitor a specific situation in the same way as a human being, such as:

- manufacturing operations;
- conditioning of tools;
- inventory control;
- working in progress;
- identification of parts, tools, pallets, etc.

Nowadays, sensing in a factory can be considered more efficient, because it implies less supervision, namely in unmanned manufacturing. As the complexity of manufacturing processes increases, it becomes necessary to acquire additional types of information obtained through sensors. [8]

Process sensors are used to control the manufacturing process and their measurements allow a better understanding of the process, generating improvements. The connection between profit and process measurement is illustrated in Figure 3.

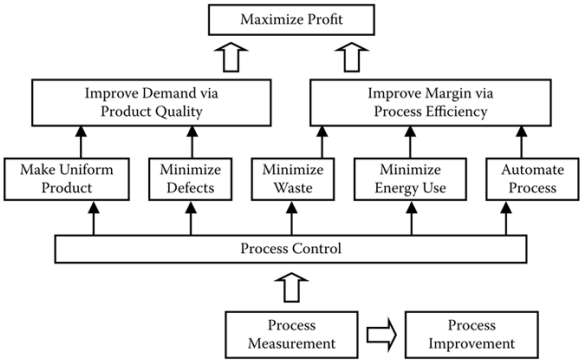



Fig. 3 - Crucial measurement process to operation and profitability [Source: [8]]

### 2.1.1 Type of sensors

According to [9], "there are many types of sensors, according to each type of condition variable. The reasons to choose a sensor are diverse, as are the type of output signal, the range of output values, the environmental conditions, the physical dimensions, and so on. Sensor signals may be digital or analog."

This author also states that the sensor signals can be digital or analog. Digital sensors can have an interface to communicate with other devices. In turn, analog sensors no longer have this capability, which is why an analog-to-digital converter is needed to enable communication with other devices.

There are several different types of sensors, with different types of characteristics, being each of them designed to measure the variables of interest. We can characterize the sensors as active (with contact) or passive (without contact).

According to [6], [4] and [10], active sensors need to be connected with an external source of excitation. They generate an electric current when the external physical environment changes. It requires physical contact with the factor to be measured, disturbing its state.

The same authors explain that passive sensors generate their own electrical output signal without the need for external voltages or currents. They change their resistive, capacitive or inductive characteristics along with their physical parameters. An external power source is required to induce an electrical output. It does not require direct contact with the variable to be measured, aiming not cause a disturbance.

In Figure 4 can be seen some examples of sensors and their characteristics.

| PROPERTY | SENSOR | ACTIVE/PASSIVE | OUTPUT |
|---|---|---|---|
| Temperature | Thermocouple | Passive | Voltage |
| | Silicon | Active | Voltage/Current |
| | RTD | Active | Resistance |
| | Thermistor | Active | Resistance |
| Force/Pressure | Strain Gage | Active | Resistance |
| | Piezoelectric | Passive | Voltage |
| Acceleration | Accelerometer | Active | Capacitance |
| Position | LVDT | Active | AC Voltage |
| Light Intensity | Photodiode | Passive | Current |

Fig. 4 - Typical output sensors [Source: [6]]

Currently, there are transducers capable of taking measurements of practically all existing physical quantities (Figure 5). For example, for the temperatures measurement, there are thermocouples, thermoresistors, thermistors and the semiconductor junction, which convert the temperature of the medium with which they are in contact to a proportional analogue signal; for flow measurement, there are, among others, turbine flow meters, which generate a rectangular wave signal whose frequency depends on the flow speed; load cells are available for measuring voltages; for pressure measurement, there are several types of pressure transducers; etc. [11], [5], [10].

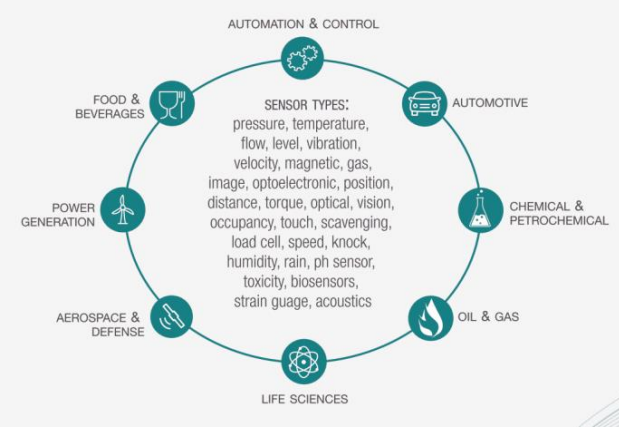



Fig. 5 - Different types of sensors and their markets [Source: [10]]

[4] characterize the different types of sensors as follows:

- *Mechanical Sensors*
  These sensors measure changes in a device or material as a result of an input that causes its mechanical deformation. Movement, speed, acceleration, and displacement are the changes that result in mechanical deformations and that can be measured. The sensor is described as electromechanical, when the inputs are converted directly to an electrical output
- *MEMS Sensors*
  MEMS (Micro Electro Mechanical Systems) are miniaturized, three-dimensional mechanical and electrical structure sensors, usually ranging from 1 to 100 mm. They are manufactured using standard semiconductor manufacturing techniques. Its constitution consists of mechanical microstructures, microsensors, microactuators, and microelectronics, which are integrated in the same silicon chip. The use of this type of sensors, with multiple axes of low energy consumption, ultra-compact and compact, allowed a high growth of electronic devices, such as: smartphones, tablets, game console controllers, portable game devices, digital cameras, etc. In the health field, these sensors are used as blood pressure monitors, pacemakers, ventilators, and respirators. The most used MEMS sensors are accelerometers and gyroscopes.
- *Optical Sensors*
  These sensors are responsible to detect waves or photons of light. It's including light in the visible, infrared, and ultraviolet (UV) spectral regions. They can measure changes in light intensity related to the emission or absorption of light by an amount of interest. They are widely used, among others, in automated doors and gates to ensure that there are no obstacles on the way, and in industrial applications, where they measure levels of liquids and materials in tanks or on factory production lines, in order to detect the existence of objects.
- *Semiconductor Sensors*
  These are low cost sensors, with high reliability, low energy consumption, long service life, and small dimensions. They are used in several applications, such as: Gas monitoring; Pollution monitoring; Breath analysers, for measurements of alcohol content in exhaled air; Monitoring of domestic gas, such as propane; Temperature, as in electronic equipment; Magnetism, for example, magnetometers for applications with six degrees of freedom; Optical sensor, such as detectors for devices attached to load on cameras.
- *Electrochemical Sensors*
  These sensors are responsible for measuring an electrical parameter of the sample of interest. They consist of sensors or electrodes, a reference electrode and, in many cases, a counter electrode. These electrodes are, usually, placed in contact with

a liquid or solid electrolyte. In the low-range-temperature (<140° C), electrochemical sensors are used to monitor pH, conductivity, dissolved ions, and dissolved gases. For measurements at high temperatures (> 500° C), such as the measurement of exhaust gases and molten metals, solid electrolyte sensors are used. They are categorized from the measurement performed. They have low energy consumption, high sensitivity, good precision and resistance to the effects of surface poisoning. In contrast, their sensitivity, selectivity, and stability are greatly influenced by environmental conditions, especially temperature.

- *Biosensors*
  These sensors detect an analytical of interest in chemical, environmental (e.g.: air, soil and water) and biological samples (e.g.: blood, saliva and urine) and use biochemical mechanisms for this. It uses an immobilized biological material, which can be an enzyme, antibody, nucleic acid or hormone, in an independent device. The biological material used in the biosensor device is immobilized in order to maintain its bioactivity.

### 2.1.2 Sensor performance characteristics

The sensors have several characteristics that define them, and through them; it is possible to choose the best sensor to use in each case. There are several sensors that can measure temperature, pressure, acceleration, voltage, acoustic, etc.; so, their manufacturers tend to focus on basic performance parameters. According to [4], "to truly understand sensors, and how sensors that measure the same measurand can differ, it is necessary to understand sensor performance characteristics". Therefore, here it will be introduced each of their most important characteristics.

According to According to [6], [10], [4], [12] and International Metrological Vocabulary (VIM in Portuguese), [13], it is presented some of the most important characteristics of the sensors:

- *Sensitivity*
  Represents the relationship between the physical input signal and the electrical output signal. It is the ratio between a small change in the electrical signal and a small change in the physical signal. It is the change in input required to generate a change in the unit at the output. It can be expressed as the derivative of the transfer function in relation to the physical signal. Not all sensors have a linear output. If the sensor response is linear, the sensitivity will be constant over the sensor range. A good sensor has to have a lot of sensitivity, able to better detect the signals of the measurand, and a constant sensitivity that is maintained throughout the dynamic range. If the sensor response is non-linear, the sensitivity will vary within the sensor range. As an example, we can consider an accelerometer as a "high sensitivity" sensor, if it can detect a small change in vibration what would result in a large change in voltage.
- *Range or Dynamic Range*
  It represents the range of physical input signals that can be converted into electrical signals by the sensor. It describes the total range that the sensor can measure from the physical input parameters, such as light intensity, sound level or temperature, and converts them into readable electrical parameters. That is, it is the maximum capacity that the sensor can measure without distortion of the generated signal. The dynamic range can be indicated by the sensor supplier. There we can see the range over which other performance characteristics described in the data sheets are expected to apply. Typical units are: g; Kelvin; Pascal; Newton; etc.
- *Accuracy or uncertainty*
  Uncertainty is generally defined as the largest expected error between the actual and ideal output signals. According to VIM, measurement uncertainty is the non-negative parameter that characterizes the dispersion of the values attributed to a measurand, based on the information used.
  Accuracy, defined by VIM, represents the degree of agreement between indications or measured values, obtained by repeated measurements, on the same or similar objects, under specified conditions. [4] explain that it is generally considered by metrologists as a qualitative term, while "uncertainty" is quantitative. For example, one sensor may have better accuracy than another if its uncertainty is 1% when compared to the other with an uncertainty of 3%.
- *Linearity*
  Linearity represents how linear the output of the sensor values is within its dynamic range. That is, the ability of the sensor to maintain its sensitivity regardless of the signal generated. Linearity is usually specified in

terms of percentage, i.e., the percentage change in sensitivity between one value of its amplitude and another. Non-linearity can generally be a result of the conditions of the environment where the measurement is being made, such as: environmental changes, temperature, vibration, acoustic noise level, and humidity.

- *Repeatability*
  This characteristic represents the ability of a sensor to produce the same output value for the same applied input as the measurand. VIM characterizes fidelity or measurement accuracy under a set of repeatability conditions. Sensors with good repeatability will always reproduce the same values for the same measurement under the same conditions. The lack of repeatability can occur due to random fluctuations in environmental inputs or errors made by the operator. [13]
- *Noise*
  Noise is produced by the sensors that go along with the output signal. All electronic components of the measurement system produce some noise. Noise is generally distributed across the frequency spectrum. Too much noise produced by the sensor will impair the measurements made by it.
- *Resolution*
  According to VIM, the resolution is the smallest variation in the measured quantity that causes a noticeable variation in the corresponding indication, [13]. The resolution is defined as the minimum oscillation of the signal. Once these oscillations are temporal phenomena, there is some relationship between the fluctuation time scale and the minimum detectable amplitude. Depending on the type of measurement to be made, we can choose a sensor with a higher or lower resolution. Knowing that, if it is chosen a sensor whose resolution is too high for the application, it is unnecessarily expensive.
- *Bandwidth*
  It is the frequency range between the upper and lower cut off frequencies. That is, it is a frequency range where the sensor can read the physical signal produced. The sensors have finite response times to an instant change in the physical signal. Some sensors are represented as have decay times. This is the time after a step-change in the physical signal for the sensor output to decay to its original value. The reciprocal of these times corresponds to the upper and lower cut off frequencies, respectively.
- *Response time*
  It is the ability of the sensor to respond to the measurement, reacting to signal changings. It is sometimes referred to as a sensor time constant when it is subject to a step change. VIM says that the response time is "time interval between the moment when an input value of a measuring instrument or a measuring system is subjected to a sharp variation between two specified constant values and the instant when the corresponding indication is keeping within specified limits around its final value in a stable regime" [13].
  It represents the period of time required for the sensor to change its output after a change in the input value.
- *Reach*
  This characteristic represents the ability of the sensor to detect the minimum and maximum value of the input or output of the signal of the measurand. Then, there are the minimum and maximum input voltages that can be used to operate the sensor. If an input voltage is applied outside this range, the sensor may be permanently damaged.
- *Transfer function*
  A term used to refer exclusively to time-invariant linear systems. In the case of sensors, there is a relationship between the measurand and the electrical output signal. This relationship, if it is time-invariant, is called the sensor transfer function. There are sensors that are calibrated, obtaining a certified calibration curve, which guarantees this relationship between the input and output values.
  A mathematical formula (Equation 1) that describes the transfer function is usually expressed as follows:

$$S = F(X) \tag{1}$$

  Where:
  $X$ - is the measurand;
  $S$ - is the electrical signal produced by the sensor.
  It is rare to find a transfer function that can be completely described by a single formula; therefore, functional approximations of the actual transfer function are used.
- *Error*

Error is the difference between the measured value and the true value, where the true value is a reference to an absolute or agreed standard. Quoting VIM, error is the difference between the measured value of a quantity and a reference value. There are two forms of error: Systematic error - which reproduces inaccuracies that can be corrected with compensation methods, such as feedback, filtering and calibration; that is, a component of the measurement error that, in repeated measurements, remains constant or varies in a predictable way and random error, also called noise, what corresponds to a signal component that does not carry information.

- *Random error*
  Component of measurement error that, in repeated measurements, varies in an unpredictable way.

## 2.2 Metrology

This section deals with the concept of metrology and its relation with the data reliability provided by the sensors that are used in Condition Based Maintenance. According to [14], modern metrology introduces key factors such as downtime and product quality.

[15] refers that, in the field of applied sciences, measurements are not accurate, as they are always subject to errors due to various causes, human and materials. The qualification of an error to later quantify an uncertainty proves that the validity of the measurement result is doubtful. Therefore, evaluating measurement uncertainties generating errors is a very complex task.

According to the website of the Portuguese Metrology Society (SPM)[1], in the beginning, as today, measuring was taken as a unit. It is reasonable to admit that the first greatness concerned by Man was the time.

With the existence of well-defined metrics, it is possible to have coherence. For these metrics existing, there must be measuring instruments that prove the measurements are always the same, at any time and in any part of the world.

Therefore, the objective of Metrology is to guarantee the effectiveness and accuracy of the measurements made by these measuring instruments. [5]

According to the website of the Portuguese Quality Institute (IPQ)[2], "metrology, as a measurement science, provides reliable material support to the measurement system, essential in the sectors of the economy, health, safety, and the environment, constituting an essential technological infrastructure in modern societies".

As time goes on, it becomes imperative for all societies to have references for units of measurement, in order to be able to maintain consistency in everything that is carried out on a daily basis; that is, with the existence of metrics, it is necessary to acquire standards that will become references in those same metrics.

In practice, everything in people's daily lives and, mainly, in the Engineering field, exists a metric, so it is essential to have good use of metrology in order to guarantee universal uniformity.

### 2.2.1 Concepts

Metrology is the science of measurement, whose main objective is to ensure that measurement equipment reproduces right values, internationally reproducible with similar quantities, [5]. [2] characterizes metrology as the science of measurement that comprises all aspects, both theoretical and practical, related to measurement, whatever its uncertainty and the domain of science and technology to which they refer, relating the processes, instruments, location, metrologist, etc., to each other.

The main objective of metrologists is trying to find ways to measure various physical quantities in the most accurate way possible. Atom sensors, such as atom interferometers and atomic clocks, are examples of a class of instruments that can measure these quantities with very high precision. [16]

In practical terms, metrology comprises a diverse set of aspects (for example, procedures, uncertainties, errors, standards, etc.) with the purpose of determining the value of magnitude, encompassing all theoretical and practical aspects of measurement, whatever the measurement uncertainty and the field of application. It comprises all aspects, both theoretical and practical, related to measurement, whatever their uncertainty and the domain of science and technology to which they refer. [17], [18], [19].

[15] explains that uncertainty reflects the way how a quantity is measured, and the confidence given to the result. The use of instruments in measurement involves calibrations and manipulations, requiring proper procedures and calculations.

[1] http://www.spmet.pt/historia.html (Access date: 03/04/2020)

[2] http://www1.ipq.pt/pt/metrologia/apresentacao/Pages/Metrologia.aspx (Access date: 03/04/2020

According to [20] and [12], the task of metrology and metrologists is to achieve reliable measurement results that are necessary to ensure:

- the quality and efficiency of production in the industry;
- the justice in trade;
- the consumer protection;
- the health and safety of life;
- the protection of the human and animal environment.

To frame the various sets of terms related to metrology, the International Metrology Vocabulary (VIM)[3] was used as a reference, which appears in the context of world metrology and seeks the international harmonization of terminologies and definitions used in the fields of metrology and instrumentation.

### 2.2.2 Practical Areas

[20], [17], [18], [21], [22], [2], [12] explain that metrology can be divided into three fundamental areas of activity, as is shown in Figure 6.

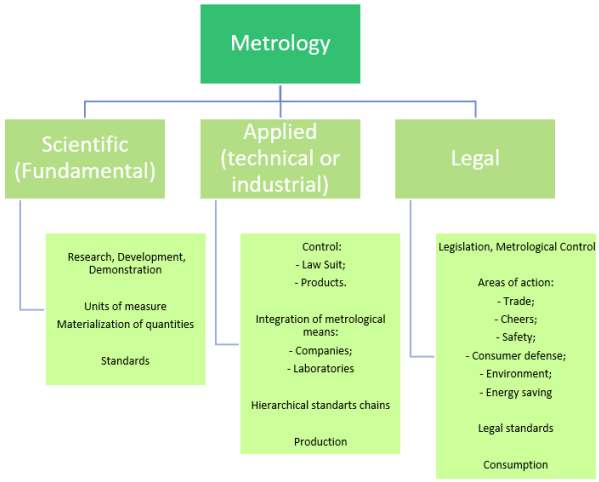



Fig. 6 - Metrology categories [Source: [18]]

In this paper, we only discuss the technical area, because it is essentially related to the productive activity (the theme under study). It exercises a certain control over processes and products that require an integration of metrology at the level of companies and laboratories; it takes into account a hierarchical chain of standards, existing in laboratories or companies/organizations, which are also traceable by primary standards that can be national or international. [23], [20], [24]

### 2.2.3 Traceability

With the objective to have international metric coherence, there must be a traceability chain. That is an uninterrupted set of comparisons that ensure the result of a measurement, or the value of a standard. Thus, it is guaranteed that the values of the standard, even being of a lower level, will agree with the higher ones.

According to VIM, the traceability is the property of a measurement result whereby that result can be related to a reference through an uninterrupted and well-documented chain of calibrations, each contributing to evaluating the measurement uncertainty.

In the context of Metrology (metrological traceability), the traceability is represented by a pyramidal shape, as can be seen in Figure 7.

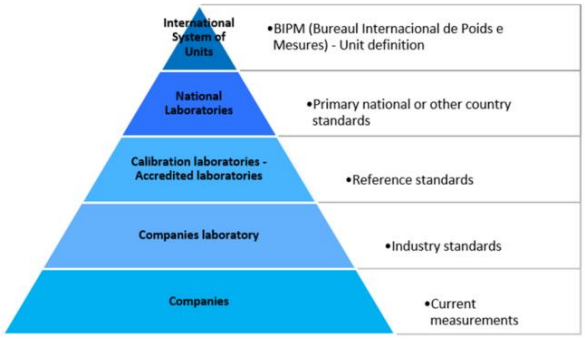



Fig. 7 - Pyramidal traceability [Source: [5]]

[25] argue that quality reliability is closely linked to the traceability chain and the availability of measurement standards. It also says that, in order to establish an appropriate traceability chain, measurement standards are required, traced back to the meter.

Traceability requires orderly and permanent records and is complemented with a succinct document related to metrology to translate the examination of each event of the procedure and what it represents. Thus, it is possible, for the user, to know the history of a process or instrument, [15]. This author also states that, through traceability, it is possible to know changes in equipment, facilitating the management of several aspects, such as:

- Varied use and proper adjustment of equipment in the workplace;
- Selection of equipment among others offered by different suppliers;
- Detection of greater or lesser accuracy (based on records).

[3] http://www1.ipq.pt/pt/metrologia/documents/vim_ipq_inmetro_2012.pdf (Access date: 07/04/2020)

At the manufacturing level, [26] show that ensuring the results of traceability of measurement during the production process is a guarantee of quality improvement. [17] says that the traceability of measurement must be guaranteed, allowing the necessary conclusions to be drawn about its metrological quality. Likewise, it is important to carry out the most relevant actions to ensure the correct indication of the measuring instruments, namely their regular calibration (industrial metrology) and the periodic verification of the measuring instruments according to legal regulations (legal metrology).
Traceability can be required by any customer in a contract, or by standards such as ISO 9000-9004 and ISO 14253. This requirement is made in order to guarantee the quality of the measurement and to protect the buyer. Consequently, any tool or equipment used in production must guarantee its measurement track. [21]

### 2.2.4 Standards

Calibration is used in order to keep the measurement and control instruments within the specified limits, and, for that purpose, a standard is used to calibrate the equipment, [27], [28]. It is necessary to have standards to be able to make a comparison between this and the equipment to be calibrated.
According to [25], "The reliability of quality assurance is closely connected with the traceability chain and the availability of measurement standards." The units of physical quantities are reproducible, with the help of reference standards and measuring devices; so, these measuring devices play an exceptionally important role in the measurement unit. [22]
According to VIM, standards are defined as an achievement of the definition of a given quantity, with a determined value and associated measurement uncertainty, used as a reference.
There is a hierarchical structure of reference standards (Figure 8) and their calibrations (Figure 9): International Standards; Primary Standards; Secondary Standards; and Work Standards. It is through this structure and consecutive calibrations that it is possible to guarantee measurements from the primary to the working standard. The standards that rank higher in the hierarchy have higher quality and accuracy in measurements. That is why higher-level standards are used to calibrate lower-level standards.

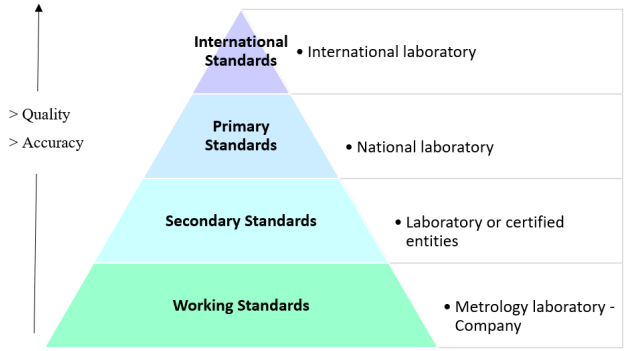



Fig. 8 - Measurement Standards Hierarchy

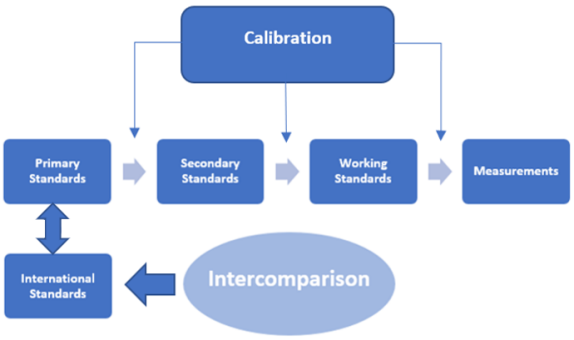



Fig. 9 - Standard hierarchy and calibration

Measurement standards traced back to the meter are used to establish a reliable traceability chain. Accredited calibration laboratories are responsible for providing calibration of work standards. This process involves a loss of precision at each stage and, in order to guarantee the production of the part within the tolerances requested at the end of the chain, the National Metrology Institutes need to calibrate as accurately as possible. [25]

## 2.3 Formal sensors calibration

In order to guarantee the reliability of the sensor data used in the industry, and to guarantee quality processes, it is necessary to carry out their calibrations.
According to [29], "The veracity of big data requires careful design of data-acquisition and calibration strategies and of feature-extraction and selection strategies, so that decision makers have clean, valid, and reliable inputs to use in making decisions."

### 2.3.1 Calibration

Calibration of reading instruments is a very important part of the asset management strategy; companies invest a lot of money in the acquisition of this type of equipment. For this reason, uncalibrated instruments, providing wrong values, can imply the companies

spent a lot of money, namely because production downtime, safety problems, production of lower quality products, etc. [30]

Juran & Godfrey [31] says that, by ensuring the reading instruments are calibrated, the company is able to confirm that the product complies with the customer's requirements. If the instrument is in error, the company can reject good parts or accept bad parts, which will imply high costs for the company, sending a message to the customer that the company does not have a basic control system.

According to Eren [30], calibration has the following benefits:

- Determines whether measurements taken before calibration were valid;
- Gives confidence in future measurements;
- Ensures consistency and compatibility with those made elsewhere;
- It leads to repeatability and reproducibility of processes;
- Provides confidence that products meet specifications;
- Increases efficiency, ensuring that measurements are correct;
- Leads to the documentation of instrument and process performance to meet quality standards;
- Frequent calibrations can provide a graphical view of equipment uncertainty over time and lead to performance reliability;
- Measurements made within international standards promote global acceptance, thereby increasing competitiveness;
- Helps the validity of compliance of measurements and processes under varying conditions.

So, [12] characterizes calibration as being one of the metrology services of measurement instruments, which is a basic tool to ensure the traceability of measurement.

Calibration is a process of comparing an "unknown" element with an equivalent or better standard. It can be included in an adjustment to correct the deviation from the value obtained from the standard, which is represented by the standard deviation, [15]. It involves the determination of the metrological characteristics of an instrument achieved through a direct comparison with standards. [32]

According to Morris & Langari [33], the calibration is a comparison between the sensor output to be calibrated and the standard sensor output in relation to the same input value (measured quantity) applied to the two sensors. It also says that the calibration is carried out over a range of inputs in order to safeguard the entire sensor measurement range.

According to Eren [30], calibration provides consistency in readings and reduces errors, validating measurements universally. It also confirms that the calibration serves to reveal the individuality of an instrument, comparing it with a reference standard.

According to Eren [34], calibration is a set of actions and procedures that aim to clarify the relationship between the values of the indicated quantities of the measuring instruments under specified conditions, assigning values to the instrument's response to the reference standards.

Through the instrument calibration, we can obtain its calibration curve. [15] explains that the calibration curve is specific for each device and converts the gross measurement into the corrected measurement.

Statistical techniques are applied to calculate the calibration curves, readings averages and standard deviations; the data collection process for the creation of the calibration curve is fundamental for the success of the calibration program.

It is obtained by subjecting the instrument to a real value of the quantity to be measured, which is made through a standard device and, accurately, reads the gross measurement provided.

According to Morris & Langari [33], the calibration must be carried out in periodic intervals of time, because the characteristics of the sensors change throughout their useful life. According to the same authors, there are several reasons that cause a change in the sensors' characteristics, such as: mechanical wear and effects of dirt, dust, smoke, chemicals and temperature changes in the operating environment; these changes in characteristics will also depend on a lot about the use of the sensor, that is, on the time that it is under the operating environment. Although the causes of these changes, its storage also can cause some deviations, due to the aging of its components. [33], [31]

Then, the calibration quantifies the change in measurements, and will ensure that the reading instruments comply with the expected specifications.

Through a good instruments calibration management, the values provided by them can be used in the optimization of processes, asset control, equipment maintenance and plant safety, leading to a better management of the plant's performance. [30]

According to Juran & Godfrey [31], it is necessary to have calibration schedules, being these done according to class of equipment, varying in order to be able to maintain the accuracy, nature and extent of the measuring instruments. The same author also explains that the schedules at the beginning are established by judgments, being changed according to the results of the checks, allowing to establish greater economic efficiency.

Morris and R. Langari [33] say that there should also always be calibration records in order to maintain efficiency and effectiveness of the measuring instruments maintenance system. Complete documentation provides a description of the measurement conditions, instruments used, calibration systems and operated procedures. There should also be calibration records for each individual instrument. To begin with this process, there must be a declaration of the measurement limits defined for each documented measurement system. These limits are established by balancing the costs of the improved precision with the customer's requirements and also in relation to the general quality level specified in the quality manual.

The existence of a calibration record is an important part of knowing the data quality. These records can be saved in software or paper format. Data such are as follows: position ID, device ID, location, serial number and work order number, that can be transferred from one form to another suitable for engineers or managers to evaluate. The data must also contain: maximum errors, pass/fail notifications, calibration date and time, calibration frequency, who performed the calibration task, and so on. [30]

These documented calibration processes are a integrating part of the quality control system audit, and the calibration systems support this quality control system. [33]

So, the timing with which it is necessary to calibrate the sensors will depend on a great extent on their use and the environmental conditions with which they work. Having all these factors involved, it becomes difficult to define the vital regularity of sensors' recalibration, based on theoretical considerations; therefore, practical experimentation to define the time periods between calibrations is important. By defining the maximum allowed measurement error, knowing the rate at which the sensor's characteristics vary, it is possible to calculate the time interval until the sensor reaches the limits of its performance. The level of measurement error that an instrument reaches shortly before recalibration is the error limit that must be quoted in the instrument's documented specifications. [33]

The instrument's calibration history records will be the main basis on which this review is made.

In relation to the management of calibration procedures, this must be done by only one person; so, this function is done efficiently and effectively. A technician trained in metrology is the determining factor in the accuracy of the measuring instrument, [31]. A good calibration management will transmit security to the customer, ensuring the accuracy of the measurements made and, in turn, guaranteeing the quality of the product. The calibration procedures related in some way to the measurements used for quality control functions are controlled by the international standard ISO 9000. [33]

Then, it is necessary to organize correctly the calibration of the instrument to be successful. Thus, there are some factors to consider [30]:

1. The type of calibration process to be employed;
2. Calibration setup and the calibration process;
3. Calibration cycles;
4. Calibration records;
5. Calibration reports and report maintenance;
6. Factors that affect calibration;
7. Use of spreadsheets;
8. Mathematical approaches;
9. Internal versus calibration outsourced;
10. Reference standards to be used;
11. Hierarchy of benchmarks;
12. Traceability;
13. ISO 9000 requirements;
14. Work patterns;
15. Measurement uncertainties;
16. Uncertainties versus errors;
17. Random and systematic errors.

Speaking now of the instruments used for calibration, the standard sensors used, usually have greater

precision than the sensors to be calibrated. Morris & Langari [33] tell that high-precision sensors are commonly used for calibrations.

The measuring instruments used for calibration at the company level, made by the instrumentation department, are called work standards. As they are instruments with relatively less use than process measurements, it can be concluded that they will keep their characteristics unchanged for much longer, although they will also always need to be calibrated, mainly due to the effects of aging. After a calibration program has been implemented for these instruments, they will be calibrated by secondary reference standards, which have greater precision. These calibrations are carried out by accredited calibration laboratories, which, at the end of the operation issue, gives a calibration certificate (mentioned in the preceding section), [33]. Tse & Morse [32] state that an instrument is "traceable" if the calibration standards can be traced back to the National Bureau of Standards.

Juran & Godfrey [31] summarize that the reference standards are used exclusively under the control of standards laboratories, composed by technicians, whose main interest is to maintain the accuracy of the calibration. Work standards are in the hands of production, inspection and test personnel, whose main interest is the control of the product and the process.

The standards used for calibration purposes should only be maintained for that purpose and should not be seen as substitute equipment for process measurements. Sensor replacements used for process measurements must be made by other sensors that have the same purpose. [33]

A calibration is usually carried out under ideal conditions; that is, charging and noise due to strange inputs are controlled, [32]. The same authors also say that calibration is normally carried out under ideal conditions because it is not possible, for a standards laboratory, to calibrate instruments for every imaginable service condition. Morris & Langari [33] add that the environmental control for the calibration of the measuring instruments must be taken care of, although sometimes it is not possible to achieve it and, even, sometimes, to be convenient to carry out the calibrations in the process plant. In these cases, the necessary corrections must be made for deviation from the calibration environmental conditions in relation to those specified.

Regarding the cost of calibration, it will depend on many factors, such as: the equipment to be calibrated; who calibrate them (internally or outsourcing); the number of instruments to be calibrated; where they will be calibrated (laboratory or plant); and the frequency of calibration, which is a major cost factor. [30]

Therefore, there are companies that can decide to acquire an accredited calibration laboratory and, the required quality only becomes economically viable for large companies that have a large number of measuring instruments in different factories. Typically, the most commonly used by small and medium-sized companies is to outsource the calibration service provided by several companies specializing in this area. [33]

In organizations, calibrations are conducted on site or in remote laboratories or sites. Large organizations may have several laboratories (called calibration laboratories) dedicated to different instruments and processes.

Laboratories are accredited by authorities in accordance to the guidelines of ISO 17025. Accreditation is the formal recognition that a particular laboratory is competent to conduct specific tests and / or calibrations. [30]

A successful laboratory calibration procedure, according to [30], requires the following basic steps:

- Selection of an appropriate benchmark with well-known values covering the range of interest;
- Conduct calibration curves (i.e., least squares adjustment) to establish the relationship between the measured and known values of the reference standard;
- Correction of measurements using calibration curves;
- Preparation of appropriate documentation of the calibration procedure, results, analysis and interpretation of the results for the customer.

Finally, a calibration originates a calibration certificate and the placement of a label. Based on this information, a user can decide whether the instrument is suitable for the application in question.

According to Morris & Langari [33], the calibration certificate must contain the following information:

- identification of calibrated equipment;
- calibration results obtained;
- measurement uncertainty;
- any limitations on the use of the calibrated equipment;
- calibration date;
- authority under which the certificate is issued.

Calibration labels, applied after the calibration has been carried out, must comply with the general requirements of ISO / IEC 17025 for the competence of testing and calibration laboratories. ISO 17025 requires the following issues [30]:

- all measuring equipment must be labelled in a safe and durable manner;
- the labels must clearly indicate the name of the calibration laboratory, calibration date, expiration date, equivalent of use and the authorized employee;
- the information on the label must be legible and durable under reasonable conditions of use and storage;
- when it is impractical to affix a label directly to an item, the label can be affixed to the instrument's container;
- temperature resistance seals can be used when needed;
- functional labels must contain reference standards.

Summing up, according to [27] and [28], the objective of a calibration system is to avoid the unreliability of the tool through the immediate detection of deficiencies. Every organization should prepare a written description of its calibration system, which should accompany the measurement of test equipment and standards, including:

- to establish realistic calibration intervals;
- to list all measurement standards;
- to establish environmental conditions for calibration;
- to ensure the use of calibration procedures for all equipment and standards;
- to coordinate the calibration system with all users;
- to ensure that the equipment is frequently checked by a periodic system or cross-checks in order to detect damage, inoperative instruments, erratic readings, and other performance degradation factors that cannot be predicted or predicted by calibration intervals;
- to provide timely and positive corrective actions;
- to establish decals, rejection tags, and records to label the calibration;
- to maintain formal records to ensure adequate controls.

The main reasons for using calibrated instruments, according to [12], are:

- to ensure that an instrument's readings are consistent with other measurements;
- to determine the accuracy of the instrument's readings;
- to establish the viability of the instrument.

Through calibration it is possible [19]:

- to assign the measured values to the indications;
- to determine corrections related to the indications;
- to determine other metrological properties, such as the effect of the influence quantities.

In short, the calibration is responsible for comparing a measuring instrument with a standard in order to determine the measurement error. Knowing this measurement error, it is possible to determine the exact value of the metric to be measured.

#### 2.3.1.1 Maintenance of calibration

According to [35], “Data veracity is about the certainty of data meaning. This feature expresses whether data reflect properly the reality or not. It depends on the way in which data is collected. It is strongly linked to the credibility of sources. For example, the veracity of the data collected from sensors depends on the calibration of sensors.”

According to [35], “Data veracity is about the certainty of data meaning. This feature expresses whether data reflect properly the reality or not. It depends on the way in which data is collected. It is strongly linked to the credibility of sources. For example, the veracity of the data collected from sensors depends on the calibration of sensors.”

[36] says that the viability of the data provided is achieved through calibration. The same author also explains that there must be a validity of the calibration for long periods of time or, else, the integration of calibration verification procedures in relation to a reference standard as an integral element of the process. Although, the sensors are pre-calibrated, some manipulations can affect their measurements. [37]

Through frequent calibration, it is possible to combat instrument error. Therefore, calibration should be performed periodically related to the type of

instrument and its stability characteristics, by using standards that are properly maintained and traceable. The records of each instrument, tests used, patterns of traced tests used, and any adjustments made, must be kept. Due to the high price of traceable standards, it may be cheaper to use specialized instrument calibration services. [36]

The calibration frequency is based on the appropriate need, such as deviation in time expressed in the service history, according to the manufacturer's specifications and regulations. [15]

With the increase of the use of sensors in the industrial processes, we also can find a vast measurement field for the equipment used. For this reason, it becomes necessary to acquire calibration management strategies. As explains [14], a complete calibration, performed on large or complex machines can take several days. Therefore, it is necessary to define the best calibration strategy for the sensors, depending on their situation. Figure 10 shows, based on time, a type of preventive calibration done at regular periods. These calibration cycles must not exceed the time necessary for the sensor to exceed its tolerance. This type of strategy is not applicable for large or complex machines, due to the time needed for its calibration, since it is done more frequently and in regular time.

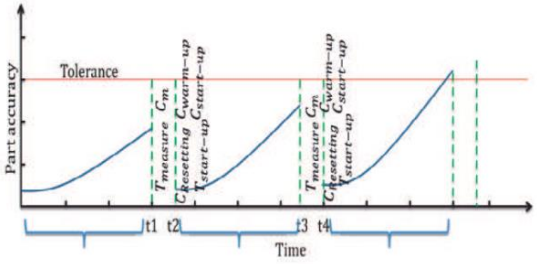



Fig. 10 - Preventive Calibration and adjustment at fixed rate [Source: [14]]

According to [27] and [28], the verification interval can be performed in terms of time (hourly, weekly, monthly), based on the amount of use (every 5,000 parts) or, in each batch. It must be based on stability, purpose, and level of use. Equipment that does not have specific calibration intervals, should be examined, at least, every six months and calibrated at intervals of, at least, one year. Sensor calibrations must be performed at a minimum of 95% of equipment with standards of the same type within tolerance when subject to regularly scheduled recalibration.

The Predictive Calibration strategy (PdC), when implemented correctly it can reduce unnecessary downtime for calibration and maintain the accuracy of the sensor (Figure 11). In this type of approach, periodic checks must be carried out, along with the application of the necessary technical knowledge, management strategies, and decision-making skills. A historical sensor verification database is built regularly, using relatively non-invasive methods, thus allowing calibration to be scheduled through a better-informed process. It is important to establish secondary KPIs associated with the measurement method; so, they indicate the poor performance of the sensor's accuracy. It is also necessary to define appropriate tolerances for the checks carried out, thus allowing to stimulate more interventions. [14]

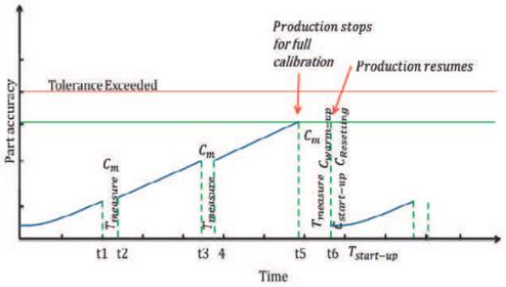



Fig. 11 - Predictive Calibration [Source: [14]]

The introduction of this strategy will bring new costs to the company, and, for this reason, it is important that these be compensated with the optimization of operational efficiency and the reduction of the total cost of downtime, related to unexpected and unplanned quality problems of the sensors. It will also reduce the overall downtime of a manufacturing facility, eliminating the sensor as the root cause of any subsequent failure in the production process. [14]

Another important calibration concept is the system calibration. The sensors can be calibrated individually, with separate calibrations, or the entire instrument chain can be calibrated, from the sensor input to the reading equipment. This will ensure greater accuracy, avoiding the accumulation of errors when multiplying the calibration of each component, taking into account any possible interaction among the components, such as low impedance correspondence, or excessive noise capture due to potential signal transmission arrangements and differentials or other factors. [36]

The interaction of the instrument and its installation can generate false readings in several ways; so, permanent instrument installations should always allow simple removal for calibration or built-in installations to allow calibration and reinstallation, without the risk of disturbing any factors that may affect readings. Economic and performance aspects need to be considered to decide the support for

checking machines and specific positional error handling policies. [14]

Finally, [27] and [28] argue that a calibration record system must be maintained in all instruments, including the following data:

- Usage history;
- Precision;
- Current location;
- Calibration interval and when due;
- Necessary calibration procedures and controls;
- Current values since the last calibration;
- Maintenance and repair history.

The same authors also say that test equipment and measurement standards must be labelled to indicate the date of the last calibration, by whom it was calibrated, and when the next calibration should take place. In this way, there is a visual indication of the maintenance status of the calibration. Both the calibration laboratory and the user of the instrument must maintain a bi-directional verification of the calibration.

### 2.3.1.2 On-line sensors calibration

As we saw above, the strategy used for maintenance is traditionally based on sensors' validation, involving their calibrations periodically, causing the reading instrument to be turned off and removed to be calibrated, causing, in some cases, very high costs. Many companies, aiming to maintain competitiveness in the market, adopt different calibration maintenance strategies, based on condition monitoring, allowing to minimize the downtime, increasing the processing availability. Condition monitoring allows early detection of failures, as an important issue in surveillance and diagnosis. This type of monitoring is called On-line Monitoring (OLM) and, according to [38], it can be implemented for instrument monitoring, equipment monitoring, or operation monitoring; however, the acronym OLM is commonly used for the extension of sensor calibration intervals.

The online monitoring systems, designed to monitor and diagnose sensors with online measurements during their measurements, were developed by researchers at the Nuclear Power Plants (NPP), providing generic online monitoring approval to reduce the calibration of the process instrumentation. In this way, it becomes possible to extend the calibration periods of the sensors, being able to improve the economic activity of the NPPs, avoiding failures that could occur during the calibration process. . [39], [40], [41], [38]

According to [42], OLM consists of estimating correct measurements that sensors must have read and constantly monitors the difference between the estimated values and the values read by the sensors.

[43] refers that "On-line monitoring evaluates the deviation of an instrument with reference to its process parameter estimate as determined by one of the predictive algorithms."

OLM is a technique non-invasive approach that is responsible to provides a more frequent assessment of instrument calibration in the real operating environment. It has the potential to mitigate problems with current calibration practices, allowing the identification of sensors that have deviations from the limits of tolerance to direct calibration activities during interruptions, [44].

[40], [45] and [43] explain that a typical OLM system collects data from reading instruments and processes them on an offline computer, which, subsequently, evaluate the individual deviation sensor channel, of the process channel estimate, as a time function. The concept online is used because the data collection is done during the operation, but it does not necessarily mean that the monitoring is carried out in real-time.

OLM evaluates the deviation of a sensor with reference to the estimation of process parameters, as determined by one of the predictive algorithms, and, in order to evaluate the sensor performance, the residual is used between the process estimation of the OLM model and the sensor output, [45]; in this way, it is possible to define the operational states of the sensor, (Figure 12).

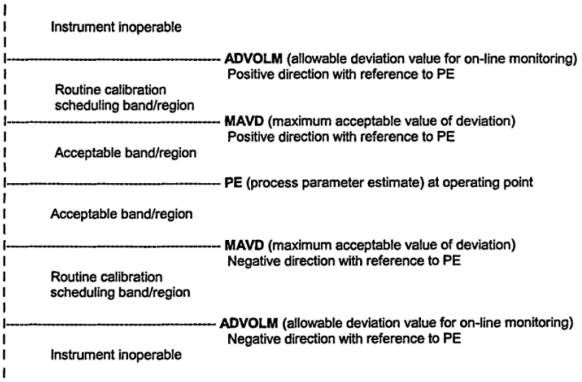



Fig. 12 - Deviation zones of sensor performance [Source: [45]]

If someone is not 95% confident that the instrument is within the MAVD (Maximum Acceptable Value of Deviation), it must be reset. If someone is not 95% confident that the instrument is within ADVOLM (Allowed Deviation Value for On-Line Monitoring), the instrument must be declared inoperative. [43]

In order to explain a basic monitoring system, we can use Figure 13 as an example, where we have a vector of sensor measurements ($x$) that is inserted into a

forecast model calculating the best estimates of the sensors (*x'*). Then, these estimates are compared to the measured values, allowing to calculate the differences called residuals (*r*). Subsequently, a decision logic module determines whether the residues are statistically different from zero, establishing the performance of each sensor. This module can also use predictive uncertainty values and deviation limits to determine the instrument's channel condition. [38], [45].

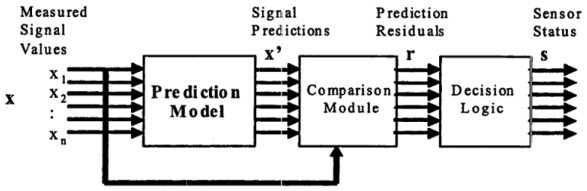



Fig. 13 - Sensor Calibration monitoring system diagram [Source: [45]]

The difference between expected (modelled) and actual (measured) behaviour, called residual, characterizes the system's deviations from normal behaviour and can be used to determine whether the sensor or system is operating in an abnormal state. [46], [47]

[38] state that OLM allows to reduce the number of unnecessary calibrations and, consequently, the reduction in the interruption time. It can also provide benefits with faster problem discovery and also allowing for more timely and convenient corrective actions than traditional calibration.

Periodic sensor calibration is expensive, time-consuming approach, and unnecessary maintenance actions that can damage the sensors. For example, when a sensor requires calibration, but it can be ignored, simply because the calibration interval has not yet passed, even if the sensor needs maintenance, this can cause unexpected stops and safety risks. Another example, and a major concern with periodic calibrations, is that performing maintenance on components that are working correctly, may cause a failure, [47], [46], [41] and [38]. OLM systems, on the other hand, monitor the condition of the sensor channel, identifying those that have been degraded to the point that justify their calibration. So, according to [38], online methods can help reduce maintenance costs, reduce the potential for calibration errors, increase instrument reliability, and, consequently, reduce equipment downtime.

[48] say that "An essential component for quality control is a low cost recalibration system which can be applied on-line during the production cycles (…)."

Schiff (2002) says that the justification for replacing traditional monitoring by online monitoring and calibration is:

- the proposed online monitoring technique should perform all necessary designated functions, better than or as good as the current traditional calibration, with the same or better reliability;
- if due to deficiencies inherent in the proposed technique, it cannot be demonstrated that it is better than or at least as good as current practice, the licensee must verify that the impact of the proposed technique on the safety of the facilities will be insignificant and the advantages of using it overcome deficiencies.

An OLM system consists of several components, the most common of which are: an offline computer, where a monitoring system exists; communication hardware and software tools for collecting process data; a history of process data; an OLM software, responsible for performing the analysis and presenting the results of the sensor calibration performance, [45]. The author explains that the first step in implementing online monitoring is the installation, testing, and verification of the data acquisition system, responsible for acquiring and storing the historical data files. In Figure 14, we can check the relative position of the typical data acquisition system in relation to the sensor. The data acquisition system generally receives data from the instrument in the form of a voltage output, whose online monitoring system resizes to the expected process units.

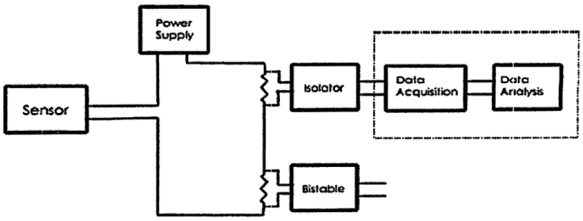



Fig. 14 - Instrument channel with On-line monitoring [Source: [45]]

The data transfer between the data acquisition system and OLM software can occur in batch mode or almost in real-time. The term batch mode means that the data files are stored somewhere and accessed by the online monitoring system at discrete time intervals, [45]. In Figure 15, we can verify the operation of a generic OLM system.

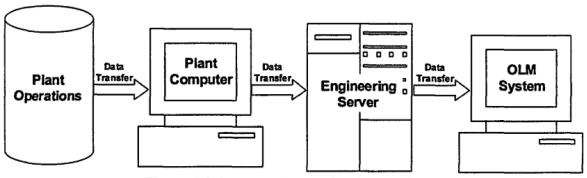



Fig. 15 - On-line monitoring system setup [Source: [45]]

So, in OLM, the sensor calibration is evaluated by comparing the measured data with the expected value of the sensor. The expected value can be measured using a variety of models, including physics-based models, neural networks, non-parametric models, etc. [44].

[45] and [47] refer that there are several empirical modelling techniques for online monitoring of the instrument channels performance, saying that these can be divided into two main categories: redundant; and non-redundant. [45] tell us that redundant modelling techniques only use the measurements of a group of redundant instrument channels to obtain the parameters' estimate, such as the simple average. Non-redundant modelling techniques, on the other hand, use a group of correlated, but not truly redundant, instrument channels to estimate the parameters.

For redundant equipment, a comparison of the instrument readings is made in order to distinguish between process deviation and instrument deviation. Redundant modelling techniques, according to [38], are techniques that use only the measurements of a group of redundant instrument channels, in order to obtain the estimate of the parameters. The author explains that the term "redundant" describes the instrument channels that measure the same process parameter in a similar operational range. Due to problems and disadvantages of redundant models, non-redundant models become the most used technique for OLM.

For non-redundant instruments, we have the empirical process' modelling. This estimate is updated frequently and compared with the output of the corresponding instruments to detect any deviation in the instrument output. This process modelling can also be used with redundant instruments, [40].

There are several modelling methods proposed to evaluate the sensor performance of redundant and non-redundant sensor groups. Figure 16 shows the summary of a selection of the most common modelling methods.

| | Redundant vs. Nonredundant | Empirical vs. First Principles | Auto-Associative vs. Inferential | Parametric vs. Nonparametric | Key References |
|---|---|---|---|---|---|
| Multivariate State Estimation Technique (MSET) | Nonredundant | Empirical | Auto-associative | Nonparametric | Singer et al. (1995); Gross et al. (1997); Singer et al. (1997); Herzog et al. (1998) |
| Neural Networks (e.g., AANN, PEANO, RBF, etc.) | Nonredundant | Empirical | Both | Parametric | Eryurek and Turkcan (1991); Black et al. (1996); Park et al. (1996); Hines et al. (1996); Wrest et al. (1996); Tsai and Chou (1996); Hines et al. (1997); Nabeshima et al. (2002); Ayaz et al. (2003); Fantoni et al. (2003); Fantoni (2005) |
| Auto-Associative Kernel Regression (AAKR) | Nonredundant | Empirical | Auto-associative | Nonparametric | Garvey et al. (2006); Shumaker and Hashemian (2006); Garvey et al. (2007) |
| Cross Calibration | Redundant | | | | Hashemian (1990) Hashemian and Petersen (1991) Hashemian (2007); Hashemian (2011) |
| Analytical Redundancy | Nonredundant | First principle-based | Inferential | Parametric | Crowe (1996); EPRI (2008b); IAEA (2008a) |
| Sensor Averaging (e.g., ICMP, ESEE, etc.) | Redundant | | | | EPRI (1993a, b, 2008b); James (1996); IAEA (2008a) |

AANN – Auto-Associative Neural Network
PEANO – Plant Evaluation and Analysis by Neural Operators
RBF – Radial Basis Function
ICMP – Instrument Calibration and Monitoring Program
ESEE – Expert State Estimation Engine

Fig. 16 - Modelling methods for OLM [Source: ([44]]

According to [45], for a modelling technique to be considered suitable for OLM, the model must:

- produce accurate results;
- produce repeatable and robust results;
- have a method for estimating forecast uncertainty.

Below, according to [45], a list of the basic steps for the development and implementation of the model is presented:

- *acquire "good" data*
  In this first stage, it must be ensured that the collected data is carefully reviewed, and its quality is guaranteed.
- *group sensors in ideal models*
  Here, the sensors to be used in each model must be selected, and this choice is more or less difficult depending on whether sensors are used for redundant or non-redundant techniques.
- *select training data*
  In this step the data is divided into training, verification, and validation data sets; the training data being used by the model to learn the relationship between the sensors, and the verification data is used to optimize the model parameters in order to reduce predictive uncertainty. The validation data is used to quantify the model's performance measures.
- *build and optimize predictive models*
  In this step, the models are optimized so, they minimize the predictive uncertainty, ensuring that they have a complexity corresponding to the complexity of the relationships to be modelled.

- *evaluate the model*
  Here, the model should be evaluated using criteria-based validation data. Here, the following must be taken into account: accuracy, which is a measure of how well the model's outputs correspond to the sensor data; robustness, which is a measure of how well a sensor forecast tracks the actual plant parameter when that sensor is drifting; overflow, which is a measure of how a drifting sensor input affects the prediction of other sensor values; and predictive uncertainty of the model.
- *uncertainty analysis*
  This step is performed after the model has been developed and optimized, and its uncertainty needs to be quantified.
- *transition to online mode*
  After the seven steps above, the OLM system can be implemented in an online or batch monitoring mode. At this stage, it may be necessary to retrain models with more up-to-date training data when changes in operational or environmental states are found.

A perfect model, as explained by [38], would be a model that make predictions of the sensor, not being significantly affected by degraded inputs and would be able to detect small faults and anomalies in the sensor.

But, as with any modelling paradigm, the predictions made have some level of associated uncertainty. Understanding and quantifying this uncertainty is an essential need to the development of an OLM system for monitoring sensor performance. [47]

The sources of uncertainty, according to [47] and [38], OLM can be categorized as:

- *process noise*
  which is the result of the normal fluctuation of the physical parameters of the process (for example, temperature, flow, pressure) over the true value of the process;
- *measurement uncertainty*
  which occurs due to several factors, including sensor accuracy, calibration accuracy (for example, calibration offset, error in converting sensor units to engineering units), environmental effects (due to temperature, vibration, pressure, etc.). These sources of uncertainty apply to the sensor;
- *electronic noise*
  where the transmission of measurements, through the instrumentation line, can induce additional noise, together with the conversion from analog to digital at the computer input;
- *modelling uncertainty*
  which results from input uncertainty (related to process noise, measurement uncertainty, and electronic noise, as described in the preceding bullet) and modelling error (resulting from model selection, model training, input selection, etc.).

The uncertainty inherent to the model's predictions affects the size of the failure that can be safely detected. [38]

In short, we have the Nuclear Regulatory Commission (NRC) conclusions for on online monitoring systems, [40]:

- the generic concept of an online monitoring technique is acceptable for online tracking of instrument performance;
- online monitoring has several advantages, including timely detection of degraded instrumentation;
- online monitoring can provide information about the direction in which the instrument's performance is going and, in this role, can be useful in preventive maintenance activities;
- although the proposed online monitoring technique, compared to the traditional calibration process, makes the results less accurate, it is considered acceptable that the precision provided by the estimation of process parameters is enough to assess the instrument's operability;
- compared to traditional calibration, due to lack of fuel, the online monitoring technique, when used as a whole, offers a greater guarantee of the instrument's operability throughout the plant's operating cycle.

### 2.3.2 Calibration Certificate

Only certified measuring instruments guarantee the reliability of the measurement results; this is only possible as long as the metrological parameters of the measuring instrument (working standards) are guaranteed by periodic calibrations, [2]. The calibration certificate records allow how the measurement traceability can be found.

According to the *Sociedade Geral de Superintendência* (SGS)[4], companies are responsible for ensuring that the equipment meets a wide range

[4] https://www.sgs.pt/pt-pt/construction/services-related-to-machinery-and-equipment/equipment-certification-and-calibration (Access date: 30/04/2020)

of international and regional regulations. Complete equipment certification and calibration services can help to ensure that products meet all relevant regulations. Calibration determines the performance of the measuring equipment. With routine calibration and adjustment of the equipment, it is possible to measure safely, ensure compliance, and avoid the costs of inaccurate measurements. It is necessary to obtain and maintain compliance through consistent procedures and comply with industry regulations and standards. The calibration is then responsible for establishing the relationship between the value of the quantity produced and the applied value and, through this process, its results are documented through a Certificate of Calibration (CC).

A current CC is, basically, a historical document, which should only be considered as a baseline control. The certificate will not guarantee that the meter is accurate at the time of measurement. This is just a statement at the moment, under certain conditions, a deviation between the indications of the device and a reference standard, [36], [15]. Therefore, before making potentially expensive decisions based on reading a transducer or instrument, it is recommended to obtain a calibration check before and after the critical measurement. Through this way it is possible to avoid breaking a machine in good condition based on a sudden drop in performance, but, conversely, it will support the decision to react if circumstances demand it. [36]

Starting with the acquisition of reading equipment, when purchasing them, there are factory certificates that serve as documentary proof of the sensor's accuracy. Industrial sensors, as they are equipment that can be subjected to very violent treatment, resulting in a significant change in their characteristics and, sometimes, in permanent damage, it is necessary that these be periodically calibrated, and that there is a calibration certificate that confirms the reliability of the measurements.

[49] refers that, "According to generally accepted practice, the calibration of measuring instruments may be carried out by an accredited calibration laboratory issuing a calibration certificate, which should contain information about the results of measurements with the associated uncertainty."

The purpose of calibration certification, according to [50], is to calibrate and reset the certified value of the reading equipment, using traceability standards for this purpose. These work standards must be traceable by a qualified and certified laboratory (having a secondary reference standard tracked by a national laboratory that has a primary reference standard), through validated and internationally recognized calibration procedures that can certify the readings. This chain of traceability ensures that the certified value of the work standard is not more than three levels removed from the primary standards of the national laboratory. The CC is the end product of a test/calibration laboratory. It is delivered to the customer in the form of a document, which leaves the laboratory as a "product". [51]

According to the website of the company Cachapuz[5], the conformity of equipment must be maintained consistently through procedures and methods that comply with technical regulations and international standards, such as: ISO 9001, ISO/IEC 17025, EURAMET, OIML, and ILAC. Calibration Certificates are a way that allows to control costs, and ensure the conformity and quality of equipment measurements.

The Calibration Certificates allow to know the main metrological characteristics of the equipment showing:

- periodic inspection of measuring equipment;
- consistency of the results;
- quality assurance of the measurements made on calibrated equipment.

It is important for companies that calibrate the sensors, that the equipment used by them is certified in the purchase and also every 12 months, not exceeding 18 months from the initial purchase or use. Certified equipment must have certification stickers, indicating the date of the last calibration and who performed the calibration and certification. The certification sticker must be clearly visible and made of resistant material. [50]

The same author also states that a new certificate is issued each time the equipment is calibrated, in order to guarantee the user's accuracy and traceability. Discarded or damaged reference equipment must be discontinued until it is verified, calibrated, or recertified.

[6] warns that the calibrations are not all the same and that in some calibration reports there may be terms such as, "nominal" or "typical", or even lack of traceability or approval of accredited stamps. This will result in a decrease in attention on the part of the manufacturer to meet a specific tolerance in these specifications, reducing scrap, since less measured specifications mean less rejection. This, although it seems beneficial for the manufacturer, who obtains

[5] https://www.cachapuz.com/servicos/certificados-de-calibracao
(Access date: 30/04/2020)

an additional profit, in fact it is not, since it will decrease the quality of its product and, consequently, the loss of confidence of the customer. Customers, in turn, also need to look beyond glossy paper and attractive graphics to ensure the integrity of the actual measured data contained in each manufacturer's calibration certificate.

Regarding the points that the CC for measuring instruments must contain: the measurement results, including measurement uncertainty and/or a declaration of conformity with an identified metrological specification, like ISO 17025: 2017 standard (General requirements for the competence of testing and calibration laboratories). Therefore, it is necessary to have an indication of the conformity of the calibrated measuring instrument with the established metrological requirements or separate metrological characteristics for the CC. It is necessary to have an assessment of the conformity about the object with the specified requirements, and the main requirement of these documents, being necessary to consider the measurement uncertainty during the conformity assessment. ([52]
According to [15] and [52], the CC of an instrument provides the deviation, being the uncertainty in that deviation called the calibration uncertainty. The sensor that user must consider in the measurement uncertainty calculation parameters includes:

- uncertainty about the calibration performed during traceability;
- uncertainty due to the accuracy of the device, if not corrected;
- uncertainty related to the deviation (fatigue) of the instrument between two calibrations;
- uncertainty related to the instrument's characteristics (reading, repeatability, etc.);
- uncertainty related to the environment, if conditions are different during calibration;
- instrumental uncertainty of the standard, its instability, changes in operating conditions, mutual influence of the standard and the sensor to be calibrated;
- the variation observed in the calibrated sensor readings.

[15] also explains that measurement uncertainties are parameters that depict the dispersion values during the measurement. In this way, the study of uncertainties aims to determine the capabilities of what measurement means.

## 3 Conclusions

Data acquisition has high importance but, even greater, is the information taken from that data, this is a valuable source for the companies' competitiveness. This what this paper addressed, namely the sensors in industries, responsible for obtaining data referring to various aspects of the companies. The paper also emphasises the importance of metrology in guaranteeing the reliability of measured values, as well as the importance of equipment certification.
The paper makes a theoretical approach about the topic under discussion, "Calibration and certification of industrial sensors - a global review", that addresses the following major subjects: Industrial Sensors; Metrology; Formal sensors Calibration; On-line sensors calibration; Calibration Certificate.
The first topic covered, "Industrial Sensors", was managed about the relevance of the reading instruments in the industry, as well as the types of existing sensors and their characteristics. It can be concluded that, these instruments are extremely important since they are responsible for translating the condition of the equipment in use, being able to extract from these the values about valuable information that could result in significant improvements in the company. It is a very important asset to follow unconditionally the conditioning maintenance.
About the second subject, "Metrology", this concept is discussed describing what it is and which is its focus on the industrial area. It can be inferred that without metrology and sensor calibration, the measurements performed would not be reliable, and may lead to unwanted and harmful results in many situations. The data obtained, without the use of metrology, from the equipment condition, is not the correct one; this can lead to maintenance measures not being implemented and, subsequently, to lead a rupture of the equipment and unexpected and consequently untimely stops, which can result in high costs for the company. It is also noteworthy that without metrological traceability, guaranteed by successive calibrations, it was not possible to trust on data measured by lower-level standards in the traceability chain.
Then, when the question is about "Formal sensors Calibration", the concept and how companies can manage this issue, in the most competent and effective way, it can be concluded that the calibration is responsible for ensuring metrological traceability and, consequently, for ensuring the veracity and effectiveness of the measurements made by the measuring instruments. It also can be found that without effective metrology system management, it can result in the use of the equipment out-of-date of

calibration state and subsequent loss of traceability to reference standards.

This paper also emphasises the importance of On-line Monitoring (OLM) aiming to maximize the equipment's availability. Through this calibration management system, it is possible to maximize, not only the use of reading instruments, but also the process equipment. In addition, it helps the asset management, allowing the anticipation of malfunction scenarios.

As final corollary, it can be said that through the calibration certification, companies can guarantee and prove the effectiveness of their measurements, using effective data, and eliminating measurement errors. So, without the support of metrology in large industries, there would be measurement errors and unreliable data, which, in turn, could result in uncontrollable risks and costs and, consequently, the loss of market confidence.

## References

[1] B. Al-najjar and I. Alsyouf, "Enhancing a company's profitability and competitiveness using integrated vibration-based maintenance : A case study," *Eur. J. Oper. Res.*, vol. 157, pp. 643–657, 2004, doi: 10.1016/S0377-2217(03)00258-3.

[2] P. Guedes, *Metrologia Industrial*, 1st Editio. Lisboa: ETEP - Edições Técnicas e Profissionais, 2011.

[3] K. K. Iniewski, *Smart Sensors for Industrial Applications*. CRC Press, 2017.

[4] M. J. McGrath and C. N. Scanaill, *Sensor Technologies: Healthcare, Wellness and Enviromental Applications*. Berkeley, CA: Apress, 2013.

[5] A. Martins, "Gestão de Laboratório de Calibração de Equipamentos de Análises de Vibrações," Instituto Superior de Engenharia de Coimbra, 2019.

[6] J. Wilson, *Sensor Technology HandBook*. Elsiver Inc., 2005.

[7] P. Wild, *Industrial Sensors and Applications for Condition Monitoring*. London: Mechanical Engineering Publications, MEP, 1994.

[8] D. Scott, *Industrial Process Sensors*, 1st Editio. CRC Press, 2008.

[9] J. M. T. Farinha, *Asset Maintenance Engineering Methodologies*, 1st Editio. Taylor & Francis Group, LCC, 2018.

[10] KeysightTechnologies, "Fundamentals of Sensor Measurements Using a Digital Multimeter," 2016. [Online]. Available: https://www.keysight.com/zz/en/assets/7018-06443/white-papers/5992-3555.pdf. [Accessed: 05-Mar-2020].

[11] L. M. V Serrano, C. J. O. P. J. Alcobia, M. L. O. S. Mateus, and M. C. G. Silva, "Sistemas de Aquisição , Processamento e Armazenamento de Dados," *Soc. Port. Metrol.*, pp. 1–15, 2004.

[12] A. B. Martins, J. Rodrigues, and J. Farinha, "Laboratório de Metrologia no apoio à Manutenção de Condição," *15.º Congr. Nac. manutenção*, 2019.

[13] IPQ, "Vocabulário Internacional de Metrologia (VIM): Conceitos Fundamentais e Gerais e Termos Associados," 2012. [Online]. Available: http://www1.ipq.pt/pt/metrologia/documents/vim_ipq_inmetro_2012.pdf. [Accessed: 16-Mar-2020].

[14] A. Shagluf, A. P. Longstaff, and S. Fletcher, "Maintenance Strategies to Reduce Downtime Due to Machine Positional Errors," *Maint. Perform. Meas. Manag.*, 2014, doi: 10.14195/978-972-8954-42-0_25.

[15] A. Grous, *Applied Metrology for Manufacturing Engineering*, 1st Editio. Hoboken, NJ, USA: John Wiley & Sons, Inc, 2011.

[16] R. Krishnakumar, "Entangling Atoms for Quantum Metrology," Stanford University, 2017.

[17] A. R. F. Pires, "Organização de um Processo de Gestão de Recursos de Monitorização e Medição," Universidade de Coimbra, 2018.

[18] P. flávio da S. Santos, "A Metrologia na Indústria Aeronáutica," Instituto Superior de Engenharia do Porto, 2015.

[19] P. Flores and F. Marques, "Sobre a metrologia," pp. 1–32, 2018.

[20] P. Cabral, C. Sousa, E. Filipe, P. Couto, and P. S. Girão, "A Metrologia como ferramenta transversal da Engenharia." Ordem Dos Engenheiros, 2015.

[21] T. Lopes, "Implementação de procedimentos de metrologia na empresa Milon," Faculdade Engenharia da Universidade do Porto, 2015.

[22] semyon G. Rabinovich, *Measurements and Uncertainties: Theory and Practice*, 3rd Editio. Springer, 2005.

[23] A. I. A. Castro, "Implementação de um Laboratório de Implementação de um Laboratório de Metrologia Elétrica," Instituto Superior de Engenharia do Porto, 2015.

[24] J. C. da S. Neto, *Metrologia e Controle Dimensional*. Elsivier, 2012.

[25] A.-K. Wiemann, M. Stein, and K. Kniel,

"Traceable metrology for large involute gears," *Precis. Eng.*, vol. 55, no. September 2018, pp. 330–338, Jan. 2019, doi: 10.1016/j.precisioneng.2018.10.001.

[26] V. V. Skliarov, P. Neyezhmakov, and A. Prokopov, "Metrological assurance and traceability for Industry 4.0 and additive manufacturing in Ukraine," in *Smart Structures and NDE for Industry 4.0*, 2018, no. March, p. 26, doi: 10.1117/12.2292821.

[27] R. Smith and R. K. Mobley, *Industrial Machinery Repair: Best Maintenance Pratices Pocket Guide*, 1st Editio. Elsevier Science (USA), 2003.

[28] R. K. Mobley, *Maintenance Fundamentals*, 2nd Editio. Elsevier Inc., 2004.

[29] K. Singhal, Q. Feng, R. Ganeshan, N. R. Sanders, and J. G. Shanthikumar, "Introduction to the Special Issue on Perspectives on Big Data," *Prod. Oper. Manag.*, vol. 27, no. 9, pp. 1639–1641, Sep. 2018, doi: 10.1111/poms.12939.

[30] H. Eren, "Calibrations in Process Control," in *Instrument Engineers' Handbook*, 4th ed., vol. 3, no. January, B. G. Liptak and H. Eren, Eds. CRC Press, 2012, pp. 141–149.

[31] J. Juran and A. Godfrey, *JURAN'S QUALITY HANDBOOK*, 5th ed. McGraw-Hill Companies, Inc, 1998.

[32] F. Tse and I. Morse, *Measurement And Instrumentation in Engineering: Principles and Basic Laboratory Experiments*. CRC Press, 2009.

[33] A. S. Morris and R. Langari, "Calibration of Measuring Sensors and Instruments," in *Measurement and Instrumentation*, Elsevier, 2012, pp. 103–114.

[34] H. Eren, "Calibration Process," in *Handbook of Measuring System Design*, no. setpember, P. H. Sydenham and R. Thorn, Eds. Chichester, UK: John Wiley & Sons, Ltd, 2005, pp. 271–277.

[35] M. Khosrow-Pour, "Big data: Concepts, Methodologies, Tools, and Applications," 2016, no. 4.

[36] A. Davies, *Handbook of Condition Monitoring: Techniques and Methodology*. Springer Netherlands, 1998.

[37] H. M. A. Fahmy, "WSNs Applications," in *Wireless Sensor Networks: Concepts, Applications, Experimentation and Analysis*, Springer, Singapore, 2016, pp. 69–213.

[38] J. W. Hines, R. Seibert, and A. Usynin, *Technical Review of On-Line Monitoring Techniques for Performance Assessment. Volume 2: Theoretical Issues*, vol. 2. United States Nuclear Regulatory Commission, 2008.

[39] S. Lee and J. Chai, "An enhanced prediction model for the on-line monitoring of the sensors using the Gaussian process regression," *J. Mech. Sci. Technol.*, vol. 33, no. 5, pp. 2249–2257, May 2019, doi: 10.1007/s12206-019-0426-7.

[40] A. J. Schiff, *Power Plant Surveillance and Diagnostics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002.

[41] M. Hoffmann, "PEANO — a tool for on-line calibration monitoring," no. January, 2004.

[42] A. A. Boechat, U. F. Moreno, and D. Haramura, "On-line calibration monitoring system based on data-driven model for oil well sensors," *IFAC Proc. Vol.*, vol. 45, no. 8, pp. 269–274, 2012, doi: 10.3182/20120531-2-NO-4020.00037.

[43] J. W. Hines, "On-Line Monitoring for Calibration Extension : an Overview and Introduction," U.S. Nuclear Regulatory Commission, 2009.

[44] J. Coble, P. Ramuhalli, R. Meyer, H. Hashemian, B. Shumaker, and D. Cummins, "Calibration Monitoring for Sensor Calibration Interval Extension Identifying Technical Gaps," in *Future of Instrumentation International Workshop (FIIW) Proceedings*, 2012, doi: 10.1109/FIIW.2012.6378348.

[45] J. W. Hines and R. Seibert, *Technical Review of On-Line Monitoring Techniques for Performance Assessment. Volume 1. State-of-the-Art*, vol. 1. U.S. Nuclear Regulatory Commission, 2006.

[46] J. B. Coble *et al.*, "A Review of Sensor Calibration Monitoring for Calibration Interval Extension in Nuclear Power Plants," no. August. Pacific Northwest National Laboratory, 2012.

[47] P. Ramuhalli *et al.*, *Uncertainty Quantification Techniques for Sensor Calibration Monitoring in Nuclear Power Plants*, no. April. Washington: Pacific Northwest National Laboratory, 2014.

[48] S. Dwivedi, Al. Verma, and J. Sneckenberger, "CAD/CAM Robotics and Factories of the Future '90," in *International Conference on CAD/CAM, Robotics, and Factories of the Future*, 1989, doi: 10.1007/978-3-662-39962-0.

[49] E. Bulska, "Metrological Traceability," in *Metrology in Chemistry*, L. N. in Chemistry,

Ed. Springer, 2018, pp. 35–51.

[50] J. E. Gallagher, "Inspection, Testing, Verification, Calibration, and Certification," in *Natural Gas Measurement Handbook*, Elsevier, 2006, pp. 387–406.

[51] M. C. Gupta, "Document control: status of calibration/test certificate," *Accredit. Qual. Assur.*, vol. 15, no. 10, pp. 591–593, Oct. 2010, doi: 10.1007/s00769-010-0682-9.

[52] I. Zakharov and P. Neyezhmakov, "Peculiarity of measurement instruments verification by results of their calibrations," in *2017 11th International Conference on Measurement*, 2017, no. Imi, pp. 19–22, doi: 10.23919/MEASUREMENT.2017.7983526.

**ACKNOWLEDGMENTS**

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowvska-Curie grant agreement 690968 project NANOGUARDAr - GA NANOGUARD2Ar - GA №690968 and the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under Project POCI-01-0145-FEDER-029494, and by National Funds through the FCT - Portuguese Foundation for Science and Technology, under Projects PTDC/EEI-EEE/29494/2017, UIDB/04131/2020, UIDP/04131/2020,

# Appendix B

applied sciences

MDPI

*Article*

# Maintenance Prediction through Sensing Using Hidden Markov Models—A Case Study

**Alexandre Martins [1,2,*], Inácio Fonseca [3], José Torres Farinha [3,4], João Reis [1,5] and António J. Marques Cardoso [2]**

1 EIGeS—Research Centre in Industrial Engineering, Management and Sustainability, Lusófona University, Campo Grande, 376, 1749-024 Lisboa, Portugal; p40500@ulusofona.pt
2 CISE—Electromechatronic Systems Research Centre, University of Beira Interior, Calçada Fonte do Lameiro, P-62001-001 Covilhã, Portugal; ajmc@ubi.pt
3 ISEC/IPC—Polytechnic Institute of Coimbra, 3045-093 Coimbra, Portugal; inacio@isec.pt (I.F.); tfarinha@isec.pt (J.T.F.)
4 CEMMPRE—Centre for Mechanical Engineering, Materials and Processes, University of Coimbra, 3030-788 Coimbra, Portugal
5 GOVCOPP—Department of Economics, Management, Industrial Engineering and Tourism, Campus Universitário de Santiago, Aveiro University, 3810-193 Aveiro, Portugal
* Correspondence: alex_bmts@hotmail.com

**Abstract:** The availability maximization is a goal for any organization because the equipment downtime implies high non-production costs and, additionally, the abnormal stopping and restarting usually imply loss of product's quality. In this way, a method for predicting the equipment's health state is vital to maintain the production flow as well as to plan maintenance intervention strategies. This paper presents a maintenance prediction approach based on sensing data managed by Hidden Markov Models (HMM). To do so, a diagnosis of drying presses in a pulp industry is used as case study, which is done based on data collected every minute for three years and ten months. This paper presents an approach to manage a multivariate analysis, in this case merging the values of sensors, and optimizing the observable states to insert into a HMM model, which permits to identify three hidden states that characterize the equipment's health state: "Proper Function", "Alert state", and "Equipment Failure". The research described in this paper demonstrates how an equipment health diagnosis can be made using the HMM, through the collection of observations from various sensors, without information of machine failures occurrences. The approach developed demonstrated to be robust, even the complexity of the system, having the potential to be generalized to any other type of equipment.

**Keywords:** Hidden Markov Models; industrial sensors; condition-based maintenance; big data; cluster analysis; principal component analysis



**Citation:** Martins, A.; Fonseca, I.; Farinha, J.T.; Reis, J.; Cardoso, A.J.M. Maintenance Prediction through Sensing Using Hidden Markov Models—A Case Study. *Appl. Sci.* **2021**, *11*, 7685. https://doi.org/10.3390/app11167685

Academic Editors: João Carlos de Oliveira Matias and Paolo Renna

Received: 15 July 2021
Accepted: 17 August 2021
Published: 21 August 2021



## 1. Introduction

Sensors are currently one of the largest sources of data, being responsible for making a direct connection between a physical phenomenon and a data acquisition system, converting signals from several types of variables (mechanical, chemical, etc.) into electrical signals. Thus, sensors are responsible for translating the equipment condition, giving outputs that can be seen as observable states. In other words, they provide data that, after being studied, can provide very useful information for companies. One of the ways in which companies maintain competitiveness and customer satisfaction is through the application of sensing, as this will allow them to carry out Condition-Based Maintenance (CBM). This is aimed at the increasing of profits, namely, due to the non-existence of unexpected stoppages in production. For instance, according to Pais et al. [1], Data Mining and Artificial Intelligence (AI) can contribute to safeguard the company's profits and the safety of people and property.

This work aims to demonstrate how it is possible to optimize the data collected from the raw signal given by the sensor and, through the transformation of observable states made by the HMM, how the equipment diagnosis can be obtained. The case study aims to diagnose the state of health of a drying press in a pulp industry, where the section task is to remove water, mainly through pressing. Further, the drying press in a pulp industry is an equipment where the occurrence of failure is harmful to the production chain, which can lead to loss of time and high costs for the company. Therefore, to diagnose this equipment, we have collected data with six sensors, which are coupled to the equipment and are responsible for measuring the current intensity, the level of the hydraulic unit, the torque, the VAT pressure, the speed of rotation, and temperature of the hydraulic unit, in order to characterize the health status of the equipment. This will be done by joining the data from the six sensors and optimizing the observable states, which will later be inserted into an HMM model, allowing the equipment to be diagnosed. The optimization of observable states goes through four phases: data preparation; features generation; Principal Components Analysis (PCA), and Clustering. With this procedure it is possible to extract more information from the signals, to reduce the computational load, and to join the observations of the six sensors, allowing the creation of only one HMM to diagnose the equipment. The objective is to characterize the health of the equipment in three states over the time of data collection: "Adequate Functioning", "Alert state", and "Equipment failure".

A theoretical basis is also made on the entire methodology used to carry out the equipment diagnosis in order to demonstrate how the HMM can be applied to assist in Condition-Based Maintenance.

As Rodrigues et al. [2] stated, there has been extensive research on maintenance optimization that has also been an important trend in the area of optimization based on maintenance simulation. The authors state that research in this context focuses on the goal of finding the best maintenance policy while minimizing the overall company costs. Like Mateus et al. [3] state, "it was clear that a well maintained property will have a longer useful life with a greater return for the organization". The main objective of this type of maintenance is to use adequate sensor signals and monitoring techniques to identify and predict the health status of machines, in order to reduce the economic loss due to degradation or failure [4]. Another aspect to consider when talking about data collection with sensors is their reliability [5]. As noted by Taylor et al. [6] and by Bunks, Mccarthy, and Al-Ani [7], a prerequisite for the implementation of effective CBM practices in the industry is an effective diagnosis and prognosis.

According to Kamlu and Laxmi [8], the CBM decision-making process based on the information obtained from the HMM is adequate to recognize the condition of the equipment. The HHM is a double stochastic process that can define, through probability, the equipment diagnosis, that is, how the machine can go from a good working state to a bad working state and what the statistics are in each state.

Bunks et al. [7] explain that the HMM has two very useful features when it comes to monitoring CBM equipment: first, they refer to the existence of computationally efficient methods to calculate probabilities (an important feature, as it promises to be an efficient tool for signal processing that can be economically implemented), and second, they mention the existence of efficient techniques that can be used to identify the system health with HMM, that is, they can be used to build equipment models based on data, aiming to identify specific characteristics in the data to be used as health indicators.

There are several examples of HMM-based approaches to monitor failure detection [4,6–15]: Yu [4] uses an unsupervised online learning scheme, where an Adaptive Hidden Markov Model (AHMM) is used to learn online the dynamic changes in the health of the equipment; Taylor et al. [6] present a method where several HMMs are used to represent a health state of a metal cutting tool; BUNKS et al. [7] show, in addition to why the HMM is a model for equipment diagnostics, its use with vibration data to diagnose helicopter gears; Arpaia et al. [10] use the HMM to detect failures for fluid machines without

adequate a priori information about failure conditions; Ocak and Lopar [12] present a bearing failure detection and diagnosis scheme based on vibration signals; Xinmin et al. [13] focus on the analysis of the bearing failure diagnosis model concentrated in an HMM model, and compare it with other approaches, namely Multilayer Perceptron (MLP) detection techniques; and Simões et al. [14] present an extensive description of the state of the art of the HMM, also describing how this model can increase the quality of the assessment of Diesel engine conditions and the efficiency of maintenance planning.

In relation to the study carried out, this paper demonstrates how by performing a multivariate analysis using several different sensors and with only one HMM it is possible to obtain a diagnosis of the equipment, in this case, a drying press. This approach is used instead of using one HMM for each sensor or multiple HMMs for each sensor, where each HMM represents an equipment health state. In this paper, the steps to be followed to optimize the HMM inputs will be demonstrated, and it will also be shown that through features generation and extracting features (using PCA), it is possible to merge the various sensors.

The paper starts with an introduction to the general theme in Section 1. In Section 2, the methodology used to optimize observable states for HMM is explained. Section 3 presents a theoretical framework for each of the steps, including the HMM model itself. Section 4 explains the case study and demonstrates all the steps taken until the equipment is diagnosed. Section 5 discusses the results and explains ideas for future work. Finally, in Section 6, the conclusions of the study are presented.

## 2. Theoretically Background

This section aims to provide knowledge of the literature review on the topics covered in this paper. That is, all the processes used to "optimize" the observable states, as well as the HMM model used to diagnose the equipment.

### 2.1. Data Preparation

According to Yin et al. [16], since the beginning of the 1980s data mining has gained increasingly more attention as a means of obtaining knowledge.

The preprocessing of raw data facilitates the stabilization of the mean and variance and aims to provide a structural, reliable, and integrated data source, and it is a very critical and complex step, which allows guaranteeing reasonable results, whether the analysis is concerned with exploratory data mining, classification, or construction of a good and robust prediction model [17–21].

Data cleaning and preparation consume approximately 80% of the total data engineering effort, as it is a process that may require many transformations and be repeated many times [21–23].

Through good data cleaning, it is possible to improve performance and quality [17,24,25], as well as to isolate characteristics of interest and eliminate elements that "bother" the theoretical models [26].

#### Data Transformation

Data formats can be transformed to meet the premises of a statistical inference procedure or to improve interoperability [27]. Data normalization is the process of transforming raw data values into another format with properties better suited for modeling and analysis. For a variable X represented by a vector $\{x_1, x_2, x_3, \ldots, x_n\}$., we can use Z-score normalization (also known as standardization).

Z-score normalization is a normalization method that transforms not only the magnitude of the data, but also the dispersion. It will convert a variable range with some mathematical heuristics, allowing all variables to have the same range. This patterning of values is not essential in machine learning algorithms, but it can make patterns in the data more visible [23]. Z-score normalization overcomes the problem of variables with different

units, as it transforms the variables; therefore, they are centered on 0 with a standard deviation of 1 [21]. Each attribute can be transformed using Equation (1):

$$x'_i = \frac{x_i - \overline{X}}{std_{dev}(X)} \quad (1)$$

where

- $x_i$ is the Z-score value of $x$;
- $\overline{x}$ is the row mean of $x$;
- $std$ is the standard deviation given by

$$std_{dev}(\overline{x}) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \quad (2)$$

### 2.2. Features Generation

The principle of predictive Feature Generation (FG) is used to maximize the exploitation of information generated exclusively from time and process data, with compact and informative representations of the obtained data [27,28].

According to Silva and Leong [29], Machine Learning (ML) algorithms can be seen as techniques to derive one or more hypotheses from a set of observations, thus being one of the ways to improve input (observations) and output (hypothesis) in selecting resources that maximize the performance of architecture.

If the transformation is chosen correctly, the transformation features can display high-information packing properties compared to the original input samples.

In summary, FG is of paramount importance in any pattern recognition task, whose objective is to discover compact and informative representations of the obtained data [30].

### 2.3. Principal Component Analysis (PCA)

Dimensionality Reduction is a technique to take a high-dimensional dataset (data objects with many features/attributes), replacing it with a much smaller dimensional dataset, preserving the similarities between the data objects [31].

Principal Components (PC) were first found by Pearson [32] and gained consideration by the statistical community when Hotelling [33] proposed them as an estimate of linear combinations of a set of random variables that retained the greatest possible variance.

Roy et al. [27], Anowar et al. [34], and Warmenhoven et al. [35] argue that it is an unsupervised linear transformation algorithm and that it uses a statistical procedure through an orthogonal transformation to map a set of observations of possibly correlated variables into a set of linear values of uncorrelated variables called Principal Components.

In short (Table 1), PCA is a method that seeks to find linear combinations of predictors, known as principal components (PCs), that capture the greatest possible variance [36]. The first PC is defined as the linear combination of predictors that captures the greatest variability of all possible linear combinations. Subsequent PCs are derived, so these linear combinations capture the greatest remaining variability, although they are not correlated with all previous PCs.

**Table 1.** PCA algorithm steps.

| **ALGORITHM—PCA** |
| --- |
| **INPUT:** $X \in R^{n*d}$ |
| **OUTPUT:** $Y \in R^{n*k}$ |
| **1: CONSTRUCT THE COVARIANCE MATRIX** $(X.X^t)$ |
| **2: APPLY LINEAR EIGEN DECOMPOSITION TO** $X.(X^T)$**TO OBTAIN EIGEN VALUES AND VECTORS;** |
| **3: SORT EIGEN VALUES IN DECREASING ORDER TO SORT EIGEN VECTORS** |
| **4: BUILD MATRIX** $W(D*k)$**WITH** $K$**TOP EIGEN VECTORS** |
| **5: TRANSFORM** $X$ **USING** $W$ **TO OBTAIN THE NEW SUBSPACE** $Y$**=**$X.W$ |

### 2.4. Clustering K-Means

Clustering is a fundamental unsupervised data mining technique that is used in a wide variety of areas for data analysis [37], where cluster analysis is done by multivariate statistical methods and the objective is to identify clusters of objects within the data, maximizing the homogeneity within each cluster and the heterogeneity between the different clusters [18,37–41].

The k-means algorithm was first proposed by Stuart Lloyd in 1957 as a technique for pulse code modulation, with a more efficient version being proposed and published in Fortran by Hartigan and Wong in 1975/79 [42,43].

It is a numerical, unsupervised, iterative, and evolutionary algorithm, which aims to find the positions of the clusters that minimize the distance from the data points to the clusters, it then partitions the $n$ observations into $k$ clusters in which the observation belongs to the cluster with the most average next [42,44].

Each cluster in the partition is defined by its member objects and its centroid, with the centroid being the point at which the sum of the distances of all objects in that cluster is minimized [43]. Then, K-means finds a centroid to correlate the data to a specific group, and the process of estimating the data belonging to a given cluster is carried out iteratively until the convergence condition is satisfied [39].

To define the cluster number $k$, the Silhouette method can be used, because, according to Menardi [45], the diagnostic used to assess the quality of a partition must be consistent with the clustering method adopted to produce that partition, being Silhouette parsing a method that fits well with the K-means Clustering algorithm.

A detailed explanation of this topic can be found in [37,45,46]. Peralta and Saeys [47] summarize by saying that the Silhouette Index (SI) is calculated for each resource and measures it through the ratio between intra-cluster and inter-cluster dissimilarity $a(k)$ and $b(k)$, respectively, as shown in Equation (3), generating a value between $-1$ and 1.

$$S(k) = \frac{b(k) - a(k)}{max\{a(k), b(k)\}} \quad (3)$$

### 2.5. Hidden Markov Models (HMM)

Hidden Markov Models (HMMs) started to be introduced around the 1970s, with the publications of Baum [48] and Baum & Petrie [49], and later became quite popular in the late 1980s with the contribution of Rabiner [50]. Currently, the HMM is a statistical modeling tool that has become popular in several areas, such as speech processing, DNA recognition, weather forecast, machine maintenance, pattern identification, and health monitoring.

HMMs are based on a doubly stochastic process, in which an underlying stochastic process that develops as a Markov chain produces an unobservable ("hidden") state, which can be inferred only through another set of stochastic processes [51]. Therefore, an HMM is a stochastic technique for modeling signals that evolve through a finite number of states. States are considered hidden and responsible for producing observations [4,6].

A typical notation used for HMM is

- $N$—Number of states

- $M$—Number of observations
- $T$—Observation sequence length
- Sequence of observations up to T is presented as:
- $O_1, O_2, O_3, \ldots, O_T$ with $O_t \in \{V_1, V_2, V_3, \ldots, V_M\}$
- Corresponding status sequence is displayed as:
  - $Q_T = \{q_1, q_2, q_3, \ldots, q_T\}$ with $q_t \in \{S_1, S_2, S_3, \ldots, S_N\}$
- $\lambda = (A, B, \pi)$ – is a representation of an HMM model $A$—Transition probability matrix between hidden states

$$A = \{a_{ij}\}, 1 \leq i, j \leq N \tag{4}$$

where:

- $a_{ij} = P$ and $q_t$ represents the hidden state at time $t$
- $B$—Emission Probability Matrix

$$B = \{b_j(k)\}(1 \leq j \leq N) \tag{5}$$

- $\pi$—initial state probability distribution

$$\pi_i = P(q_t = S_i) \tag{6}$$

where the observations are issued from each state according to the probability distribution

$$b_j(k) = P(O_k \vee q_t = S_j), \text{ it's} \tag{7}$$

An HMM has several hidden states ($N$), an initial probability value ($\pi$) for each state, a transition probability matrix ($A$) indicating the probability of transition from one hidden state to another, and a probability matrix of transition ($B$) which indicates the probability of an observation given a certain hidden state. The sum of all initial probabilities must be equal to 1, as well as the sum of all elements in a row in the transition and issue probability matrix [50].

According to Rabiner [50], there are three basic HMM problems:

*Problem 1*—This is an evaluation problem, where, given a sequence of observations $O = \{O_1; O_2; O_3; \ldots; O_T\}$ and the model $\lambda = (A, B, \pi)$, it allows to efficiently to calculate the associated probability to the $P(O \vee \lambda)$ emission sequence.
*Problem 2*—This is a decoding problem, which consists of finding the most likely sequence of hidden states given the sequence of observed emissions $O = \{O_1; O_2; O_3; \ldots; O_T\}$ and the model λ, namely, how to find a corresponding sequence of states $S = \{S_1; S_2; S_3; \ldots; S_T\}$.
*Problem 3*—This is about to know a database of sequences and how to adjust the parameters of the model $\lambda = (A, B, \pi)$ in order to maximize $P(O \vee \lambda)$.

➢ The solution for Problem 1

To solve the first problem, the forward and backward algorithms must be used. In this way, it is possible to find the forward variable $\alpha_t(i)$, defined as

$$\alpha_T(i) = P(O_1, O_2, O_3, \ldots, O_T, q_T = S_i \vee \lambda) \tag{8}$$

This variable will give the probability of the observation sequence, $O_1, O_2, O_3, \ldots, O_t$ and state $S_i$ at time $t$ given the model $\lambda$.

Through the forward and backward procedure, we also obtain the backward variable $\beta_t(i)$, defined as

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, O_{t+3}, \ldots, O_T \vee S_t = q_i, \lambda) \tag{9}$$

This variable gives the final probability of the observation sequence from $t + 1$, given the state $S_i$ at time $t$ and the model $\lambda$. This variable is not necessary for this first problem, but it will be useful to solve *Problem 3*.

➢ The solution for Problem 2

This problem aims to find the best sequence of hidden states that best fits the observed states. This is done using the Viterbi algorithm.

➢ The solution for Problem 3

The third problem is to adjust the HMM parameters to maximize the probability of the observation sequence. This is done using the Baum–Welch algorithm.

The Baum–Welch algorithm (or Baum–Welch expectation-maximization algorithm) makes use of both the direct variable $\alpha_T(i)$ and the regressive variable $\beta_t(i)$ when determining updated parameters for the HMM. Because of this, the Baum–Welch algorithm is also known as the Forward-Backward algorithm [11].

As Yu [4] refers, three basic algorithms are used in HMM: the Forward-Backward procedure, the Baum–Welch algorithm, and the Viterbi algorithm, all of which are used for learning and recognizing the parameters of the model $\lambda = (A, B, \pi)$.

## 3. Global Framework

Rabiner [50] states that "Real-world processes generally produce observable outputs which can be characterized as signals." These observable states will be used as input to the HMM model which will later give us the hidden states which, in turn, will correspond to the equipment state. To improve these inputs to the model, a sequence of methodologies is performed to perform the "optimization" of the observable states (Figure 1).

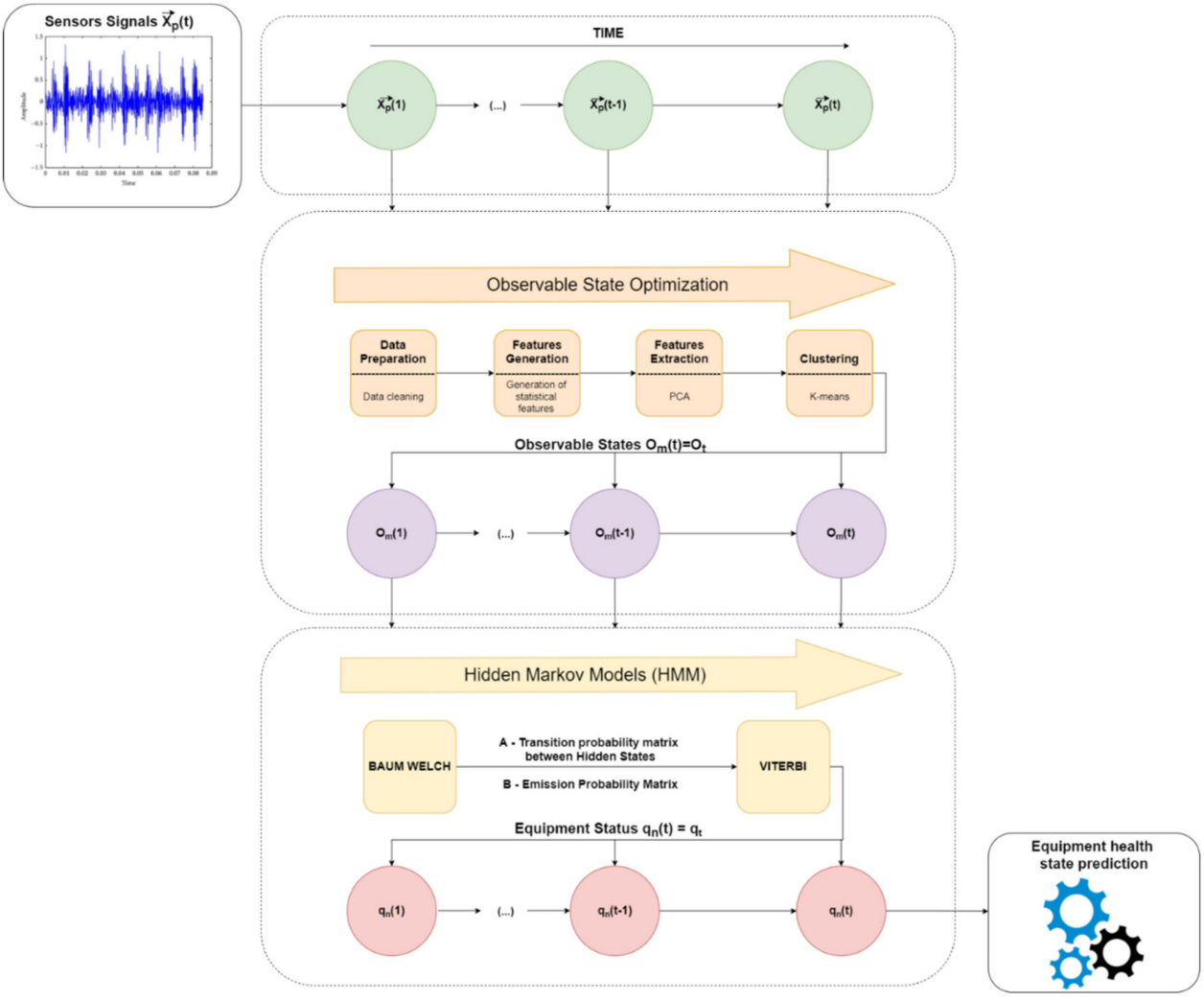



**Figure 1.** Methodology for observable states optimization to insert into the HMM.

As we can see in Figure 1, the sensor data will be modified to obtain better algorithm efficiency, both in calculation time and global performance.

First, the data are collected by sensors over time, which, later, are managed aiming to optimize the observable states. After data collection, these go to the preprocessing stage, where a data preparation is carried on, to improve the quality of the data set.

At the end of data preparation, the data are grouped in time intervals and, after, replaced by statistical values representative of each interval. Based on this generation of characteristics, it is possible to extract more information from the data, as the statistical resources, based on the time domain, providing good performance to characterize trends and changes in data [52].

The data are then normalized and inserted into a dimensional reduction model using Principal Component Analysis (PCA). Sizing features through Z-Score standardization is an important preprocessing step to resize features according to a standard normal distribution, with a mean zero and a standard deviation one. In our case, it is important to use the Z-Score standardization to implement PCA method because we are interested in the components that maximize the variance. If one feature varies less than another, because of their respective scales, the PCA defines the direction of the maximum variation that more closely matches the axis, which varies more if those features are unscaled. It is important to standardize the data to make the algorithm more sensitive to the changes that may occur in different measurements. Therefore, standardization is performed because this model requires resources to be on the same scale aiming to find directions that maximize the variation. In this process, we will only obtain the principal components (PCs) for the study and reduction of the dimensionality of the dataset. The observations are now defined in a new space, redefining the axes through the PCs instead of the original variables. In other words, the characteristics of the data will be reduced to reduce the computational load. In this study, the PCA was chosen; it is a very consolidated and easy method to use. It is an unsupervised model that permits to find the most significant coordinate system of the data to find the strongest characteristics of the samples. Additionally, it is not time-consuming, it reduces well the overfitting, and can be used as a noise removal and data compression technique [27,34,38,53]. The elimination of the large part of the experimental noise is possible—the noise originates random errors, which are not correlated with the information contained in the data matrix—improving the numerical stability of the model [36]. Finally, a clustering is performed, which will group similar data and differentiate from data from other data groups. In other words, clustering will group the observations as homogeneously as possible within each cluster and as heterogeneously as possible among the different clusters. In this way, observations of the equipment in good working order will be grouped in a cluster, observations of the equipment in failure will be grouped in another, and so on. Each cluster formed will, in turn, be used as an observable state that will enter the HMM. The method used in this study will be Clustering k-means, which does the grouping using Euclidean metrics. k-means is the algorithm chosen to carry out our study; because it is a model that uses the unsupervised learning method, it has a fast enough convergence speed and does not require large computational resources; it is an easy algorithm to implement and apply, even to large data sets [33,34].

In addition, it is an algorithm that can be used after refocusing the data from the source of the metrics framework involved in the center of the cloud (process done by the PCA). This causes the data to be centered in relation to an origin where it is positioned, being relatively equidistant from the center and positioned in the several quadrants in the $Rp$ space ($p$ being the PCA's Principal Components), which will allow a good performance of the K-means.

Finally, the HMM will give us, through the hidden states, the diagnosis of the equipment.

## 4. Case Study

This case study focuses on the fault diagnosis status of equipment (i.e., a drying system press) in a Portuguese paper pulp industry. The objective is, through the data provided by the company, to be able to make a diagnosis of the equipment state in order to know if it is in "Good Operation", "Alert Status", or "Faulty Equipment", using the HMM model. This case study's specification adds value about evaluating the equipment diagnosis, namely, because it does not need a priori information concerning the failure conditions. To carry out the study, the MatLab software and its functions were used.

### 4.1. Data Preparation

The data used for the study were collected by six sensors (responsible for measuring current intensity, hydraulic unit level, torque, VAT pressure, rotation speed, and hydraulic unit temperature), which collect an observation every minute. The data collected are from August 2017 to October 2020. Thus, the data used were acquired over three years and three months (Table 2).

**Table 2.** Used sensors: Number of data collected.

| Equipment | Sensors | Metric | Nº of Data by Day (Each Sensor) | Nº of Data by Year (Each Sensor) | TOTAL of Data (Each Sensor) |
|---|---|---|---|---|---|
| **Drying Presses** | Sensor−1 | Hydraulic unit Temperature | 1440 | $1440 * 365 = 525600$ | $(1440 * 153) + (525600 * 2) + (1440 * 305) =$ |
| | Sensor−2 | Hydraulic unit level | | | |
| | Sensor−3 | Rotation speed Press | | | |
| | Sensor−4 | Torque Press | | | |
| | Sensor−5 | Current intensity with endless extraction drive | | | |
| | Sensor−6 | VAT Pressure vat press | | | |

Table 2 shows that there is a dataset with 10,264,320 observations. However, as we referred above in the theoretical framework, there are always errors in sensor data collection and, therefore, not all data in this dataset are valid. According to Van Den Broeck et al. [54], data cleansing should be based on knowledge of technical errors and expected ranges of normal values. As such, through a data collection confidence index, it was possible to verify whether the collected values were valid or not. This confidence index is a protection of the data acquisition system, which through just two values (0 or 100) indicates the quality of the data. Regarding the implementation of this system, there are not many details as it was installed by a tertiary company. If the index had a value of 100, it meant that the data had been well collected and they were valid, and when the index had a value of zero, it meant that something had happened with the data collection and, as such, they were considered invalid. Therefore, it was decided to eliminate all points whose confidence index was equal to zero.

In addition to this data cleaning, it was also decided to make an imputation of the total mean of the data signal, whose values from the current intensity, torque, VAT pressure, and rotation speed sensors were below a certain threshold. In this way, we guarantee that we can remove the equipment shutdown and to replace them by the respective average of each signal. Therefore, the equipment shutdowns will not affect the prediction of the model. The value imputation was performed, which is the most used, according to the most usual referred in literature, as can be seen, for example, in [23,31,55]; the most general opinion supports this approach when the data of a variable is missing. Based on this approach,

some care was considered: when a slice was replaced in one of the signals, the others were all replaced by their respective mean, increasing the integrity of the data set. No further data cleaning is performed as the following data processing models themselves complement the "filtering" of the dataset.

In Figure 2, we can graphically compare the data set with all the values, with the valid values and without pulp drying press downtimes.

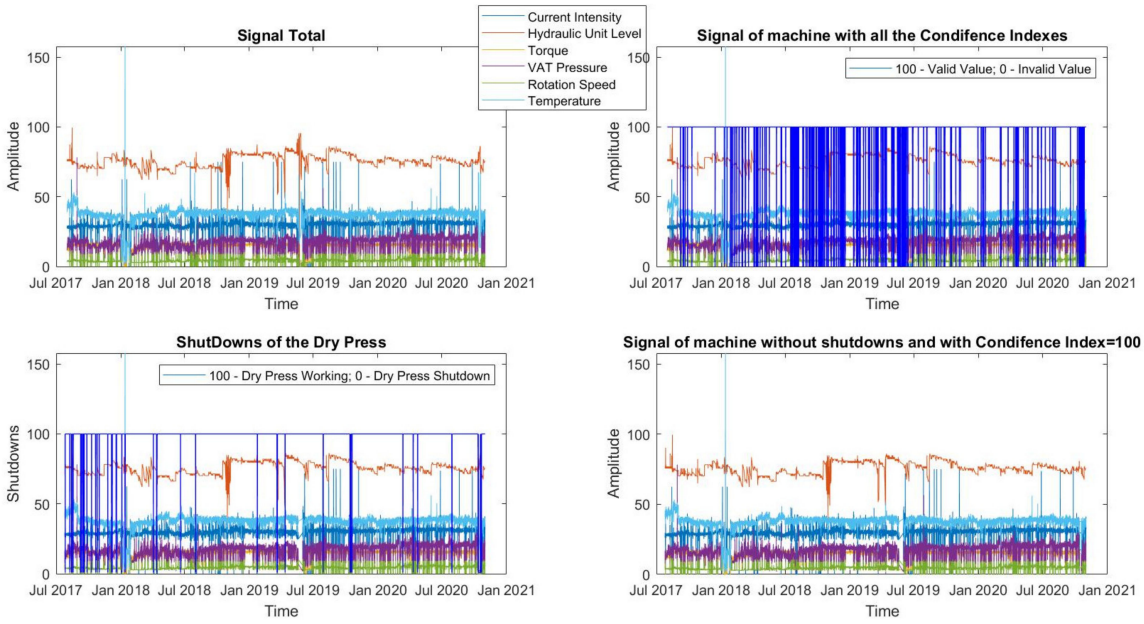



**Figure 2.** Data cleaning of the signals with through Confidence Indexes and shutdowns of the equipment.

Through the confidence index, 69,840 data were eliminated, leaving a data set of 1,640,880 per sensor. Regarding equipment shutdowns, these were replaced by the average of each signal and, therefore, the value of 1,640,880 remained unchanged.

As is well known, outliers can be classified into two different subgroups: natural and interesting ones, and those caused by defective instruments. The first group will contribute with data that may improve the analysis or construction of the subsequent model, while the second will contribute with errors that make the information extraction less accurate [21,23,56]. Therefore, in this case, as we are carrying out a study of the failure diagnosis of drying presses, we chose not to remove what could be the outliers, as in this case, because these values may represent equipment malfunctions and, by consequence, an invalid number.

### *4.2. Features Generation*

Six-hour time windows were created to represent four shifts of the equipment's workday, where each window has 360 data. Thus, 4558 time windows were additionally created with 360 data each. The objective is to create statistical characteristics of each window to better characterize the signal and to make a dimensional reduction of the data. Ten different characteristics were obtained from each data window to better define the signal and understand its behavior over time. The characteristics generated were 1. Average; 2. Standard deviation; 3. Variance; 4. Kurtosis; 5. Skewness; 6. Coefficient of variation; 7. Maximum; 8. Minimum; 9. Mode; 10. Median. These 10 characteristics were chosen in order to have a limit of resources for the study to verify if they are sufficient to obtain a diagnosis. The features are all performed in the time domain, as none of the variables need frequency analysis. Furthermore, as stated in [52], "statistical resources based on the time domain provide a high performance to characterize trends and changes

in signals". The corresponding equations of the considered set of characteristics can be found in the references [52,57]. In Figure 3, we can see the features generated for each variable.

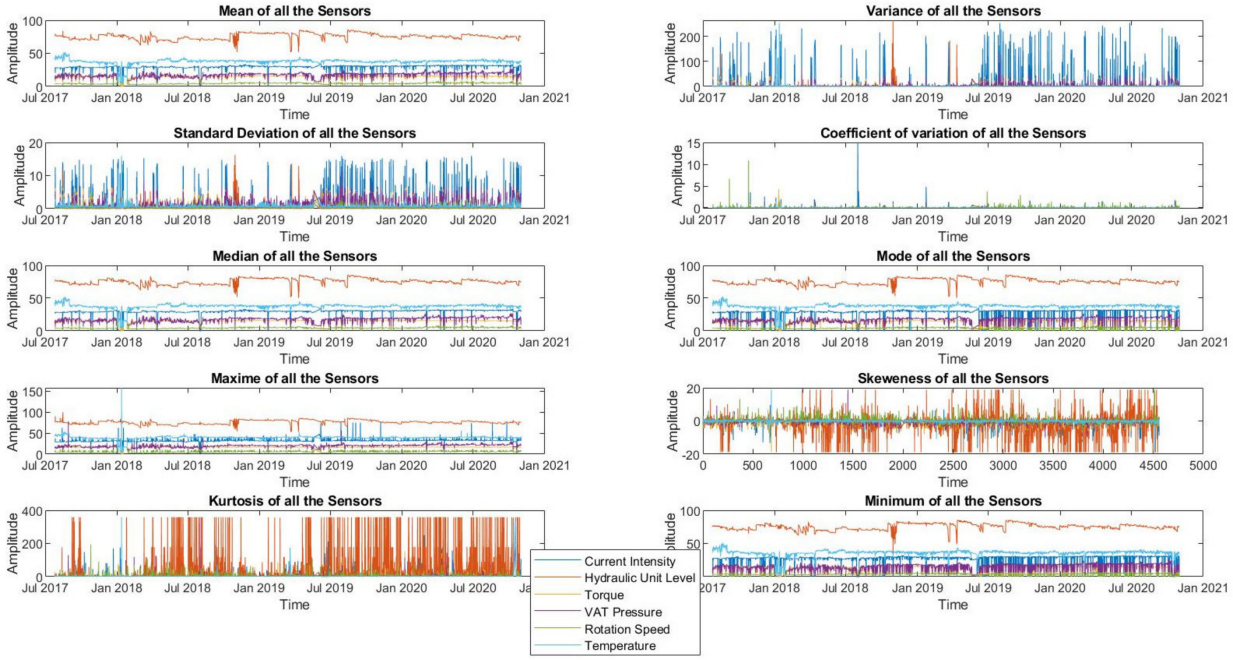



**Figure 3.** All features for each one of the signals sensors.

Based on these 10 characteristics of each sensor, 60 variables were generated from a total of 4558 data. That is, we are left with 4558 mean values for each sensor, 4558 variance values for each sensor, etc. Therefore, we get an array of a dimension of $4558 * 60$.

### *4.3. PCA*

From the matrix built above, we made a feature extraction and reduced the size of the matrix caused by feature generation using the Principal Component Analysis (PCA) tool for the study of the principal components. From the matrix $4558 \times 60$, a reduction was made, starting to have a matrix with only ten main components, $4558 \times 10$. As already described, the first dimensions are those that contain the greatest variation of the data. In Table 3, taken from the MatLab PCA function, we can see the percentage variability of each of the main components and how it decreases in each PCA.

**Table 3.** Percentage variability explained by the main components.

| Principal Components | % of Preserved Data |
|---|---|
| 1 | 26.8299362662924 |
| 2 | 11.3444018431582 |
| 3 | 10.5209646496898 |
| 4 | 7.43400606597570 |
| 5 | 6.75048105033166 |
| ( . . . ) | ( . . . ) |

Therefore, the top ten components were chosen as this number of components preserves ~82% of the data. For a graphical analysis of the variability of the data on each PCA, we used a Pareto diagram (Figure 4).

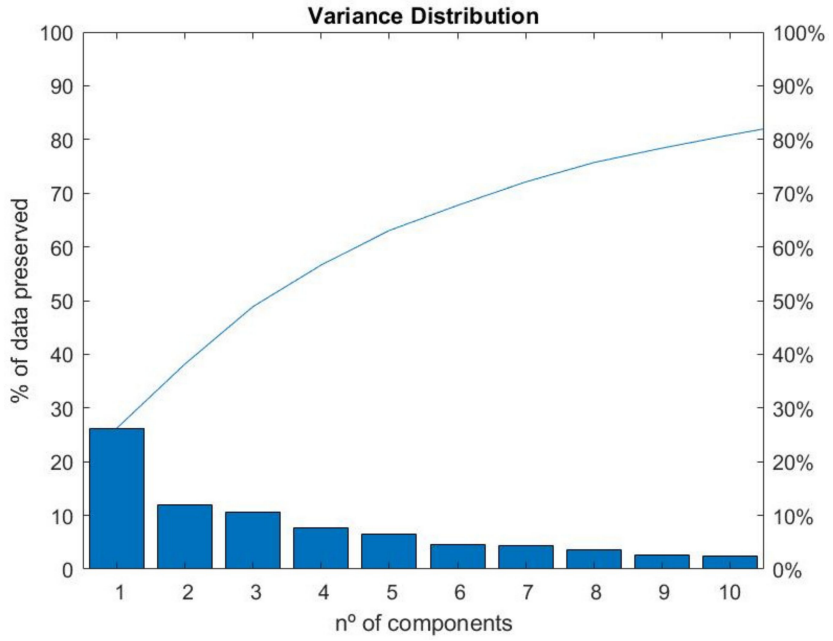



**Figure 4.** Pareto diagram to represent the percentage of data preserved in each Principal Component.

In addition to the feature extraction through PCA, a study can still be done where we can identify which features contributed most to each of the main components, through the coefficients of the main component. This can be very well represented using a heat map represented in Figure 5.

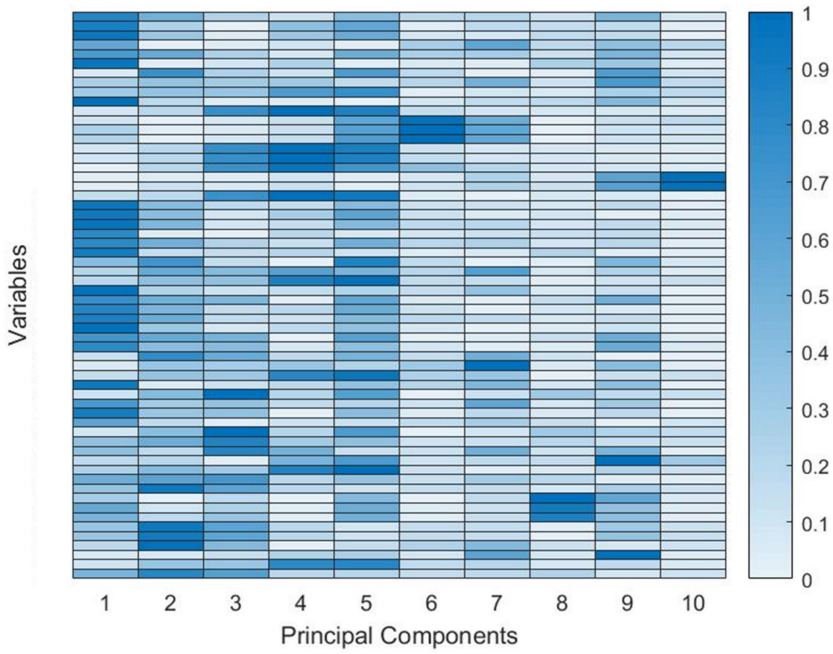



**Figure 5.** Contribution of each of the features in each of the Principal Components.

In the coefficient matrix, each coefficient column contains coefficients for a principal component and the columns are in descending order of component variance. Therefore, we represented in the lines the 60 variables extracted for each one of the sensors and, the 10 Main Components were represented in the columns. Therefore, the first 10 lines will

represent all the features extracted for the first sensor and so on, respecting the order of sensors and features described above. In this way, it is possible to verify which variables most contribute to each of the PCA, and the ones that contribute the most are the darker blue ones. As PCA number one is the one with the most variability of data, we can deduce that the features that most influence this component are the most important.

### *4.4. Clustering*

Through the matrix points extracted from the PCA, a grouping was made to identify the groups that most identify themselves. Each group that identifies itself is seen as an observation of the equipment; in other words, an observation that matches a normal data equipment behavior will be grouped into a cluster that groups observations with that behavior and malfunction characteristics, that will be grouped into another cluster, and so on. The algorithm used to perform Clustering is K-means, therefore, it is necessary to define, a priori, the number of clusters that will be performed. This criterion will be done through a silhouette analysis where the silhouette criterion is used to assess the ideal number of clusters; for these data, it was found that the ideal would be three clusters (Figure 6).

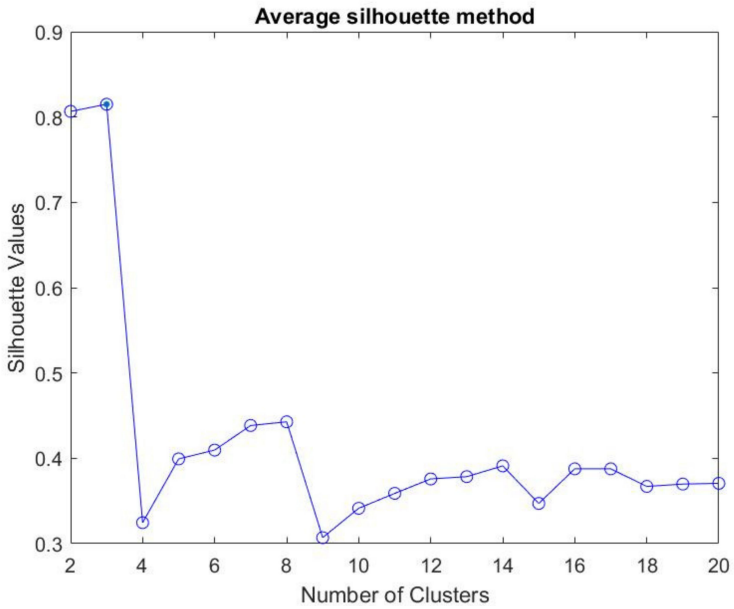



**Figure 6.** Identification of the ideal number of clusters for the sample under study through silhouette analysis.

Each cluster will then be classified as the observable state in the study period. A descending ordering of clusters was also done. In this way, we guarantee that cluster 1 is the one with the most points, cluster 2 is the second cluster with the most points, and so on. Based on this, cluster three corresponds to the observation with the fewest points and is what happens less frequently and, therefore, the rarest to happen, which leads to deduce that an observation occurs when the equipment is not in good condition (this will be demonstrated in the HMM section). In Figure 7, we can see how each cluster looks over time. As each cluster is in an observable state of the equipment, we can look at the graph and see the development of the observable states of the equipment over the study period.

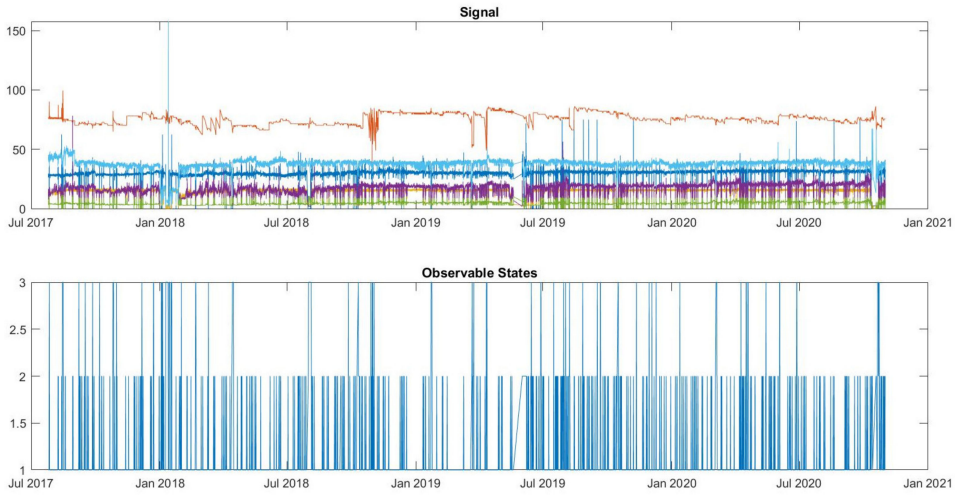



**Figure 7.** Observable states of equipment over time with K-means clustering.

In order to verify if the groups chosen by K-means make sense, we can compare the cluster development with the direct observations of the reading equipment. As we saw above, cluster 3 will be the most unlikely to happen, and it should be represented when there are observations in the less common data, as we can see in the Figure 7, that is what is happening. This leads to the inference that K-means did a good data grouping.

### *4.5. HMM—Equipment Diagnostics*

In the HMM model, there are observable states (defined in the previous section) and hidden states (equipment condition). For now, we have chosen three states to characterize the equipment health status: the 1st representing the good functioning of the equipment, the 2nd identifying an alert state, and the 3rd state representing the equipment malfunction.

The accuracy of data classification used by the HMM is nothing more than the quantity of correctly validated estimates for all types of events or status classes of a system divided by the total monitoring of the condition of a system, after mounting the model [14]. Thus, through the outputs taken from Clustering, the data were divided into training data and test data to validate the model. Thus, through the outputs taken from Clustering, the data were divided into training data and test data to assess the model. According to the common practice in machine learning, 70% of the data were used to train the HMM model and 30% were used to test the performance of the system in unknown data. Considering that the data are actual time series, the first 70% of the samples were used for training and the latter 30% were used for testing.

Then, through the observable training states, based on the Baum–Welch algorithm, we were able to train the HMM to obtain the transition and emission matrices:

Transition matrix:

$$\begin{bmatrix} Stage & State1 & State2 & State3 \\ State1 & 0.9113 & 0.0857 & 0.0031 \\ State2 & 0.5080 & 0.4627 & 0.0294 \\ State3 & 0.0615 & 0.3291 & 0.6094 \end{bmatrix}$$

Emission Matrix:

$$\begin{bmatrix} Stage & Obs1 & Obs2 & Obs3 \\ State1 & 0.9892 & 0.0108 & 0.0000 \\ State2 & 0.3133 & 0.6431 & 0.0436 \\ State3 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$$

Note that the division of data into training data and test data was done in a temporal way, where 70% of training data goes from August 2017 to mid-October 2019, with the remaining months for 30% of test data. In order to find out if this was a good way to divide the data, a comparison was made between the matrices trained by the training data and the matrices trained with the totality of data. Through the Root Mean Square Error metric (Equation (10)), we obtained the following results:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{ij}^{n}\left(a_{Data_{Total}}(i,j) - a_{Data_{Train70\%}}(i,j)\right)^2} \quad (10)$$

$$RMSE_{Transition}=0.0709$$

$$RMSE_{Emission}=0.0289$$

Through this metric, we verified that the division performed in a temporal way was a good choice and, from here, we can also take the model to be trained that does not need all the data.

Having trained the model, through the obtained matrices, it is now possible to generate observable states and to compare them with 30% of the test data. Then, a comparison is made between the sequence of observable states calculated by the model already trained, and the sequence of observable states that actually exit Clustering (test data). Through the test data, it becomes possible to calculate the probability, $y$, of such sequences overlapping [58]. In other words, based on the collected data, it is possible to specify the probability of the estimated states corresponding to the real states of the system (Equation (11)).

$$y = \left(\sum(HMM_{Observations} = Data_{Test})/n\right) * 100 \quad (11)$$

As the HMM works based on probabilities, 10,000 observation sequences were generated with a number equal to the test data, in order to calculate the respective accuracy and, later, to perform the average, in order to obtain the accuracy of the model in question, which was equal to 78.05%.

Having done the training of the model, it is already possible to obtain an HMM scheme for this case study (Figure 8), both with the transition probabilities between hidden states with the probability of emission from hidden states to observable states.

Now, once the parameters of the HMM model are known and embedded in the characteristic matrices, the conditions for using the "Viterbi" algorithm are met, which will allow the characterization of the most likely sequence of hidden states over time, which will give the diagnosing of the equipment's health.

Then, based on Figure 9, it is possible to see the evolution of the failure state of the drying press over time. To better understand the equipment diagnosis (Hidden States), the observable states extracted from the Clustering and the "raw" signal of each sensor is also represented in Figure 9.

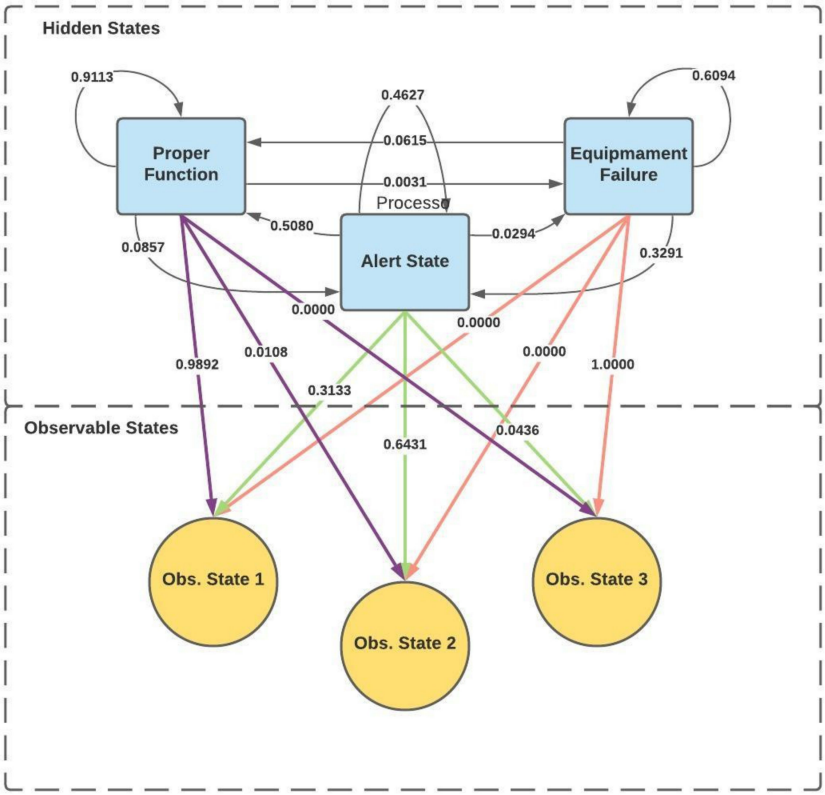



**Figure 8.** HMM scheme for transition and emission.

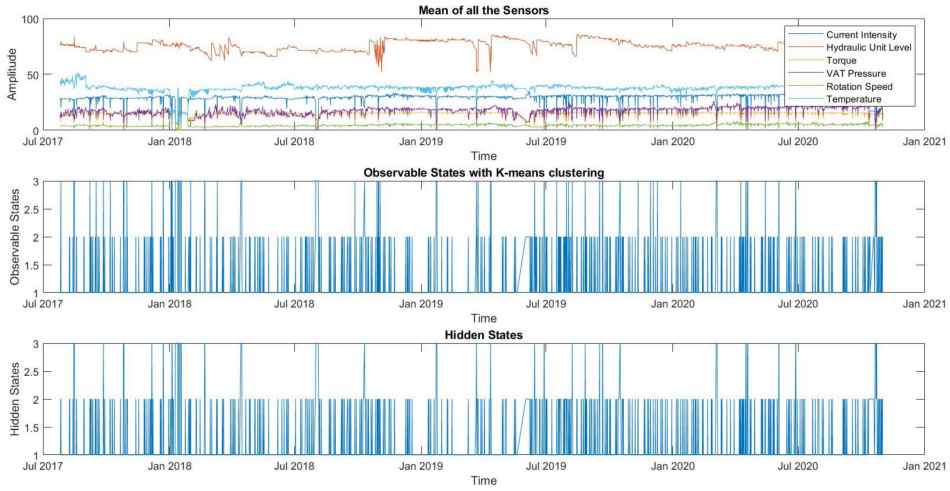



**Figure 9.** HMM Hidden States using the observable K-means Clustering.

## 5. Discussion

Observing the matrices, we evidenced, through the transition matrix, that hidden state 1 has a 91.13% probability of remaining in the same state and that it is unlikely to evolve to another state. Thus, as hidden state 1 represents the equipment in order of "good functioning", the ideal will always be to remain in this state. Hidden states 2 and 3, on the other hand, have some tendency to remain in the same state or to develop to the previous state, which can happen due to maintenance actions. Regarding the emission matrix, we found that the hidden state 1 occurs in a higher percentage when observing the observable state 1 and it never happens when observing the observable state 3. The opposite happens

with the hidden state 3. This makes sense because, as it was described along with this paper, observable state 1 was defined as the observable state that has the highest number of points, and by contrast, observable state 3 occurs the least number of times and, therefore, it is the one that is observed in the highest percentage, when the equipment is defective. Thus, hidden state 3 only occurs when observable state 3 is observed.

Apparently, the equipment had several alert states that return to a good working condition, in principle through (preventive) maintenance. As for the equipment failure status, this happened several times over the three years, which may signify that the drying press maintenance must be improved.

This paper tested several observable state "optimization" tools to insert into the HMM algorithm to perform the equipment diagnosis. As explained above, there is not a priori information about the equipment failure conditions being not possible, through the Working Orders (WO), to verify the effectiveness of the model. However, through graphical analysis, we can see, through Figure 9, that hidden state 3 happens in situations where the direct observations of the sensors deviate from the expected pattern.

Regarding the contribution of this model, it allows an online diagnosis of equipment from sensors, making possible the detection of failures without prior information about the equipment's condition. In addition, it is easy to adapt the number of features, depending on the equipment being sensed. Furthermore, the model can be generalized to any number of sensors and any equipment.

To improve the model in the future, we intend to make a diagnosis and prognosis of the equipment in an operation based on an online algorithm. For that, an Artificial Neural Network class, such as Multi-Layer Perceptron (MLP), may be used to improve the HMM performance. The objective of MLP is to predict which is the next observable state of the three observable states defined to, through the Viterbi algorithm, translating this observable state into a hidden state. Thus, it will be possible to say in advance what the health state of the equipment may be. Furthermore, to retrain the HMM model with the new observable states that may happen, a classification model, such as the Support Vector Machine (SVM), may be used, whose objective is to direct the new observable data read, for the respective cluster. In this way, through the new observable states that are happening, it is possible, through the Baum–Welch algorithm, to go retraining the HMM model. In addition, this study is expected to be extended to the company's five drying presses.

## 6. Conclusions

This paper presented a diagnostic method based on Hidden Markov Models.

The objective of the research described was to present an approach to diagnose failures in drying press equipment in a pulp industry, through HMM. The paper provides us with a theoretical basis for an application of a fault detection method through the optimization of observable states later applied to an HMM model. The added value of this failure detection method is due to the model making its prediction without prior information about the equipment conditions. The method was experimentally validated through a case study. For this, six variables were considered, through the collection of observations from six sensors, constituting a set of data with approximately three years and three months and with observations collected minute by minute.

The method was based on the optimization of observable states inserted in the HMM model, where data preparation is done first, followed by resource generation, where it generated 10 different statistical features, by which the objective was to better characterize the signals and do a reduction of the data. Then, a PCA is made and, finally, a Clustering, where each cluster represents the observable states to be used to train the HMM model. Then, a training and testing procedure was carried out on the HMM model to validate the method's ability to detect a failure.

Through the research carried out, it can be inferred that the method presented, based on HMM, can be used to diagnose the equipment health state supported on CBM strategies.

Based on this, companies have greater support for decision-making about equipment reliability, helping them be more competitive in the market.

Based on the previous case study, further research will be carried out with the aim of extending the time for equipment failure prediction as well as for Online Calibration Monitoring (OML). In this way, it will be possible to allow the extension or elimination of periodic sensor calibration intervals, allowing to improve the safety and reliability of observations through greater precision and reliability of the sensors used.

**Author Contributions:** Conceptualization, J.T.F., I.F., J.R. and A.J.M.C.; methodology, J.T.F. and I.F.; software, A.M. and I.F.; validation, J.T.F., I.F. and J.R.; formal analysis, J.T.F., I.F. and J.R.; investigation, A.M. and I.F.; resources, J.T.F., I.F., J.R. and A.J.M.C.; writing—original draft preparation, A.M.; writing—review and editing, J.T.F., I.F. and J.R.; project administration, J.T.F. and A.J.M.C.; funding acquisition, J.T.F. and A.J.M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowvska-Curie grant agreement 871284 project SSHARE and the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under Project POCI-01-0145-FEDER-029494, and by National Funds through the FCT—Portuguese Foundation for Science and Technology, under Projects PTDC/EEI-EEE/29494/2017, UIDB/04131/2020, and UIDP/04131/2020.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Pais, E.; Farinha, J.T.; Cardoso, A.J.M.; Raposo, H. Optimizing the Life Cycle of Physical Assets—A Review. *WSEAS Trans. Syst. Control* **2020**, *15*, 2224–2856.
2. Rodrigues, J.; Costa, I.; Farinha, J.; Mendes, M.; Margalho, L. Predicting Motor Oil Condition Using Artificial Neural Networks and Principal Components Analysis. *Eksploatacja i Niezawodnosc Maint. Reliab.* **2020**, *22*, 440–448. [CrossRef]
3. Mateus, B.; Torres, J.; António, F.; Cardoso, M. Production Optimization versus Asset Availability—A Review. *WSEAS Trans. Syst. Control* **2020**, *15*, 320–332.
4. Yu, J. Adaptive hidden Markov model-based online learning framework for bearing faulty detection and performance degradation monitoring. *Mech. Syst. Signal Process.* **2017**, *83*, 149–162. [CrossRef]
5. Martins, A.B.; Farinha, J.T.; Cardoso, A.M. Calibration and Certification of Industrial Sensors—A Global Review. *WSEAS Trans. Syst. Control* **2020**, *15*, 394–416. [CrossRef]
6. Taylor, P.; Baruah, P.; Chinnam, R.B. HMMs for diagnostics and prognostics in machining processes. *Int. J. Prod. Res.* **2005**, *43*, 1275–1293.
7. Bunks, C.; McCarthy, D.; Al-Ani, T. Condition-based maintenance of machines using hidden markov models. *Mech. Syst. Signal Process.* **2000**, *14*, 597–612. [CrossRef]
8. Kamlu, S.; Laxmi, V. Condition-based maintenance strategy for vehicles using hidden Markov models. *Adv. Mech. Eng.* **2019**, *11*, 1–13. [CrossRef]
9. Camci, F.; Chinnam, R.B. Health-State Estimation and Prognostics in Machining Processes. *IEEE Trans. Autom. Sci. Eng.* **2010**, *7*, 581–597. [CrossRef]
10. Arpaia, P.; Cesaro, U.; Chadli, M.; Coppier, H.; De Vito, L.; Esposito, A.; Gargiulo, F.; Pezzetti, M. Fault detection on fluid machinery using Hidden Markov Models. *Measurement* **2020**, *151*, 107126. [CrossRef]
11. Bjerkeseth, M. Using Hidden Markov Models for Fault Diagnostics and Prognostics in Condition Based Maintenance Systems. Master's Thesis, University of Agder, Kristiansand, Norway, Grimstad, Norway, 2010.
12. Ocak, H.; Lopar, K. A new bearing fault detection and diagnosis scheme based on hidden markov modeling of vibration signals. In Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, Salt Lake City, UT, USA, 7–11 May 2001; Volume 5, pp. 3141–3144.
13. Xinmin, T.; Baoxiang, D.; Yong, X. Bearings Fault Diagnosis B ased on HMM and Fractal Dimensions Spectrum. In Proceedings of the 2007 International Conference on Mechatronics and Automation, Harbin, China, 5–8 August 2007; pp. 1671–1676.
14. Simões, A.; Viegas, M.; Farinha, T.; Fonseca, I. The State of the Art of Hidden Markov Models. *Qual. Reliab. Eng.* **2017**, *33*.
15. Wang, F.; Tan, S.; Yang, Y.; Shi, H. Hidden Markov Model-based Fault Detection Approach for Multimode Process. *Ind. Eng. Chem. Res.* **2016**, *55*, 1–35. [CrossRef]
16. Yin, X.; Zhou, Z.; Li, N.; Chen, S. An Approach for Data Filtering Based on Rough Set Theory. *Lect. Notes Comput. Sci.* **2001**, *2118*, 367–374.

17. Prasad, G.S.; Reddy, N.V.S.; Acharya, U.D. Knowledge Discovery from Web Usage Data: A Survey of Web Usage Pre-processing Techniques, in Information Processing and Management, BAIP 2010. In *Communications in Computer and Information Science*; Springer: Berlin/Heidelberg, Germany, 2010; Volume 10, pp. 505–507.
18. Banhatti, A.G.; Deka, P.C. *Effects of Data Pre-Processing on the Prediction Accuracy of Arti fi cial Neural Network Model in Hydrological Time Series*; Springer International Publishing: Switzerland, 2016; ISBN 9783319401959.
19. Cho, W.; Choi, E. Big data pre-processing methods with vehicle driving data using MapReduce techniques. *J. Supercomput.* **2017**, *73*, 3179–3198. [CrossRef]
20. Rinnan, A.; Norgaard, L.; Berg, F.; Thygesen, J.; Bro, R.; Engelsen, B. Data Pre-Processing. In *Infrared Spectroscopy for Food Quality Analysis and Control*; Academic Press: Cambridge, MA, USA, 2009; ISBN 9780123741363.
21. Abdallah, Z.S.; Webb, G.I. Data Preparation. In *Encyclopedia of Machine Learning and Data Mining*; Springer: Boston, MA, USA, 2017.
22. Zhang, S.; Zhang, C.; Yang, Q. Data preparation for data mining. *Appl. Artif. Intell.* **2003**, *17*, 375–381. [CrossRef]
23. Nisbet, R.; Miner, G.; Yale, K. Data Understanding and Preparation. In *Handbook of Statistical Analysis and Data Mining Applications*; Academic Press: Toronto, ON, Canada, 2018; pp. 55–82. ISBN 9780124166325.
24. Ganti, V.; Sarma, A. Data Cleaning Cleaning: A Pratical Perspective. Ozsu, M., Ed.; Morgan & Claypool: San Rafael, CA, USA, 2013; ISBN 9781608456789.
25. Do, H.H.; Broder, A.; Dean, J.; Henzinger, M.R.; Deshmukh, K.; Sarawagi, S. Bulletin of Technical Committee on Data Engineering. In *Special Issue on Data Cleaning*; Lomet, D., Ed.; IEEE Computer Society: Washington, DC, USA, 2000; Volume 23.
26. Cogley, T. Data Filters. In *The New Palgrave Dictionary of Economics*; Palgrave Macmillan: London, UK, 2008.
27. Roy, S.; Sharma, P.; Nath, K.; Bhattacharyya, D.; Kalita, J. Pre-Processing: A Data Preparation Step. *Encycl. Bioinforma. Comput. Biol.* **2018**, *463*, 1–9.
28. Han, Z.; He, Q.; Davier, M. Von Predictive Feature Generation and Selection Using Process Data From PISA Interactive Problem-Solving Items: An Application of Random Forests. *Front. Psychol.* **2019**, *10*, 1–15. [CrossRef]
29. De Silva, A.M.; Leong, P.H.W. *Grammar-Based Feature Generation for Time-Series Prediction*; Janusz, K., Ed.; Springer: London, UK, 2015; ISBN 9789812874108.
30. Theodoridis, S.; Koutroumbas, K. Feature Generation I: Data Transformation and Dimensionality Reduction. In *Pattern Recognition*; Springer: Cham, Switzerland, 2009; pp. 323–409.
31. Sakr, S.; Zomaya, A. *Encyclopedia of Big Data Technologies*; Sakr, S., Zomaya, A., Eds.; Springer: Cham, Switzerland, 2019; ISBN 9783319775241.
32. Pearson, K. On lines and planes of closest fit to systems of points in space. *Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [CrossRef]
33. Hotelling, H. Relations Between Two Sets of Variates. *Biometrika* **1936**, *28*, 321–377. [CrossRef]
34. Anowar, F.; Sadaoui, S.; Selim, B. Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Comput. Sci. Rev.* **2021**, *40*, 100378. [CrossRef]
35. Warmenhoven, J.; Bargary, N.; Liebl, D.; Harrison, A.; Gunning, E.; Hooker, G. PCA of Waveforms and Functional PCA: A Primer for Biomechanics. *J. Biomech.* **2020**, *116*, 110106. [CrossRef] [PubMed]
36. Kuhn, M.; Johnson, K. Data Pre-Processing. In *Applied Predictive Modeling*; Springer Science&Business Media: New York, NY, USA, 2013; ISBN 9781461468493.
37. Nidheesh, N. A Hierarchical Clustering algorithm based on Silhouette Index for cancer subtype discovery from genomic data. *Neural Comput. Appl.* **2019**, *32*, 11459–11476. [CrossRef]
38. Townes, F.W.; Hicks, S.C.; Aryee, M.J.; Irizarry, R.A. Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model. *Genome Biol.* **2019**, *20*, 1–16. [CrossRef] [PubMed]
39. Alguliyev, R.M.; Aliguliyev, R.M.; Sukhostat, L.V. Parallel batch k-means for Big data clustering. *Comput. Ind. Eng.* **2021**, *152*, 107023.
40. Wendler, T.; Gröttrup, S. Cluster Analysis. In *Data Mining with SPSS Modeler: Theory, Exercises and Solutions*; Springer International: Switzerland, 2016; pp. 587–712. ISBN 9783319287096.
41. Haoran, L.; Fazhi, H.; Yilin, C. Learning dynamic simultaneous clustering and classification via automatic differential evolution and firework algorithm. *Appl. Soft Comput. J.* **2020**, *96*, 106593.
42. Mary, S.; Selvi, T. A Study of K-Means and Cure Clustering Algorithms. *Int. J. Eng. Res. Technol.* **2014**, *3*, 1985–1987.
43. Subbalakshmi, C.; Krishna, G.R.; Mohan, S.K.; Rao, P.V. A Method to Find Optimum Number of Clusters Based on Fuzzy Silhouette on Dynamic Data Set. *Procedia Comput. Sci.* **2015**, *46*, 346–353. [CrossRef]
44. Llet, R.; Ortiz, M.C.; Sarabia, L.A.; Sánchez, M.S. Selecting variables for k -means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Anal. Chim. Acta* **2004**, *515*, 87–100. [CrossRef]
45. Menardi, G. Density-based Silhouette diagnostics for clustering methods. *Stat. Comput.* **2011**, *21*, 295–308. [CrossRef]
46. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
47. Peralta, D.; Saeys, Y. Robust unsupervised dimensionality reduction based on feature clustering for single-cell imaging data. *Appl. Soft Comput. J.* **2020**, *93*, 106421. [CrossRef]
48. Baum, L.E. An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes. *Inequalities* **1972**, *3*, 1–8.

49. Baum, L.; Petrie, T. Statistical Inference Probabilistic Function Of Finite State Markov Chains. *Ann. Math. Stat.* **1966**, *37*, 1554–1563. [CrossRef]
50. Rabiner, L. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE* **1989**, *77*, 257–286. [CrossRef]
51. Joshi, J. Prediction of weather states using Hidden Markov model. In Proceedings of the International Conference on Cryosphere and Climate Change, Hainan, China, 10–12 November 2012; 2012; pp. 1–7.
52. Saucedo-Dorantes, J.J.; Jaen-cuellar, A.Y.; Delgado-prieto, M.; Romero-troncoso, R.D.J.; Osornio-Rios, R.A. Condition monitoring strategy based on an optimized selection of high-dimensional set of hybrid features to diagnose and detect multiple and combined faults in an induction motor. *Measurement* **2021**, *178*, 109404. [CrossRef]
53. Merola, G.; Abraham, B. Dimension Reduction Methods Used in Industry. In *Handbook of Statistics*; Springer: Cham, Switzerland, 2003; Volume 22, pp. 995–1039.
54. Van Den Broeck, J.; Cunningham, S.; Herbst, A. Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities. *PLoS Med.* **2005**, *2*, e267.
55. Dasu, T.; Johnson, T. Data Quality: Techniques and Algorithms. In *Exploratory Data Mining and Data Cleaning*; John Wiley & Sons: Hoboken, NJ, USA, 2003; p. 224.
56. ROADIDEA. Data Filtering Methods. In *7th Community Framework Programme for Research and Technological Development*; Technical Report; European Commission: Brussels, Belgium, 2008; pp. 1–42.
57. Dhamande, L.S.; Chaudhari, M.B. Compound gear-bearing fault feature extraction using statistical features based on time-frequency method. *Measurement* **2018**, *125*, 63–77. [CrossRef]
58. Simões, A. *Manutenção Condicionada Às Emissões Poluentes Em Autocarros Urbanos*; Instituto Superior Técnico: Lisboa, Portugal, 2011.

# Appendix C

# Prediction Maintenance based on Vibration Analysis and Deep Learning – A case study of a drying press supported on a Hidden Markov Model

Alexandre Martins[a,b], Inácio Fonseca[c], José Torres Farinha[c,d], João Reis[a,e] and António J. Marques Cardoso[b]

[a]*Industrial Engineering and Management, Faculty of Engineering, Lusófona University and EIGeS, Campo Grande, Lisbon, 1749-024, Portugal*

[b]*CISE – Electromechatronic Systems Research Centre Portugal, Univ. Beira Interior, Covilhã, Portugal*

[c]*ISEC/IPC - Polytechnic Institute of Coimbra, R. Pedro Nunes, Coimbra, 3030-199, Portugal*

[d]*CEMMPRE - Centre for Mechanical Engineering, Materials and Processes, University of Coimbra, Coimbra, Portugal*

[e]*Department of Economics, Management, Industrial Engineering and Tourism, GOVCOPP, Campus Universitário de Santiago, Aveiro University, Aveiro, 3810-193, Portugal*

ARTICLE INFO

*Keywords*:
Maintenance
Industrial Sensors
Hidden Markov Models
Deep Neural Networks
Big Data
Clustering
Principal Components Analysis

ABSTRACT

The main objective of this paper is to describe a methodology that was developed to support maintenance decision-making methods based on equipment condition. Condition-Based Maintenance allows to increase equipment availability and maximize investments. This is mainly due to the prevention of unexpected equipment downtime. By avoiding turning on/off industrial equipment, production flows are more efficient, allowing manufacturers to improve the quality of the end-product. The industry aims more and more to correspond satisfying customer expectations. We argue that the methodology developed in this paper adds value to the existing literature, namely because the fact that it is possible to anticipate the state of an equipment without a large amount of support data. In other words, although one could find information gaps regarding the occurrence of failures, it was possible to accurately assess the state of the equipment. This approach is robust, as it can be used in distinct equipment with different sensors, making this methodology generalizable for Condition-Based Maintenance. The paper presents the validation of the preceding through a case study on drying presses in the paper industry. To do so, three states were adopted, namely: "Proper function"; "Alert state"; and "Equipment failure". The methodology follows a series of steps, going through the collection of values from vibration sensors, imputation of values using Deep Artificial Neural Networks through on-line sensors, until reaching the last stage of classification carried out by the Hidden Markov Model. Through optimized observations from the previous steps, it was possible to define the hidden states through the Viterbi algorithm, which corresponds to the health states of the equipment. Additionally, it was

possible to demonstrate that the proposed methodology can accurately characterize the condition states of the equipment based on the data obtained and can be generalized to other types of equipment.

## 1. Introduction

### 1.1. Framework

Implementing a maintenance system in a company is an integral part of asset management [1]. If successfully implemented, a maintenance system is supposed to significantly reduce production costs, mainly due to the prevention of unexpected equipment downtime and, consequently, the maximization of its availability. "These types of cost associated with an asset's life cycle must be seen as investments, because they originate a return that is indexed to the quality of the investment - initial and during the asset's life - and accrue value to the asset throughout its life cycle" [2]. Avoiding unnecessary switching on/off of equipment, industrial engineers can expect better end-product quality and customer satisfaction. This is due not only to better product quality, but also to reach delivery deadlines. The above actions aim to reduce the company's costs and maximize profits. In that regard, maintenance today is no longer seen as an additional cost, but as an investment. There are numerous methods of maintenance that can be used depending on the equipment's criticality. When referring to critical equipment, it is vital to implement Condition-Based Maintenance (CBM) to maintain ongoing "surveillance" to ensure production takt time. Predictive condition management, according to Rodrigues et al.[3], is one of the most essential ways to maximizing equipment availability. CBM aids in the support of maintenance decisions based on data gathered through condition monitoring, with three primary steps: data gathering, data processing, and maintenance decision making [4]. To do so, it is necessary to have sensors responsible for directly connecting the physical phenomenon to be studied in the equipment and the data acquisition system. Sensors will convert mechanical quantities into electrical signals [5]. Martins et al.[5] conducted previous study that gave some light on the theory relating to the trustworthiness of data emitted by sensors, as well as its usefulness in supporting the CBM, as well as the need to develop understanding in this area.

Thus, one of the most used and reliable physical quantities to take information about the health status of an equipment component is the vibration signal, especially when we refer to rotating components [6], which is the case of our study. According to Pawlik et al.[7], monitoring systems that use vibration signals are often used in industry to diagnose

alex_bmts@hotmail.com (A. Martins); inaciosaf@gmail.com (I. Fonseca); torresfarinha@gmail.com (J.T. Farinha); p40500@ulusofona.pt (J. Reis); ajmc@ubi.pt (A.J.M. Cardoso)

ORCID(s): 0000-0001-9947-0894 (A. Martins); 0000-0003-0167-7489 (I. Fonseca); 0000-0002-9694-8079 (J.T. Farinha); 0000-0002-8504-0065 (J. Reis); 0000-0001-8737-6999 (A.J.M. Cardoso)

the equipment condition. Its success can be ascribed in large part to the availability of signal processing techniques in the time domain, frequency domain, and time-frequency domain, which can be used to distinguish engineering aspects (feature extraction, selection, and manipulation) [8]), which is a vital point for the model used in this paper. Among several existing papers dealing with vibration analysis applied to CBM, we can highlight the followings: Saucedo-Dorantes et al.[9] suggested a condition monitoring methodology based on the estimation and optimization of a collection of high-dimensional hybrid characteristics to identify and evaluate the presence of multiple and combined faults in an induction motor. For this the author used for this analysis the stator vibrations and currents using techniques from different domains; Dhamande and Chaudhari[10] use vibration analysis in different domains to explore more complex situations of compound bearing failure; Pranesh et al.[11] use vibration signals, acquired by a piezoelectric accelerometer, to diagnose a hydraulic brake failure, using machine learning. Akpudo and Hur[8] developed a new hybrid feature classifier for industrial physical Assets' vibration-based Fault Detection and Isolation (FDI).

Since the unexpected downtime of equipment has a significant contribution to production and costs, it becomes necessary to develop techniques to improve maintenance technologies based on condition with Artificial Intelligence (AI) algorithms to help the decision-making process concerning the real-time state. AI and complex algorithms are critical for gathering and analyzing data in the physical world, namely in machines. The knowledge derived from the analysis of huge amounts of data is critical for supporting the decision-making process, with the goal of enabling machines with the ability to recognize potential problems on their own [4]. As a result, the goal of this research is to propose a classification methodology for the health state of a roll in a paper drying press utilizing vibration signals using the HMM algorithm. One of the barriers of the case study was the low amount of data collected, which can impair the model's prediction. As Rodrigues et al.[3] refer, the success of the HMM implementation depends on the monitoring frequency. To overcome this obstacle, Neural Networks (NN) are used to perform imputation of vibration values. NNs can acquire, store, and use all the knowledge gained in the past to extract patterns and detect trends unnoticed by humans. NN can be broadly categorized into Artificial NN (ANN) and Deep NN (DNN), used in this paper. Neural Networks can be compared to the human nervous system. They, too, collect information from their surroundings through a learning process, and this information is stored in synaptic weights, also known as interneuron connection intensities [12, 13]. So, NN correspond to algorithms that are based on biological nervous systems, simulating the way the human brain works, constituted by computational "neurons", which operate in parallel to describe nonlinear and interaction effects [12, 14, 15, 16, 17, 18, 19]. The basic functions of all neural networks are receiving external data, deciding whether these will be activated and considered or discarded as negligible, analyzing the error minimization by iterating the data, thus obtaining an output or performance for the whole trial [15]. The relationship between input and output is established by adding the weights assigned to neurons and changing the sum of polarization terms to obtain the output,

which is determined by the strength of the connection by which the input represents the output; the process of establishing the relationship being input output, is called training. To eliminate errors between projected and desired outputs, the network adjusts its weights [12]. An ANN may have one or more hidden layers between the input and the output, although ANNs with multiple hidden layers is known as Deep Neural Network (DNN), thus exhibiting higher levels of feature extraction capability relative to each hidden layer added [16, 17]. Thus, deep learning is an area derived from artificial intelligence that stems from the study of neural networks, which can be defined as one with several hidden layers. Unlike superficial artificial neural networks, they can directly use the original data as input and learn characteristics, layer by layer, thus obtaining a more efficient feature extraction [18, 20]. The term "deep" refers to the more complicated layer and node structure, which considerably increases the amount of weights and bias terms, resulting in a more abstract, high-dimensional feature mapping from the input and output data [19]. DNN is made up of a collection of neurons structured in a series of many layers, with each neuron receiving the activations of the neurons in the preceding layers as input. The network's neurons work together to execute a complicated nonlinear mapping from input to output that is learned from data by adjusting the weights of each neuron using an error backpropagation technique [21].

After imputing values through a DNN algorithm, these data will go through a variety of steps to classify the diagnostic status through the HMM. Further, it will be necessary to carry out a characteristic generation phase in two vibration domains: time domain and frequency domain, where static characteristics that best describe the signal over time will be removed. After the features have been generated, Principal Component Analysis will be used to reduce the number of dimensions (PCA). Through the PCA outputs, Clustering is used to group the most similar data in a respective cluster and thus characterize the observable states that will be input to the HMM. Then, through the HMM, the equipment's health status diagnosis will be carried out. HMM has already proven, as demonstrated in several papers, that it can be used for CBM, as can be verified in the paper of [22].

### 1.2. Objectives and Contributions of the paper

This paper aims to present a methodology for classifying the health status of a drying press used in the paper industry. The proposed methodology in this paper innovates on the following aspects:

- in the classifying of the status of equipment without having previous information or technical knowledge about it - in other words, it is a methodology that, through various Machine Learning (ML) tools, it is able to detect patterns that are outside the regular operational operation of the equipment - in this way, it is possible to detect and predict anomalies in the equipment;
- based on Deep Artificial Intelligence, it was also added the possibility to manage time series data to better diagnose the equipment condition;

- the developed methodology is generic, then it does not require technical knowledge or prior information about the equipment, and can be used online to provide real-time input and support a condition predictive maintenance strategy;
- additionally, the developed algorithm is generic and can be used in different equipment with different sensors.

### 1.3. Paper structure

The present paper is structured as follows: an introduction (Section 1) where the leading research topics are introduced (condition-based maintenance; vibration signals for maintenance; application of Neural Networks; health diagnostic classification through HMM and all the inherent methodology). Section 2 provides a brief explanation of each method used in the study methodology. Section 3 presents the methodology used on the work, aiming to explain the procedure from the imputation of values in the vibration data through DNN as well as the HMM classification methodology. Section 4 presents the case study, which consists on carrying out a health diagnosis of a roll of a drying press in the paper industry. Section 5 comprises a discussion of results where an evaluation of the methodology implemented is made, and the results are unveiled. Finally, in section 6, a conclusion of the entire work is made.

## 2. Background

### 2.1. Multi-layer perceptron (MLP)

The Multi-Layer Perceptron (MLP) is a supervised learning method that works with non-linear problems thanks to its multi-layer construction. It is one of the most useful feedforward networks [23, 24]. Being a DNN, MLP is divided into three levels: input layer; hidden layer; and output layer, having the same structure as a single-layer neural network but with more hidden layers [15]. Each neuron is fully connected, and each neuron has weights that are used to calculate whether the input data has the information we need by adjusting the weights between elements [14, 23, 25]. So, MLP has an input layer of neurons that act as receptors, one or more hidden layers of neurons that compute the data and go through iterations, and the output layer that predicts the output [26]. Each hidden layer is composed by several artificial neurons, being the output of an artificial neuron obtained through the Equation 1, [17, 19, 27, 28]:

$$f\left(\sum_{i=1}^{\mathrm{p}}(W_i X_i + b)\right) \tag{1}$$

Where:

- $f$ is the activation function (e.g., sigmoidal function, exponential linear unit, hyperbolic tangent function, or rectified linear unit);

- $p$ is the number of inputs;
- $W$ is the input weight;
- $X$ it's the input;
- $b$ represents the polarization value.

During the training phase, weights and biases are calibrated using accessible input-output data sets. During training, two operations occur: activation propagation (feedforward), sending the information forward, and error propagation (backpropagation) [13]. During the first step, each neuron generates an output signal based on Equation 1 that is compared to the target, and the mistakes are propagated back through the network during the second stage. The weights and deviations are adjusted based on an optimizing algorithm selected to minimize the predefined loss function [13]. The same author also explains that several iterations are necessary during the training process until a convergence criterion is satisfied [13]. Hyperparameters relating to model structure, such as the number of inputs, layers, number of neurons per layer, activation function, and so on, must be carefully selected and changed, since they have a significant impact on training efficiency and performance. "(...) a good choice of hyperparameters is usually necessary to make them work well on real-world problems, and tricks are often used to make the most efficient use of these methods and to extend their capabilities" [21]. In Fig. 1, we can see a schematic of an MLP.

### 2.2. Features Generation (FG)

With compact and useful representations, Feature Generation (FG) is used to maximize the exploitation of the information created from the data [22]. In this approach, FG may be viewed as one of the methods for improving the input (observations) and output (hypothesis) of a Machine Learning model in order to maximize its architecture performance [29]. "If the transformation is chosen correctly, the transformation features can display high-information packing properties compared to the original input samples" [22]. To better understand the degradation process of a component using vibration signals, feautures must be generated from the raw vibration signal [30], which can build multiple resources in both the time and frequency domains [11]. Frequency analysis has become a fundamental tool for processing vibration signals [31]. To perform FG in the frequency domain, the Fast Fourier Transform (FFT) is usually used, which, as explained by Cheng et al.[30], is one of the most commonly utilized techniques in the analysis of the frequency domain of the signal and is responsible for converting the monitoring signal into a frequency spectrum.

### 2.3. Principal Components Analysis (PCA)

Principal Components Analysis (PCA) is used to project feature vectors in orthogonal space for robustness and computing efficiency [32]. That is, it is responsible for making a dimensionality reduction to replace a high-dimensional dataset (resource-heavy), with a much smaller dimensional dataset with just a few Principal Components (PC) [22]. In

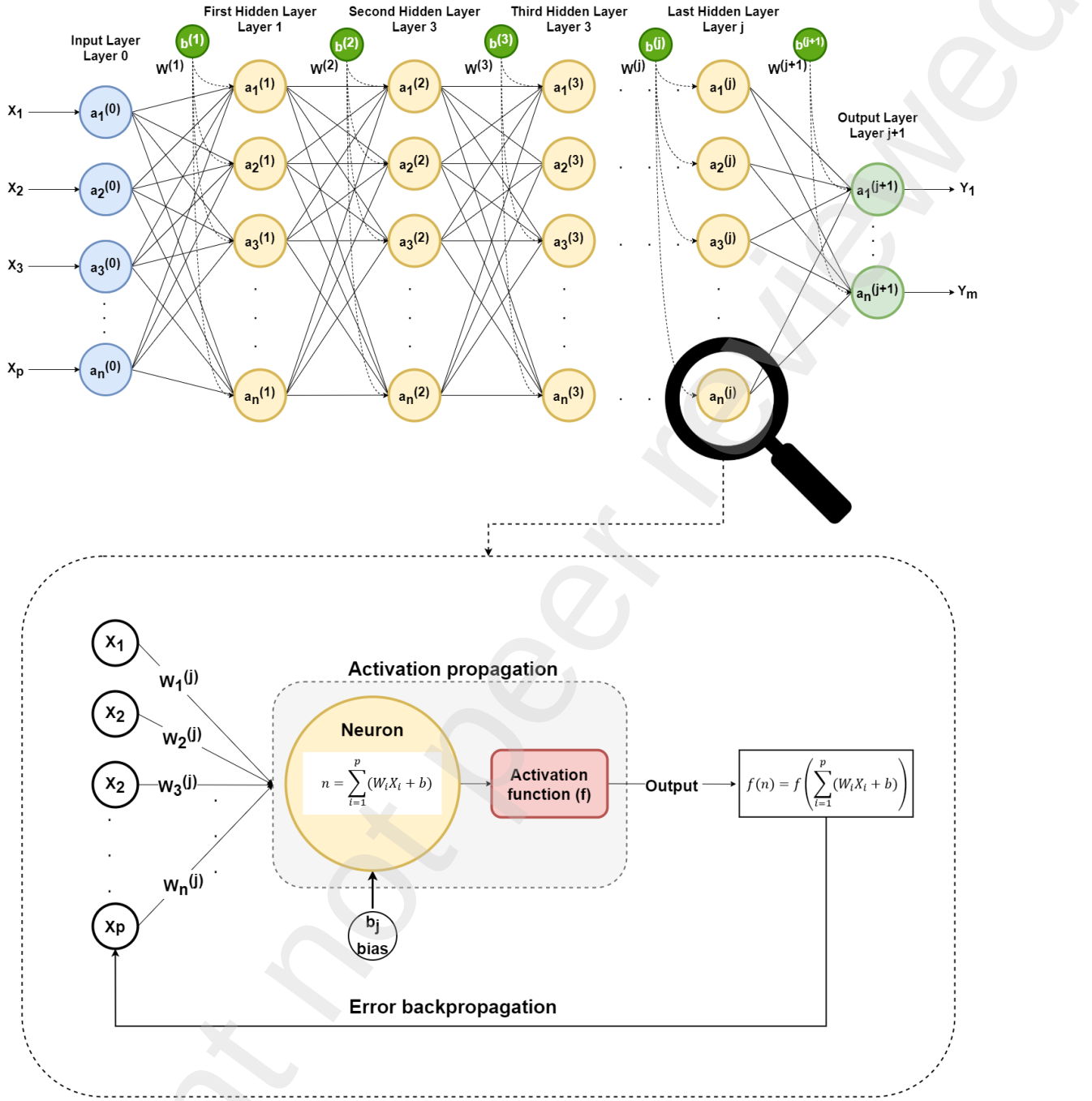



**Figure 1:** Schematic of the functioning of a Multi-Layer Perceptron Neural Network

an unsupervised linear transformation procedure, an orthogonal transformation is used to turn a set of observations of possibly correlated variables into a set of linear values of uncorrelated variables known as Principal Components [33], [34]. The PCA method consists of finding linear combinations of predictors: the PCs capture the most significant variability in the data. Extracted PCs refer to variables that can explain the maximum variation (variance) in all data. The first Principal Component (PC 1) is removed to have the most significant variability of all possible linear combinations data, being of

utmost importance. The second Principal Component (PC 2) is extracted in an orthogonal connection to the existing PC to reflect data variations that PC1 was unable to fully explain, and so on [17]. According to the author, the minimum number of PCs is based on 85% or more of the cumulative contribution rate.

As stated in Table 1, the PCA method is implemented in five steps [17, 22].

**Table 1**
Implementation sequence of the model - Principal Component Analysis

| |
|---|
| Input: $X \in R^{(n*d)}$ |
| 1 – Calculate the covariance of data sets $(X.X^T)$; |
| 2 - Calculate eigenvectors and eigenvalues applying Linear Eigen decomposition $X.(X^T)$; |
| 3 – Sort Eigen Values in decreasing order to sort Eigen Vectors; |
| 4 – Build a Matrix $W(D * k)$ with KTop Eigen Vectors; |
| 5 – Transform $X$ using $W$ to obtain the new subspace |
| $Y = X.W$; |
| Output: $X \in R^{(n*k)}$ |

### 2.4. Clustering – K-means

Clustering is an unsupervised data mining technique used in a wide range of data analysis areas where multivariate statistical methods allow for cluster analysis. The goal is to identify clusters of objects within the data while maximizing homogeneity within each cluster and heterogeneity between different clusters [22]. In short, Clustering is a procedure that aims to explore data relationships and patterns, grouping objects according to their similarity; this is, similar objects will belong to the same cluster [35, 36].

The k-means algorithm will be used as a Clustering approach in this paper. It's an iterative and evolutionary method for finding cluster positions that minimize the distance between data points and clusters, and then partitioning the n observations into clusters with the closest mean [37]. That is, it is a clustering algorithm used to group data into K clusters, using the averaging for this [38], usually comparing the data using the Euclidean distance (Equation 2) [39].

$$d_{euclidean}(p1, p2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{2}$$

where:

- $p1(x_1, y_1)$ and $p2(x_2, y_2)$ are the two points in Euclidean space;
- $x$, $y$ are the Euclidean vectors from the origin of space.

The member and centroid objects define each cluster in the division. The centroid is the point at which the sum of all the distances between the objects in this cluster is the

smallest [37]. Thus, the clustering process consists of the initial definition of the centroids and the assignment of events (attribute vectors) that are inserted as input to the cluster with the nearest centroid [35]. As objects are inserted and assigned to clusters, centroids are updated until they have little or no changes, reaching the convergence condition [35, 40]. The ideal number of clusters is a fundamental problem in clustering partitioning in the K-means method, and it is necessary to know in advance to obtain a good result [38]. To choose the value of k clusters that best fits the data under study, and since cluster analysis is not part of supervised learning and is performed as part of exploratory data analysis, there are several model evaluation metrics. Talei et al.[39] claim that the best results are obtained when the inter-cluster distance is decreased, and the intra-cluster distance is maximized.

### 2.5. Hidden Markov Models (HMM)

HMM is a stochastic process with two levels of uncertainty (doubly stochastic process), being an extension of the Markov Chains, producing inferential statistical information about a sequence of states. These states are finite and are considered hidden, where for each hidden state, an observation is related [22, 41]. To put it another way, HMM is a stochastic technique for modeling signals that develop through a finite number of hidden states and produce observations. The state of the model can only be discerned through the observation sequence and cannot be examined directly [32]. An HMM model has a set of parameters, $\lambda = (A, B, \pi)$, where:

- $A = \{a_{i,j}\}, 1 \leq i, j \leq N$ - represents probability of transition from one hidden state to another;
  - The transition distribution $a_{i,j}$ is defined as:

$$a_{i,j} = P(S_{t+1} = q_j \mid S_t = q_i), 1 \leq i, j \leq N \tag{3}$$

- $B = \{b_j(k)\}, (1 \leq j \leq N)$ - defines the probabilities with which the different classes of emissions occur for each of the hidden states $j$, with $j = 1, 2, 3, \ldots, N$, where observations are emitted from each state according to the probability distribution:

$$b_j(k) = P(O_t = \nu_k \mid S_t = q_j), 1 \leq k \leq M \tag{4}$$

- $\pi$ - represents the initial probability of each of the hidden states:

$$\pi = \pi(S_t = q_i), 1 \leq i \leq N \tag{5}$$

where:

- $N$ - represents the number of hidden states, where the individual hidden states are represented as $S$ and the state at time $t$ as $q_t$. Corresponding to the status sequence, it is displayed as, $S_T = S_1, S_2, S_3, \ldots, S_t$ with $S_t \in Q = \{q_1, q_2, q_3, \ldots, q_N\}$

- $O$ – is the vector of observations, where $M$ is the number of observations per state and $T$ is the length of the observation sequence. The sequence of observations up to $T$ is presented as, $O = \{O_1, O_2, O_3, \ldots, O_T\}$ with $O_t \in \{\nu_1, \nu_2, \nu_3, \ldots, \nu_M\}$

In short, the variable $S_t(S_t \in Q)$ indicates the hidden state, and $O_t(O_t \in \nu)$ is the observation at time t. The model parameters, $\lambda = (A, B, \pi)$, specify three components: hidden state transition probability matrix $A$; observation probability matrix $B$; initial state distribution $\pi$.

To maximize the train of HMM model, we must identify the parameters that maximize the probability of the observation sequence. we must find $\lambda = (A, B, \pi)$ that maximize $P(O|\lambda)$. The Baum-Welch algorithm is used for this. The Baum-Welch algorithm makes use of the forward variable $\alpha_T(i)$ and the regressive variable $\beta_t(i)$, which correspond respectively to the forward and backward algorithms, where we can define the forward variable $\alpha_T(i)$ as: $\alpha_T(i) = P(O_1, O_2, O_3, \ldots, O_T, q_T = S_i|\lambda)$. This variable will contain the probability of the observation sequence, $O_1, O_2, O_3, \ldots, O_T$, as well as the model's state $S_i$ at a particular time $t$, given the model $\lambda$. The backward variable $\beta_t(i)$ is $\beta_t(i) = P(O_{t+1}, O_{t+2}, O_{t+3}, \ldots, O_T|S_t = q_i, \lambda)$. Given the state $S_i$ at time $t$ and the model $\lambda$, this variable yields the final probability of the observation sequence of $t + 1$. Because of this, The Forward-Backward algorithm is another name for the Baum-Welch algorithm.

The Viterbi method is used to discover the most likely sequence of hidden states $S = \{S_1, S_2, S_3, \ldots, S_T\}$, given the series of observed emissions $O = \{O_1, O_2, O_3, \ldots, O_T\}$ and the model $\lambda$. The algorithm corresponds to finding $S^* = \arg\max_S P(S \mid O, \lambda)$.

The entire theory of the HMM model can be seen in more detail in papers [42, 43, 44].

## 3. Methodology

The objective of this work is, through observations collected by vibration sensors, to characterize the operating state of a component of production equipment, namely a compression roller of a drying press. Thus, the methodology adopted for the classification and health diagnosis of the component is composed of three phases (Figure 2), being each one composed by different methods, which complement to each other's.

*Phase 1*

After collecting and preparing observations made by several sensors, we started by carrying out phase 1, whose objective is to create an imputation of temporal vibration observations. This phase is necessary since the vibration observations were collected in a period superior to those of other sensors coupled to the same equipment. In this way, through multivariate analysis performed through deep neural networks (Figure 3), it is possible to make an imputation of vibration data filling the missing temporal space, converting it into a vector with the same number of temporal observations as the sensors.

We first start by resampling the multivariate observations to get a collection size equivalent to the collections of vibration observations. Once this is done, we can train a deep neural network, more specifically an MLP network, to obtain vibration observations.

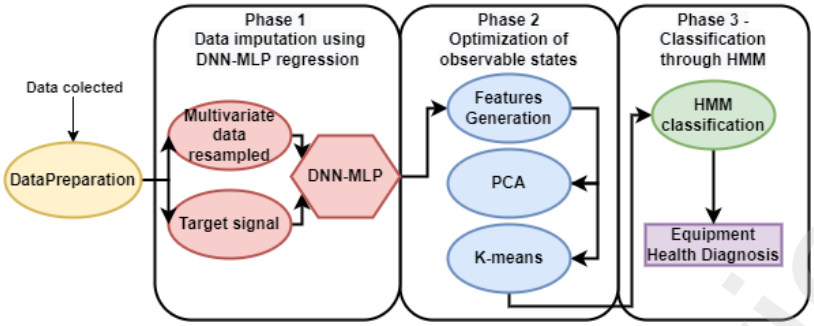



**Figure 2:** Methodology used to classify the health status of a component of a production equipment

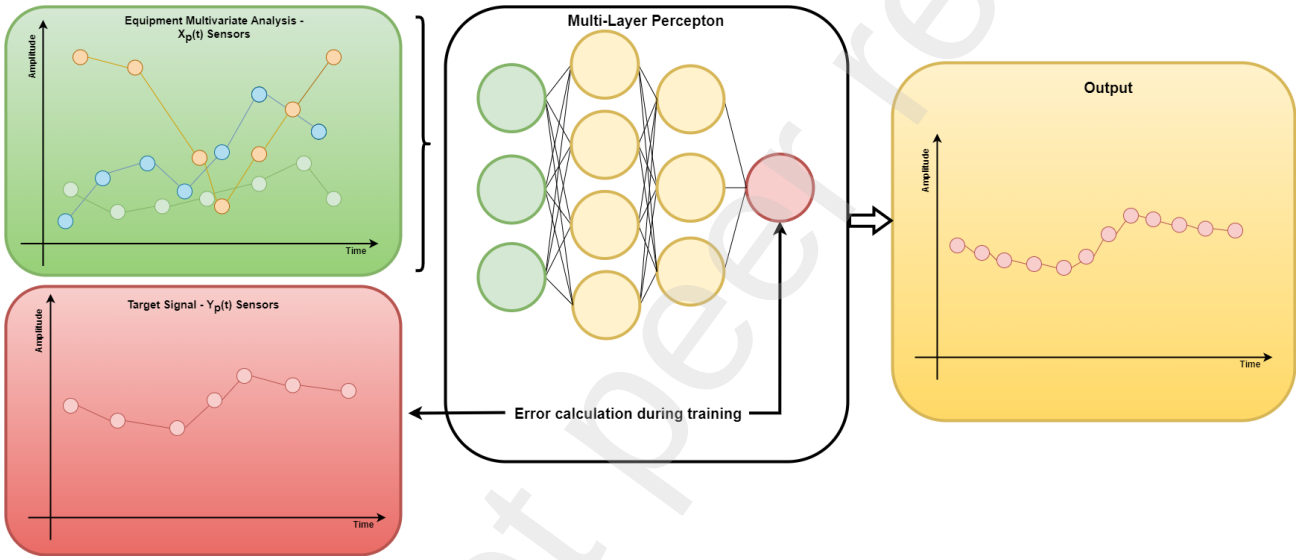



**Figure 3:** Phase1 - Scheme for imputing values through an MLP neural network

MLP has become the most frequent, ubiquitous, basic, and necessary for most tasks among numerous categories of NNs [15]. When compared to more sophisticated models, MLP has various advantages, such as: ease of implementation, high-quality models, and a relatively short training period [23]. "One of the main problems in designing a deep neural network is to tune the hyperparameters and architecture of the network" [45]. To train the model and to choose the best hyperparameters, we will use a python library named KerasTuner. It is a scalable and user-friendly hyperparameter optimization library that addresses the problems associated with hyperparameter search [46]. Bayesian Optimization, Hyperband, and Random Search are the three search algorithms that can help to find the optimal hyperparameter values for the models. In this way, the model's hyperparameters can be chosen through the library, thus ensuring better confidence in the MLP. Having obtained a good model with low errors, we ran the calculated model to get a vector of data collections equal to that performed by the multivariate analysis before the resampling. Thus, we were

able to attribute new temporal vibration observations in the empty time space and have a more significant amount of data to study the following phases.

*Phase 2*

After Phase 1, we already have a vector with enough time observations to start Phase 2. This phase is called the observable state optimization phase. It goes through a series of methodologies (Figure 4) whose objective is to reduce the number of amplitude observations that will enter Phase 3. That is, Phase 1 is responsible for increasing the temporal observations on the x-axis, and Phase 2 is responsible for decreasing the amplitude observations on the y-axis.

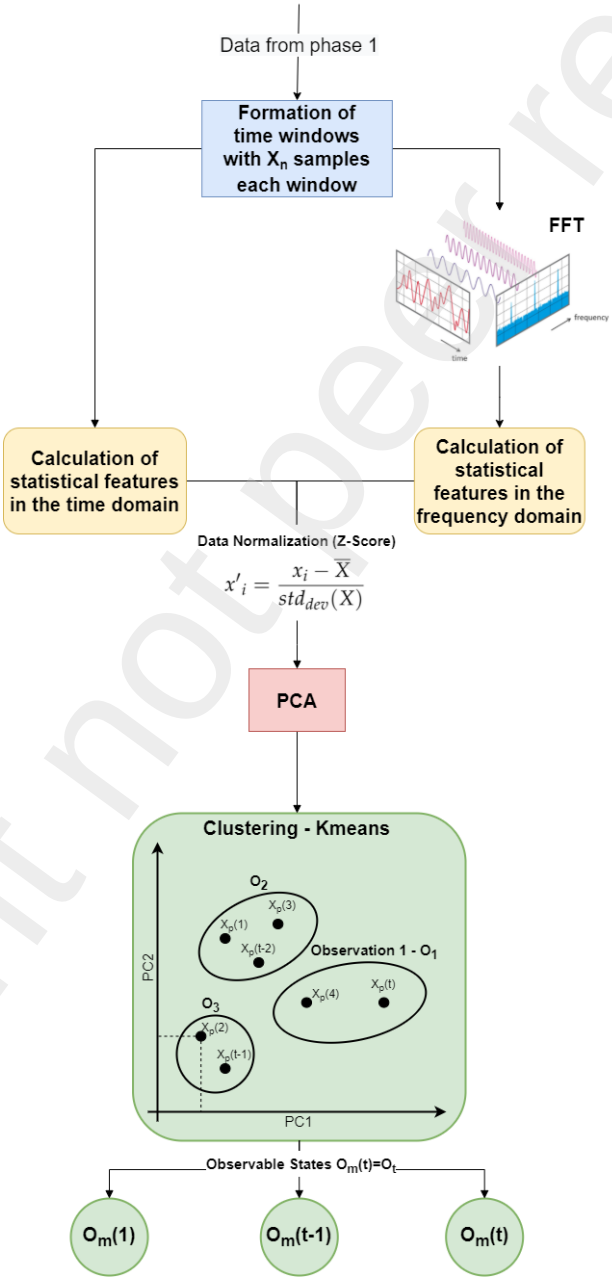



**Figure 4:** Phase2 - Methodology for Optimization of Observations

Increasing temporal observations, it is essential since the first task of Phase 2 aims to create temporal blocks with $x_n$ values for each block, with $n = 1, 2, ..., N$, where N is the number of signal samples. If we had collections made in larger temporal spaces, the temporal blocks created here would have fewer $x_n$ values and would not give such reliable information. Through each time block, the objective is to extract a series of characteristics both in the time domain and in the frequency domain better to characterize the behavior of the signal over time, obtaining more critical information contained in the signal. "A time and frequency domain vibration signal is generally used by the analyst for detecting faults in a system" [10]. As Cheng et al.[30] explained, the time domain degradation feature is one of the simplest and most effective methods of vibration signal analysis, using calculations through statistical analysis of the vibration signal.

Vibration signal frequency analysis is also widely used to diagnose equipment failures [47]. The frequency-domain signal to obtain information about the roll conditions is obtained using the Fast Fourier Transform (FFT) analyzer. The most common frequency domain analysis approach is FFT analysis, which converts time waveform data into frequency components [48].

The statistical parameters of the time and frequency domains were calculated, as described below, regarding Table 2 and Table 3, where the mathematical equations corresponding to the statistical characteristics in both domains are summarized. Table 2 represents the mathematical equations of the proposed set for the time domain characteristics, where $x_n$ is the time domain signal series occupied in each time block. Table 3 represents the mathematical equations of the frequency domain characteristics where $y_m$ is the Fourier transform for $m = 1, 2, ., M$, where M is the number of lines in the spectrum and $f_m$ is the frequency value of $m^t h$ line of the spectrum. According to some papers, there are several characteristics that can be used to best represent the signal for fault detection [9, 10, 32, 49].

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

After calculating the characteristics in the time domain and in the frequency domain, data standardization is now performed. Since data grouping is based on the difference between value points, it is essential to have all features on the same scale; otherwise, high values will be grouped independently of their defaults [39]. To remove the volume difference in the data, we use the Z-score method. Z-Score standardization is a key preprocessing step that allows you to scale features using a traditional normal distribution with a mean of zero and a standard deviation of one. Because we are interested in the components that maximize variance, it is critical to adopt Z-Score standardization to improve PCA method that will be used in our situation. We assume that one feature varies less than the other due to scale differences. In that case, the PCA defines the direction of maximum variation that most closely approximates the axis, which varies more if these features that are not to scale [22]. As a result, standardizing the data is critical in order for the algorithm to be more sensitive to

**Table 2**
Mathematical equations for Time Doamin-Based statistical features.

| Parameter | Mathematical Equation | Parameter | Mathematical Equation |
|---|---|---|---|
| Mean | $T_1 = \frac{\sum_{n=1}^{N} x(n)}{N}$ | A Factor | $T_{12} = \frac{T_5}{T_2.T_3}$ |
| Standard Deviation | $T_2 = \sqrt{\frac{\sum_{n=1}^{N}(x(n)-T_1)^2}{N-1}}$ | B Factor | $T_{13} = \frac{T_7.T_8}{T_2}$ |
| Variance | $T_3 = \frac{\sum_{n=1}^{N}(x(n)-T_1)^2}{N-1}$ | SRM | $T_{14} = \left(\frac{\sum_{n=1}^{N}\sqrt{\lvert x(n)\rvert}}{N}\right)^2$ |
| RMS | $T_4 = \sqrt{\frac{\sum_{n=1}^{N}(x(n))^2}{N-1}}$ | SRM Shape Factor | $T_{15} = \frac{T_{14}}{T_1}$ |
| Absolute Maximum | $T_5 = max\lvert x(n)\rvert$ | Latitude Factor | $T_{16} = \frac{T_5}{T_{14}}$ |
| Coefficient of Skewness | $T_6 = \sqrt{\frac{\sum_{n=1}^{N}(x(n)-T_1)^3}{(N-1).T_2^3}}$ | Fifth Moment | $T_{17} = \sqrt{\frac{\sum_{n=1}^{N}(x(n)-T_1)^5}{(N-1).T_2^5}}$ |
| Kurtosis | $T_7 = \sqrt{\frac{\sum_{n=1}^{N}(x(n)-T_1)^4}{(N-1).T_2^4}}$ | Sixth Moment | $T_{18} = \sqrt{\frac{\sum_{n=1}^{N}(x(n)-T_1)^6}{(N-1).T_2^6}}$ |
| Crest Factor | $T_8 = \frac{T_5}{T_4}$ | Median | $T_{19} = medianx(n)$ |
| Margin Factor | $T_9 = \frac{t_5}{T_3}$ | Mode | $T_{20} = modex(n)$ |
| RMS Shape Factor | $T_{10} = \frac{T_4}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$ | Minimum | $T_{21} = minx(n)$ |
| Impulse Factor | $T_{11} = \frac{T_5}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$ | | |

changes in different measurements. Because this model demands that the characteristics be on the same scale in order to determine directions that maximize variation, standardization is done. The Dimensional Reduction process, carried out through the PCA algorithm, will only obtain the Principal Components (PCs). The observations are now described in a new space, with the axes being redefined using PCs rather than the original variables. PCA is an unsupervised model that allows you to identify the most effective coordinate system of

**Table 3**
Mathematical equations of statistical characteristics calculated from the Frequency Domain.

| Parameter | Mathematical Equation | Parameter | Mathematical Equation |
|---|---|---|---|
| Mean | $F_1 = \frac{\sum_{m=1}^{M} y(m)}{M}$ | D Factor | $F_8 = \sqrt{\frac{\sum_{m=1}^{M} f_m^4 y(m)}{\sum_{m=1}^{M} f_m^2 y(m)}}$ |
| Variance | $F_2 = \frac{\sum_{m=1}^{M} (y(m)-F_1)^2}{(M-1)}$ | E Factor | $F_9 = \frac{\sum_{m=1}^{M} f_m^2 y(m)}{\sqrt{\sum_{m=1}^{M} y(m) \sum_{m=1}^{M} f_m^4 y(m)}}$ |
| Third Moment | $F_3 = \frac{\sum_{m=1}^{M} (y(m)-F_1)^3}{M(\sqrt{F_2})^3}$ | G Factor | $F_{10} = \frac{F_6}{F_5}$ |
| Forth Moment | $F_4 = \frac{\sum_{m=1}^{M} (y(m)-F_1)^4}{M(\sqrt{F_2})^2}$ | Third Moment 1 | $F_{11} = \frac{\sum_{m=1}^{M} (f_m-F_5)^3 y(m)}{M F_6^3}$ |
| Grand Mean | $F_5 = \frac{\sum_{m=1}^{M} f_m y(m)}{\sum_{m=1}^{M} y(m)}$ | Forth Moment 1 | $F_{12} = \frac{\sum_{m=1}^{M} (f_m-F_5)^4 y(m)}{M F_6^4}$ |
| Standard Deviation 1 | $F_6 = \sqrt{\frac{\sum_{m=1}^{M} (f_m-F_5)^2 y(m)}{M}}$ | J Factor | $F_{13} = \frac{(F_7+F_8)}{F_1}$ |
| C Factor | $F_7 = \sqrt{\frac{\sum_{m=1}^{M} f_m^2 y(m)}{\sum_{m=1}^{M} y(m)}}$ | | |

the data to find the strongest characteristics of the samples. It is a very simple and easy-to-use tool. It is also quick, minimizes overfitting efficiently, and can be used as a noise reduction and data compression technique [22, 33, 50]. After the PCA was completed, we reduced multiple features into a few PCs, which allowed us to minimize a major portion of the experimental noise, which could create random mistakes unrelated to the information contained in the data matrix, thereby enhancing the numerical stability of the model [22].

After performing the dimensional reduction and PC characterization, the composite values in these new orthogonal axes will be grouped. Clustering is regarded as one of the most important unsupervised learning algorithms. It is a starting point for a complicated knowledge extraction procedure that deals with database partitioning in unexpected conditions [35]. The objective of Clustering will be to group similar data in the same cluster that will differ from other clusters, grouping observations within each cluster homogeneously

and heterogeneously separating the clusters from each other. Thus, equipment health data will stay in the same cluster while malfunction data will go to a different cluster. In this way, each cluster will be seen as an observable state, where the cluster with good functioning data will be observed. This can be characterized by the amount of data in each cluster, ordering the clusters in descending order concerning the amount of data of each has. Thus, we assume that cluster 1, which will have the most data, will be considered as observation 1; the cluster 2 will be the second cluster with the highest amount of data and, in turn, will be the observation 2, and so on. The last cluster will be considered the observation that may represent the equipment malfunction since it is the one with the least data and, consequently, happens less often. The k-means Clustering method will be used to define the clusters, which clusters use Euclidean metrics. This choice is justified because it is an algorithm widely evaluated by the scientific community, unsupervised, it has a sufficiently fast convergence speed, and does not require extensive computational resources, in addition to being easy to implement [35]. These qualifications allow the exploitation of a vast and reliable database for comparison and analysis. It's also an algorithm that can be applied after reorienting the source data from the cloud core's metrics framework (process done by PCA). This leads the data to be centered in regard to an origin, being relatively equidistant from the center and placed in the different quadrants of the Rp space (p denoting the Principal Components of the PCA), allowing the K-means to work well [22].

*Phase 3*

After performing the Clustering and defining the optimized observations over time, we now move on to phase three. This is the classification phase, which aims to classify the health status of the component through the observations obtained from Phase 2. The classification algorithm used is the Hidden Markov Models (HMM) model (Fig. 5), which will be responsible for diagnosing the equipment through the hidden states of the model.

First, the observed states are divided into training data and test data. Through the training observations and the Baum-Welch algorithm of the HMM, it is possible to train the model and obtain its parameters, $\lambda = \{A, B, \pi\}$. Having obtained the model parameters, the HMM can generate a sequence of observable states and compare them with the test observations to obtain Accuracy. Following the model's metric analysis, we can use the Viterbi technique to determine the maximum probability estimate of the most likely sequence of hidden states that correspond to the component health states using the whole observable states and HMM parameters.

## 4. Case of Study

The case study is based on determining the health status of a pressure roller of a drying press in the paper industry, using values collected by vibration sensors over a period of about one year (Start date: 2020-11-05 and End Date: 2021-08-21). In this regard, the objective will be to classify, during the study period, the state of the equipment through a "traffic light" classification, having three states:

1. Good working states, displayed by green color;

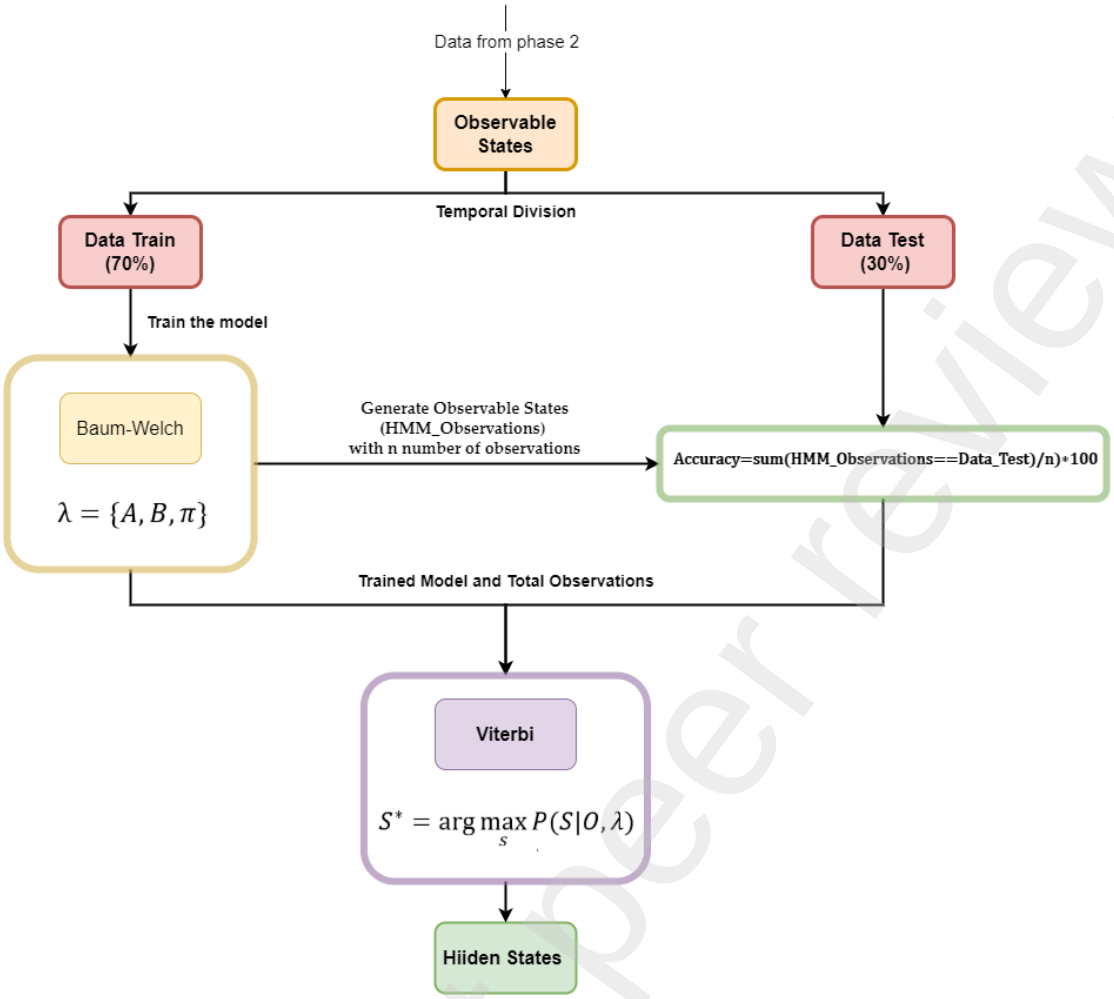



**Figure 5:** Phase 3 - Methodology for classifying the health status of the component through the HMM

2. Alert states, shown by yellow color;
3. Malfunction states, shown by red color.

### 4.1. Data preparation

Vibration sensors are composed of four types: vibra-m acc; vibra-m op acc; vibra-facc; vibra-f op acc. Taking into account the selected methodology, data from the vibration sensors were collected every 4 hours (Figure 6), which corresponds to little data, as it is necessary to create time windows of data to create characteristics in the time and frequency domains. In this regard, values from other sensors also coupled to the same equipment were used, which were collected every five minutes. The complementary sensors were composed of six variables (Figure 7): Electrical Current; Hydraulic unit level; Torque Press; VAT Pressure; Rotation speed Press; Hydraulic unit Temperature.

Since there are always errors when collecting data from the sensors, a data preparation was carried out, where all incoherent values (+inf; -inf; Nan) were replaced by the average of the respective signal. This was done both in the vibration sensors and in the sensors

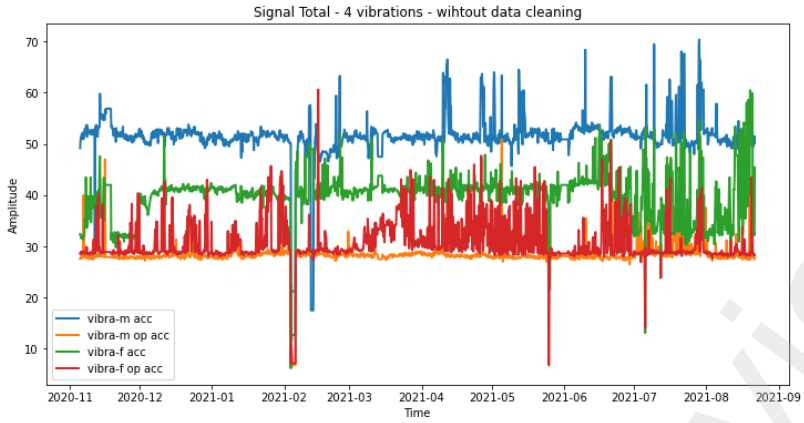



**Figure 6:** Four Vibration sensors with data collected every four hours

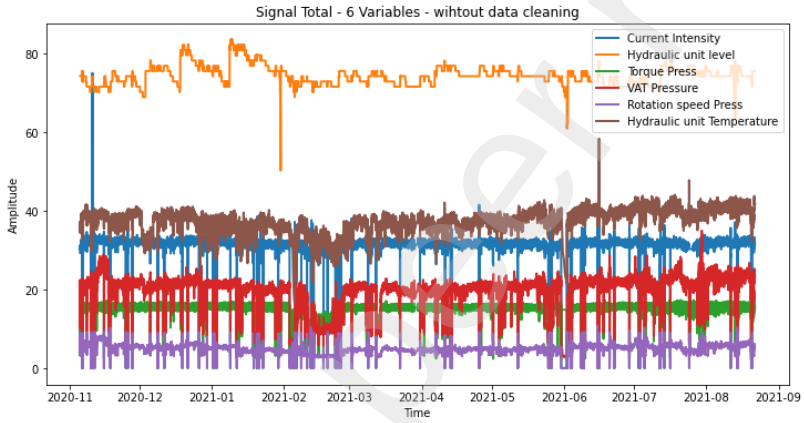



**Figure 7:** Six Complementary sensors with data collected every five minutes

complementary to the study. In addition to data cleaning, it was determined to perform an imputation of the mean (in all signals) in cases when the data, such as Current Intensity, Torque Press, VAT Pressure, and Rotation Speed Press, were below a respective value at the same time. (Figure 8). This was done with the objective of eliminating what could be the equipment shutdowns since these stops could impair the prediction of the adopted methodology.

In Figure 8 we can see represented when the equipment is turned off (in blue). No more data cleaning was added, since the outliers present here may not represent the poorly collected values, but the equipment malfunction. In addition, the "filtering" of the data set will be complemented by the entire methodology used in the next steps.

### 4.2. Data imputation using MLP

After this preparation, whose objective was to increase the integrity of the data, the complementary sensors were resized so that they only have data collected every four hours (Figure 9). This is done with the aim of using these values to create an MLP model that

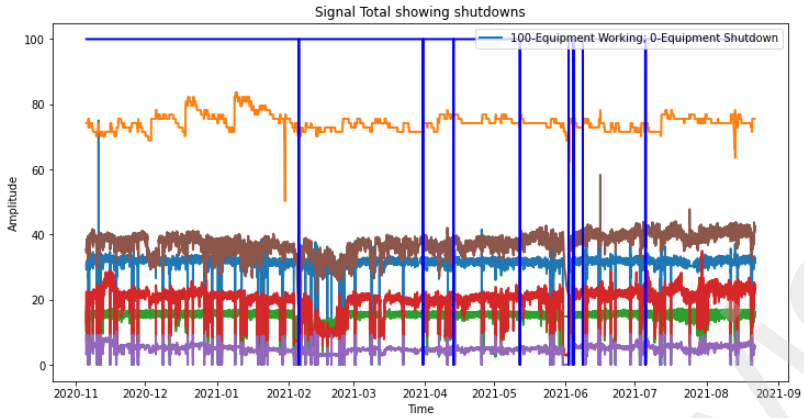



**Figure 8:** Representation of equipment shutdowns

can predict the values of the vibration sensors, through the complementary sensors. Once this is done, it will be possible, through the created MLP model and the real data of the six variables (collected every five minutes) to create vibration signals data in every five minutes.

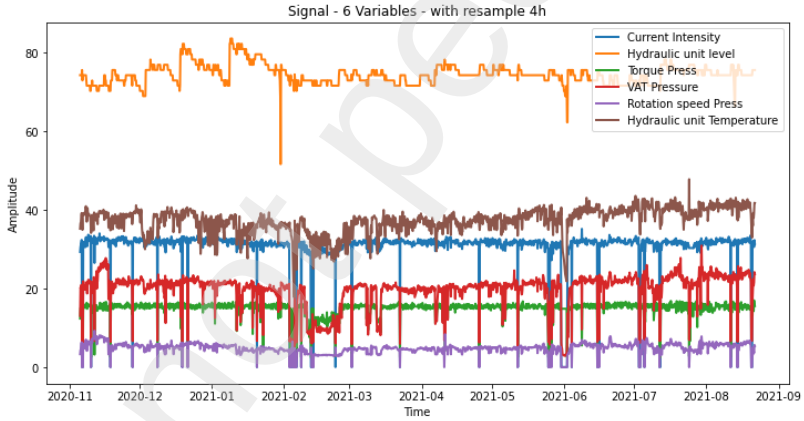



**Figure 9:** 6 complementary variables with resizing to values each four hours

Once the resample is done, we can train an MLP neural network to obtain the vibration observations. The six input variables were used for each of the vibration signals used as the output of the model. After the resample, the data were split into training and test data, 70% for training data and 30% for testing. To train the model and choose the best hyperparameters, the Python library KerasTuner was used and, through the Random search algorithm, and the smallest error of Mean Absolute Error (MAE), it was determined the best structure to be used for each vibration signal. The number of layers, the number of neurons per layer, the Activation Function, and the Learning Rate were all chosen using this library.

It is possible to see, based on Table 4, which MLP structures ought to be chosen for each vibration signal.

**Table 4**
MLP network structure for each of the vibration signals

| Hyperparameters | vibra-m acc | vibra-m op acc | vibra-f acc | vibra-f op acc |
|---|---|---|---|---|
| Number of Layers: | 8 | 9 | 3 | 4 |
| 1st Layers<br>Activation function | 288 Neurons<br>tanh | 128 Neurons<br>tanh | 352 Neurons<br>tanh | 160 Neurons<br>relu |
| 2nd Layer<br>Activation function | 160 Neurons<br>tanh | 160 Neurons<br>tanh | 160 Neurons<br>relu | 160 Neurons<br>tanh |
| 3rd Layer<br>Activation function | 320 Neurons<br>tanh | 192 Neurons<br>tanh | 384 Neurons<br>tanh | 192 Neurons<br>relu |
| 4th Layer<br>Activation function | 416 Neurons<br>relu | 64 Neurons<br>tanh | | 352 Neurons<br>relu |
| 5th Layer<br>Activation function | 128 Neurons<br>tanh | 384 Neurons<br>relu | | |
| 6th Layer<br>Activation function | 448 Neurons<br>relu | 256 Neurons<br>relu | | |
| 7th Layer<br>Activation function | 224 Neurons<br>relu | 384 Neurons<br>tanh | | |
| 8th Layer<br>Activation function | 352 Neurons | 512 Neurons<br>relu | | |
| 9th Layer<br>Activation function | | 32 Neurons<br>relu | | |
| Learning rate | 0.0001 | 0.001 | 0.001 | 0.001 |
| Score - MAE | 1.8357 | 0.9170 | 2.9033 | 2.5250 |

After knowing which MLP structure best suit each data, a prediction of the vibration signals was made. In the Figures 10-14, we can see in a graphic way the real data of the vibration signals, and the data predicted by the MLP model.

Through the graphical analysis, we can verify that the model was able to make good predictions of the vibration data and, as such we can now use these respective models, to create the vibration values with intervals of five minutes, as can be seen in Figure 15. The vibration data predicted for observations every five minutes will be the data that will continue the study to give the health status of the equipment pressure roller.

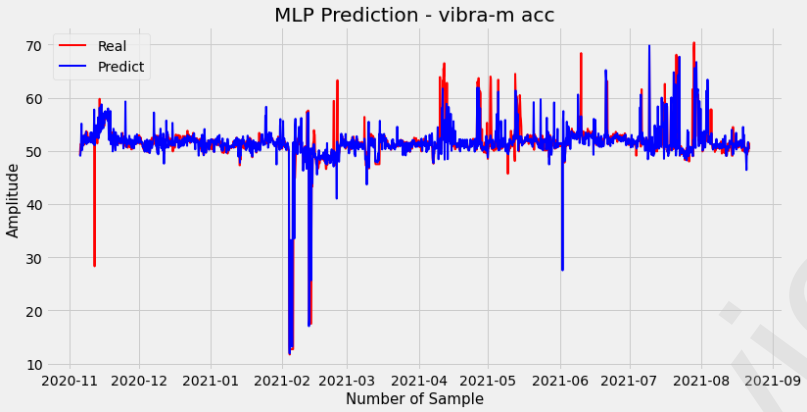



**Figure 10:** MLP prediction vs Actual values for the Vibra-m acc signal

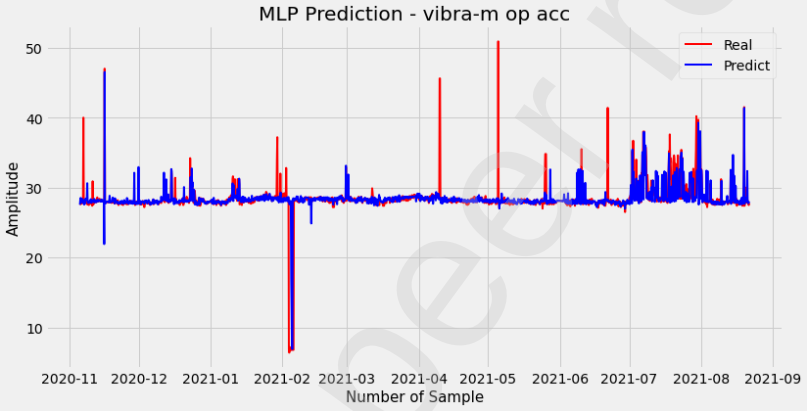



**Figure 11:** MLP prediction vs Actual values for the Vibra-m op acc signal

### 4.3. Features Generation

Now, we have a superior dataset, with a greater number of observations per day then, we can create time windows of one day. The data in the study time performed with observations every five minutes have 83,329 samples. In this way, creating time windows of one day, there are 290-time windows where each one has 288 observations. This is already a considerable number to be able to generate statistical characteristics in each of the time windows. As explained above, statistical characteristics will be created both in the time domain and in the frequency, domain using the characteristics described in Table 2 e 3. Therefore, in each time window, 21 characteristics will be removed in the time domain and 13 in the domain frequency, thus building a matrix of $290 * 140$. After creating the characteristics matrix, to remove the volume difference in the data, the values will pass a Z-Score standardization. The numbers begin to take on a conventional normal distribution, with a mean of zero and a standard deviation of one (Figure 16).

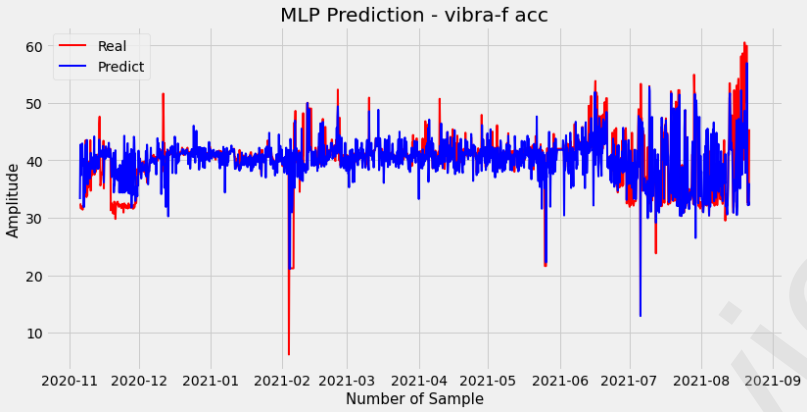



**Figure 12:** MLP prediction vs Actual values for the Vibra-f acc signal

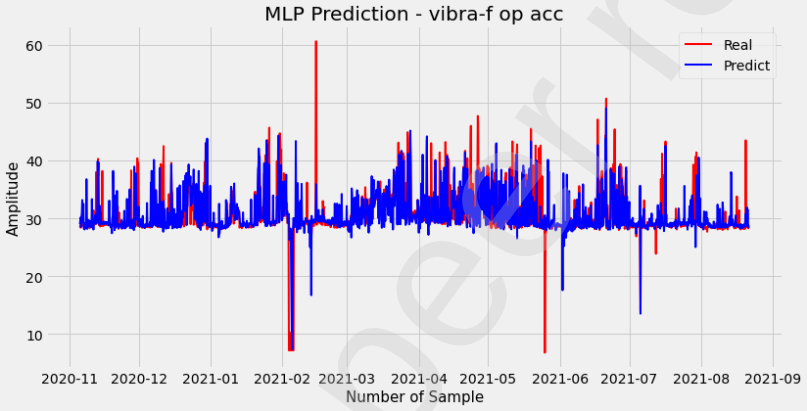



**Figure 13:** MLP prediction vs Actual values for the Vibra-f op acc signal

### 4.4. Principal Components Analysis

This step aims to make a dimensional reduction, allowing to reduce the 140 generated features to just 10 Principal Components. In this way, we no longer have a 290*140 matrix; now, we have a 290*10 matrix. 10 PCs were chosen because this is the number of components that preserves about 85% of the data variability, as we can see through a graphical analysis of the data variability using a Pareto diagram (Figure 17). In this way, we make a dimensional reduction and can manage to preserve most of the information in the data.

As described in the theoretical framework, the first PCs are those that contain the greatest variation in the data. Figure 18 shows how data from the first three PCs are distributed in these three dimensions and what percentage of data variability of each of the PCs has.

### 4.5. K-means

Having done the dimensional reduction and orthogonal redistribution through PCA, we will now use the matrix extracted from the previous step and cluster the data. The data were grouped through K-means Clustering, where each cluster will correspond to a new

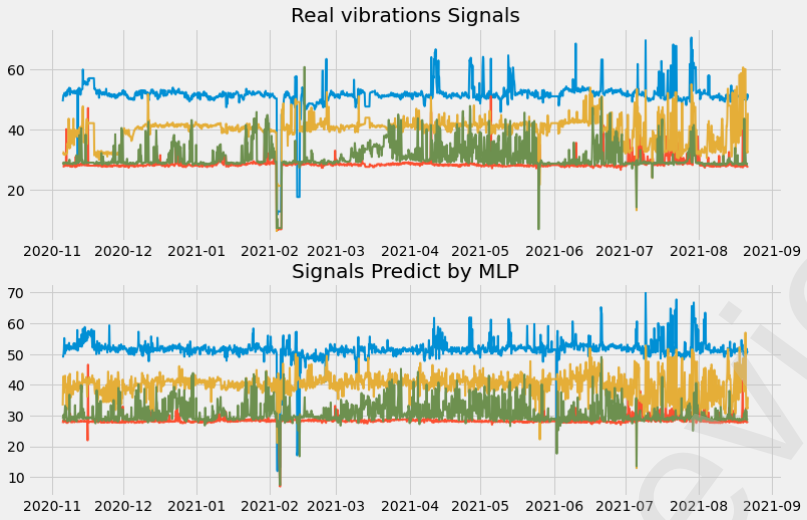



**Figure 14:** All Actual Vibration Data Vs MLP Predicted Vibration Data

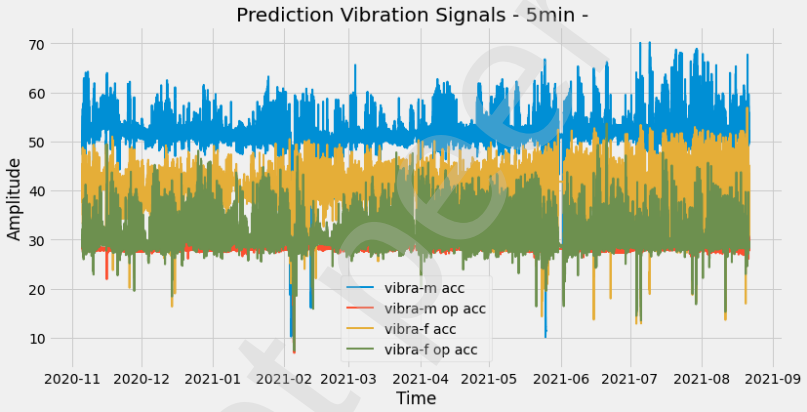



**Figure 15:** Five-minute vibration data predicted by MLP

observable state for future input into the HMM classification algorithm. Since the algorithm used for this study was K-means, it is necessary to determine in advance the number of clusters that will be used. The K-means algorithm is the most popular clustering algorithm minimizing Sum of Squares Error (SSE) [51]. For this reason, the elbow method is used (Figure 19). This is one of the most used to select the number of Clusters, using the SSE vs the number of Clusters. SSE is an internal measurement index that is used to measure the quality of a clustering structure [52].

It can be seen in Figure 19 that the value $k = 7$ creates an elbow, where the value of SSE starts to decrease much more slowly. This means that from the point $k = 7$ the clusters start to be divided into smaller clusters without significantly improving the SSE [53]. For this study, the relevant number of clusters is seven. It is possible to demonstrate how the data are divided into the first three dimensions extracted by the PCA (Figure 20).

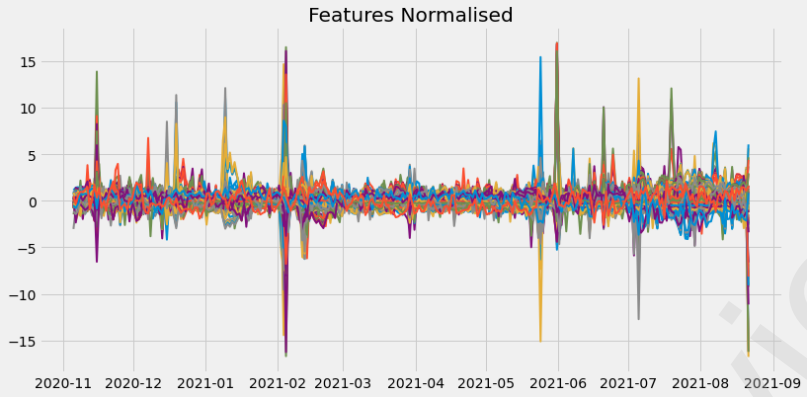



**Figure 16:** Normalized generated features values

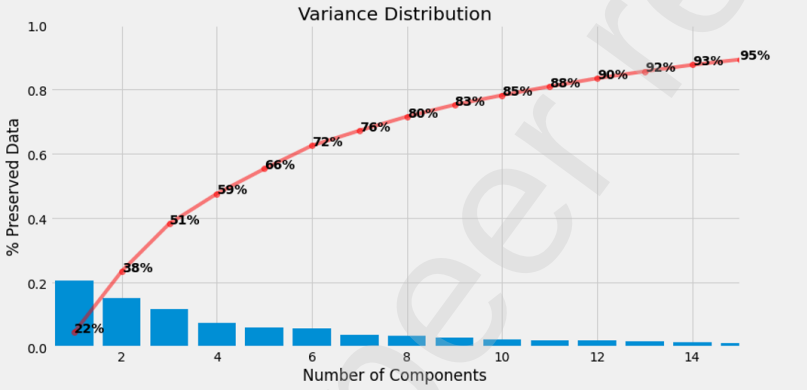



**Figure 17:** Pareto diagram with the percentage of data variability for each principal component

The seven clusters now represent seven observable states. The clusters were sorted in descending order, with the first cluster having the most data and so on. In this way, we guarantee that observation seven is the one that occurs less often and therefore the most unlikely to happen. This could mean equipment malfunction, but this will be sorted by the HMM in the next step. Observations can be represented over time, as shown in Figure 21.

### 4.6. Hidden Markov Models

The classification of the equipment health status is done in this last step, through the HMM. For this, the optimized observable states from the k-means clustering, being used three hidden states, which will represent respectively: the $1^{st}$ State of Good Operation; the $2^{nd}$ State of Alert; the $3^{rd}$ State of Malfunction. To perform the classification and to determine the hidden states, it is necessary to go through a training phase of the model, performed by the Baum-Welch algorithm, and then determining the temporal sequence of hidden states that best "translate" the observable states, performed through the Viterbi algorithm. To perform the model training, we first split the observable states into training data (70%) and test data (30%), as is common practice in Machine Learning. Because the observations were recorded in real time, the first 70% of the samples were used for training

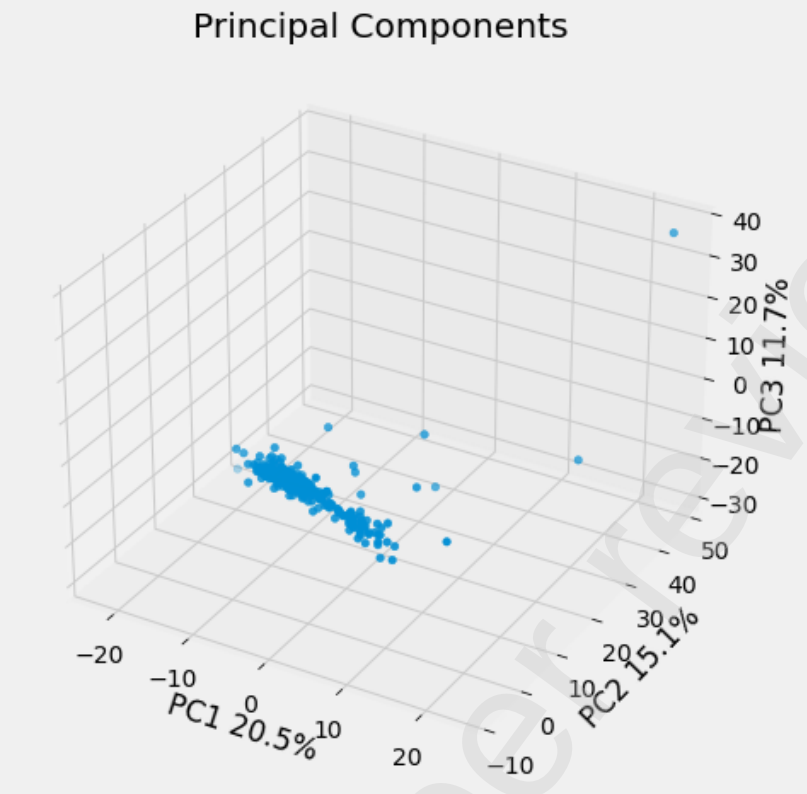



**Figure 18:** Distribution of data in the first three Principal Components

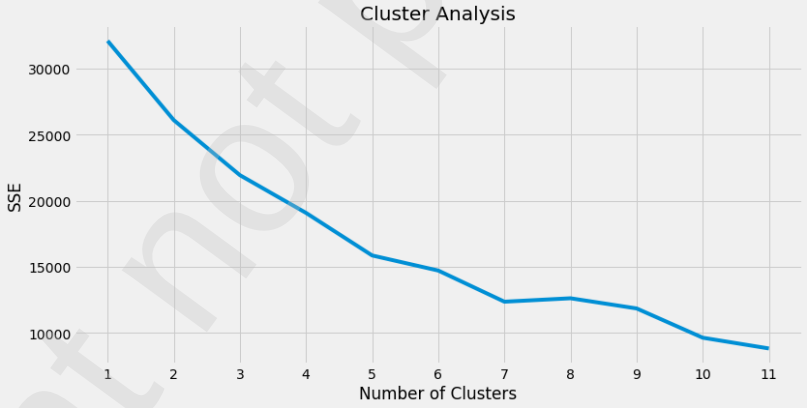



**Figure 19:** SSE vs Number of Clusters - Method to determine the number of clusters

and the remaining 30% for testing. After building the model, it is feasible to measure the correctness of the HMM model, as mentioned above in the methodology, by dividing the number of successfully validated estimates for all types of events or status classes of a system by the entire monitoring of the condition of a system [22, 43].

After the model has been trained using 70% of the training data, it is possible to produce observable states using the HMM parameters and compare them to the remaining 30% of

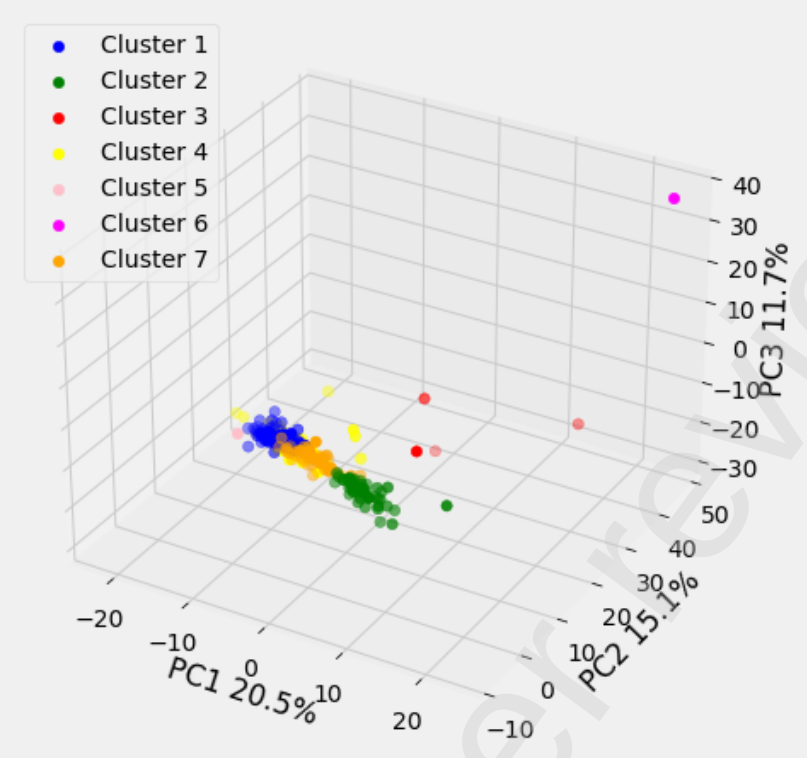



**Figure 20:** The 7 Clusters represented in the first 3 Principal Components

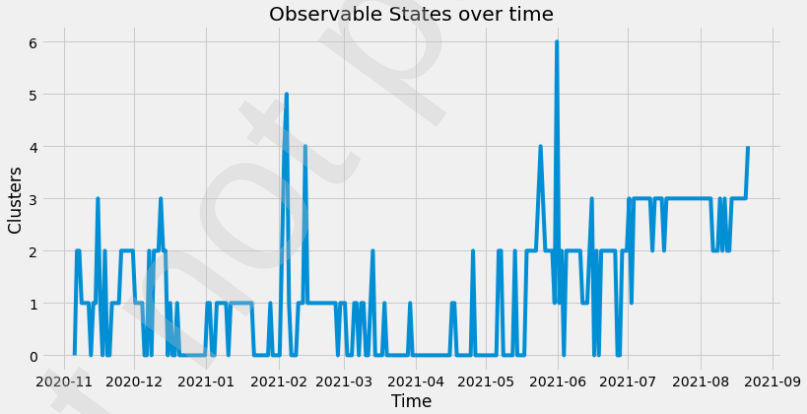



**Figure 21:** Time series of clusters over study time

the test data. In this way, it is possible to determine the superposition of sequences of real and predicted observable states and to determine the accuracy (Equation 6).

$$Accuracy = \frac{\sum(HMM_{Observation} = Data_{Test})}{n_{samples}} * 100 \quad (6)$$

Because the HMM's parameters $\lambda$ are based on probabilities, 10,000 observation sequences with a number equal to the test data were created, and the respective accuracy

was measured. Subsequently, the average of all accuracy was calculated. For these data in question, with a sequence of seven observable states (represented above) and three hidden states, the model obtained an accuracy of only 32.26%. To understand this low Accuracy, tests were carried out with other HMM models where the number of hidden states was varied and tests, where the temporal form of the signal of observations from clustering, were modified. In this case, we divide the signal into 40%, 30% and 30% temporally and then exchange the 30% temporal blocks with each other, leaving the observations represented as shown in Figure 22.

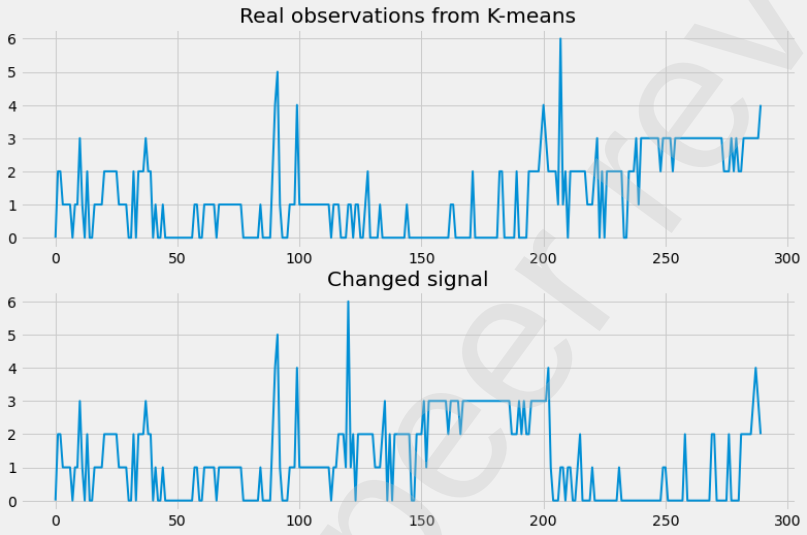



**Figure 22:** Actual Observations and altered Observations

The objective is to understand if the size of the signal influences the Accuracy of the model since in the final phase of the test data, we have observations that did not occur until then. Thus, we obtained the following results (Table 5):

Through the results obtained and shown in Table 5, we can see that the accuracy increases when we change the sign and place certain observations in the training data. This means that the model gets better training because it trains with all the observations. In addition to the accuracy, the log-likelihood was determined, which determines the ability of the model to adapt to the observations and, consequently, the ability to use the Viterbi algorithm. According to Arpaia et al.[54], "log-likelihood is calculated to verify if each sequence of data fits the model". The log-likelihood values are all in the same order of magnitude. In this way, we can admit that, although the model used has a low Accuracy, it is able to adapt the observations and to determine the hidden states. To obtain a better accuracy, it will be necessary to obtain a larger set of data, collected over a longer period, where all possible observations for the equipment under study occur.

Based on this principle, we are going to train the HMM model with all the observations and three hidden states, that will represent the health status of the equipment. After training,

**Table 5**
Accuracy and Log-Likelihood for each of the times series of observations and different number of hidden states

| | Real Observation from K-means | Changed Signal |
|---|---|---|
| Nº. of Hidden States | 3 | |
| Accuracy | 0.32264367816091954 | 0.48747126436781607 |
| Log-Likelihodd | -248.86347140364788 | -285.5781288565171 |
| Nº. of Hidden States | 4 | |
| Accuracy | 0.2695402298850575 | 0.6603448275862069 |
| Log-Likelihodd | -245.69196936616808 | -282.9993414180359 |
| Nº. of Hidden States | 5 | |
| Accuracy | 0.27206896551724136 | 0.7096551724137931 |
| Log-Likelihodd | -244.29181521910746 | -278.9731023761346 |
| Nº. of Hidden States | 6 | |
| Accuracy | 0.37091954022988505 | 0.7295402298850575 |
| Log-Likelihodd | -242.7765380922869 | -278.67744793543966 |
| Nº. of Hidden States | 7 | |
| Accuracy | 0.3671264367816092 | 0.7045977011494253 |
| Log-Likelihodd | -242.77622604230996 | -278.7458634574436 |

we performed the Viterbi algorithm to determine the state of the equipment over the time under study (Figure 23).

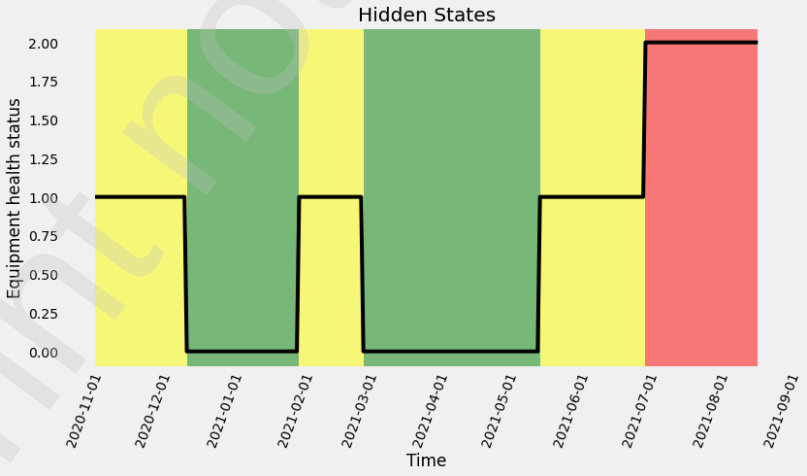



**Figure 23:** Health status of the equipment over the study period

## 5. Discussion

As can be seen in Figura 24, several steps were taken to obtain the health status of the equipment under study.

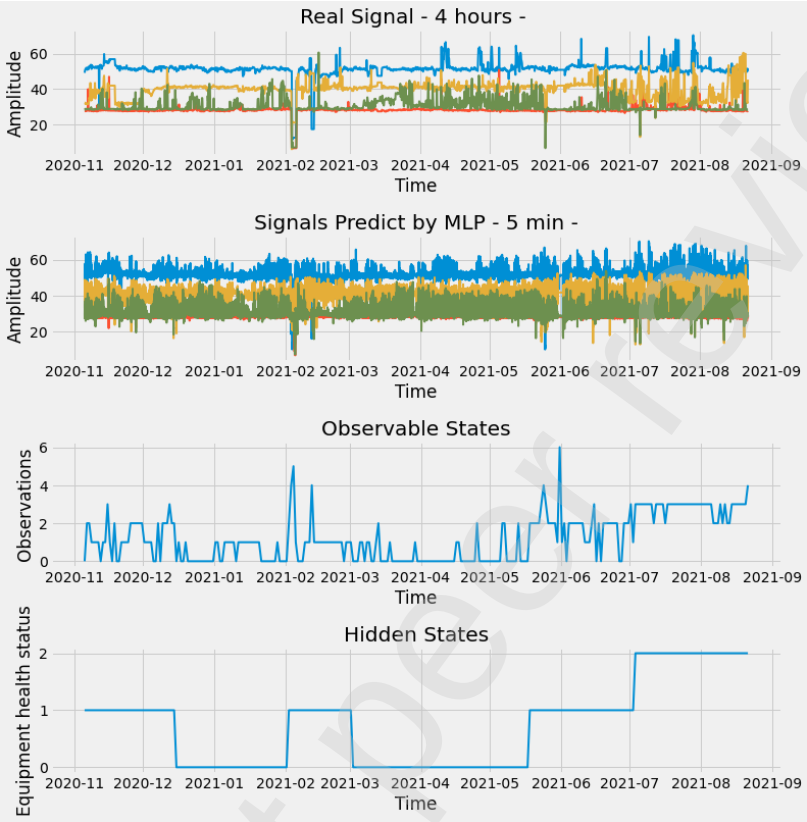



**Figure 24:** Graphic representation of each of the steps of the methodology used

Through the hidden states, we can verify that there were some alert states. We have two alert states that have returned to a good working state, possibly due to routine maintenance performed. But, in the last months of the study, we can see a state of alert that was not intervened and later developed into a state of equipment failure, where it may be necessary to do a corrective maintenance. If we compare the development of hidden states with the first graph, which represents the real signal collected every 4 hours, we can see that the appearance of hidden states 2 and 3, it appears when there are significant changes in the pattern of sensor values. It is also possible to verify that, in the final months of the study, all sensors begin to have large changes in values, which is when there is a development from hidden state 2 to hidden state 3. Thus, through a graphical analysis, we can also infer that the classification of the health status of the equipment can be correct and that it can detect changes in the behavior patterns in the equipment.

Several tools were used, both for predicting values and for "optimizing" the observable state to insert into the HMM classification algorithm. This is a methodology that can be used in investigations where there is no prior knowledge of the equipment's failure

conditions. In addition, it can be adaptable to any type of equipment and sensors. It will allow for online diagnosis of equipment that uses sensors, allowing failures to be detected without prior information on the equipment status. In this methodology, some of the possible tools were used in each of the stages. Depending on the type of data, there may be other tools that are better suited to the study. As such, it will be necessary to carry out a study where different tools are compared and the combination that best suits the data under study is made. In the first phase, a comparison of classes of Artificial Neural Networks can be made, such as: MLP; Long short-term memory (LSTM); Convolutional Neural Network (CNN); etc. In the second phase, for dimensional reduction, the following can be compared: PCA; Linear Discriminant Analysis (LDA); Independent Component Analysis (ICA); etc. For Clustering, the following ones: K-means; Density-Based Spatial Clustering of Applications with Noise (DBSCAN); Gaussian Mixture Models (GMM); etc. By carrying out a study with the greatest possible number of combinations between these study tools, it will be possible to obtain the models that best fit the data under study. This paper is only demonstrating the possible methodology to classify the health status of equipment or component without having a priori information about it.

A method for doing diagnostic and prognosis in an operation based on an online algorithm will also be incorporated in the future to improve the methodology. The objective will be to add an Artificial Neural Network class, like some of the ones mentioned above, to make a prediction of the next observable states that come from clustering. In this way, it will then be possible, through the HMM's Viterbi algorithm, to determine the equipment health status for the future. In addition, it will be necessary to perform a retraining methodology of the HMM model in case that new observable states appear. For this, it will be necessary to carry out an adaptive methodology that warns if any value has a very large discrepancy from the clusters formed previously and it is necessary to create a new observable state.

It is also intended that this study will be extended to other drying presses of the company.

## 6. Conclusion

This paper presented a study where a methodology was implemented to determine the health diagnosis of a compression roller of a drying press in a paper industry. To do so, several methods of Machine Learning were used and HMM for classification of the equipment condition. Using vibration sensors, our objective was to determine the health status of the component without obtaining prior information. Through a case study we were able to validate the method, which was the added value of the research. In that regard, six sensors were used, where data were input through a Deep Neural Network to increase their set. To do so, complementary sensors were used, coupled to the same equipment and that collected values at smaller time intervals. After completing this step, an observable state optimization step was performed. Statistical characteristics were first created in the time and frequency domains. In the next step, a PCA was used to make a dimensional reduction, followed by a Clustering k-means to group similar data into groups. These groups became the observable states that were inputs to the HMM, which, later classified them into three

hidden states. The three hidden states represent the health status of the compression roller, as follows: $1^{st}$ State of good functioning; $2^{nd}$ State of Alert; $3^{rd}$ State of malfunction.

We can conclude that the adopted methodology can be used to perform equipment diagnostics. Therefore, this methodology can support CBM's maintenance strategies and increase equipment reliability, making companies more competitive.

In addition, we verified the need for the HMM to have a large volume of data that contains all possible observations. In this way, the accuracy of the model can be improved.

In the context of future research, we aim at improving the prediction of equipment failure, both in terms of production and in measurement. To this end, studies are being carried out for Online Calibration Monitoring (OML) in measuring equipment. The objective will be to increase their availability and to ensure that the data collected by these devices are reliable to be used in CBM strategies.

## References


[1] E. Pais, J. T. Farinha, A. J. Cardoso, H. Raposo, Optimizing the life cycle of physical assets – A review, WSEAS Transactions on Systems and Control 15 (2020) 417–430.

[2] J. Pais, H. D. Raposo, J. Farinha, A. J. M. Cardoso, P. A. Marques, Optimizing the life cycle of physical assets through an integrated life cycle assessment method, Energies 14 (2021) 24.

[3] J. Rodrigues, J. T. Farinha, A. M. Cardoso, Predictive maintenance tools – a global survey, WSEAS Transactions on Systems and Control 16 (2021) 96–109.

[4] J. Fernandes, J. Reis, N. Melão, L. Teixeira, M. Amorim, The Role of Industry 4 . 0 and BPMN in the Arise of Condition-Based and Predictive Maintenance : A Case Study in the Automotive Industry, Applied Sciences 11 (2021) 16.

[5] A. B. Martins, J. T. Farinha, A. M. Cardoso, Calibration and Certification of Industrial Sensors – a Global Review, WSEAS TRANSACTIONS ON SYSTEMS AND CONTROL 15 (2020) 394–416.

[6] R. Ahmad, S. Kamaruddin, An overview of time-based and condition-based maintenance in industrial application, COMPUTERS INDUSTRIAL ENGINEERING 63 (2012) 135–149.

[7] P. Pawlik, K. Kania, B. Przysucha, The Use of Deep Learning Methods in Diagnosing Rotating Machines Operating in Variable Conditions, Energies 14 (2021) 4231.

[8] U. E. Akpudo, J.-W. Hur, D-dCNN: A Novel Hybrid Deep Learning-Based Tool for Vibration-Based Diagnostics, Energies 14 (2021) 5286.

[9] J. J. Saucedo-Dorantes, A. Y. Jaen-cuellar, M. Delgado-prieto, R. D. J. Romero-troncoso, R. A. Osornio-rios, Condition monitoring strategy based on an optimized selection of high-dimensional set of hybrid features to diagnose and detect multiple and combined faults in an induction motor, Measurement 178 (2021) 109404.

[10] L. S. Dhamande, M. B. Chaudhari, Compound gear-bearing fault feature extraction using statistical features based on time-frequency method, Measurement 125 (2018) 63–77.

[11] H. Pranesh, K. Suresh, S. S. Manian, R. Jegadeeshwaran, G. Sakthivel, T. Alamelu Manghi, Vibration-based brake health prediction using statistical features – A machine learning framework, Materials Today: Proceedings 46 (2021) 1167–1173.

[12] D. Goyal, B. S. Pabla, S. S. Dhami, K. Lachhwani, Optimization of condition-based maintenance using soft computing, Neural Computing and Applications 28 (2017) 829–844.

[13] J. Chen, M. B. Alawieh, Y. Lin, M. Zhang, J. Zhang, Y. Guo, D. Z. Pan, Powernet: SOI Lateral Power Device Breakdown Prediction With Deep Neural Networks, IEEE Access 8 (2020) 25372–25382.

[14] D. Magare, S. Labde, M. Gofane, V. Vyawahare, Imputation of missing data in time series by different computation methods in various data set applications, ITM Web of Conferences 32 (2020) 03010.

[15] M. E. Shaik, M. M. Islam, Q. S. Hossain, A review on neural network techniques for the prediction of road traffic accident severity, Asian Transport Studies 7 (2021) 100040.

[16] A. Thakkar, K. Chaudhari, A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions, Expert Systems with Applications 177 (2021) 114800.

[17] D. Han, S. Kwon, Application of Machine Learning Method of Data-Driven Deep Learning Model to Predict Well Production Rate in the Shale Gas Reservoirs, Energies 14 (2021) 3629.

[18] G. Nguyen, S. Dlugolinsky, M. Bobák, V. Tran, Á. López García, I. Heredia, P. Malík, L. Hluchý, Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey, Artificial Intelligence Review 52 (2019) 77–124.

[19] P. M. R. Bento, J. A. N. Pombo, M. R. A. Calado, S. J. P. S. Mariano, Stacking Ensemble Methodology Using Deep Learning and ARIMA Models for Short-Term Load Forecasting, Energies 14 (2021) 7378.

[20] J. Du, Y. Xu, Hierarchical deep neural network for multivariate regression, Pattern Recognition 63 (2017) 149–157.

[21] G. Montavon, W. Samek, K.-R. Müller, Methods for interpreting and understanding deep neural networks, Digital Signal Processing 73 (2018) 1–15.

[22] A. Martins, I. Fonseca, J. Farinha, J. Reis, A. Cardoso, Maintenance prediction through sensing using hidden markov models—a case study, Applied Sciences (Switzerland) 11 (2021) 1–20.
[23] F. Haghighat, Predicting the trend of indicators related to Covid-19 using the combined MLP-MC model, Chaos, Solitons and Fractals 152 (2021).
[24] A. Z. Shirazi, M. Hatami, M. Yaghoobi, S. Javad, S. Chabok, An Intelligent Approach to Predict Vibration Rate in a Real Gas Turbine, Intelligent Industrial Systems 2 (2016) 253–267.
[25] S.-W. Tan, S.-W. Huang, Y.-Z. Hsieh, S.-S. Lin, The Estimation Life Cycle of Lithium-Ion Battery Based on Deep Learning Network and Genetic Algorithm, Energies 14 (2021) 4423.
[26] M. Desai, M. Shah, An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN), Clinical eHealth 4 (2021) 1–11.
[27] S. G. Kumbhar, R. G. Desavale, N. V. Dharwadkar, Fault size diagnosis of rolling element bearing using artificial neural network and dimension theory, Neural Computing and Applications 33 (2021) 16079–16093.
[28] J. Rynkiewicz, General bound of overfitting for MLP regression models, Neurocomputing 90 (2012) 106–110.
[29] A. M. D. Silva, P. H. W. Leong, Grammar- Based Feature Generation for Time-Series Prediction, Springer, London, 2015. URL: https://link.springer.com/book/10.1007/978-981-287-411-5. doi:10.1007/978-981-287-411-5.
[30] Y. Cheng, K. Hu, J. Wu, H. Zhu, C. K. M. Lee, A deep learning-based two-stage prognostic approach for remaining useful life of rolling bearing, Applied Intelligence (2021).
[31] L. S. Dhamande, M. B. Chaudhari, Detection of Combined Gear-Bearing Fault in Single Stage Spur Gear Box Using Artificial Neural Network, Procedia Engineering 144 (2016) 759–766.
[32] T. Qiu, M. Liu, G. Zhou, L. Wang, K. Gao, An Unsupervised Classification Method for Flame Image of Pulverized Coal Combustion Based on Convolutional Auto-Encoder and Hidden Markov Model, Energies 12 (2019) 2585.
[33] F. Anowar, S. Sadaoui, B. Selim, Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE), Computer Science Review 40 (2021) 1–13.
[34] J. Warmenhoven, N. Bargary, D. Liebl, A. Harrison, M. A. Robinson, E. Gunning, G. Hooker, PCA of waveforms and functional PCA : A primer for biomechanics, Journal of Biomechanics 116 (2021) 110106.
[35] V. Ferreira, A. Pinho, D. Souza, B. Rodrigues, A New Clustering Approach for Automatic Oscillographic Records Segmentation, Energies 14 (2021) 1–18.
[36] Z. Zhang, A. Cheshmehzangi, S. P. Ardakani, A Data-Driven Clustering Analysis for the Impact of COVID-19 on the Electricity Consumption Pattern of Zhejiang Province, China, Energies 14 (2021) 1–22.
[37] C. Subbalakshmi, G. R. Krishna, S. K. Mohan, P. V. Rao, A Method to Find Optimum Number of Clusters Based on Fuzzy Silhouette on Dynamic Data Set, Procedia - Procedia Computer Science 46 (2015) 346–353.
[38] F. Aksan, M. Jasinski, T. Sikorski, D. Kaczorowska, J. Rezmer, V. Suresh, Z. Leonowicz, P. Kostyla, J. Szymanda, P. Janik, Clustering Methods for Power Quality Measurements in Virtual Power Plant, Energies 14 (2021) 1–20.
[39] H. Talei, D. Benhaddou, C. Gamarra, H. Benbrahim, M. Essaaidi, Smart Building Energy Inefficiencies Detection through Time Series Analysis and Unsupervised Machine Learning, Energies 14 (2021) 1–21.
[40] R. M. Alguliyev, R. M. Aliguliyev, L. V. Sukhostat, Parallel batch k-means for Big data clustering, Computers Industrial Engineering 152 (2021) 107023.
[41] F. Lu, J. Jiang, J. Huang, X. Qiu, An Iterative Reduced KPCA Hidden Markov Model for Gas Turbine Performance Fault Diagnosis, Energies 11 (2018) 1807.
[42] L. R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of the IEEE 77 (1989) 257–286.
[43] A. Simões, J. M. Viegas, J. T. Farinha, I. Fonseca, The State of the Art of Hidden Markov Models for Predictive Maintenance of Diesel Engines, Quality and Reliability Engineering International 33 (2017) 2765–2779.
[44] L. R. Rabiner, B. H. Juang, An Introduction to Hidden Markov Models, IEEE ASSP Magazine 3 (1986) 4–16.
[45] A. Kara, A data-driven approach based on deep neural networks for lithium-ion battery prognostics, Neural Computing and Applications 33 (2021) 13525–13538.
[46] Keras, Keras Tuner, ???? URL: https://keras.io/keras_tuner/.
[47] D. Goyal, B. S. Pabla, Condition based maintenance of machine tools — A review, CIRP Journal of Manufacturing Science and Technology 10 (2015) 24–35.
[48] B. Guo, S. Song, A. Ghalambor, T. R. Lin, An Introduction to Condition- Based Maintenance, in: B. Guo, S. Song, A. Ghalambor, T. R. Lin (Eds.), Offshore Pipelines, second edi ed., 2014, pp. 257–297. URL: https://doi.org/10.1016/B978-0-12-397949-0.00017-0https://www.sciencedirect.com/science/article/pii/B9780123979490000170. doi:10.1016/B978-0-12-397949-0.00017-0.
[49] J. Yu, Adaptive hidden Markov model-based online learning framework for bearing faulty detection and performance degradation monitoring, Mechanical Systems and Signal Processing 83 (2017) 149–162.
[50] F. W. Townes, S. C. Hicks, M. J. Aryee, R. A. Irizarry, Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model, Genome Biology 20 (2019) 1–16.
[51] W. Kwedlo, A clustering method combining differential evolution with the K -means algorithm, Pattern Recognition Letters 32 (2011) 1613–1621.
[52] R. J. Kuo, Y. R. Zheng, T. Phuong, Q. Nguyen, Metaheuristic-based possibilistic fuzzy k -modes algorithms for categorical data clustering, Information Sciences 557 (2021) 1–15.
[53] V. D. T. Thuy, T.-d. Do, P. N. Nguyen, A Data Mining-Based Approach for Exploiting the Characteristics of University Lecturers, in: Industrial Conference on Data Mining, July, 2016. URL: https://www.researchgate.net/publication/316890659%0AA. doi:10.

1007/978-3-319-41561-1_4.

[54] P. Arpaia, U. Cesaro, M. Chadli, H. Coppier, L. De Vito, A. Esposito, F. Gargiulo, M. Pezzetti, Fault detection on fluid machinery using Hidden Markov Models, Measurement 151 (2020) 107126.

# Appendix D

*energies*

MDPI

*Article*

# Predicting the Health Status of a Pulp Press Based on Deep Neural Networks and Hidden Markov Models

**Alexandre Martins [1,2,*], Balduíno Mateus [1,2], Inácio Fonseca [3], José Torres Farinha [3,4], João Rodrigues [1,2], Mateus Mendes [3] and António Marques Cardoso [2]**

1 EIGeS—Research Centre in Industrial Engineering, Management and Sustainability, Lusófona University, Campo Grande, 376, 1749-024 Lisboa, Portugal
2 CISE—Electromechatronic Systems Research Centre, University of Beira Interior, 62001-001 Covilhã, Portugal
3 Instituto Superior de Engenharia de Coimbra, Polytechnic of Coimbra, 3045-093 Coimbra, Portugal
4 Centre for Mechanical Engineering, Materials and Processes—CEMMPRE, University of Coimbra, 3030-788 Coimbra, Portugal
* Correspondence: alex_bmts@hotmail.com

**Abstract:** The maintenance paradigm has evolved over the last few years and companies that want to remain competitive in the market need to provide condition-based maintenance (CBM). The diagnosis and prognosis of the health status of equipment, predictive maintenance (PdM), are fundamental strategies to perform informed maintenance, increasing the company's profit. This article aims to present a diagnosis and prognosis methodology using a hidden Markov model (HMM) classifier to recognise the equipment status in real time and a deep neural network (DNN), specifically a gated recurrent unit (GRU), to determine this same status in a future of one week. The data collected by the sensors go through several phases, starting by cleaning them. After that, temporal windows are created in order to generate statistical features of the time domain to better understand the equipment's behaviour. These features go through a normalisation to produce inputs for a feature extraction process, via a principal component analysis (PCA). After the dimensional reduction and obtaining new features with more information, a clustering is performed by the K-means algorithm, in order to group similar data. These clusters enter the HMM classifier as observable states. After training using the *Baum–Welch* algorithm, the *Viterbi* algorithm is used to find the best path of hidden states that represent the diagnosis of the equipment, containing three states: state 1—"State of Good Operation"; state 2—"Warning State"; state 3—"Failure State". Once the equipment diagnosis is complete, the GRU model is used to predict the future, both of the observable states as well as the hidden states coming out from the HMM. Thus, through this network, it is possible to directly obtain the health states 7 days ahead, without the necessity to run the whole methodology from scratch.

**Keywords:** maintenance; diagnosis; prognosis; deep neural network; hidden Markov models; machine learning



**Citation:** Martins, A.; Mateus, B.; Fonseca, I.; Farinha, J.T.; Rodrigues, J.; Mendes, M.; Cardoso, A.M. Predicting the Health Status of a Pulp Press Based on Deep Neural Networks and Hidden Markov Models. *Energies* **2023**, *16*, 2651. https://doi.org/10.3390/en16062651

Academic Editor: Daniel Morinigo-Sotelo

Received: 13 January 2023
Revised: 2 March 2023
Accepted: 9 March 2023
Published: 11 March 2023



## 1. Introduction

The mechanical systems of companies' production equipment suffer degradation and the remaining useful life (RUL) shortens as the equipment components deteriorate over time [1]. With the current evolution of production equipment, its complexity has also increased. Thus, modern industrial systems are often exposed to various failure modes [2]. An unexpected stoppage of equipment can cause great economic losses to a company and/or even put the health of workers at risk.

### 1.1. Maintenance

According to Kumar et al. [3], it is estimated that 85% of the total life-cycle cost of equipment is determined by decisions made during its operation, where maintenance actions are included. Therefore, it becomes necessary to obtain a maintenance system

adaptable to the criticality of the equipment [4]. "Chronologically, maintenance strategies have evolved from the naive breakdown or run-to-failure maintenance to preventive maintenance, with first static time-based preventive maintenance, and then condition-based preventive maintenance (CBM)" [5]. Thus, maintenance can be divided into three major categories: corrective maintenance (CM), time-based maintenance (TBM) and condition-based maintenance (CBM) [6]. CM is a maintenance strategy that acts on the system only to repair or replace the failed components. It is a maintenance policy that can cause unexpected equipment downtime, production stoppages and low safety and is only used for low-criticality equipment [7]. TBM and CBM are seen as types of proactive maintenance since they are done before the failure occurs. TBM is based on reliability parameters of the equipment components and its maintenance is usually done periodically and before the failure occurs, which can lead to excessive maintenance. On the other hand, CBM depends on the condition of the system being monitored, and interventions only occur when they are really necessary. It has many advantages over other strategies [8].

### *1.2. Condition-Based Maintenance—CBM*

CBM is a strategy that is based on condition monitoring of equipment over its operations, enabling just-in-time maintenance [8]. According to Leoni et al. [9], one of the main advantages over preventive maintenance strategies is that it can be scheduled only based on the condition of the equipment, with no need to plan interventions based on reliability parameters. Lee et al. [10] mention that although implementing a CBM strategy is more difficult and costly, it ultimately leads to less wasted equipment life, reducing costs associated with production, parts stock, number of personnel, tools, etc. It also increases equipment availability and decreases the number of unexpected stops, improving failure prevention and simultaneously reducing the operational cost [11–13]. According to Zhang et al. [14], a CBM policy makes better use of system information, reduces the risk of failures and unnecessary maintenance. Therefore, CBM is a maintenance methodology, which, once implemented, is more cost effective than traditional methodologies, which helps to improve equipment reliability, reduce operating costs, improve safety and reduce the frequency and severity of equipment failures [12,15–20]. In addition to the points enumerated above, with the advancement of technology and the arrival of industry 4.0., companies are becoming more and more sensorized, and it becomes evident that CBM is the maintenance that is most suitable and can achieve the best results [17,21–26]. CBM is used in smart factories through the use of IoT systems, CPS (cyberphysical systems), sensor technology and AI technologies. In this way, the collection and processing of large quantities of data (big data) can be done in real time [9,10,27,28].

CBM is the maintenance responsible for detecting, through condition monitoring, the first signs of anomalies in [29] equipment. In this way, it is possible to determine when maintenance is required based on the actual condition of the equipment. Through condition monitoring, it is possible to inspect and assess the current state of a piece of equipment through data collected in real time from different types of sensors (e.g., vibration, noise, temperature, etc.) [1,16,30,31]. The better the adequacy of equipment degradation modelling as well as the better the monitoring accuracy, the better the CBM strategy becomes [11]. Moreover, the reliability of the collected data is very important for this type of maintenance, hence the need to have calibrated sensors that make reliable measurements [32,33].

Summarizing, CBM can be divided into three phases [9,10,18,34]:

(i) Data acquisition, where useful data are collected, usually from sensors;

(ii) Data processing, where noise components are filtered out of the physical observations and subsequently, a data analysis is performed;

(iii) Maintenance decision-making.

### *1.3. Diagnosis and Prognosis*

CBM can be divided into two types of decision: failure diagnostics, which is the ability to detect a cause of failure; failure prognostics, which is responsible for detecting a failure that may occur in the future [35]. The prognosis of the health status of a piece of equipment has become an unavoidable concept in the context of today's industry, where intelligent manufacturing and industrial big data provide input solutions for maintenance [36,37]. Then, through a CBM strategy, it is also possible to make a prognosis of the state of the equipment, having the indication of possible problems and/or failures that could happen in the future [38]. Effective prognostic techniques that can anticipate future conditions are integrated into an advanced predictive maintenance framework that is incorporated into the condition monitoring system [1]. We then speak of predictive maintenance (PdM), whose aim is to predict the condition diagnosis of the equipment.

### *1.4. Predictive Maintenance PdM*

PdM is a CBM strategy, but instead of diagnosing the equipment, it performs a real-time prognosis [5]. Thus, as the authors pointed out, PdM is more efficient at planning maintenance actions than a CBM policy. PdM improves productivity, product quality and overall manufacturing efficiency [24]. According to Harald et al. [39], predictive maintenance can reduce machine downtime by 30% to 50% and extend machine life by 20% to 40%. Therefore, it is possible to think of predictive maintenance as being more accurate than condition-based maintenance [40]. The aim of CBM is to identify deviations or substantial modifications that are usually signs of a development failure. Consequently, condition-based monitoring serves as a crucial pillar of predictive maintenance [40].

By monitoring a machine's operating conditions in real time, it is possible to detect key patterns for future failure prediction, thus predicting maintenance actions. According to Yam et al. [38], an "alarm" can be given when values outside the normal operating range are predicted, allowing system operators to take appropriate action to check machine conditions and repair faults before a more critical failure, hence the need for companies to acquire an equipment-condition-based maintenance strategy with a focus on prediction. Then, a predictive maintenance strategy is also based on equipment condition monitoring but integrates prognostics to effectively conclude on the state of equipment health in the future [1]. According to Oakley et al. [17], predictions based on information extracted from values collected by sensors have helped to improve decision-making regarding the maintenance of equipment.

### *1.5. HMM-GRU to Perform Maintenance*

Normally, the phenomena of the degradation evolution of equipment components have a stochastic nature and can be described by stochastic modelling processes [19,34,41]. These stochastic models are modelled based on probability and statistical theories and can be related to component degradation and failure occurrences [41]. In this paper, we use a doubly stochastic process, the hidden Markov model (HMM), to characterise the state of equipment degradation. The objective of the HMM is to classify the health state of the equipment. To do this, we use the data collected by the sensors responsible for monitoring the equipment. These types of data are characterized by time series data, since they are collected over time in a continuous manner. Time series data are the most collected type of data in this new era of industry 4.0. These observations collected over time can be analysed and used as a tool that prevents unexpected equipment failures [42]. Thus, it is essential to use machine learning (ML) tools, which is a field of AI that extracts key patterns from collected temporal data through different paradigms such as: supervised learning, semisupervised learning, unsupervised learning and reinforcement learning. According to Kiangala and Wang [42], deep learning is a branch of ML consisting of different methods such as: artificial neural network (ANN), convolution neural networks (CNN), long short-term memory (LSTM), etc. In this paper, we will use the Gated Recurrent Units (GRU) tool, which is like an LSTM network with a forgetting gate. This is currently a well-recognized

prognostic tool and is the tool responsible for making the prediction of future data later classified by the HMM to obtain the health status of the equipment, thus performing a prognosis and supporting a PdM methodology.

### *1.6. Related Work*

There are several examples of papers supporting PdM, which use DNN to make predictions of the future, as well as other papers that use HMM to classify the state of a system. Mateus et al. [43] made a comparison about a predictive system using an LSTM network or a closed recurrent unit (GRU) for a multivariate data set. Antunes et al. [44] proposed in their paper a variation of the exponential smoothing technique for short-term forecasting and an artificial neural network for long-term forecasting. Mateus et al. [45] presented in his article predictive models using an LSTM network to predict future equipment status based on data from an industrial paper press. Zhang et al. [46] proposed an approach to perform a prognosis of rotating equipment's health using wavelet transform (WT), a principal component analysis (PCA) and artificial neural networks (ANN) to classify the failure and predict the condition of components, equipment and machines. Martins et al. [47] showed how it was possible to classify the health condition of equipment via an HMM with multivariate analysis. Yu [48] proposed an adaptive HMM method that evolved over time to detect equipment failures and component degradation monitoring. Arpaia et al. [49] presented a fault detection method using HMM for fluid machinery without a priori information about its failure conditions. Mateus et al. [50] presented a case study where they determined the future behaviour of the data collected by sensors coupled to industrial equipment; the authors used time series models and deep learning.

### *1.7. Contributions of the Paper*

In our case, the article stands out for integrating a diagnosis methodology that uses an HMM with a forecasting tool, the GRU. In this way, the methodology explained in this article demonstrates how a diagnosis and prognosis of the production equipment's health status can be made in an online mode. The added value of the methodology is that it is generic and can be used in any equipment with different sensors. Furthermore, it is a methodology that detects unusual patterns in the equipment's operation without prior information using unsupervised ML tools. Furthermore, the GRU predicts optimised observable states after going through a process of feature generation, PCA and clustering. That is, the GRU does not make the prediction directly on the data collected by the sensors but on a data set optimized by ML processes. Moreover, a future prediction is made with the GRU algorithm directly on the hidden states classified by the HMM. Thus, the methodology does not only make a future prediction about the observable data that need to be classified but also a prediction of the classification itself performed by the HMM.

### *1.8. Paper Structure*

This article is divided as follows: Section 1 gives an introduction with the objective to explain the context of the article, where maintenance concepts are described and where the tools used to accomplish the diagnosis and prognosis of equipment are introduced, as well as related works. In Section 2, a theoretical framework of each of the tools used in this article is presented. Section 3 talks about the methodology used and explains how it can be performed both to make a diagnosis of the equipment using the HMM, as well as a prognosis using the GRU. Section 4 describes the case study carried out in a production equipment used in the paper industry, where the objective was to realize 3its diagnosis and prognosis. In Section 5, a discussion of the results is provided, where an evaluation of the methodology is performed to understand if it works. Finally, in Section 6, the conclusion of the work is drawn.

## 2. Background

### *2.1. Principal Component Analysis (PCA)*

A principal component analysis (PCA) is an unsupervised learning method of feature extraction and dimensional reduction (moving p-dimensional data to a lower-dimensional m-dimensional linear subspace), retaining the original features of the data and selecting their key properties [42,51–53]. It analyses a data table in which observations are described by several intercorrelated quantitative dependent variables and is widely used due to its ability to extract interpretable information by efficiently removing redundancies [54,55]. It is typically used to perform the dimensional reduction of large sets of time series observations [56], moving from representing possibly correlated variables to a new set of orthogonal, uncorrelated variables and preserving the highest percentage of information [40,46,55]. In this way, it allows a rapid assessment of any relationships between variables [54]. In other words, it is a method of projecting large dimensional measurements towards a minimum dimensional space and preserving maximum variance [57] by compressing sensory data according to their spatial and temporal correlations [58]. Then, the PCA produces linear combinations of the original variables to generate new axes, known as principal components (PCs), with the first PC having as high a variance as possible, possessing the greatest variability in the data, and each subsequent component in turn having as high a variance as possible under the constraint that it is not correlated with the previous components [46,54]. In other words, the PCA is a linear transformation that rearranges the data into a new coordinate system, in which the first PC is defined as the coordinate that shows the greatest variation in the data when projected in that direction; the second PC is the coordinate that presents the second largest variation, and so on for the other components [54].

### *2.2. K-Means Algorithm*

The K-means algorithm is a clustering, nonhierarchical, unsupervised learning method in ML, from the branch of multivariate statistical analysis, where the number of clusters K is determined, and the observations closest to the cluster centre are included in that same cluster [59–63]. Clusters are formed so that the distribution of samples among clusters maximises intragroup cohesion, i.e., the distance of observations from their centroid, increasing similarity within the same cluster, and dissimilarity between different clusters [63–66]. K-means clustering is used to perform the classification of unlabelled data in which the specific response variables are unknown [67]. In this process, the similarity between observations in the same cluster increases and the similarity with data from other clusters decreases [67,68].

The K-means algorithm uses the distance between data points as the standard measure of data similarity, usually using the Euclidean distance [60,69] (which is the one used in this paper). That is, it minimises the square sum of the distances from each data point to its assigned grouping centre [70]. The Euclidean distance equation is represented below in Equation (1) [60,61,69,71].

$$dist(x_i, x_j) = \sqrt{\sum_{d=1}^{D} (x_{i,d} - x_{j,d})^2} \quad (1)$$

where $x_i$, $x_j$—are two sequence points and $D$ represents the dimension.

The K-means approach can be quite sensitive to the initial value chosen for the number of clusters (k), which, if improperly defined, can significantly affect the result of the clustering process, the number of iterations needed to reach algorithm convergence, as well as the accuracy and complexity of the clustering algorithm [59,60,68]. Therefore, to minimise the influence of the initial choice of $k$ and define the optimal number of clusters, it is common to use decision support methods. The Elbow method is a technique used for this purpose, in which the cost function is calculated for different values of $k$ during the clustering process [61]. Usually, it is plotted a graph of the cost function as a function of

different values of k is plotted and, from this graph, it is possible to identify the "elbow" point that indicates the optimal number of clusters.

$$E_k = \sum_{r=1}^{k} \frac{1}{nr} Dr \tag{2}$$

where $k$ denotes the size of the cluster, $nr$ represents the number of data points in the cluster and $D_r$ is the sum of the distances between all points within the cluster.

### *2.3. Gated Recurrent Unit (GRU)*

The GRU was introduced by Cho et al. [72] and is based on LSTM, with a relatively simpler structure requiring fewer parameters for its formation, having only two gates (Figure 1) [73–79]: the reset gate is a mechanism that can be used to help with model encryption, allowing one to determine the amount of past information that can be forgotten; the update gate, on the other hand, is responsible for combining the entry and forgetting gates in an LSTM model, allowing one to determine the degree of the previous hidden state that will be used to update the current state. Both the reset gate and the update gate are mechanisms used to solve the leak gradient problem in neural networks, since they allow the manipulation of the information in intermediate layers without losing relevant information for future predictions [76]. According to the authors, what makes these mechanisms special is their ability to maintain long-term memories, without removing relevant information for future predictions, which enables a better model performance.

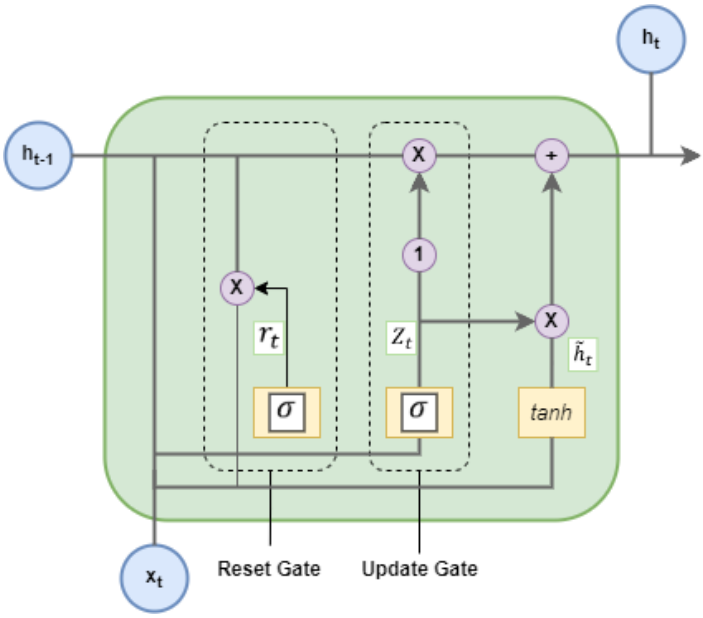



**Figure 1.** Recurrent Gated Neural Network.

Then, the GRU has an update gate $z$ and a reset gate $r$ to simplify the memory block structure of the original LSTM network. The input of the GRU network being $x_t$, the formula to calculate the next output and state value in GRU is [73,75–77,80,81]:

1. The algorithm starts with the calculation of the update gate $z_t$ for time step $t$ (Equation (3)):
$$z_t = \sigma(W_z.[h_{t-1}, x_t]); \quad (3)$$
when connecting the network unit, the value $x_t$ is multiplied by its respective weight $W_z$, as well as the value $h_{t-1}$, which contains the information of the previous units in $t-1$. The results of these multiplications are then summed, and an activation function *sigmoid* is applied to normalise the result between 0 and 1. In this way, the relevant information is kept and irrelevant information is filtered out.
2. The reset gate allows the model to determine how much past information should be forgotten, thus controlling how much information is retained (Equation (4)):
$$r_t = \sigma(W_r.[h_{t-1}, x_t]) \quad (4)$$
3. In this step, new memory content is introduced that uses the reset gate to store the most important information from the past. After obtaining the toggle signal, the toggle reset activation function is used to obtain the reset data and combine them with the *tanh* activation function, resulting in $\tilde{h}_t$.
$$\tilde{h}_t = tanh(W_h.[r_t * h_{t-1}, x_t]) \quad (5)$$
*tanh* is capable of controlling the range of output values between $-1$ and 1. It is possible to observe that the input data are incorporated and the hidden information is regulated by the *tanh* activation function.
4. Finally, the vector $h_t$ is calculated to contain the relevant information of the current unit and transmit it to the next stage of the network, determining what should be kept from the previous stages $h_{t-1}$. The end result, $h_t$, contains the current unit and previous step information that is relevant to the final output (Equation (6)):
$$h_t = (1 - z_t) \times h_{t-1} + z_t * \tilde{h}_t \quad (6)$$

Through direct training, it can be inferred that the state of the reset gate $r_t$ controls the combination between the input $x_t$ and the previous state $h_{t-1}$. On the other hand, the $z_t$ update gate determines the use of current and previous-time information.

### *2.4. Hidden Markov Models (HMMs)*

An HMM is a statistical modelling technique widely used to model sequential data such as time series. Its dynamic Bayesian network structure is relatively simple, but it can capture complex patterns of temporal dependence between observable variables and latent (unobservable) variables [82]. It is developed on the basis of the Markov chain, which is a discrete memoryless random process responsible for describing the relationship between the sequence of states of the next moment with the current one [83–85]. An HMM is an evolution of a Markov chain that requires two stochastic processes, adding a random relationship between the sequence of states and the observation vector, and where the sequence of states cannot be directly observed [83,84,86–89]. Then, an HMM is a probabilistic time series model, doubly stochastic, which includes the transition of hidden states and emitting observations [90]. The hidden state transition, which follows Markov chains, is the actual state within the system, mapped by observable states, which are directly observed and have a correlation with the hidden states [90–93].

A typical HMM can be expressed by $\lambda = (N, M, \pi, A, B)$ [82–84,90,91,93,94], where:

- $N$ represents the number of hidden states, where a certain state $q_t \in S, S = S_1, S_2, ..., S_N$;
- $M$ is the number of observable states, where the observation at time $t$ corresponds to the hidden state $q_t$ and is represented by $O_t \in V, V = V_1, V_2, ..., V_M$;
- $\pi$ is the initial distribution of the hidden states, where $\pi_i = P(q_1 = S_i)$;
- $A$ is the matrix of transition probabilities between hidden states, where $A = a_{ij}{}_{N*M}$ and where $a_{ij}$ equals:

$$a_{ij} = P(q_{i+1} = S_j | q_i = S_i), 1 \leq i, j \leq N \quad (7)$$

- $B$ is the emission matrix, where the probability that the $j$th hidden state generates the $i$th observable state is represented, where $B = b_{jk}{}_{N*M}$. $b_{jk}$ is represented by:

$$b_{jk} = P(O_t = V_k | q_t = S_t), 1 \leq j \leq N, 1 \leq k \leq M \quad (8)$$

There are three problems that need to be solved to use an HMM in an integrity assessment and health-status prediction of equipment [83,93]:

- Evaluation

Using the model $\lambda = (\pi, A, B)$ and an observation sequence $O = O_1, O_2, ..., O_T$, the probability of the observation sequence $O$ is determined by the forward–backward algorithm. This algorithm evaluates the model with the most suitable observation.

- Training

It consists of training the model with the HMM, where given a sequence of observations, the model parameters are re-evaluated to maximise the likelihood. The Baum–Welch algorithm is a learning technique for tuning the parameters of an HMM model based on observed data, establishing the relationship between the parameters of the old and new HMM model, and continuing to iterate until convergence is achieved. Specifically, the HMM elements, $\lambda = (\pi, A, B)$, are identified such that the likelihood, $P(O|\lambda)$, of the observation sequence O given the $\lambda$ model is maximised [87,92].

- Decoding

The ideal state sequence is created in order to ensure that the likelihood will reach its highest value given the model, $\lambda$, and the observation sequence, $O$. Viterbi's method solve the HMM decoding problem, inferring the HMM's most likely series of hidden states, $S$ [91].

## 3. Methodology

In order to explain the procedure and methodology used in the case study presented in this article, we can follow Figure 2, where the whole process used is schematized.

First, several data were collected over time through several sensors, $X_p(t)$, attached to the equipment. These data subsequently went through a data cleaning phase whose goal was to eliminate/replace everything that decreased the integrity of the set to analyse. After that, they passed to an optimization phase of observable states, with the objective of improving the observations that mapped the hidden states of the HMM. In this way, the optimization of observable states started with a feature creation of the data set, to obtain more information about the data collected from the equipment, since continuous and variable features over time can provide a prediction of possible failures [51]. By comparing the extracted features to the original signal, the feature extraction sought to gather more precise data that increased the accuracy of the performance assessment [1]. In order to generate features and consequently reduce the dimension of the data, a time-window processing method [51] was used. Time windows with different intervals can be created depending on the study to be performed. A feature generation method was applied in the time domain, where several features were created in each of the time windows. The features chosen for the characterization of the equipment failure behaviour were chosen

according to the articles [47,95] as they showed that they related well to the detection of deviations in equipment behaviour.

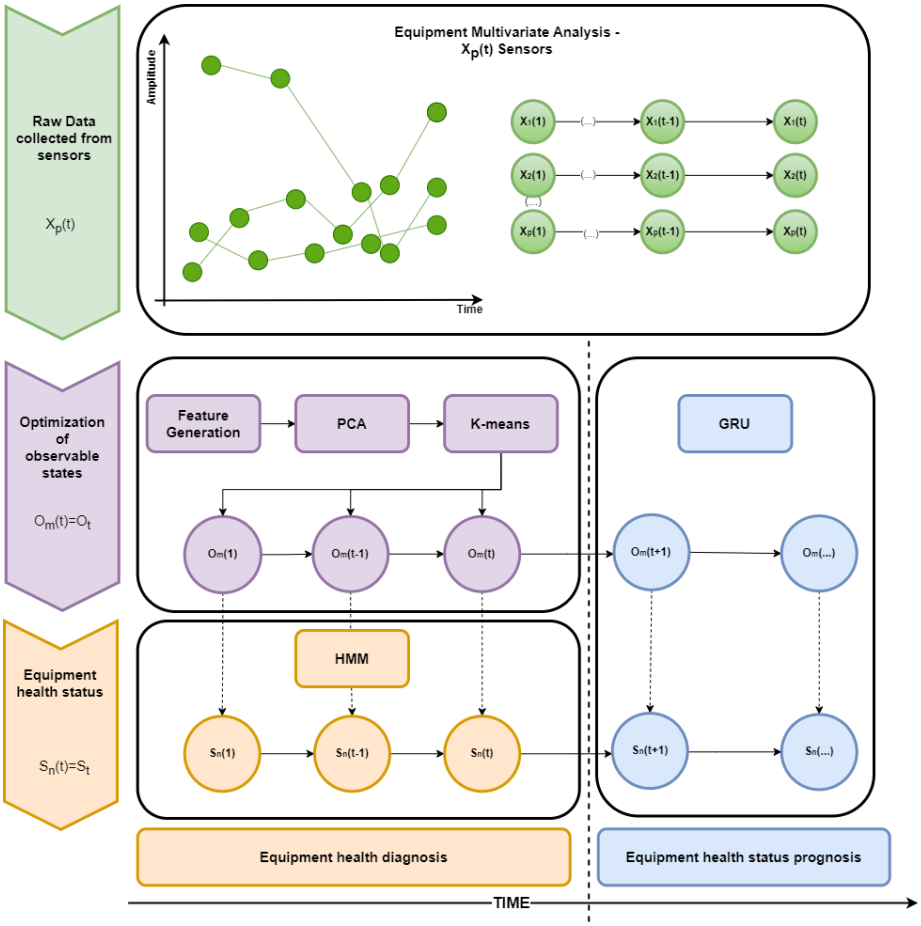



**Figure 2.** Methodology used to make a diagnosis and prognosis of the equipment through HMM-GRU.

After creating the time-domain features, a feature selection and dimensional reduction process, namely PCA, was used. Here, the aim was to work only with important features, the principal components, as well as to solve the curse of dimensionality by reducing the number of features and thus increasing the speed. From the various existing data reduction methods, PCA-based data compression is the most widely used technique and has superior performance in terms of reconstruction error [58]. From the point of view of equipment health diagnosis, the reduction of class representation speeds up the decision phase [57]. Moreover, as [96] explained, highly correlated features lead to overfitting, and the PCA technique is applied to remove the highly correlated features based on the correlation matrix, thus increasing the prediction accuracy. Without increasing the computational complexity, it considers potential correlations between the answer variables [97], transforming the data into uncorrelated features that assist in converting the data from a high-dimensional to a low-dimensional space, retaining the maximum amount of information [98]. As there were large differences between the ranges of the variables that were provided as inputs to the PCA, a normalisation of the data was first performed. Through normalisation, the amplitude of the initial continuous variables contributes equally to the analysis. Features with larger amplitudes do not overlap features with smaller amplitudes, thus not leading to one-sided results [54]. Therefore, the data were first transformed into comparable scales. This was performed using a *Zscore* normalisation (Equation (9)), where the mean was

subtracted and divided by the standard deviation for each value of each feature, causing all features to be standardized to a zero mean and one standard deviation.

$$Z_{Score} = \frac{x_i - \bar{X}}{std_{dev}(X)} \quad (9)$$

After the normalisation, the data were fed to the PCA that was responsible for transforming the orthogonal orientations moving them in the direction of the bigger variability of the data. These new vectors, known as principal components, can be thought of as new axes that offer the best perspective for visualizing and analysing data in order to make the differences between findings more obvious. The original answer variables are converted into uncorrelated principal components after the original data are mapped to a new vector [97]. It is also a widely used technique when, as is the case with the methodology in this paper, it is necessary to check the clustering trend of the data. PCA, in an effort to preserve all pertinent information, employs projection methods from high-dimensional spaces to lower-dimensional subspaces [99].

Once the PCA phase was over, the new variables, the principal components, were input into the clustering process, which is a multivariate analysis technique to judge the degree of similarity between objects in order to classify them [59,70]. This was done over time in order to understand the clusters that were forming. The clusters were sorted in descending order with cluster 1 having the most data, cluster 2 having the second most data and so on. Cluster 1 was the cluster that appeared most often over time since it was the one with the most data and the last cluster was the rarest. To perform this step, K-means clustering was used. K-means clustering is an unsupervised learning algorithm, used to highlight the intrinsic properties and laws of the data [69]. Among the various existing clustering types, K-means clustering was chosen because it is [65,69] relatively simpler with an easy implementation and fast convergence, it has a strong interpretation ability and it can handle a large number of observations efficiently. According to [100], K-means clustering is the most useful tool for data mining, summarization, probability density estimation and many other essential tasks. Due to the findings' clarity and high scalability, it has recently become one of the most widely used algorithms in data analysis [101]. It is suitable for the reduction of large-scale original failure scenarios [66]. To cluster data, first, it defines the number of clusters, $k$, and then the data sets are assigned to each of the $k$ clusters. The goal is to minimise the square sum of the distances from each data point to its assigned cluster centre. The distance between the data points and the centre of the clusters is calculated in order to assign the data to the nearest clusters. This is repeated for several iterations until convergence is reached [70]. With this analysis, we only had observations optimized by clusters where each cluster represented one observation. These new observations served as the HMM's inputs to train the model and conduct equipment diagnosis. The HMM fitted well with the detection method whose spectrum states were unknown but the receiver could be determined [93]. In this case, the collected and processed observations came from the sensors and the hidden states that represented the health state of the equipment. The objective of the HMM was, through the observable states, to determine which hidden states best applied to each observation over time. To this end, three hidden states were defined that represented the diagnosis of the health state of the equipment: hidden state 1 represents the "State of good functioning"; state 2 represents the "Alert state"; state 3 represents the "State of bad functioning". To use the HMM as a classifier, the HMM parameters whose sizes are determined in advance must be determined by training [90]. Thus, first, the Baum–Welch algorithm was used to train and update the initial parameters of the HMM, $\lambda$, that could explain the observation sequence. That is, the HMM parameters were identified such that the probability, $P(O|lambda)$, of the observation sequence O given the *lambda* model was maximised [92]. Based on the observed state sequence, the HMM training first calculated the maximum probability of the model parameters [91]. After obtaining the parameters of the HMM, $\lambda$ , finally, the diagnosis of the equipment was performed through the Viterbi algorithm

that indicated which hidden states best applied over time. That is, the Viterbi algorithm used dynamic programming to find the maximum likelihood path and finally performed a prediction on the HMM [91]. It is a dynamic programming algorithm to find the most likely sequence of hidden states called the Viterbi path that results in a sequence of observed events [85]. The HMM was used to make the diagnostic classification of the equipment, since it is a model suitable for continuous dynamic signal processing and its function is able to discover the hidden state with a higher probability, through a sequence of observations (in our case, coming from the clustering) [83]. According to the author, in statistical learning theory, an HMM is most efficient in pattern recognition processing. HMMs have often been used for recognising changing behaviours of dynamic features of a system [90], modelling time-series-based phenomena due to their computational efficiency and because they can be used to build data-driven models that provide characteristic indicators [92] and modelling nonstationary and complex random physical processes of machine condition deterioration; the hidden Markov model (HMM) is able to perform both monitoring and diagnosis [102].

Once the diagnosis of the equipment was made, it was possible to move on to a prognosis with the aim of predicting the condition of the equipment in the future, using a DNN algorithm for the time series prediction. According to the literature, deep learning has become an active and promising area of research, and the most widely used deep learning algorithms are the recurrent neural network (RNN), the long short-term memory (LSTM), the convolution neural network (CNN), and the gated recurrent unit (GRU) [81]. In this paper, the GRU recurrent neural network was used, which is a simplification of the LSTM architecture, being able to train more quickly because there are fewer parameters to modify. Furthermore, it can also avoid gradient leakage issues [103]. According to recent research, recurrent units are simpler, necessitating the use of RNNs with smaller memory requirements and less demanding training algorithms. GRU uses the so-called update and reset gate, which are the two vectors that determine what information should be passed to the output, addressing the gradient leakage issue of a standard RNN [81]. According to the author, the GRU's unique quality is that they can be taught to retain information over the long term without forgetting it or deleting information that is unrelated to the prediction. The goal here was to use this neural network to make a prediction of a few days both on the optimised observable states of the K-means clustering and directly on the hidden states of the HMM. In this way, it was possible to see in which of the cases we obtained better values from the GRU model. Since the predictions were made on smaller optimized states, the GRU network was chosen since it showed better results on smaller and less frequent datasets [104,105]. The advantage of GRU cells is that they are as powerful as LSTM cells, even for small datasets [105].

## 4. Case of Study

### *4.1. Data Preparation*

For this study, we used data collected from a company in the paper industry. More specifically, data were acquired from a pulp-drying press, whose objective is to remove moisture from the pulp. This is a very important process in the production flow, so this equipment needs a type of PdM maintenance. For this, six sensors attached to the equipment were used, collecting observations continuously (every 5 min) over time. Each of the sensors collected values of different magnitudes, as follows: current, hydraulic level, torque, pressure, rotation speed and temperature. With these magnitudes, it was possible to obtain a better picture of the state of the health of the equipment. The data had three years of history, and we used 83,329 data points collected for each of the six sensors (Figure 3).

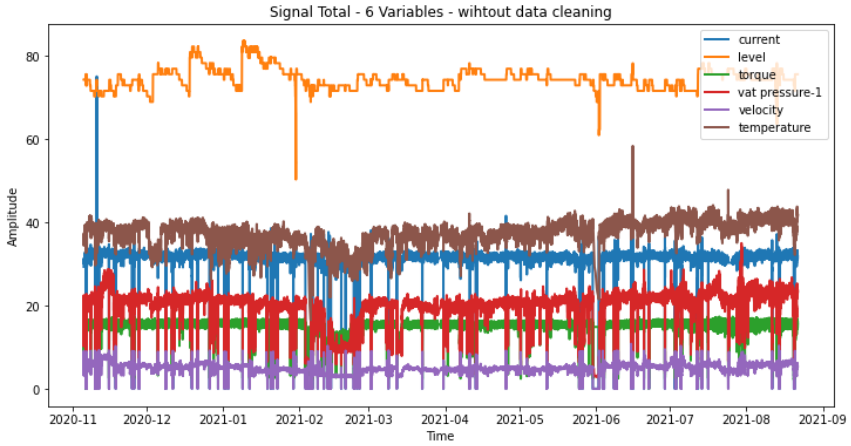



**Figure 3.** Amplitude of each of the variables under study over time without data preparation.

The quality of the data collected is an aspect to be taken into account, since it may present some flaws in automated data collection processes. This causes inaccurate or incorrect data to be collected. To find meaningful information from big data, it is essential to perform a preprocessing of the data [106]. This step is of utmost importance to ensure reasonable results, whether it is analysis with exploratory data mining, classification or building a good and robust predictive model. Thus, data cleaning was performed to remove data incoherence and increase data integrity. To do this, a program was created that replaced duplicate data, nonexistent data and zeros with the average of their sign. It was chosen not to remove what could be outliers, since they could represent a real malfunction of the equipment and therefore add value to the prediction. We also replaced the equipment stoppages (by the respective average of each signal), since these could be confused with malfunctioning and could reduce the effectiveness of the prediction. A program was created that, when detecting that current, torque, pressure and speed were below a certain threshold at the same time, would be seen as a shutdown of the equipment (Figure 4). The data used for the study are represented in Figure 5.

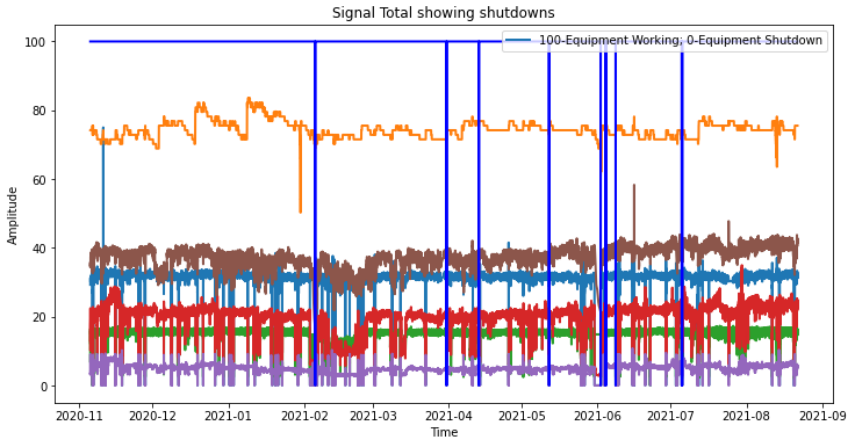



**Figure 4.** Equipment stoppages over time.

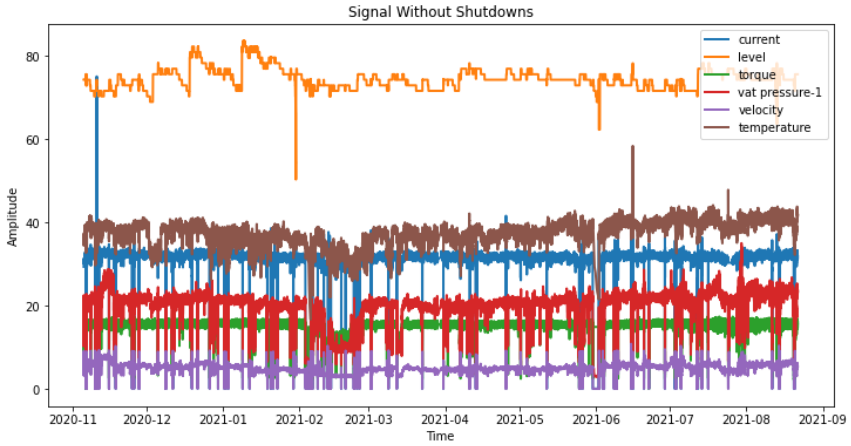



**Figure 5.** Final signal used for the equipment health status study.

### *4.2. Feature Generation*

After increasing the integrity of the data, they were divided into temporal windows with the aim of creating various time domain features (Table 1).

**Table 1.** Mathematical equations for time domain-based statistical features.

| Parameter | Mathematical Equation | Parameter | Mathematical Equation |
|---|---|---|---|
| Mean | $T_1 = \frac{\sum_{n=1}^{N} x(n)}{N}$ | A Factor | $T_{12} = \frac{T_5}{T_2.T_3}$ |
| Standard deviation | $T_2 = \sqrt{\frac{\sum_{n=1}^{N} (x(n)-T_1)^2}{N-1}}$ | B factor | $T_{13} = \frac{T_7.T_8}{T_2}$ |
| Variance | $T_3 = \frac{\sum_{n=1}^{N} (x(n)-T_1)^2}{N-1}$ | SRM | $T_{14} = \left(\frac{\sum_{n=1}^{N} \sqrt{\lvert x(n)\rvert}}{N}\right)^2$ |
| RMS | $T_4 = \sqrt{\frac{\sum_{n=1}^{N} (x(n))^2}{N-1}}$ | SRM shape factor | $T_{15} = \frac{T_{14}}{T_1}$ |
| Absolute maximum | $T_5 = max\lvert x(n)\rvert$ | Latitude factor | $T_{16} = \frac{T_5}{T_{14}}$ |
| Coefficient of skewness | $T_6 = \sqrt{\frac{\sum_{n=1}^{N} (x(n)-T_1)^3}{(N-1).T_2^3}}$ | Fifth moment | $T_{17} = \sqrt{\frac{\sum_{n=1}^{N} (x(n)-T_1)^5}{(N-1).T_2^5}}$ |
| Kurtosis | $T_7 = \sqrt{\frac{\sum_{n=1}^{N} (x(n)-T_1)^4}{(N-1).T_2^4}}$ | Sixth moment | $T_{18} = \sqrt{\frac{\sum_{n=1}^{N} (x(n)-T_1)^6}{(N-1).T_2^6}}$ |
| Crest factor | $T_8 = \frac{T_5}{T_4}$ | Median | $T_{19} = medianx(n)$ |
| Margin factor | $T_9 = \frac{t_5}{T_3}$ | Mode | $T_{20} = modex(n)$ |
| RMS shape factor | $T_{10} = \frac{T_4}{\frac{1}{N}\sum_{n=1}^{N} \lvert x(n)\rvert}$ | Minimum | $T_{21} = minx(n)$ |
| Impulse factor | $T_{11} = \frac{T_5}{\frac{1}{N}\sum_{n=1}^{N} \lvert x(n)\rvert}$ | | |

The data were then divided into 6 h windows, in order to cover four daily operating shifts. In total, 1158 temporal windows were created, where in each one there was a set of 72 data. Then, for each window, 21 characteristics were taken for each one of the six sensors, producing a matrix of 1158 × 132.

### *4.3. PCA*

The created features went through the PCA method in order to reduce the dimension and also generate new features, the PCs, to use later in the K-means algorithm. In this way, we increased computational speed and worked only with characteristics that were really important for the study, which, although few in number, preserved most of the information. To do this, a $z - score$ normalisation of the data was first performed, since they had different amplitudes (Figure 6).

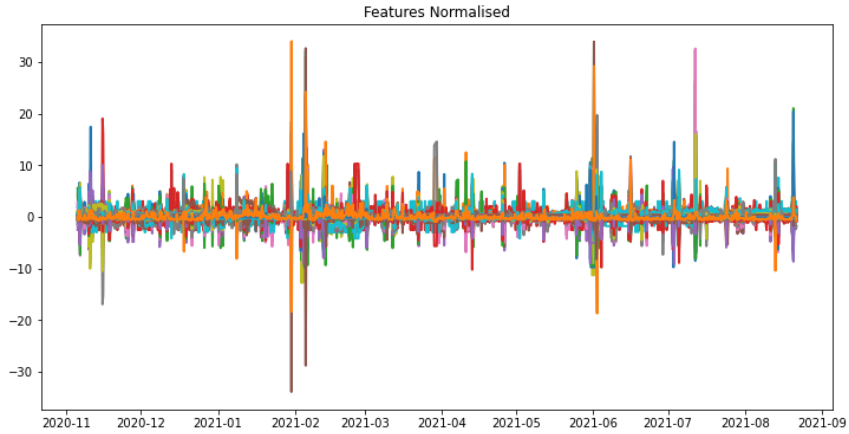



**Figure 6.** All features normalised over time.

The normalised features were then fed into the PCA. Through the study of eigenvectors and eigenvalues, we verified that 10 PCs preserved about 85% of the data variance, as we can see in the Pareto chart represented in Figure 7.

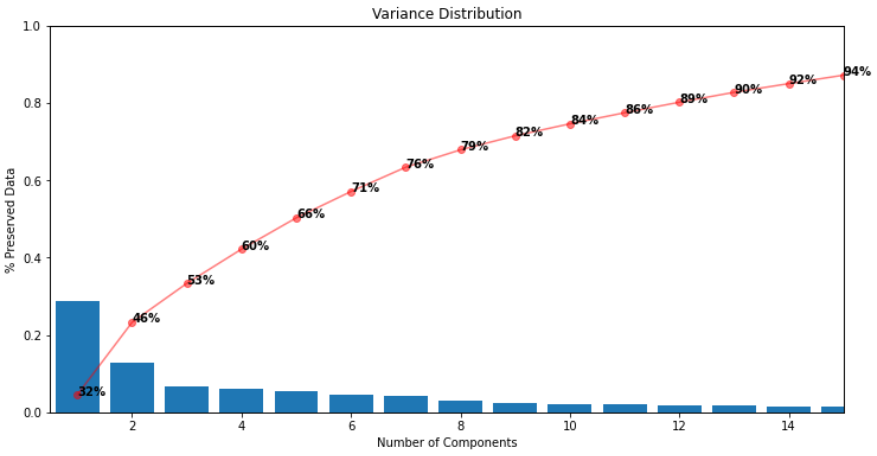



**Figure 7.** Pareto graph with the percentage of preserved information vs. the number of PCs.

### *4.4. K-Means Clustering*

The matrix with the new characteristics, the PCs, resulting from the PCA were sent to the clustering process where the objective was to group the data points that most resembled each other in the same group. To do so, K-means clustering was used. The K-means clustering algorithm followed the following procedure [33,59,65,71,101,107]:

Step 1. Determine the initial number of clusters $k$.

Step 2. Randomly select the initial $k$ centroids $c_j, j = 1, 2, ..., k$ in the observations.

Step 3. Calculate the distance between observation and the initial centroid, and assign the observed object to the cluster closest to the result, using Equation (1)

Step 4. Define a new centroid based on the average of the cluster variables (Equation (10)).

$$c_j = \frac{1}{N_j} \sum_{x_i \in S_j} x_i \tag{10}$$

Step 5. Repeat Step 3 using the new centroid until the observed objects are not relocated to another cluster.

This is an iterated process, which iteratively moves the centroids to minimise the total variance within the cluster [67], with two conditions for terminating the iteration [61,65,69]: the specified number of iterations is reached; the cluster centre no longer changes. This paper used the second form. As found in the theoretical framework, to start the K-means method, it is necessary to indicate the number of clusters, $k$. This was done using the elbow method, one of the most used methods to select the number of clusters, using for this the error sum of squares (SSE) vs. the number of clusters (Figure 8).

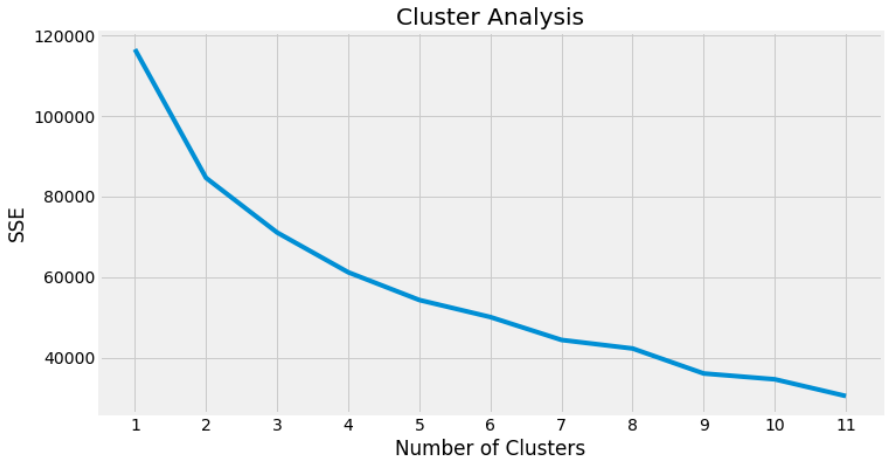



**Figure 8.** Elbow SSE method vs. no. of clusters.

Through the graph of the elbow method, we can see that from $k = 4$ to $k = 6$, an elbow started to be created, where the SSE values decreased more slowly. This showed that the number of clusters increased without significantly improving the SSE value. As the elbow method graph alone was inconclusive, we also conducted a silhouette study (Figure 9) to support the decision of the number of clusters.


**Figure 9.** Silhouette method with silhouette coefficient vs. no. of clusters.

Starting from the elbow graph, where it was unclear which of $k = 4$, $k = 5$ or $k = 6$ to choose, through the silhouette graph, we concluded that $k = 4$ was the optimal choice for this dataset. In this research, the silhouette index was used to evaluate the clustering algorithm and choose the number $k$ of clusters. The silhouette index is applied in cases of exclusive partitioned clustering and takes into consideration measures of coherence and separation of events in a cluster [108]. The silhouette function calculates the average silhouette coefficient of all samples based on the average intracluster distance and the average distance from the nearest cluster for each sample. Its index ranges from $[-1, 1]$, with high silhouette values reflecting good solutions for clustering processes.

Knowing the number of clusters, $k = 4$, the k-means clustering was performed and the distribution of the clusters is shown in temporal form (Figure 10). The clusters are sorted in descending order of the number of points, with the first cluster having the most values.

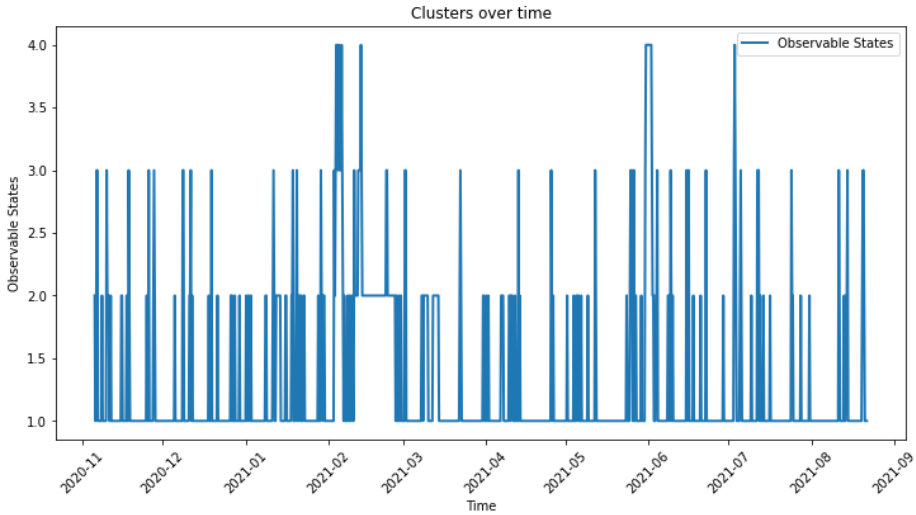



**Figure 10.** Cluster-optimised observations over time.

### 4.5. HMM

After the clustering phase, we applied a classification phase, where the diagnosis of the equipment was made through three hidden states: hidden state 1 represented the "good working state", state 2 represented the "alert state", and state 3 was the "malfunctioning state". Thus, for the HMM, a doubly stochastic method, the observable states were represented by the clusters defined by the K-means clustering and the hidden states were the health state of the equipment. We started by using the observable states to train the model and obtain its parameters. This was performed using the *Baum–Welch* algorithm, which employs a special case of the expectation–maximization algorithm to find local maxima of $P(O|\lambda)$ [92]. The HMM training first estimated the maximum likelihood of the model parameters based on the observed state sequence. The *Baum–Welch* algorithm was then used to calculate the transition probability $A = a_{ij}$, the observation probability $B = b_{jk}$, and the initial observation probability $\pi_i$, from which the updated algorithm was used to predict the hidden states, and the final algorithm resulted in the prediction of the equipment's health state. The accuracy was used to evaluate the model. For this, the observations were divided in a temporal manner into training and test data, where 70% were used for training and 30% for testing. The division was made in a temporal manner since the diagnosis reported by the *Viterbi* algorithm represented the hidden states that best fitted the observations over time. After the model was trained with 70% of the data, it generated observable states with the same number of samples as the test data, in order to determine the accuracy (Equation (11)).

$$Accuracy = \frac{\sum (HMM_{GeneratedObservations} = Data_{Test})}{n_{Samples}} * 100 \quad (11)$$

As the parameters of the HMM were based on probabilities, observable states were generated 10,000 times and the accuracy was taken as the average accuracy over the 10,000 runs. For this specific case, an accuracy of approximately 72% was obtained.

Once a good value for the accuracy of the model was obtained, the corresponding parameters were used to determine the sequence of hidden states. For this purpose, the *Viterbi* algorithm was used. The *Viterbi* algorithm finds the most probable sequence of hidden states resulting from the sequence of clusters. Through Figure 11, we can verify the evolution of the health status of the equipment throughout the study time.

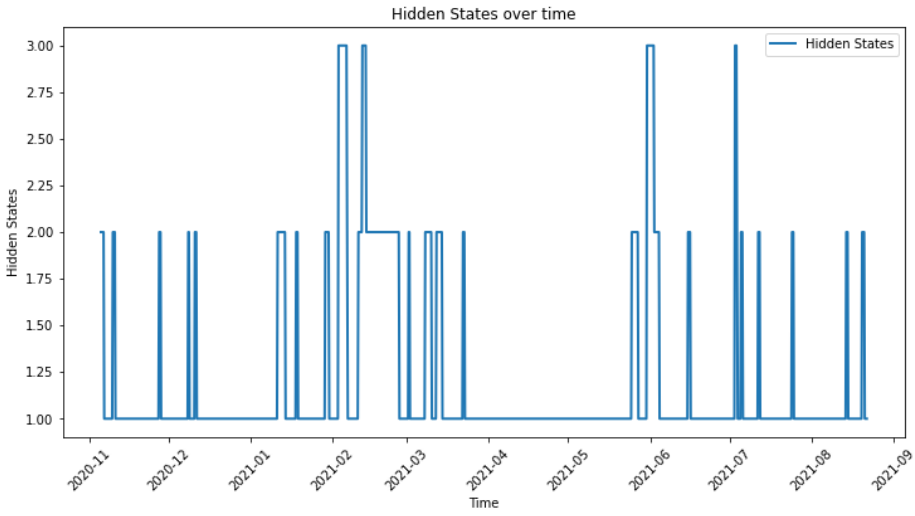



**Figure 11.** Hidden states/diagnosis of equipment over time.

### *4.6. Prognostic with GRU Model*

Once the equipment diagnosis was made, the object in this phase was to make a prediction, using the GRU network. The network was prepared to predict 7 days in the future both the observable states that were input into the HMM, as well as the hidden states of the classifier model.

Using a recurrent neural network with an encoder and decoder structure, we made a prediction of the observable states over a period of 7 days into the future, corresponding to one week. Figure 12 presents the prediction of the 7 days with a five-unit GRU recurrent neural network, with a delay window of 3 days. The structure of the network featured a *relu* activation function in the first layer and a *relu* function in the second layer. Figure 13 represents the prediction after processing the data using the function of Equation (12) for a better visualization of the model prediction results. The purpose of that operation was to scale the predicted data, $x_n$, to the same values as those of the observable states.

$$ScaleValues = round([x_n - min(x_n)] * [\frac{n_{States} - 1}{max(x_n) - min(x_n)}]) + 1 \quad (12)$$

To validate the accuracy of the model, we used the errors $MAPE$ (Equation (13)), $RMSE$ (Equation (14)), $MAE$ (Equation (15)) and $R^2$ (Equation (16)), which showed, respectively, the following values of 5.49, 0.33, 0.086 and 0.68, showing a good evaluation of the model.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - y_i|}{x_i} \quad (13)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}} \quad (14)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \quad (15)$$

$$R^2 = 1 - \sum_{i=1}^{n}\frac{(x_i - y_i)^2}{(x_i - \overline{y})^2} \quad (16)$$

where $x_i$ is the actual value, $y_i$ is the value predicted by the model, and $n$ is the total number of observations.

Using the same neural network with the same architecture, changing only the function to a sigmoid in the last layer, with the same delay window, it was possible to perform a prediction of the hidden states for the next 7 days as shown in Figure 14, which, after scaling, persisted, as shown in Figure 15. The $MAPE$, $RMSE$, $MAE$ and $R^2$ errors of the model evaluation had the values 2.79, 0.22, 0.035 and 0.71, respectively.

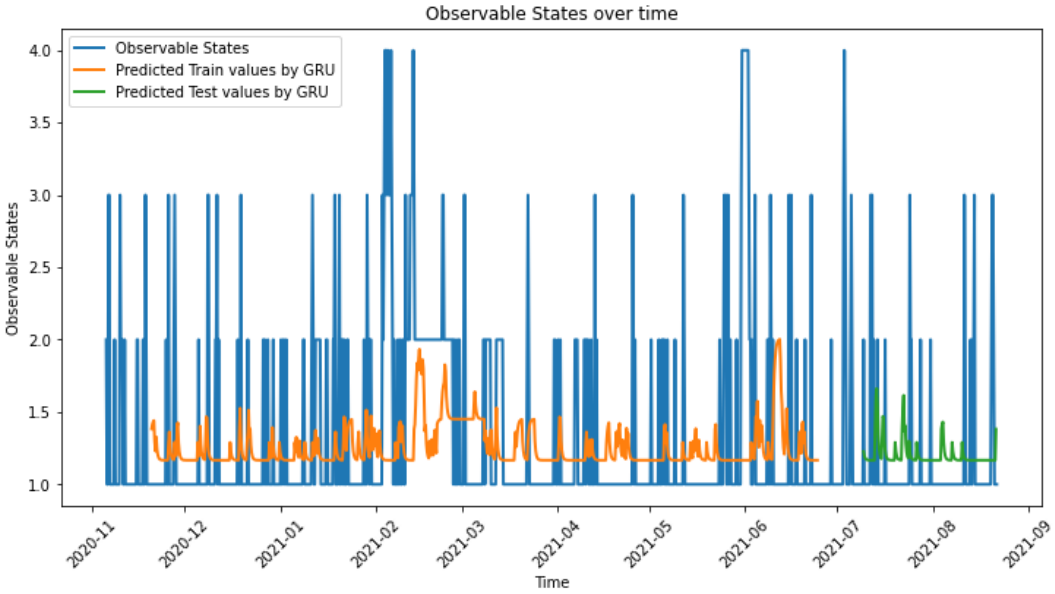



**Figure 12.** Predicted training and test values for the observable states.

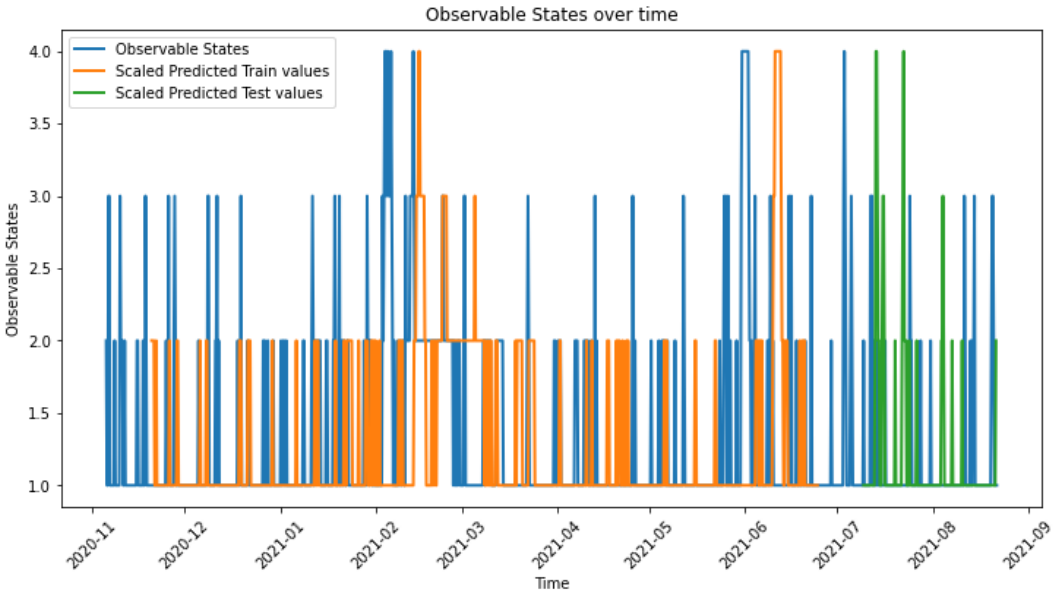



**Figure 13.** Scaled training and test predicted values for the observable states.

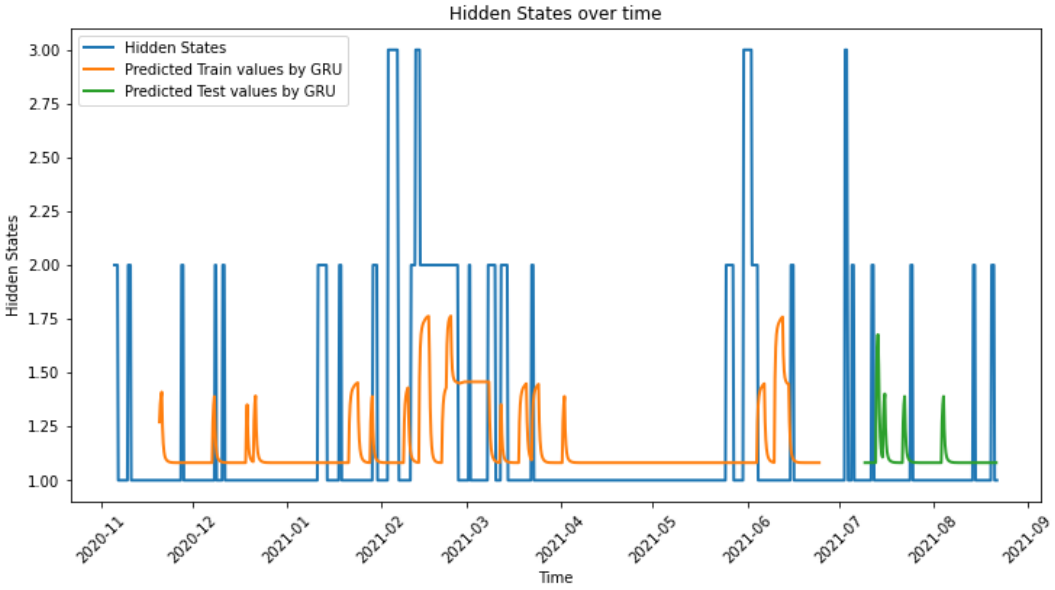



**Figure 14.** Scaled training and test predicted values for the observable states.

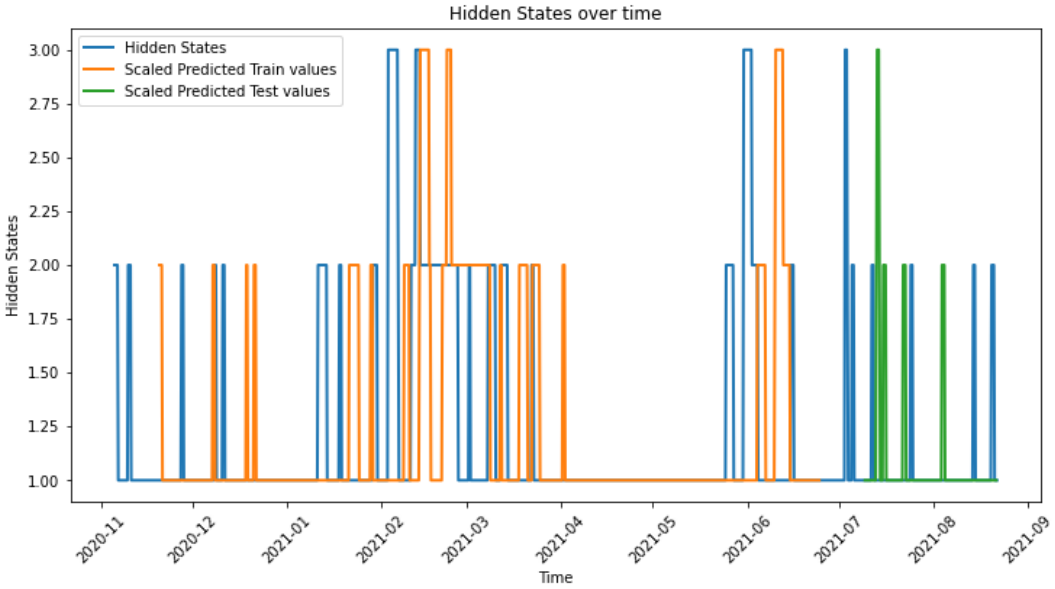



**Figure 15.** Scaled training and test predicted values for the observable states.

## 5. Discussion

Starting with the diagnostic analysis of the health state of the production equipment, we verified through Figure 11 that the HMM was able to detect with an accuracy close to 72% several states 2 and 3 throughout the study time. We can see in Figure that failure state 3 appeared four times over time, indicating equipment failure. The well-functioning state, state 1, was almost constant, showing that the equipment was mostly in good condition. Several alert states appeared that were quickly solved in order not to reach the equipment failure state.

We also verified through Figures 5 and 11, that state 3, as predicted by the HMM, happened when the values collected by the sensors escaped from the normal pattern of behaviour. This was verified mainly in the periods February 2021 and March 2021, where

we saw that the temperature and pressure sensors had a big drop in their measured values. This indicated that the HMM was able to classify the health status of the equipment without having any previous information about its operation or anomalies. Furthermore, this is a nonsupervised methodology, capable of being used in any type of equipment with different types of sensors. It is also capable of performing fault detection in real time, and through the application of a DNN, the GRU, it is also capable of performing equipment prognostics.

The recurrent GRU neural network according to the literature review presents advantages in predicting data of reduced size. This paper showed that the proposed model, a five-unit GRU network, presented a good accuracy of the observed and hidden states of the machine. In this way, it was possible to anticipate a fault detection 7 days ahead with a high degree of accuracy. Normally, the future prediction is done on data collected by sensors, and then it is necessary to run the whole fault-detection methodology. It should be noted that, in our methodology, the prediction was made directly on clusters, optimized observable states, and also directly on hidden states classified by the HMM. In the prediction made on the clusters, it was necessary to use the HMM classifier to obtain the operating states of the equipment. When the prediction was made on the HMM states, we obtained directly the health status of the equipment 7 days ahead. The GRU's capacity to prevent information overlap allows it to perform better with smaller quantities of training data despite having a less complex architecture [103]. In terms of faster computation times and superior results, the GRU also demonstrates the ability to outperform LSTM. Thus, through our methodology, it is possible to directly detect the states of the equipment, through 7-day prognostics, in a faster way, to obtain information in real time and be able to act more quickly with regard to unexpected breakdowns.

## 6. Conclusions

This paper showed a methodology capable of making a diagnosis and prognosis of the state of health of production equipment. First, data cleaning was conducted, followed by the generation of statistical features in the time domain. After the feature generation, a data normalisation and dimensional reduction were performed, through a PCA, to obtain new features with more information and improve compactness. Then, K-means clustering was used to group similar data into groups and create the observable states that were input to the HMM. The HMM, in turn, was responsible for classifying the observable states into hidden states that represented the state of health of the equipment, numbered from one to three, where state 1 represented "Good Operation", state 2 represented "Alert State" and state 3 "Failure State". After the detection of health states over time, it was possible to make a prognosis directly on these states, or on the observable states obtained from the clustering, 7 days ahead. For that, a GRU was used, which obtained good results with these types of data. This is a generic methodology, which can be used in different types of equipment, with different types of sensors. It works without prior information about the behaviour of the equipment and can work in real time.

We conclude that through this methodology, it is possible to improve the quality of CBM and PdM maintenance, thus improving the production flow and consequently, the company's profits.

In future work, other types of models will be applied to each of the steps in order to check which combination obtains the best results. That is, in dimensional reduction, other algorithms will be used, such as linear discriminant analysis (LDA), independent component analysis (ICA), etc.; in the clustering, the Gaussian mixture model (GMM), density-based spatial clustering of applications with noise (DBSCAN), among others. Moreover, the realization of the prediction will use other DNNs that can get better results as well as a prediction in bits through one-hot encode. Other types of statistics that can detect the behaviour of the equipment will also be added. Finally, a classification algorithm will also be implemented to input new data, collected by the sensors, in order to understand in which cluster they best fit. This will lead to the implementation of a metric that will call for

a new training session for the entire methodology if the data disperse significantly from the previously created clusters.

**Author Contributions:** Conceptualization, A.M., J.T.F., I.F. and A.M.C.; methodology, A.M., I.F. and J.T.F.; software, A.M., B.M. and I.F.; validation, A.M., J.T.F., B.M., I.F., M.M., J.R. and A.M.C.; formal analysis, A.M., J.T.F. and I.F.; investigation, A.M., B.M. and I.F.; resources, A.M., J.T.F. and A.M.C.; writing—original draft preparation, A.M.; writing—review and editing, J.T.F., B.M. and I.F.; project administration, J.T.F. and A.M.C.; funding acquisition, J.T.F. and A.M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** Our research, leading to these results, has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowvska-Curie grant agreement 871284 project SSHARE and the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under project POCI-01-0145-FEDER-029494, and by national funds through the FCT—Portuguese Foundation for Science and Technology, under projects PTDC/EEI-EEE/29494/2017, UIDB/04131/2020, and UIDP/04131/2020.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Restrictions apply to the availability of these data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| (ANN) | Artificial neural network |
| (CBM) | Condition-based maintenance |
| (CM) | Corrective maintenance |
| (CNN) | Convolution neural networks |
| (CPS) | Cyberphysics systems |
| (DBSCAN) | Density-based spatial clustering of applications with noise |
| (DNN) | Deep neural network |
| (GMM) | Gaussian mixture model |
| (GRU) | Gated recurrent unit |
| (HMM) | Hidden Markov model |
| (ICA) | Independent component analysis |
| (IoT) | Internet of things |
| (LDA) | Linear discriminant analysis |
| (LSTM) | Long short-term memory |
| (MAE) | Mean absolute error |
| (MAPE) | Mean absolute percentage error |
| (ML) | Machine learning |
| (PCA) | Principal components analysis |
| (PCs) | Principal components |
| (PdM) | Predictive maintenance |
| (RMSE) | Root-mean-square error |
| (RUL) | Remaining useful life |
| (TBM) | Time-based maintenance |
| (WT) | Wavelet transform |

## References


1. Zhang, M.; Amaitik, N.; Wang, Z.; Xu, Y.; Maisuradze, A.; Peschl, M.; Tzovaras, D. Predictive Maintenance for Remanufacturing Based on Hybrid-Driven Remaining Useful Life Prediction. *Appl. Sci.* **2022**, *12*, 3218. [CrossRef]
2. Hu, J.; Sun, Q.; Ye, Z.S. Condition-Based Maintenance Planning for Systems Subject to Dependent Soft and Hard Failures. *IEEE Trans. Reliab.* **2021**, *70*, 1468–1480. [CrossRef]
3. Kumar, P.; Sushil, C.; Basu, K.; Chandra, M. Quantified Risk Ranking Model for Condition-Based Risk and Reliability Centered Maintenance. *J. Inst. Eng. Ser. C* **2017**, *98*, 325–333. [CrossRef]
4. Pais, J.; Raposo, H.D.; Farinha, J.; Cardoso, A.J.M.; Marques, P.A. Optimizing the life cycle of physical assets through an integrated life cycle assessment method. *Energies* **2021**, *14*, 6128. [CrossRef]
5. Huynh, K.T.; Grall, A.; Berenguer, C. A Parametric Predictive Maintenance Decision-Making Framework Considering Improved System Health Prognosis Precision. *IEEE Trans. Reliab.* **2019**, *68*, 375–396. [CrossRef]
6. Chuang, C.; Ningyun, L.U.; Bin, J.; Yin, X. Condition-based maintenance optimization for continuously monitored degrading systems under imperfect maintenance actions. *J. Syst. Eng. Electron.* **2020**, *31*, 841–851. [CrossRef]
7. Zhu, Z.; Xiang, Y. Condition-based maintenance for multi-component systems: Modeling, structural properties, and algorithms. *IISE Trans.* **2021**, *53*, 88–100. [CrossRef]
8. Koochaki, J.; Bokhorst, J.A.C.; Wortmann, H.; Klingenberg, W. The influence of condition-based maintenance on workforce planning and maintenance scheduling. *Int. J. Prod. Res.* **2013**, *51*, 2339–2351. [CrossRef]
9. Leoni, L.; Carlo, F.D.; Abaei, M.M.; Bahootoroody, A. A hierarchical Bayesian regression framework for enabling online reliability estimation and condition-based maintenance through accelerated testing. *Comput. Ind.* **2022**, *139*, 103645. [CrossRef]
10. Lee, S.M.; Lee, D.; Kim, Y.S. The quality management ecosystem for predictive maintenance in the Industry 4.0 era. *Int. J. Qual. Innov.* **2019**, *5*, 1–11. [CrossRef]
11. Liu, B.; Do, P.; Iung, B.; Xie, M. Stochastic Filtering Approach for Condition-Based Maintenance Considering Sensor Degradation. *IEEE Trans. Autom. Sci. Eng.* **2020**, *17*, 177–190. [CrossRef]
12. Kolhatkar, A.; Pandey, A. Predictive maintenance methodology in sheet metal progressive tooling: A case study. *Int. J. Syst. Assur. Eng. Manag.* **2021**, 1–10. [CrossRef]
13. Raposo, H.; Farinha, J.T.; Pais, E.; Galar, D. An integrated model for dimensioning the reserve fleet based on the maintenance policy. *WSEAS Trans. Syst. Control.* **2021**, *16*, 43–65. [CrossRef]
14. Zhang, J.; Zhao, X.; Song, Y.; Qiu, Q. Joint optimization of condition-based maintenance and spares inventory for a series—Parallel system with two failure modes. *Comput. Ind. Eng.* **2022**, *168*, 108094. [CrossRef]
15. Soltanali, H.; Khojastehpour, M.; Farinha, T.; Pais, J.E. An Integrated Fuzzy Fault Tree Model with Bayesian Network-Based Maintenance Optimization of Complex Equipment in Automotive Manufacturing. *Energies* **2021**, *14*, 7758. [CrossRef]
16. Li, Y.; Peng, S.; Li, Y.; Jiang, W. A review of condition-based maintenance: Its prognostic and operational aspects. *Front. Eng. Manag.* **2020**, *7*, 323–334. [CrossRef]
17. Oakley, J.L.; Wilson, K.J.; Philipson, P. A condition-based maintenance policy for continuously monitored multi-component systems with economic and stochastic dependence. *Reliab. Eng. Syst. Saf.* **2022**, *222*, 108321. [CrossRef]
18. Shin, M.K.; Jo, W.J.; Cha, H.M.; Lee, S.H. A study on the condition based maintenance evaluation system of smart plant device using convolutional neural network. *J. Mech. Sci. Technol.* **2020**, *34*, 2507–2514. [CrossRef]
19. Peng, S.; Feng, Q.M. Reinforcement learning with Gaussian processes for condition-based maintenance. *Comput. Ind. Eng.* **2021**, *158*, 107321. [CrossRef]
20. Hsu, J.Y.; Wang, Y.F.; Lin, K.C.; Chen, M.Y.; Hsu, J.H.Y. Wind Turbine Fault Diagnosis and Predictive Maintenance Through Statistical Process Control and Machine Learning. *IEEE Access* **2020**, *8*, 23427–23439. [CrossRef]
21. Staden, H.E.V.; Boute, R.N. The effect of multi-sensor data on condition-based maintenance policies. *Eur. J. Oper. Res.* **2021**, *290*, 585–600. [CrossRef]
22. Kenda, M.; Klob, D.; Bra, D. Condition based maintenance of the two-beam laser welding in high volume manufacturing of piezoelectric pressure sensor. *J. Manuf. Syst.* **2021**, *59*, 117–126. [CrossRef]
23. Kumar, G.; Jain, V.; Gandhi, O.P. Availability analysis of mechanical systems with condition-based maintenance using semi-Markov and evaluation of optimal condition monitoring interval. *J. Ind. Eng. Int.* **2018**, *14*, 119–131. [CrossRef]
24. Adu-amankwa, K.; Attia, A.K.A.; Janardhanan, M.N.; Patel, I. A predictive maintenance cost model for CNC SMEs in the era of industry 4.0. *Int. J. Adv. Manuf. Technol.* **2019**, *104*, 3567–3587. [CrossRef]
25. Tsuji, K.; Imai, S.; Takao, R.; Kimura, T.; Kondo, H. A machine sound monitoring for predictive maintenance focusing on very low frequency band. *SICE J. Control. Meas. Syst. Integr.* **2021**, *14*, 27–38. [CrossRef]
26. Rodrigues, J.; Martins, A.; Mendes, M.; Farinha, T.; Mateus, R.; Cardoso, A.J. Automatic Risk Assessment for an Industrial Asset Using. *Energies* **2022**, *15*, 9387. [CrossRef]
27. Ingemarsdotter, E.; Lena, M.; Jamsin, E.; Sakao, T.; Balkenende, R. Challenges and solutions in condition-based maintenance implementation—A multiple case study. *J. Clean. Prod.* **2021**, *296*, 126420. [CrossRef]
28. Liu, Y.; Yu, W.; Dillon, T.; Fellow, L.; Rahayu, W.; Li, M. Empowering IoT Predictive Maintenance Solutions With AI: A Distributed System for Manufacturing Plant-Wide Monitoring. *IEEE Trans. Ind. Inform.* **2022**, *18*, 1345–1354. [CrossRef]
29. Lin, C.Y.; Hsieh, Y.M.; Cheng, F.T.; Huang, H.C.; Adnan, M. Time Series Prediction Algorithm for Intelligent Predictive Maintenance. *IEEE Robot. Autom. Lett.* **2019**, *4*, 2807–2814. [CrossRef]

30. Soltanali, H.; Khojastehpour, M.; Pais, J.E.; Farinha, J. Sustainable Food Production: An Intelligent Fault Diagnosis Framework for Analyzing the Risk of Critical Processes. *Sustainability* **2022**, *14*, 1083. [CrossRef]
31. Ghasemi, A.; Yacout, S.; Ouali, M.S. Optimal condition based maintenance with imperfect information and the proportional hazards model. *Int. J. Prod. Res.* **2007**, *45*, 989–1012. [CrossRef]
32. Martins, A.B.; Farinha, J.T.; Cardoso, A.M. Calibration and Certification of Industrial Sensors—A Global Review. *WSEAS Trans. Syst. Control.* **2020**, *15*, 394–416. [CrossRef]
33. Martins, A.; Fonseca, I.; Farinha, T.; Reis, J.; Cardoso, A. Online Monitoring of Sensor Calibration Status to Support Condition-Based Maintenance. *Sensors* **2023**, *23*, 2402. [CrossRef]
34. Wu, Z.; Guo, B.I.N.; Tian, X.; Zhang, L. A Dynamic Condition-Based Maintenance Model Using Inverse Gaussian Process. *IEEE Access* **2020**, *8*, 104–117. [CrossRef]
35. Nguyen, K.A.; Do, P.; Grall, A. Condition-based maintenance for multi- component systems using importance measure and predictive information. *Int. J. Syst. Sci. Oper. Logist.* **2014**, *1*, 228–245. [CrossRef]
36. Zhang, W.; Yang, D.; Wang, H. Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. *IEEE Syst. J.* **2019**, *13*, 2213–2227. [CrossRef]
37. Pais, E.; Farinha, J.T.; Cardoso, A.J.; Raposo, H. Optimizing the life cycle of physical assets—A review. *WSEAS Trans. Syst. Control.* **2020**, *15*, 417–430. [CrossRef]
38. Yam, R.C.M.; Tse, P.W.; Li, L.; Tu, P. Intelligent Predictive Decision Support System for Condition-Based Maintenance. *Adv. Manuf. Technol.* **2001**, *17*, 383–391. [CrossRef]
39. Harald, R.; Schj, P.; Wabner, M.; Frie, U. Predictive Maintenance for Synchronizing Maintenance Planning with Production. *Adv. Manuf. Autom. VII* **2018**, *451*, 439–446. [CrossRef]
40. Popescu, T.D.; Aiordachioaie, D.; Culea-Florescu, A. Basic tools for vibration analysis with applications to predictive maintenance of rotating machines: An overview. *Int. J. Adv. Manuf. Technol.* **2022**, *118*, 2883–2899. [CrossRef]
41. Chen, C.; Zhu, Z.H.; Member, S.; Shi, J. Dynamic Predictive Maintenance Scheduling Using Deep Learning Ensemble for System Health Prognostics. *IEEE Sens. J.* **2021**, *21*, 26878–26891. [CrossRef]
42. Kiangala, K.S.; Wang, Z. An Effective Predictive Maintenance Framework for Conveyor Motors Using Dual Time-Series Imaging and Convolutional Neural Network in an Industry 4.0 Environment. *IEEE Access* **2020**, *8*, 121033–121049. [CrossRef]
43. Mateus, B.C.; Mendes, M.; Farinha, J.T.; Assis, R.; Cardoso, A.M. Comparing LSTM and GRU Models to Predict the Condition of a Pulp Paper Press. *Energies* **2021**, *14*, 6958. [CrossRef]
44. Antunes, J.; Torres, J.; Mendes, M.; Mateus, R.; Cardoso, A. Short and long forecast to implement predictive maintenance in a pulp industry. *Eksploat. Niezawodn. Maint. Reliab.* **2022**, *24*, 33–41. [CrossRef]
45. Mateus, B.C.; Mendes, M.; Farinha, J.T.; Cardoso, A.M. Anticipating future behavior of an industrial press using lstm networks. *Appl. Sci.* **2021**, *11*, 6101. [CrossRef]
46. Zhang, Z.; Wang, Y.; Wang, K. Intelligent fault diagnosis and prognosis approach for rotating machinery integrating wavelet transform, principal component analysis, and artificial neural networks. *Int. J. Adv. Manuf. Technol.* **2013**, *68*, 763–773. [CrossRef]
47. Martins, A.; Fonseca, I.; Farinha, J.; Reis, J.; Cardoso, A. Maintenance prediction through sensing using hidden markov models—A case study. *Appl. Sci.* **2021**, *11*, 7685. [CrossRef]
48. Yu, J. Adaptive hidden Markov model-based online learning framework for bearing faulty detection and performance degradation monitoring. *Mech. Syst. Signal Process.* **2017**, *83*, 149–162. [CrossRef]
49. Arpaia, P.; Cesaro, U.; Chadli, M.; Coppier, H.; De Vito, L.; Esposito, A.; Gargiulo, F.; Pezzetti, M. Fault detection on fluid machinery using Hidden Markov Models. *Measurement* **2020**, *151*, 107126. [CrossRef]
50. Mateus, B.; Mendes, M.; Farinha, J.; Martins, A.; Cardoso, A. Data Analysis for Predictive Maintenance Using Time Series and Deep Learning Models—A Case Study in a Pulp Paper Industry. In *Proceedings of IncoME-VI and TEPEN 2021*; Springer International Publishing: Cham, Switzerland, 2023; pp. 11–25. [CrossRef]
51. Chen, H.; Hsu, J.Y.; Hsieh, J.Y.; Hsu, H.Y.; Chang, C.H.; Lin, Y.J. Predictive maintenance of abnormal wind turbine events by using machine learning based on condition monitoring for anomaly detection. *J. Mech. Sci. Technol.* **2021**, *35*, 5323–5333. [CrossRef]
52. Tipler, S.; Alessio, G.D.; Haute, Q.V.; Parente, A.; Contino, F.; Coussement, A. Predicting octane numbers relying on principal component analysis and artificial neural network. *Comput. Chem. Eng.* **2022**, *161*, 107784. [CrossRef]
53. Gu, Y.K.; Zhou, X.Q.; Yu, D.P.; Shen, Y.J. Fault diagnosis method of rolling bearing using principal component analysis and support vector machine. *J. Mech. Sci. Technol.* **2018**, *32*, 5079–5088. [CrossRef]
54. Booker, N.K.; Knights, P.; Gates, J.D.; Clegg, R.E. Applying principal component analysis (PCA) to the selection of forensic analysis methodologies. *Eng. Fail. Anal.* **2022**, *132*, 105937. [CrossRef]
55. Kamari, A.; Schultz, C. A combined principal component analysis and clustering approach for exploring enormous renovation design spaces. *J. Build. Eng.* **2022**, *48*, 103971. [CrossRef]
56. Lim, Y.; Kwon, J.; Oh, H.S. Principal component analysis in the wavelet domain. *Pattern Recognit.* **2021**, *119*, 108096. [CrossRef]
57. Babouri, M.K.; Djebala, A.; Ouelaa, N.; Oudjani, B.; Younes, R. Rolling bearing faults severity classification using a combined approach based on multi-scales principal component analysis and fuzzy technique. *Int. J. Adv. Manuf. Technol.* **2020**, *107*, 4301–4316. [CrossRef]
58. Zhu, T.; Cheng, X.; Cheng, W.; Tian, Z.; Li, Y. Principal component analysis based data collection for sustainable internet of things enabled Cyber—Physical Systems. *Microprocess. Microsystems* **2022**, *88*, 104032. [CrossRef]

59. Park, J.H.; Kang, Y.J. Evaluation index for sporty engine sound reflecting evaluators' tastes, developed using K-means cluster analysis. *Int. J. Automot. Technol.* **2020**, *21*, 1379–1389. [CrossRef]
60. Yang, Y.; Liao, Q.; Wang, J.; Wang, Y. Application of multi-objective particle swarm optimization based on short-term memory and K-means clustering in multi-modal multi-objective optimization. *Eng. Appl. Artif. Intell.* **2022**, *112*, 104866. [CrossRef]
61. Voronova, L.I.; Voronov, V.; Mohammad, N. Modeling the Clustering of Wireless Sensor Networks Using the K-means Method. In Proceedings of the International Conference on Quality Management, Transport and Information Security, Information Technologies (IT QM IS), Yaroslavl, Russia, 6–10 September 2021; pp. 740–745. [CrossRef]
62. Sinaga, K.P.; Yang, M.S. Unsupervised K-Means Clustering Algorithm. *IEEE Access* **2020**, *8*, 80716–80727. [CrossRef]
63. Sinaga, K.P.; Hussain, I.; Yang, M.S. Entropy K-Means Clustering With Feature Reduction Under Unknown Number of Clusters. *IEEE Access* **2021**, *9*, 67736–67751. [CrossRef]
64. Niño-adan, I.; Landa-torres, I.; Portillo, E.; Manjarres, D. Influence of statistical feature normalisation methods on K-Nearest Neighbours and K-Means in the context of industry 4.0. *Eng. Appl. Artif. Intell.* **2022**, *111*, 104807. [CrossRef]
65. Lakshmi, K.; Visalakshi, N.K.; Shanthi, S. Data clustering using K-Means based on Crow Search Algorithm. *Sādhanā* **2018**, *43*, 1–12. [CrossRef]
66. Ni, Y.; Zeng, X.; Liu, Z.; Yu, K.; Xu, P.; Wang, Z.; Zhuo, C.; Huang, Y. Faulty feeder detection of single phase-to-ground fault for distribution networks based on improved K-means power angle clustering analysis. *Int. J. Electr. Power Energy Syst.* **2022**, *142*, 108252. [CrossRef]
67. Liao, Q.Z.; Xue, L.; Lei, G.; Liu, X.; Sun, S.Y.; Patil, S. Statistical prediction of water fl ooding performance by K-means clustering and empirical modeling. *Pet. Sci.* **2022**, *19*, 1139–1152. [CrossRef]
68. Yoo, J.H.; Park, Y.K.; Han, S.S. Predictive Maintenance System for Wafer Transport Robot Using K-Means Algorithm and Neural Network Model. *Electronics* **2022**, *11*, 1324. [CrossRef]
69. Han, P.; Wang, W.; Shi, Q.; Yue, J. A combined online-learning model with K-means clustering and GRU neural networks for trajectory prediction. *AdHoc Netw.* **2021**, *117*, 102476. [CrossRef]
70. Chen, L.; Shan, W.; Liu, P. Identification of concrete aggregates using K-means clustering and level set method. *Structures* **2021**, *34*, 2069–2076. [CrossRef]
71. Visalaxi, S.; Sudalaimuthu, T. Endometrium Phase prediction using K-means Clustering through the link of Diagnosis and procedure. In Proceedings of the 8th International Conference on Signal Processing and Integrated Networks (SPIN), Amity University, Noida, India, 26–27 August 2021; pp. 1178–1181. [CrossRef]
72. Cho, K.; Merrienboer, V.; Gulcehre, B.; Bahdanau, C.; Bougares, D.; Schwwenk, F.; H. Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734. [CrossRef]
73. Ke, K.; Hongbin, S.; Chengkang, Z.; Brown, C. Short-term electrical load forecasting method based on stacked auto- encoding and GRU neural network. *Evol. Intell.* **2019**, *12*, 385–394. [CrossRef]
74. Lin, S.; Fan, R.; Feng, D.; Yang, C.; Wang, Q.; Gao, S. Condition-Based Maintenance for Traction Power Supply Equipment Based on Decision Process. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 175–189. [CrossRef]
75. Benhaddi, M.; Ouarzazi, J. Multivariate Time Series Forecasting with Dilated Residual Convolutional Neural Networks for Urban Air Quality Prediction. *Arab. J. Sci. Eng.* **2021**, *46*, 3423–3442. [CrossRef]
76. Liu, X.; Lin, Z.; Feng, Z. Short-term offshore wind speed forecast by seasonal ARIMA—A comparison against GRU and LSTM. *Energy* **2021**, *227*, 120492. [CrossRef]
77. Gugnani, V.; Kumar, R. Analysis of deep learning approaches for air pollution prediction. *Multimed. Tools Appl.* **2022**, *81*, 6031–6049. [CrossRef]
78. Veeramsetty, V.; Reddy, K.R.; Arjun, M.S.; Gaurav, M. Short-term electric power load forecasting using random forest and gated recurrent unit. *Electr. Eng.* **2022**, *104*, 307–329. [CrossRef]
79. Wang, Y.; Zou, R.; Liu, F.; Zhang, L.; Liu, Q. A review of wind speed and wind power forecasting with deep neural networks. *Appl. Energy* **2021**, *304*, 117766. [CrossRef]
80. Wang, S.; Chen, J.; Wang, H.; Zhang, D. Degradation evaluation of slewing bearing using HMM and improved GRU. *Measurement* **2019**, *146*, 385–395. [CrossRef]
81. Becerra-rico, J.; Aceves-fernández, M.A.; Esquivel-escalante, K.; Pedraza-ortega, J.C. Airborne particle pollution predictive model using Gated Recurrent Unit (GRU) deep neural networks. *Earth Sci. Inform.* **2020**, *13*, 821–834. [CrossRef]
82. Zhao, Z.; Wang, Y.; Feng, M.; Peng, G.; Liu, J.; Jason, B.; Tao, Y. Autoregressive State Prediction Model Based on Hidden Markov and the Application. *Wirel. Pers. Commun.* **2018**, *102*, 2403–2416. [CrossRef]
83. Sun, L.; Li, Y.; Du, H.; Liang, P.; Nian, F. Fault Diagnosis Method of Low Noise Amplifier Based on Support Vector Machine and Hidden Markov Model. *J. Electron. Test.* **2021**, *37*, 215–223. [CrossRef]
84. Li, Y.; Li, H.; Chen, Z.; Zhu, Y. An Improved Hidden Markov Model for Monitoring the Process with Autocorrelated Observations. *Energies* **2022**, *15*, 1685. [CrossRef]
85. Jandera, A.; Skovranek, T. Customer Behaviour Hidden Markov Model. *Mathematics* **2022**, *10*, 1230. [CrossRef]
86. Lin, T.; Wang, M.; Yang, M.; Yang, X. A Hidden Markov Ensemble Algorithm Design for Time Series Analysis. *Sensors* **2022**, *22*, 2950. [CrossRef] [PubMed]

87. Yao, Y.; Zhao, X.; Wu, Y.; Zhang, Y.; Rong, J. Clustering driver behavior using dynamic time warping and hidden Markov model. *J. Intell. Transp. Syst.* **2021**, *25*, 249–262. [CrossRef]
88. Liu, H.; Wang, K.; Li, Y. Hidden Markov Linear Regression Model and Its Parameter Estimation. *IEEE Access* **2020**, *8*, 187037–187042. [CrossRef]
89. Li, W.; Ji, Y.; Cao, X.; Qi, X. Trip Purpose Identification of Docked Bike-Sharing From IC Card Data Using a Continuous Hidden Markov Model. *IEEE Access* **2020**, *8*, 189598–189613. [CrossRef]
90. Park, S.; Lim, W.; Sunwoo, M. Robust lane-change recognition based on an adaptive hidden Markov model using measurement uncertainty. *Int. J. Automot. Technol.* **2019**, *20*, 255–263. [CrossRef]
91. Shang, Z.; Zhang, Y.; Zhang, X.; Zhao, Y.; Cao, Z.; Wang, X. Time Series Anomaly Detection for KPIs Based on Correlation Analysis and HMM. *Appl. Sci.* **2021**, *11*, 1353. [CrossRef]
92. López, C.; Naranjo, Á.; Lu, S.; Moore, K. Hidden Markov Model based Stochastic Resonance and its Application to Bearing Fault Diagnosis. *J. Sound Vib.* **2022**, *528*, 116890. [CrossRef]
93. Feng, Y.; Xu, W.; Zhang, Z.; Wang, F. Continuous Hidden Markov Model Based Spectrum Sensing with Estimated SNR for Cognitive UAV Networks. *Sensors* **2022**, *22*, 2620. [CrossRef]
94. Soleimani, M.; Campean, F.; Neagu, D. Integration of Hidden Markov Modelling and Bayesian Network for fault detection and prediction of complex engineered systems. *Reliab. Eng. Syst. Saf.* **2021**, *215*, 107808. [CrossRef]
95. Martins, A.; Fonseca, I.; Torres, F.J.; Reis, J.; Cardoso, A.J.M. Prediction Maintenance Based on Vibration Analysis and Deep Learning—A Case Study of a Drying Press Supported on Hidden Markov Model. SSRN. 2022. Available online: https://ssrn.com/abstract=4194601 (accessed on 3 January 2023).
96. Gokilavani, N.; Bharathi, B. Test case prioritization to examine software for fault detection using PCA extraction and K-means clustering with ranking. *Soft Comput.* **2021**, *25*, 5163–5172. [CrossRef]
97. Fard, N.; Xu, H.; Fang, Y. A unique solution for principal component analysis-based multi-response optimization problems. *Int. J. Adv. Manuf. Technol.* **2016**, *82*, 697–709. [CrossRef]
98. Gang, A.; Bajwa, W.U. A linearly convergent algorithm for distributed principal component analysis. *Signal Process.* **2022**, *193*, 108408. [CrossRef]
99. Sancho, A.; Ribeiro, J.C.; Reis, M.S.; Martins, F.G. Cluster analysis of crude oils with k-means based on their physicochemical properties. *Comput. Chem. Eng.* **2022**, *157*, 107633. [CrossRef]
100. Reddy, S.; Nethra, S.; Sherer, E.A.; Amritphale, A. Prediction of the number of COVID-19 confirmed cases based on K-means-LSTM. *Array* **2021**, *11*, 100085. [CrossRef]
101. Li, Y.; Chu, X.; Tian, D.; Feng, J.; Mu, W. Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm. *Appl. Soft Comput.* **2021**, *113*, 107924. [CrossRef]
102. Yang, W.; Chen, L. Machine condition recognition via hidden semi-Markov model. *Comput. Ind. Eng.* **2021**, *158*, 107430. [CrossRef]
103. Lee, Z.; Ean, W.; Najah, A.; Marlinda, A.; Malek, A. A systematic literature review of deep learning neural network for time series air quality forecasting. *Environ. Sci. Pollut. Res.* **2022**, *29*, 4958–4990. [CrossRef]
104. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. NIPS 2014 Deep Learning and Representation Learning Workshop. 2014. Available online: http://xxx.lanl.gov/abs/1412.3555 (accessed on 3 January 2023).
105. Gruber, N.; Jockisch, A. Are GRU Cells More Specific and LSTM Cells More Sensitive in Motive Classification of Text? *Front. Artif. Intell.* **2020**, *3*, 40. [CrossRef]
106. Cho, W.; Choi, E. Big data pre-processing methods with vehicle driving data using MapReduce techniques. *J. Supercomput.* **2017**, *73*, 3179–3195. [CrossRef]
107. Guo, J.; Wang, L.; Fukuda, I.; Ikago, K. Data-driven modeling of general damping systems by K-means clustering and two-stage regression. *Mech. Syst. Signal Process.* **2022**, *167*, 108572. [CrossRef]
108. Ferreira, V.; Pinho, A.; Souza, D.; Rodrigues, B. A New Clustering Approach for Automatic Oscillographic Records Segmentation. *Energies* **2021**, *14*, 6778. [CrossRef]

# Appendix E

sensors

MDPI

*Article*

# Online Monitoring of Sensor Calibration Status to Support Condition-Based Maintenance

**Alexandre Martins [1,2], Inácio Fonseca [3], José Torres Farinha [3,4], João Reis [1,*] and António J. Marques Cardoso [2]**

[1] EIGeS—Research Centre in Industrial Engineering, Management and Sustainability, Lusófona University, Campo Grande 376, 1749-024 Lisboa, Portugal
[2] CISE—Electromechatronic Systems Research Centre, University of Beira Interior, Calçada Fonte do Lameiro, 62001-001 Covilhã, Portugal
[3] Instituto Superior de Engenharia de Coimbra, Polytechnic of Coimbra, 3045-093 Coimbra, Portugal
[4] Centre for Mechanical Engineering, Materials and Processes—CEMMPRE, University of Coimbra, 3030-788 Coimbra, Portugal
* Correspondence: p40500@ulusofona.pt

**Abstract:** Condition-Based Maintenance (CBM), based on sensors, can only be reliable if the data used to extract information are also reliable. Industrial metrology plays a major role in ensuring the quality of the data collected by the sensors. To guarantee that the values collected by the sensors are reliable, it is necessary to have metrological traceability made by successive calibrations from higher standards to the sensors used in the factories. To ensure the reliability of the data, a calibration strategy must be put in place. Usually, sensors are only calibrated on a periodic basis; so, they often go for calibration without it being necessary or collect data inaccurately. In addition, the sensors are checked often, increasing the need for manpower, and sensor errors are frequently overlooked when the redundant sensor has a drift in the same direction. It is necessary to acquire a calibration strategy based on the sensor condition. Through online monitoring of sensor calibration status (OLM), it is possible to perform calibrations only when it is really necessary. To reach this end, this paper aims to provide a strategy to classify the health status of the production equipment and of the reading equipment that uses the same dataset. A measurement signal from four sensors was simulated, for which Artificial Intelligence and Machine Learning with unsupervised algorithms were used. This paper demonstrates how, through the same dataset, it is possible to obtain distinct information. Because of this, we have a very important feature creation process, followed by Principal Component Analysis (PCA), K-means clustering, and classification based on Hidden Markov Models (HMM). Through three hidden states of the HMM, which represent the health states of the production equipment, we will first detect, through correlations, the features of its status. After that, an HMM filter is used to eliminate those errors from the original signal. Next, an equal methodology is conducted for each sensor individually and using statistical features in the time domain where we can obtain, through HMM, the failures of each sensor.

**Keywords:** sensors; calibration; condition-based maintenance; online calibration status; HMM; K-means; PCA; features generation



**Citation:** Martins, A.; Fonseca, I.; Farinha, J.T.; Reis, J.; Cardoso, A.J.M. Online Monitoring of Sensor Calibration Status to Support Condition-Based Maintenance. *Sensors* **2023**, *23*, 2402. https://doi.org/10.3390/s23052402

Academic Editor: Jun Wu

Received: 20 January 2023
Revised: 7 February 2023
Accepted: 20 February 2023
Published: 21 February 2023



## 1. Introduction

### 1.1. The Importance of Sensors in CBM

Industrial maintenance is currently seen as an investment that dramatically decreases a company's production costs. For companies with highly critical equipment, where an unexpected stoppage can cost a very high daily monetary loss, it is necessary to implement a Condition-Based Maintenance (CBM) policy. In this type of maintenance, the equipment is monitored by several sensors, responsible for translating the physical behavior of the

equipment into electrical signals, amenable to reading [1]. Thus, sensors play a key role, allowing intelligent decisions, prediction of future conditions, etc. [2]. Through the collection of values continuously, in online mode, it is possible to determine the health status of the equipment in real-time. This requires the use of methodologies and algorithms that obtain information from the collected data [3]. Using Artificial Intelligence (AI) and Machine Learning (ML) methods, it is possible to detect patterns in the data that provide information about the behavior of the equipment.

In this area, there are some fundamental works: Li et al. [4] use AI and ML methods to classify patterns for the detection of False Data Injection Attacks (FDIA) in Cyber-Physical Systems (CPS) in order to improve security cybernetics of intelligent networks; Antunes et al. [5] use ML methodologies and propose a condition-based maintenance system of a wood chip pump system; Mateus et al. [6] use neural networks, such as Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU), to predict univariate and multivariate data in a maintenance system based on the conditions of the equipment used in the paper industry. Simoes et al. [7] use the HMM for a diesel engine maintenance system; Kou et al. [8] propose a condition-based monitoring and maintenance method for Smart Offshore Wind Farms using several ML tools.

### *1.2. Industrial Metrology to Support CBM*

In recent years, the increasing globalization of the market and the distributed production of highly complex technical systems have significantly increased the demand for reliable and accurate systems [9–11]. The need to acquire data constantly and online is increasing; so, reliability and accuracy in data collection are increasingly relevant [12], especially with regard to CBM. Having a large volume of data collected in the industry is valuable as long as it is accurate [13]. The author also presents smart metrology as a new approach based on reliability, which is presented as a solution. The lack of metrological traceability is a major obstacle faced [2]. Therefore, to use the information acquired by the collected data, it is necessary to trust the data. For this, metrology plays a very important role in this industrial process. Metrology generates information and knowledge [14], with its acquisition subject to measurement and information transfer able to derive from this knowledge consequences for the development of know-how and, finally, understanding for the management of process improvements and manufacturing products [14]. Industrial metrology can be seen as the basis of good CBM maintenance. It is responsible for performing the calibration of sensors, being concerned with measurement performed by them. With the advancement of Industry 4.0, there is an increasing pressure to improve the integration, interoperability, and accessibility of measurement information in industrial metrology and related activities with measurement information for operations. This contributes to establishing a reliable Internet of Things (IoT) or Cyber-Physical Systems (CPS) environment, where measurement data need to be accurate, reliable, and easily accessible [12]. Because of this, sensors need to be calibrated in accredited laboratories, compared with standards from higher levels of the traceability chain, and obtain a calibration certificate that ensures the reliability of the collected data [1]. Calibration Certificates (CC) are essential to maintain the accuracy of measuring instruments and guarantee the quality of products and services [15]. Calibration is a part of metrology that evaluates the quality and accuracy of measurements. Over time, measuring equipment can show errors in their results due to misuse or external factors such as environmental and operating conditions [13]. These deviations can also be caused by dirt or other materials contaminating [15] components. As explained by the author, these conditions result in inaccurate measurements and uncertainty in the results; to reduce these effects, any measuring instrument needs to be calibrated [15]. Calibration is conducted through metrological traceability. Traceability of measurements based on recognized standards is essential for data comparability [16]. For measurements to be considered metrologically accurate, they must be related to their units through a properly documented chain of calibrations. Each calibration contributes to the measurement uncertainty. Without a clear and complete record of these traceable calibrations, measurements

cannot be safely compared [17]. Typically, calibration occurs on a periodic basis, thus performing a preventive calibration, where often sensors undergo calibration without it being necessary (over-calibration) or else are working out of calibration because the time has not yet come for them to be calibrated (inefficient calibration).

As a consequence of the industry's competitive need to be highly efficient and quality conscious, manufacturing metrology is evolving from traditional engineering metrology, performed on a periodic basis, to automatic inspection methods by online mode [18]. This requires a type of calibration based on the condition of the sensor. In this way, the sensor undergoes calibration only when it is needed, making calibration management based on the condition of the sensor. For this, it is also necessary to obtain status information from the reading equipment through their own collected data.

### 1.3. Online Calibration Monitoring (OLM)

Online monitoring of sensor calibration status is assessed when the sensor needs to be removed from the equipment to be calibrated to an equivalent or higher standard (either locally or at an accredited laboratory). This type of performance monitoring is a condition-based methodology, offering an alternative approach to traditional calibration status maintenance performed at regular intervals or by checking the condition of the readout by conducting periodic checks. This can cause a sensor requiring calibration to be overlooked simply because the calibration interval has not yet passed or the sensor used for verification has a drift in the same direction as the sensor being monitored, which causes the need for calibration to not be detected. It should also be noted that much of the calibration monitoring effort is currently carried out in the verification of sensors that do not need any maintenance [1,19]. The extensive use of verification procedures as calibration procedures is undeniable, as is the fact that measurement standards errors are neglected during verification [20]. Whereas, with an online calibration status monitoring system, there may be a reduction in unnecessary field calibrations, which can reduce the associated labor costs; reduce the potential for incorrect calibration; and, if it is done in an oil or nuclear company, may reduce the radiation exposure of personnel as it uses the data from the sensor itself and monitoring can be performed under normal operational operation. Further, in this way, it becomes unnecessary to use redundant sensing to protect very important components.

### 1.4. Methodology Developed for OLM

Through the collected data, we can obtain information about the status of the production equipment as well as the reading equipment. The difficulty lies in distinguishing the different information from the same set of data, since a high reading can either be due to malfunctioning production equipment or by reading equipment that are out of calibration. The objective of this paper is to present a methodology capable of—through the same set of data—extracting information about the health state of the production equipment, as well as of the reading equipment. To solve metrology delays, consistent reading deviations, and sudden changes in deviations, the method was developed using a methodology of optimized observations, through ML processes, to provide input to the Hidden Markov Models (HMM) classifier, which after filtering errors of the production equipment, will determine calibration errors inherent to the sensors. So, this methodology adds value based on a set of data, where it is possible to collect health status information, both from the reading equipment attached to the production equipment, as well as from the production equipment itself. Furthermore, the methodology can be used online to obtain information in real-time. It is possible to evaluate the condition of the equipment, even in operation, without having to switch off to analyze it. Based on this, it also is no longer necessary to perform periodic checks made to the sensors that take a long time and high costs (due to the need for manpower). It is also possible to reduce the use of redundant sensors for the same component. Through pattern detection and classification performed by the HMM, it is possible to detect behaviors of the equipment without previous information about

them—that is, without knowing which data represent malfunction or good operation of the equipment, the AI and ML methodology can learn autonomously and without being supervised. The methodology can also be used in any type of equipment and/or sensor, making it generic for industrial support in maintenance and metrology. Through this method, maintenance and calibrations are only performed when necessary. This increases the availability of the equipment (both the production ones and the reading ones) and, consequently, an increase in the company's profits.

### *1.5. Related Work*

To present an overview to understand the evolution and current status of studies related to the subject in question, this chapter presents and analyzes the main relevant works published recently in the area. Articles presented propose techniques for identifying faults in sensors, as well as approaches to issues such as missing data, assessment of data reliability, and data prediction. Methods that use correlation methodologies are also demonstrated.

Lai et al. [21] propose a technique for identifying faults in sensor nodes using correlation theory to prevent fault data injection attacks, identifying these nodes based on spatial correlation and events. Tipireddy et al. [22] present virtual sensors to temporarily replace faulty physical sensors, allowing the safe postponement of recalibration, using a Gaussian model to process data from redundant and nearby sensors. Hines and Rasmussen et al. The authors of [23] discuss forecast range estimation methods for three nonlinear empirical modeling strategies (artificial neural networks, partial least squares neural networks, and local polynomial regression), applied to operational data from a nuclear power plant to monitor sensor calibration. Rao et al. [24] present a fault prediction method based on spatial correlation using the Vector Space Model (VSM) to identify reliable or faulty nodes. Berjab et al. [25] suggest a new method of extracting sensor relationships based on cross-correlation; it combines information from spatiotemporal correlations and multivariate attributes to determine whether the sensor has abnormalities or actual events. Lee and Chai et al. [26] propose a modification to Gaussian Process Regression (GPR) to improve the estimation of the sensor conditions in the online monitoring system of nuclear power plants. Fu et al. [27], define a strategy for detecting faults in Wireless Sensor Networks (WSNs), called the Trend-Correlation-based Fault Detection strategy (TCFD); the strategy detects damaged sensor nodes by analyzing the trend correlation and the mean value of neighboring nodes. Li et al. [28] present a method to identify serious failures in structural integrity monitoring sensors; the method uses a generalized likelihood ratio and correlation coefficient to evaluate each sensor in the network and detect faults through multiple hypothesis testing. Rajesh and Chaturvedi et al. [29] address in their article the problem of missing data in wireless sensor networks; the correlation between different data modalities is used to recover missing data and predict data; three classical estimates (Pearson, Spearman, and Kendall-tau) as well as four robust estimates of correlation coefficients are used to determine the correlation between modalities on data characteristics. Karmakar et al. [30] propose a new model that evaluates the reliability of IoT sensor data representing temporal correlation. Biswas and Samanta et al. [31] describe an algorithm to detect faults in sensor nodes; the algorithm is based on the Pearson correlation coefficient and the Support Vector Machine (SVM) algorithm.

## 2. Methodology

### *2.1. Signal Simulation*

To explain and test the developed methodology, a multivariate signal was created with four different non-redundant variables (Figure 1), i.e., sensors that measure different physical phenomena. An equipment error was simulated where all the sensors reacted to the changes in the equipment's behavior. After that, maintenance was performed, and the measured values returned to the equipment's normal operational behavior. Next, an error of one of the sensors was simulated, where the value of one of the sensors was increased

without the others suffering any changes, thus simulating a sensor drift. This leads to the conclusion that it is not a production equipment error but a sensor error, since there was no reaction from the other three sensors.

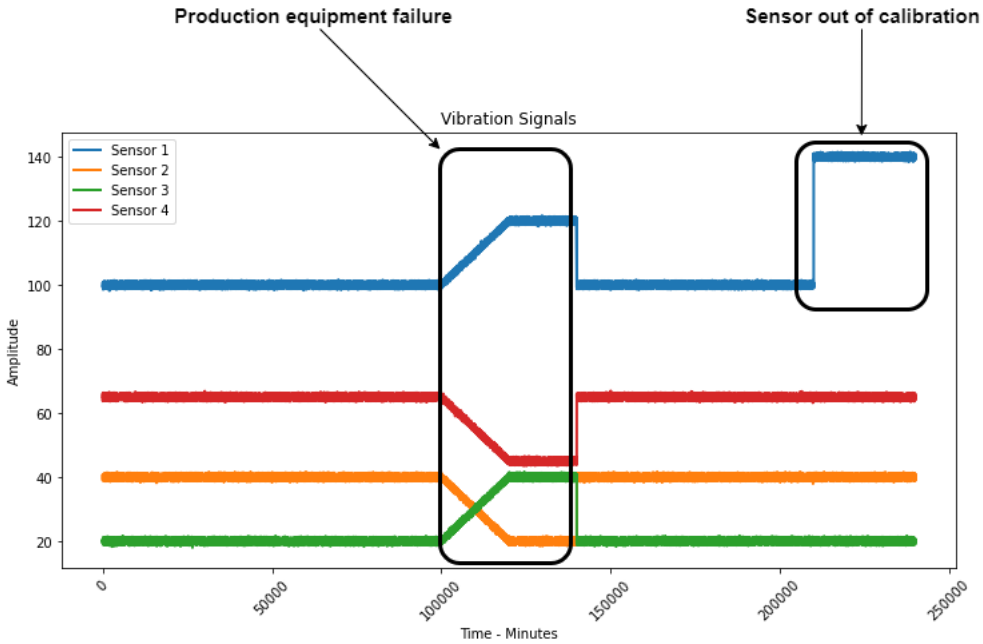



**Figure 1.** Simulated signal with production equipment error and deviation in one of the sensors.

It is necessary to select suitable sensor groups to be applied in a model, since the performance of the model is based on the correlation between the sensors [32]. For this method to work, it is necessary to have a large set of sensors or else a set of sensors that are correlated with changes in the behavior of the production equipment. According to Coble et al. [33], modeling non-redundant sensor clusters requires that the sensors in a model share related information, which can be identified by linear correlations or physical understanding. These modeling methods can also be applied to redundant sensor groups. In a heterogeneous environment with more than one sensor, the variables tend to be correlated [29]. Data from several heterogeneous sensors tend to show a strong correlation in space and time, which can be used to improve the detection of anomalies [25] and, in addition, improve the performance of each individual sensor [34].

Making a correlation study between the sensors, we can verify (Figure 2) that sensors present high correlation values only when there is a failure in the production equipment, keeping low values in normal operation or even when there is a deviation in one of the sensors.

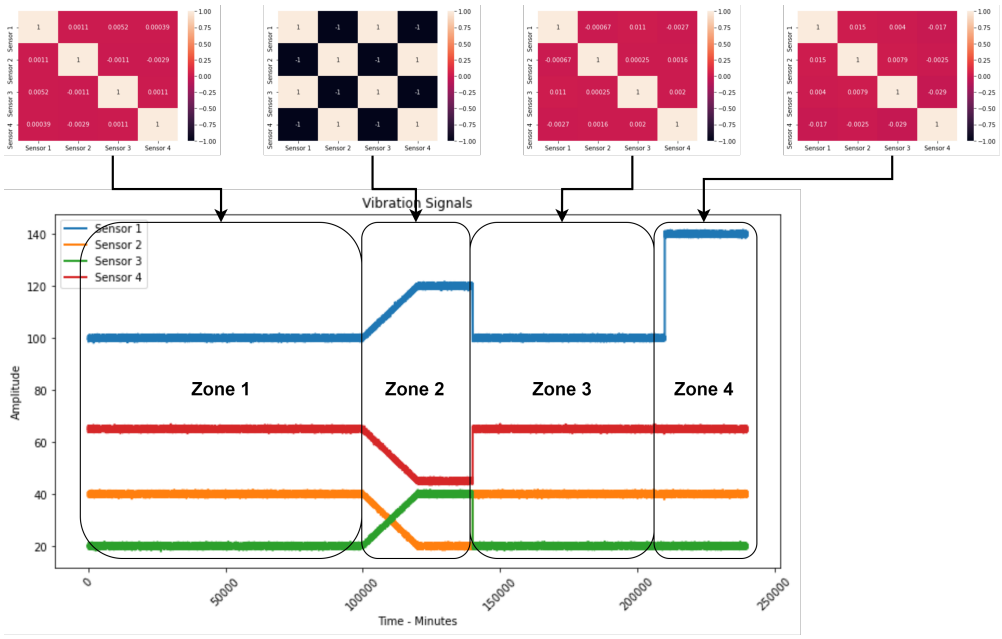



**Figure 2.** Correlations shown in each of the signal phases.

### *2.2. Generation of Correlation Features*

The first stage of any Machine-Learning-based model is feature selection, which depends on the problem under study [30]. The goal here is to use dispersed correlation values to detect faults in production equipment. To generate the features, the data collected from a sensor are divided into blocks [30]. Then, to be able to obtain the correlation states over time, temporal windows are created, where, for each one, we can choose the number of samples per window. Before creating the temporal windows, subtraction and ratio were performed on all the sensors to increase the relationship between them. They were tested without this method and the results obtained were not good. After increasing the relationship between the sensors, with $X = x_1, x_2, \ldots, x_n$ representing the dataset for each window, the temporal windows were created. In this example, temporal windows with a dataset of 288 samples per window were chosen, giving a total of 833 windows for this dataset with a sampling of 240.000 samples (Table 1). If the data were collected every 5 min, this would represent a 1-day time window.

**Table 1.** Time windows with data subtracted and ratios between each sensor pair.

| | **Col 1**<br>**Sensor 1/Sensor 2** | **Col 2**<br>**Sensor 1/Sensor 3** | **(...)** | **Col 24**<br>**Sensor 4/Sensor 3** |
|---|---|---|---|---|
| 1st Chunk | 60.63244195683154<br>59.64240130685382<br>(...)<br>59.262459386540584 | 79.78937757904451<br>79.86135304565789<br>(...)<br>79.87646380056816 | (...) | 3.0538616746523806<br>3.062736422782496<br>(...)<br>3.4567110848650304 |
| 2nd Chunk | 59.82199978785137<br>60.19302454746067<br>(...)<br>60.122575793590606 | 78.56797868615018<br>78.74172104561583<br>(...)<br>80.59289809317407 | (...) | 3.0538616746523806<br>3.062736422782496<br>(...)<br>3.4567110848650304 |
| (...) | (...) | (...) | (...) | (...) |
| 834th Chunk | 100.12391557061736<br>99.36665172104591<br>(...)<br>100.67188973084929 | 119.97019459713032<br>119.88307859672575<br>(...)<br>120.40164877024282 | (...) | 3.210611636252731<br>3.3163521246018797<br>(...)<br>3.2317746352961993 |

Continuous and time-varying features can provide a prediction of possible failures [35]. So, after creating the temporal windows, it is possible to create correlation features between each of the columns for each day (Table 2). In the case of multiple nearby sensors, the performance can be improved by performing cross-correlation between all sensor combinations in pairs [34].

As explained by Alqahtani et al. [36], correlation is a widely used mathematical technique for measuring the relationship between two or more variables by describing how they vary together. It is a common similarity evaluation technology [37]. Correlation in space and time is common in many physical phenomena, where spatial correlation refers to an association of measurements of two variables at a specific moment in time $t$ [21,31]. According to these authors, there are several possible measures of correlation, where the correlation coefficient can be viewed as a degree of linearity between X and Y variables. The degree of correlation describes the response of the structure built between two positions that reflects their level of correlation. The parameter ranges are from $-1$ to 1, where the closer the correlation is to the value 1, the more strongly the structural responses between two positions are correlated; when the correlation value is 0, it corresponds to more weakly correlated structural responses between two positions [38,39].

For this methodology, three types of correlations are used:

- Pearson Correlation

Pearson correlation is a technique that measures the covariance and degree of correlation between two estimates of input sets $X$ and $Y$ [39]. It is widely used in feature selection research [40] and is defined by the following equation (Equation (1)) [29,31,38]:

$$COO[X, Y] = \frac{cov[X, Y]}{\sqrt{var[X], var[Y]}} \quad (1)$$

where

$cov[X, Y]$ is the covariance of $X$ and $Y$;
$var[X]$ is the variance of the random variable $X$.

- Spearman Correlation

Spearman's correlation is an alternative measure to Pearson's correlation, which assesses monotonous relationships, not just linear ones. Instead of using the actual values of the observations of two variables $X$ and $Y$, Spearman's correlation uses the corresponding ranks, $rg(X_i)$ and $rg(Y_i)$, to measure the similarity between the observations [29,41]. It can be represented by Equation (2):

$$rs = \frac{cov(rg(X), rg(Y))}{\sqrt{\sigma_{rg}(X)\sigma_{rg}(Y)}} \quad (2)$$

where

$cov(rg(X), rg(Y))$ is the covariance of the variables in ranks;
$\sigma_{rg}(X)$ and $\sigma_{rg}(Y)$ are the standard deviations of the variables in ranks.
Another popular formula to represent the Spearman correlation is given by Equation (3):

$$rs = 1 - \frac{6\sum_{i=1}^{n} d_1^2}{n(n^2 - 1)} \quad (3)$$

where

$d_i = rg(X_i) - rg(Y_i)$;
$n$ is the number of observations.

- Kendall Correlation

Any pair of observations $(x_i, y_i)$ and $(x_j, y_j)$, where $i \neq j$, is concordant if the ratings of both elements agree with each other, i.e., if $x_i > x_j$ and $y_i > y_j$. They are discordant

if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant nor discordant. Kendall's $\tau$ coefficient is defined as Equation (4) ([29]):

$$\tau = \frac{n_1 - n_2}{n(n-1)/2} \tag{4}$$

After creating the time windows, the three correlation measures are calculated for all pairs represented in Table 1. This means that, for each time window, there will be a combination of the correlation measures for all sensor subtraction and ratio pairs. After making all the correlations, Table 2 and the graph shown in Figure 3 were obtained, representing all the correlation characteristics obtained over time.

**Table 2.** Correlations between each pair of sensors in each of the time windows.

| | **Col 1**<br>**Corr. Pearson Col1-Col2** | **(...)** | **Col 828**<br>**Corr. Spearman Col13-Col1** | **(...)** | **Col 1656**<br>**Corr. Kendall Col24-Col23** |
|---|---|---|---|---|---|
| 1st Chunk | 0.53820 | (...) | 0.90987 | (...) | 0.08561 |
| 2nd Chunk | 0.49547 | (...) | 0.88902 | (...) | 0.14987 |
| (...) | (...) | (...) | (...) | (...) | (...) |
| 834th Chunk | 0.43642 | (...) | 0.84575 | (...) | 0.05570 |

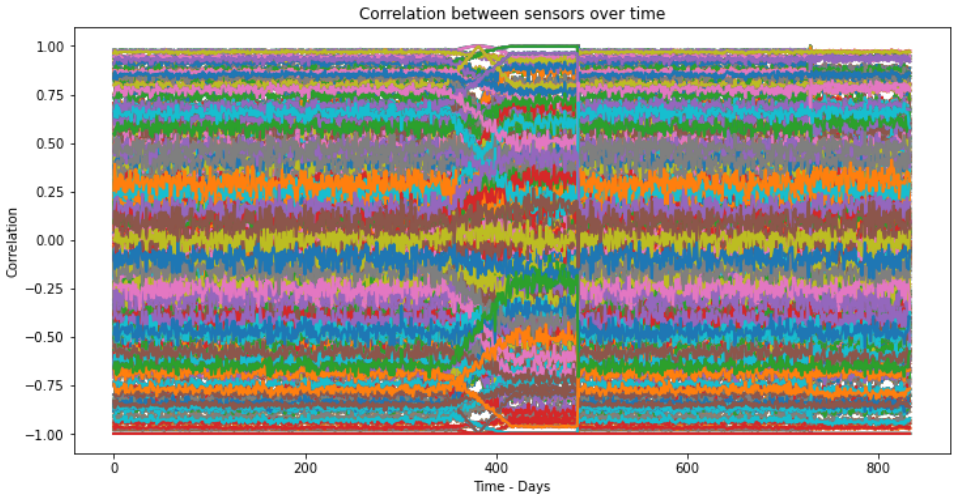



**Figure 3.** Correlations between each sensor pair over signal time.

### *2.3. Normalization*

To ensure better results, normalization attempts are important [42]. There is a need to normalize the data to make the magnitude and time scale uniform [25]. To obtain consistent results, usually, ML models use a normalization mechanism before training [43]. The normalization of the data through the Z-score (Equation (5)) aims to transform the magnitude and dispersion of the data so that the mean is 0 and the standard deviation is 1—that is, $Z \sim N(0.1)$. This method helps to make patterns in the data more visible, as it converts a range of variables so they all have the same range, transforming not only the magnitude of the data but also the scatter [44].

$$Z_{Score} = \frac{x_i - \bar{X}}{\sigma(X)} \tag{5}$$

where

$\bar{X}$ is the mean of the dataset $X$;

$\sigma(X)$ is the standard deviation of $X$.

Through normalization, the amplitudes of the initial continuous variables also contribute to the analysis, with features with larger amplitudes not overlapping, which could lead to biased results [43]. As the authors explain, to avoid this problem, it is recommended to scale each feature in the same value range, which, in turn, also increases the training speed. Normalization is one of several data transformation techniques, which has the effect of reducing the parameters to a common range, providing a measure that allows the relative importance of any factor or interaction to be identified more clearly [45].

As can be seen through Figure 4, all correlations for all sensor pair combinations are normalized with mean zero and standard deviation 1. This causes two-phase peaks to be created. This means that values closer to 0 will be more concentrated and values farther from 0 will be less frequent. After performing normalization, it is already possible to better understand the behavior of the data. The normalization of the dataset will also help to improve the K-means clustering technique [46], which will be used later.

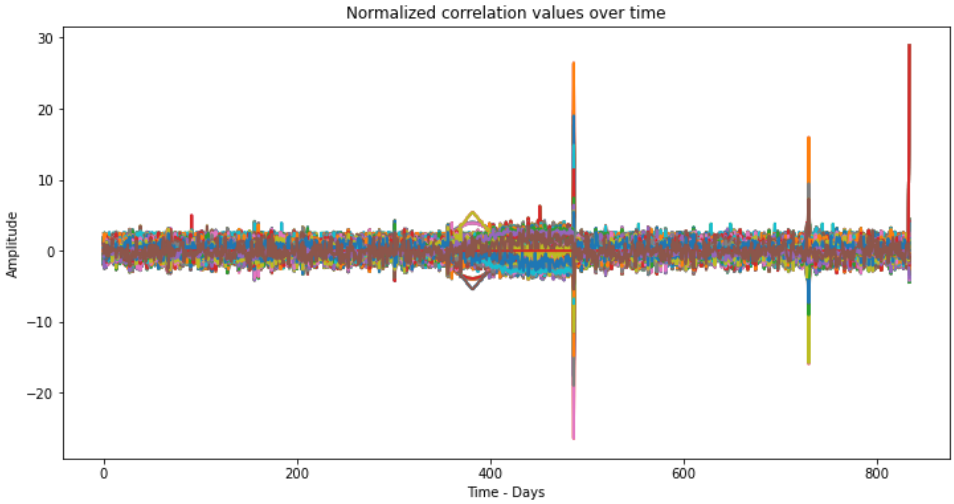



**Figure 4.** Correlation features normalized by Z-score over study time.

### *2.4. Dimensional Reduction through Principal Components Analysis (PCA)*

After normalization, a dimensionality reduction technique, Principal Component Analysis (PCA), is used, which can be considered a fault detection method based on multivariate statistical analysis [28]. It is a feature extraction and dimensionality reduction method in Machine Learning [35] and is one of the most widely used methods for dimensionality reduction of data from a multidimensional space [40,47]. It can reduce the dimensionality of high-dimensional data and remove noise by dimensionality reduction [48]. To perform dimensionality reduction and feature extraction, the data undergo orthonormal rotations of the coordinate system. In this way, we will increase the processing speed of the algorithm as we will use new variables that have more important information, thus increasing the prediction ability. It is a technique that transforms several potentially correlated variables into a set of uncorrelated variables, the principal components (PCs). The first principal components are responsible for explaining most of the information present in the data, with the number of PCs being limited to the number of original variables [49]. The PCA process aims to find a new coordinate system of the dataset centered on the mean, whose axes are perpendicular and have a maximum variance in descending order [50]. So, $PC_1$ is responsible for having the highest data variability, while each subsequent PC has the highest possible variance, under the constraint of being orthogonal to the previous PCs [47].

As Zhang et al. [51] explain, the algorithm can be described as follows: in the feature space of dimension $N$, we find a direction that maximizes the variance of the data. We then use that direction as our first principal direction and project the data onto the $N-1$

dimension space, removing the principal direction. This process is repeated $M$ times, where $M \leq N$, to obtain a transformation of the data in the main dimensions.

The optimization process in feature selection is based on new components representing correlated features. These components are generated as columns in the matrix $X'$, making it $m \times n'$. To preserve the original feature information, it is common that $n' = n - 1$. The desired default dimension for the lower-dimensional space is $n' = n - 1$, where $n$ represents the dimension of the original dimensional space.

The cumulative variance is used to assess how much of the original data information is retained in each principal component. According to Figure 5, it is possible to see, after calculating the cumulative variance, that the first 10 PCs contain about 95% of the total variance of the original data.

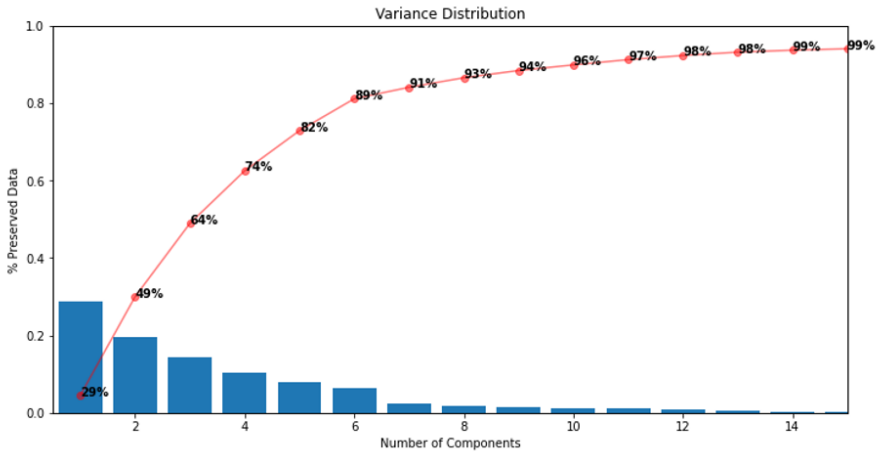



**Figure 5.** Pareto chart with the percentage of information for each PC.

It was possible to preserve about 95% of the raw information by shrinking the matrix from $X = [x_{ij}]_{(834\times1656)}$ to $X' = [x'_{ij}]_{(834\times10)}$.

We can also verify, through Figure 6, the movement over time of the points of each of the first 10 PCs coming from the orthogonal reorientation performed by the PCA. It is already possible to better distinguish the areas that are really coming out of the normal operating standards of the equipment.

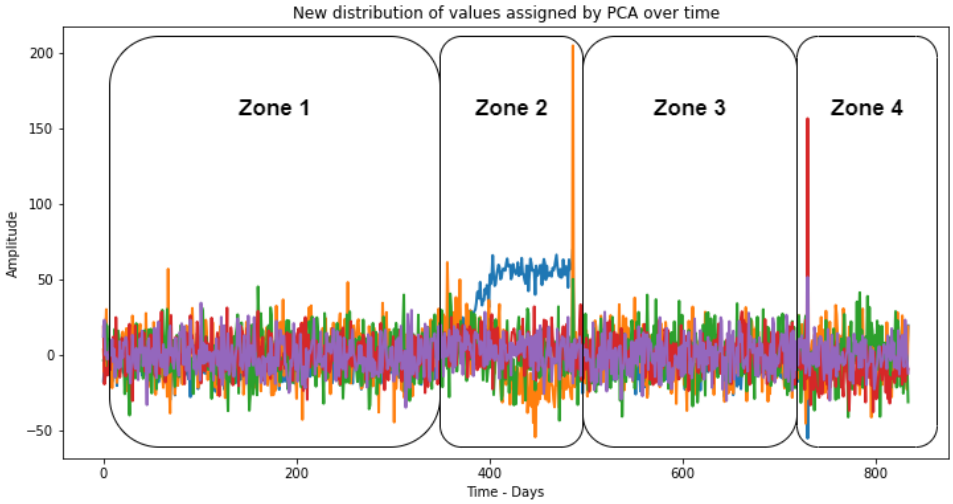



**Figure 6.** Data movement of the first 10 PCs processed by PCA.

When projecting the data of Figure 4 into 10 PCs, the two-phase peaks were lost because the PCs are directions that maximize the variation of the data and do not necessarily preserve the original characteristics of the data. Since the two-phase peaks were not related to the directions of maximum variation of the data, they were lost when projecting the data into the principal components.

In addition, PCA aims to reduce the dimensionality of the data—that is, to represent the data with fewer principal components without losing much information. This means that some of the less expressive peaks can be lost in the projection process to the principal components and, hence, represent the PCs through time.

### *2.5. Clustering through K-Means*

According to Thrun and Ultsch et al. [52], the use of the PCA algorithm is frequent when there is a high number of variables since it allows the application of a preliminary reduction of these variables; next, it is common to use K-means clustering based on early PCs that preserve a significant amount of information.

The cluster analysis [53] achieves the following:

- Examines the underlying structure of the data;
- Identifies patterns and categories in the data in order to establish the similarity between the points;
- Performs dimensionality reduction, with the aim of grouping and simplifying the data in an understandable way.

In the K-means clustering algorithm, data points are grouped within a cluster based on similar shared characteristics [48,54], with good equipment functioning data being in the same cluster and bad equipment functioning data in a distinct cluster.

K-means is one of the first proposed clustering methods and assumes that each sample is linked to only one group, assigned to the one closest to [55]. It is an unsupervised technique that is widely used to identify similarities between objects based on distance measurements suitable for small datasets [53]. It has several advantages including brevity, simplicity, efficiency, speed, and less computational power, which make it the most widely used clustering algorithm [53,56–58].

The K-means algorithm has the main objective of grouping similar data points and revealing the structure underlying the data [59]. This is achieved by fixing a defined number of clusters (k) to be used in the analysis. To each one of the clusters is assigned a centroid, which has a location in the center of the cluster. After k is chosen, each data point is allocated to the nearest cluster by summing the squared distances of the Euclidean distances among the items and the centroid (Equation (6)), minimizing intra-cluster variation [59].

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \tag{6}$$

Here, $x_i$ is the $i$th data point of cluster $k(C_k)$ and $\mu_k$ is the mean value of the points in cluster $k$. The Total Within-Cluster Variation Equation (7) describes the clustering quality, in which the Sum of Squared Errors (SSE) is used as a measure. The lower the SSE, the higher the quality of the cluster [58].

$$TotalWithinClusterVariation = \sum_{k=1}^{k} W(C_k) \tag{7}$$

Then, as explained by Peng et al. [48] and Borlea et al. [60], K-means is used to process a dataset $D = x_1, x_2, \ldots, x_n \in \Re^d$, where $x$ is a dataset record defined as $X_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^T \in \Re^d, i = 1 \ldots n$. $d$ is the dimension of a dataset record and $T$ stands for matrix transpose. The algorithm divides the dataset D into a set of k predefined numbers of clusters $C_(j), j = 1 \ldots k$. Each cluster $C_j$ is composed by a center of mass called centroid and defined as $C_j = [C_{j1}, C_{j2}, \ldots, C_{jd}]^T \in \Re^d, j = 1 \ldots k$. The total number of points assigned to

each cluster is $n$, with the cluster expression $C_j = (c_j, n_{C_j})$. The centroid array is defined as $c = [c_1^T, c_2^T, \ldots, c_k^T]^T \in \Re^d k$, which represents the centroids of all existing clusters. The main objective of the algorithm is to minimize the intra-cluster variance (Equation (8)).

$$c^* = \arg\min_{c \in \Re^{dk}} V(c), \; V(c) = \sum_{j=1}^{k} \sum_{\substack{i=1 \\ x_i \in C_j}}^{n_{c_j}} ||x_i - c_j|| \tag{8}$$

where

$c_j$ is the centroid of the cluster $C(j), j = 1 \ldots k$;
V is the objective function or the criterion;
$c^*$ is the optimal arrangement of centroids.
The K-means clustering process has the following steps [54,56,58–61]:

1. Specify the number of clusters (k);
2. Randomly select k data points as initial centroids;
3. Assign the dataset $x_i$ to the nearest centroid $c_j$ using the Euclidean distance (Equation (9));

$$d_{x_i, C_j} = ||x_i - c_j|| = \sqrt{(x_{i1} - c_{j1})^2 + (x_{i2} - c_{j2})^2 + \ldots + (x_{id} - c_{jd})^2} \tag{9}$$

4. Next, all data points are redistributed using the previous process to find the next clusters. The process continues like this until the elements in each cluster are no longer changed.

Due to the initial cluster center and the clustering criterion function of similarity measure, it easily converges to the local minimum, selecting different initial clustering centers to lead with different clustering results [62,63]. As the authors say, the correct selection of the initial clustering center in the K-means algorithm has great influence on the quality of clustering results. To calculate the optimal number of clusters k, there are different methods [59] and, for this methodology, we will resort to the Sum of Squared Error (SSE) method, or the elbow method, as it is known because of the graph it forms. SSE is one of the most popular cluster evaluation methods [58]. We use different numbers of k and calculate the total based on the sum of squares for each value of k and plot these in Figure 7, where k is represented in the plot as a fold (elbow) location, which we consider the optimal $k$ number. The elbow method is the most widely used and comprises four steps [57]:

- Perform a centroid-based clustering variance of each clustering result, e.g., sum of squared errors algorithm, such as K-means, for each $k \in \Re$;
- Calculate the (SSE) for K-means;
- Plot the results on a graph;
- Select the elbow curve on the graph.

The amount of clusters is represented on the x-axis by the value $k$; after viewing the corresponding graph, it is noted that the reduction in the sum of squared errors (SSE) becomes negligible with the increase in the value of $k$.

So, the aim of the K-means algorithm is to minimize the sum of squared errors of the criterion [58]. Thus, through the elbow graph illustrated in Figure 7, it is possible to observe that the total distance of the sum of squares decreases as the value of k increases. However, from $k = 4$, the additional clusters cause only an insignificant reduction in the sum of squares. Thus, the ideal number of clusters can be considered as 4.

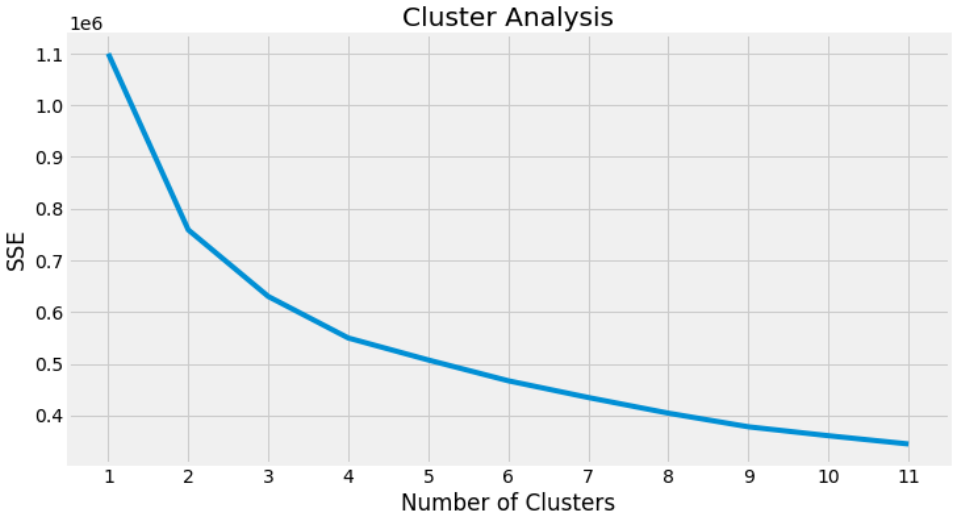



**Figure 7.** SSE Analysis to determine nº of clusters.

Having chosen the number $k$ of clusters, we can now proceed with the K-means algorithm itself, which will fulfill the steps mentioned above. After grouping the data in clusters, each cluster $C_j$ will be represented over time (Figure 8). Thus, clusters $C_1, C_2, C_3, C_4$ will be seen as new optimized observations of the values read by the four sensors. Note that an increasing ordering of clusters is chosen so that the first cluster is the one with the highest number of points and the last cluster has the lowest number of points, $n_{C_1} > n_{C_2} > n_{C_3} > n_{C_4}$.

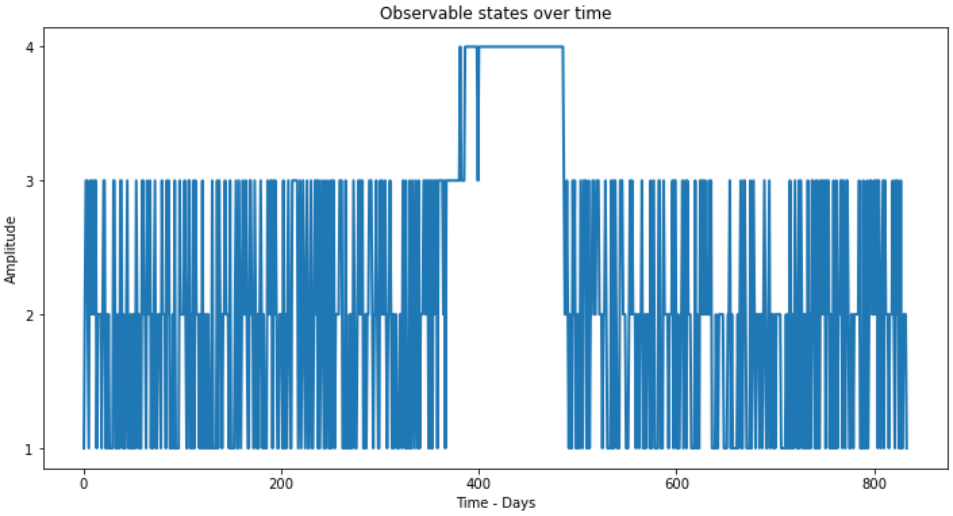



**Figure 8.** Cluster (optimal observable states) over study time.

### *2.6. Behavior Classification of Production Equipment Using HMM*

The Hidden Markov Model (HMM) is a doubly stochastic process, which has hidden states and observable states [64].

A typical HMM model can be explained by the parameters $\lambda = (N, M, \pi, A, B)$, where [7,65,66] $N$ represents the number of hidden states, $S = S_1, S_2, .., S_N$; $M$ represents the number of observable states, $C, = C_1, C_2, ..., C_M$; $A$ represents the transition matrix that specifies the transition probabilities between the hidden states, $A = a_{ijN*N}$; $B$ is the emission matrix, which specifies the probabilities of observing a given observable state given that the system is in a given hidden state, $B = b_{jkN*M}$; $\pi$ is the initial probability vector, which specifies the initial probabilities of being in each hidden state, $\pi_i = P(q_1 = S_i)$.

For simplicity, the parameters of the HMM model can be represented by the notation $\lambda = (\pi, A, B)$.

After computing the optimized time series observations, which merge multivariate information from sensors attached to the equipment, we can now use the clusters as observable states to provide input to the HMM model. The observations at instant t are represented by $O_t \in C, = C_1, C_2, C_3, C_4$; already, the hidden states will represent the operating states of the production equipment within a certain state $q_t \in S, S = S_1, S_2, \ldots, S_N$. The hidden states are a Markov Chain, which is obtained through the observable states—that is, through the observable states (collected and optimized by ML processes), we can deduce the status of the equipment. In this case, we will choose three hidden states, $S = S_1, S_2, S_3$, to represent the operation of the equipment, where state $S_1$ is the one with the most points and state $S_3$ is the one with the fewest points. If $n_{S_1} > n_{S_2} > n_{S_3}$, it can be deduced that the first state, which happens more often, will represent the good functioning state and states 2 and 3 will represent alert and failure states. For HMM, there are three basic problems that need to be solved [7,64,67]:

- The evaluation problem—which computes the probability of the observed fusion outcome sequence $O = O_1, O_2, ..., O_T$, given the model $\lambda = (\pi, A, B)$. This is performed using the forward–backward algorithm.
- The training problem—which adjusts the model parameters, $\lambda = (\pi, A, B)$, to maximize the probability of the observed sequence, i.e., given a chain of observable states, which model $\lambda$ best fits, $P(O|\lambda)$. This is performed using the Baum–Welch algorithm.
- The prediction problem—calculates the most probable hidden state sequence according to the observation sequence and the model parameters. Through the model $\lambda$ and the observation sequence $O$, it is possible to detect the best hidden state sequence $S$. It can be solved by the Viterbi algorithm.

Thus, through the observable states coming from the clustering phase, in the first step, the training is applied through the Baum–Welch method, where it is possible to obtain the model parameters $\lambda$. Then, through the observations and the model parameters, it is possible to determine the sequence of hidden states (Figure 9) using the Viterbi algorithm.

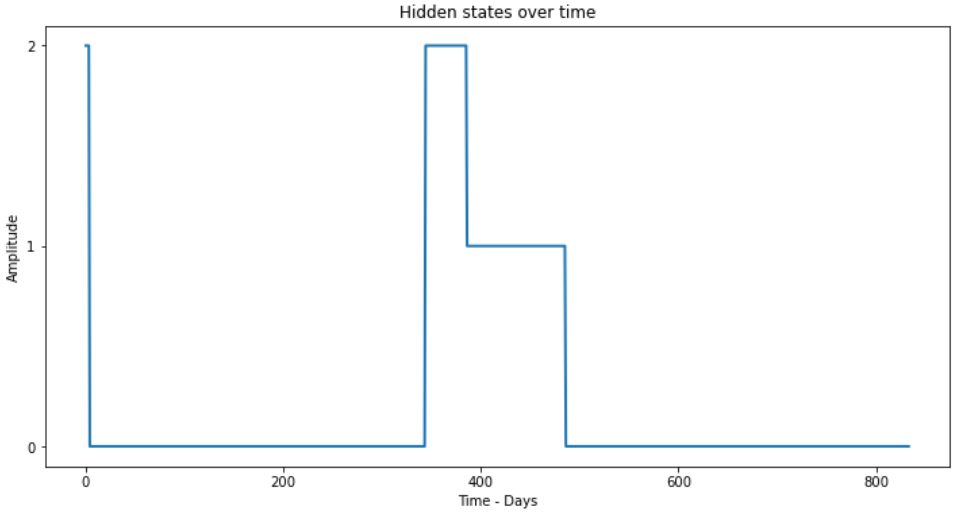



**Figure 9.** Hidden States (health states of the production equipment) over the study time.

### 2.7. HMM Filter

As can be seen in Figure 10, the HMM detects hidden states different from 0 at the time when the equipment malfunction is simulated, showing that this methodology works and detects only the errors of the production equipment. It starts with state 3 due to the initiation problems of the Viterbi algorithm and these initial values can be ignored.

Once the malfunctioning states of the production equipment are detected, these are eliminated in the original signal using a filter with the values of the hidden states of the HMM, i.e., any point in the original signal that coincides at the same time with any hidden state other than the first state—which means, being coincident with $S_2$ or $S_3$—is eliminated (Figure 10). In this way, we eliminate the errors of the original signal production equipment, leaving only errors that may arise from the reading equipment (Figure 11).

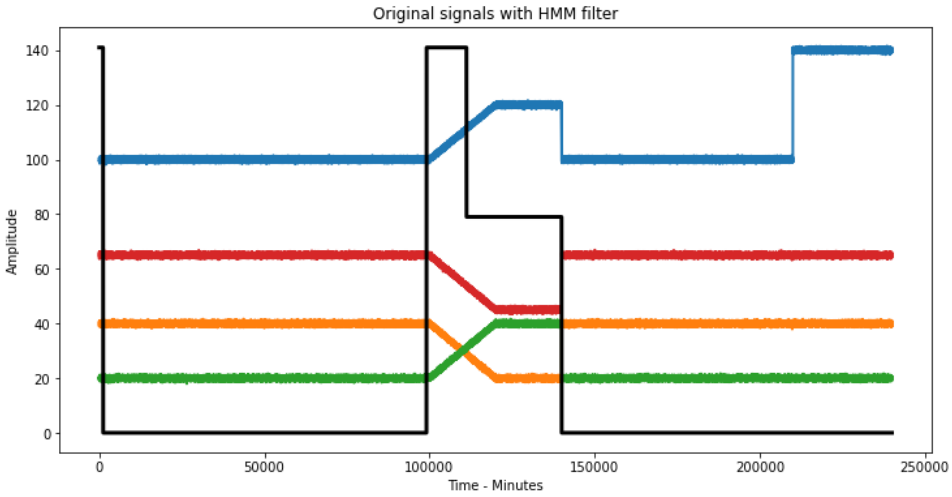



**Figure 10.** Original signal with overlapping HMM states.

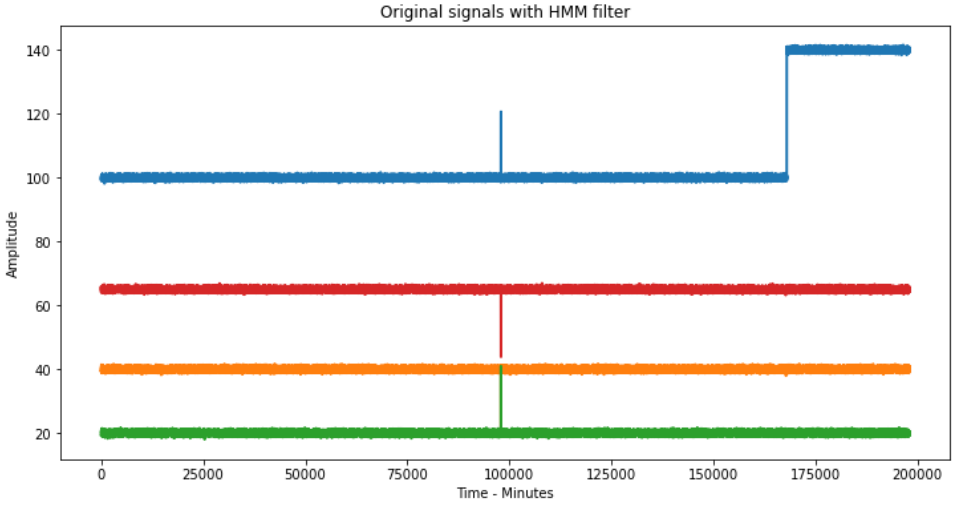



**Figure 11.** Original signal after HMM filter.

### *2.8. Classification of Sensor Behaviour*

As stated by Boechat et al. [68], an assumption is that the sensor deviations are not correlated with each other, despite the correlations among the process variables. Through the methodology presented above, we can demonstrate just that. So, now, we have only one signal with sensor errors. In this way, the objective is to use the same steps of the above methodology with the difference that now, instead of creating features of correlation among sensors, we will apply the method sensor-to-sensor using statistical features in the time domain (Table 3). As stated by Saucedo-Dorantes et al. [69], statistical features based on the time domain provide a good performance basis to characterize patterns and behavioral changes in equipment. Several features were used (Table 3), which were taken from other papers whose aims were to detect faults over time [44,69,70]. An individual study for each sensor is performed with the aim of understanding how it develops over time.

**Table 3.** Mathematical equations for time-domain-based statistical features.

| Parameter | Mathematical Equation | Parameter | Mathematical Equation |
|---|---|---|---|
| Mean | $T_1 = \frac{\sum_{n=1}^{N} x(n)}{N}$ | A Factor | $T_{12} = \frac{T_5}{T_2 . T_3}$ |
| Standard Deviation | $T_2 = \sqrt{\frac{\sum_{n=1}^{N} (x(n)-T_1)^2}{N-1}}$ | B Factor | $T_{13} = \frac{T_7 . T_8}{T_2}$ |
| Variance | $T_3 = \frac{\sum_{n=1}^{N} (x(n)-T_1)^2}{N-1}$ | SRM | $T_{14} = \left( \frac{\sum_{n=1}^{N} \sqrt{\lvert x(n) \rvert}}{N} \right)^2$ |
| RMS | $T_4 = \sqrt{\frac{\sum_{n=1}^{N} (x(n))^2}{N-1}}$ | SRM Shape Factor | $T_{15} = \frac{T_{14}}{T_1}$ |
| Absolute Maximum | $T_5 = max\lvert x(n) \rvert$ | Latitude Factor | $T_{16} = \frac{T_5}{T_{14}}$ |
| Coefficient of Skewness | $T_6 = \sqrt{\frac{\sum_{n=1}^{N} (x(n)-T_1)^3}{(N-1).T_2^3}}$ | Fifth Moment | $T_{17} = \sqrt{\frac{\sum_{n=1}^{N} (x(n)-T_1)^5}{(N-1).T_2^5}}$ |
| Kurtosis | $T_7 = \sqrt{\frac{\sum_{n=1}^{N} (x(n)-T_1)^4}{(N-1).T_2^4}}$ | Sixth Moment | $T_{18} = \sqrt{\frac{\sum_{n=1}^{N} (x(n)-T_1)^6}{(N-1).T_2^6}}$ |
| Crest Factor | $T_8 = \frac{T_5}{T_4}$ | Median | $T_{19} = medianx(n)$ |
| Margin Factor | $T_9 = \frac{t_5}{T_3}$ | Mode | $T_{20} = modex(n)$ |
| RMS Shape Factor | $T_{10} = \frac{T_4}{\frac{1}{N} \sum_{n=1}^{N} \lvert x(n) \rvert}$ | Minimum | $T_{21} = minx(n)$ |
| Impulse Factor | $T_{11} = \frac{T_5}{\frac{1}{N} \sum_{n=1}^{N} \lvert x(n) \rvert}$ | | |

After applying the methodology to each sensor individually and with the statistical features in the time domain, we obtain the results of the hidden states given by the HMM, as can be seen in Figure 12.

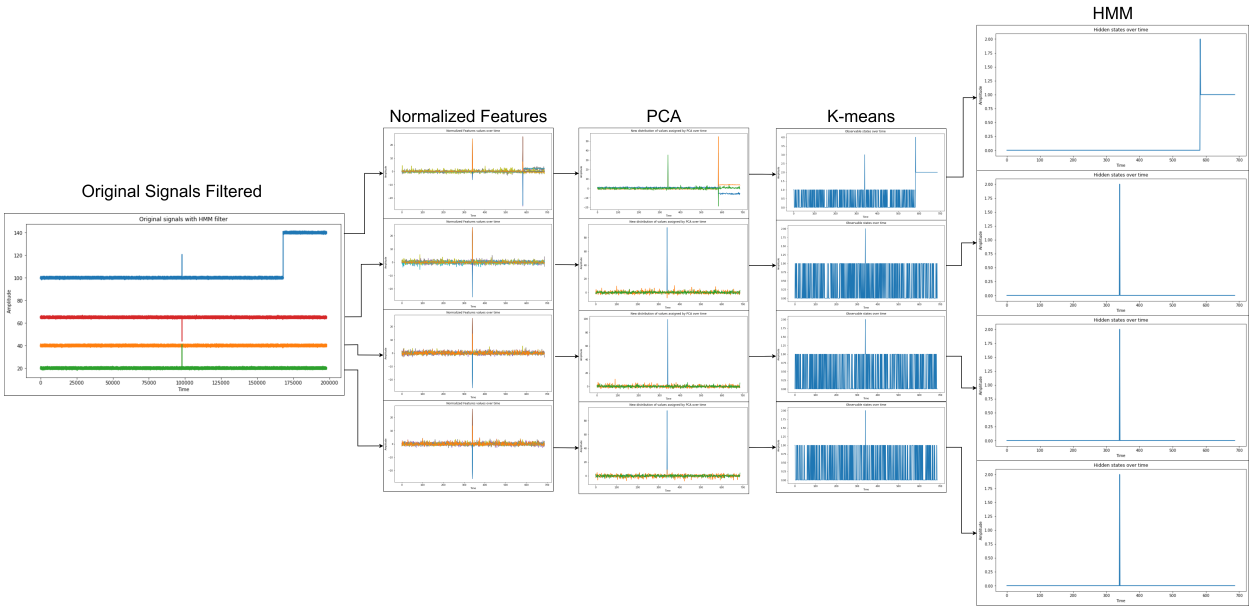



**Figure 12.** Methodology used to detect sensor errors individually.

## 3. Discussion of Results

Through the analysis of the results obtained, we can verify that, in the first phase, the HMM is able to classify the faults of the production equipment, without detecting the sensor drift (Figure 9). It detects state 3 in an initial phase that happens when the equipment starts to present the failure. Then, it moves to state 2, where it remains until the maintenance is performed. We can verify that the chosen correlation features will ensure that the methodology detects only equipment failures since they only give relevance to correlated behavioral changes of the sensors. The selection of important features in the data requires sufficient communication of domain expert knowledge. Thus, the choice of similarity measures and feature extraction techniques are critical to the success of clustering approaches, as this significantly affects the quality of results obtained when processing data to identify relevant patterns [36]. In this way, it is possible to extract only the health status information from the production equipment.

Having only the status of the production equipment, it is possible to use this information to filter the original data. After that, the HMM is able to detect only the sensor errors. Through Figure 12, we can verify that, in the first sensor, a hidden state different from state 1 is only detected at the end of the study period, which is when the sensor represents a deviation. When a hidden state different from state 1 is displayed for some time, we can conclude that there is a deviation in the sensor under study. As for the other sensors, which did not present any deviation, the HMM only stays different from state 1 for a very brief instant, which means it has no relevance. Therefore, we cannot admit that this is a sensor error. This only happens because, when the HMM filter is made based on the faults of the production equipment, there is a small part of the error that is not eliminated. It is quickly despised, because it is present in the final classification made by the HMM to determine deviations from the sensors.

This methodology has the added value of being able to obtain two different types of information from the same set of data. When a sensor rise is presented, it is difficult to know whether it is caused by the production equipment or by the reading equipment. This highlights the importance of a method that can extract the correct information to be used. So, we can admit that this methodology will improve CBM actions since it also detects sensor errors. Thus, it is possible to ensure that the data are reliable to understand the true information that the data are transmitting. Furthermore, this methodology can be used in different equipment and sensors with added value because it does not need

previous information about the equipment failures. It is an unsupervised methodology where, through AI and ML, the algorithm can provide information about the health state of equipment to the production engineer.

In future work, it is intended to increase the range of algorithms to be used in each of the steps to understand the best possible combination for the methodology. Other types of strategies will be used in normalization, as well as in dimensional reduction and clustering. Furthermore, a study of other types of correlation will be conducted, such as distance correlation. Temporal correlations will also be applied to evaluate the correlations between temporal windows t and t+1, with a single sensor. The hidden states coming from the HMM model will be used to build a supervised classifier, such as Support Vector Machine (SVM). This classifier will be used to direct the new data collected by the sensors into the hidden states of the HMM, allowing to immediately determine the health status of the equipment without having to go through the entire methodology. For this, it will be necessary to train the methodology so that it can teach the supervised classifier. In addition, we will have an adaptive methodology, in which, through a distance metric, an alert will be issued if new data collected appears too distant from the rest. This would mean that there is a new behavior in the equipment that has never been seen before, which means that the methodology needs to be updated. Besides this, it is intended to apply this methodology in a real factory aiming to demonstrate its importance in real situations.

For practical cases, it will be necessary to perform an initial cleanup to remove incorrect or inconsistent data such as zeros, NaN (Not a Number), or infinities. An algorithm will also be created to eliminate equipment downtime. In addition, depending on the application, additional filtering may be required to remove noise in the data. There are several filtering techniques that can be chosen, depending on the study. Since the goal is to study failure cases, outliers can be seen as important data. Therefore, a quartile filter performed on time windows will be created. In this way, it will be possible to filter out noise without losing valuable information for the study. Moreover, as Martins et al. [44] explain, the methodology itself uses tools that help with filtering data throughout the process.

## 4. Conclusions

The methodology goes through several steps, where the first is the correlation between the sensors over time, using temporal windows. The subtraction is made alongside the ratio between each sensor, aiming to increase the relationship between them; then, three types of correlations are made: Pearson, Spearman, and Kendall. Based on this, we obtain several features after passing through the Z-score normalization, which undergo a new orthogonal variation through the PCA, to extract a new set of variables that is reduced but with more information. This feature extraction process is also responsible for increasing the prediction quality of K-means, which aims to group the most similar values, thus creating a set of clusters that have new optimized observations. These observations feed the HMM classifier, which, through three hidden states (1—Smooth operation; 2—Warning; 3—Failure), is responsible for detecting the health state of the production equipment. Knowing the health states of the equipment, through the hidden states of the HMM, a filter is performed on the original data, where the values coincide with the different states from the well-functioning state, which permits filtering the data when the machine is in good working order. After that, the sensors are studied one by one, through the same methodology, but now making a generation of statistical characteristics in the time domain to evaluate the individual behavior of each sensor. Finally, again through the HMM and the three hidden states, it is possible to define the status of each sensor. If states 2 or 3 remain for some time, it is necessary to take the sensor for calibration because it is having a deviation. We can conclude that, through this methodology, it is possible to detect failures in the production equipment as well as deviations in the sensors that need to be calibrated. The authors will continue working to improve the implementation in practical cases.

**Author Contributions:** Conceptualization, A.M., J.T.F., I.F., and A.J.M.C.; methodology, A.M., I.F. and J.T.F.; software, A.M. and I.F.; validation, A.M., J.T.F., I.F., and A.J.M.C.; formal analysis, A.M.,

J.T.F. and I.F.; investigation, A.M. and I.F.; resources, A.M., J.T.F. and A.J.M.C.; writing—original draft preparation, A.M.; writing—review and editing, J.T.F., I.F. and J.R.; project administration, J.T.F. and A.J.M.C.; funding acquisition, J.T.F. and A.J.M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** Our research, leading to these results, has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowvska-Curie grant agreement 871284, project SSHARE; the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under project POCI-01-0145-FEDER-029494; and by national funds through the FCT—Portuguese Foundation for Science and Technology, under projects PTDC/EEI-EEE/29494/2017, UIDB/04131/2020, and UIDP/04131/2020.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Restrictions apply to the availability of these data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| AI | Artificial Intelligence |
| CBM | Condition-Based Maintenance |
| CC | Calibration Certificates |
| CPS | Cyber-Physical Systems |
| HMM | Hidden Markov Models |
| IoT | Internet of Things |
| ML | Machine Learning |
| OLM | Online Calibration monitoring |
| PCA | Principal Component Analysis |
| PCs | Principal Components |
| SVM | Support Vector Machine |
| TCFD | Trend-Correlation-based Fault Detection |
| VSM | Vector Space Model |
| WSNs | Wireless Sensor Networks |

## References

1. Martins, A.B.; Farinha, J.T.; Cardoso, A.M. Calibration and Certification of Industrial Sensors—A Global Review. *WSEAS Trans. Syst. Control.* **2020**, *15*, 394–416. [CrossRef]
2. Chaudhary, K.P.; Dubey, P.K.; Gahlot, A.; Dahiya, A. Advances in Sensors and Measurements for Metrological Applications. *MAPAN J. Metrol. Soc. India* **2021**, *36*, 211–213. [CrossRef]
3. Mateus, B.; Mendes, M.; Farinha, J.; Martins, A.; Cardoso, A. Data Analysis for Predictive Maintenance Using Time Series and Deep Learning Models—A Case Study in a Pulp Paper Industry. In *Proceedings of IncoME-VI and TEPEN 2021*; Springer: Cham, Switzerland, 2023; pp. 11–25. [CrossRef]
4. Li, Y.; Wei, X.; Li, Y.; Dong, Z.; Shahidehpour, M. Detection of False Data Injection Attacks in Smart Grid: A Secure Federated Deep Learning Approach. *IEEE Trans. Smart Grid* **2022**, *13*, 4862–4872. [CrossRef]
5. Antunes, J.; Torres, J.; Mendes, M.; Mateus, R.; Cardoso, A. Short and long forecast to implement predictive maintenance in a pulp industry. *Eksploat. Niezawodn. Maint. Reliab.* **2022**, *24*, 33–41. [CrossRef]
6. Mateus, B.C.; Mendes, M.; Farinha, J.T.; Assis, R.; Cardoso, A.M. Comparing LSTM and GRU Models to Predict the Condition of a Pulp Paper Press. *Energies* **2021**, *14*, 6958. [CrossRef]
7. Simões, A.; Viegas, J.M.; Farinha, J.T.; Fonseca, I. The State of the Art of Hidden Markov Models for Predictive Maintenance of Diesel Engines. *Qual. Reliab. Eng. Int.* **2017**, *33*, 2765–2779. [CrossRef]
8. Kou, L.; Li, Y.; Zhang, F.; Gong, X.; Hu, Y.; Yuan, Q.; Ke, W. Review on Monitoring, Operation and Maintenance of Smart Offshore Wind Farms. *Sensors* **2022**, *22*, 2822. [CrossRef]
9. Huhne, M.; Krystek, M.; Odin, A. Traceability of Measurement Results in Industrial Metrology. In *Simposio de Metrología*; Physikalisch-Technische Bundesanstalt: Braunschweig, Germany, 2004; pp. 1–6.
10. Aswal, D.K. *Metrology for Inclusive Growth of India*, 1st ed.; Springer: Singapore, 2020; p. 1084. [CrossRef]

11. Pais, J.; Raposo, H.D.; Farinha, J.; Cardoso, A.J.M.; Marques, P.A. Optimizing the life cycle of physical assets through an integrated life cycle assessment method. *Energies* **2021**, *14*, 6128. [CrossRef]
12. Sousa, J.; Pedro, J.; Machado, J. A generic interface and a framework designed for industrial metrology integration for the Internet of Things. *Comput. Ind.* **2022**, *138*, 103632. [CrossRef]
13. Kaya, M.C.; Nikoo, M.S.; Schwartz, M.L.; Oguztuzun, H. Internet of Measurement Things Architecture: Proof of Concept with Scope of Accreditation. *Sensors* **2020**, *20*, 503. [CrossRef]
14. Kunzmann, H.; Pfeifer, T.; Schmitt, R.; Schwenke, H.; Weckenmann, A. Productive Metrology—Adding Value to Manufacture. *CIRP Ann.* **2005**, *54*, 155–168. [CrossRef]
15. Gadelrab, M.S.; Abouhogail, R.A. Towards a new generation of digital calibration certificate: Analysis and survey. *Measurement* **2021**, *181*, 109611. [CrossRef]
16. Mester, C. The role of national metrology institutes, the International System of Units and the concept of traceability. In Proceedings of the First International Colloquium on Smart Grid Metrology, Split, Croatia, 24–27 April 2018; pp. 1–5. [CrossRef]
17. Hall, B.D. An opportunity to enhance the value of metrological traceability in digital systems. In Proceedings of the 2019 II Workshop on Metrology for Industry 4.0 and IoT (MetroInd4.0 IoT), Naples, Italy, 4–6 June 2019; pp. 16–21. [CrossRef]
18. Jones, B. Sensors in industrial metrology. *J. Phys. Sci. Instrum.* **1987**, *20*, 1113. [CrossRef]
19. Kim, M.; Kang, S.; Lee, J.; Cho, H.; Cho, S.; Su, J. Computers & Industrial Engineering Virtual metrology for copper-clad laminate manufacturing. *Comput. Ind. Eng.* **2017**, *109*, 280–287. [CrossRef]
20. Levin, S.F. General problems of metrology and measurement technique calibration and verification of measuring instruments: Conceptual transformation. *Meas. Tech.* **2022**, *64*, 8–18. [CrossRef]
21. Lai, Y.; Tong, L.; Liu, J.; Wang, Y.; Tang, T.; Zhao, Z.; Qin, H. Identifying malicious nodes in wireless sensor networks based on correlation detection. *Comput. Secur.* **2022**, *113*, 102540. [CrossRef]
22. Tipireddy, R.; Lerchen, M.; Ramuhalli, P.; Northwest, P. Virtual sensors for robust on-line monitoring (OLM) and diagnostics. In Proceedings of the NPIC & HMIT 2017, San Francisco, CA, USA, 11–15 June 2017.
23. Hines, J.W.; Rasmussen, B. Online Sensor Calibration Monitoring Uncertainty Estimation. *Nucl. Technol.* **2017**, *151*, 281–288. [CrossRef]
24. Rao, M.S.; Rao, D.N.; Reddy, P.C.; Shree, V.U. Fault prediction based on spatial correlation analysis using VSM in distributed sensor network. *Mater. Today Proc.* **2021**, 2214–7853. [CrossRef]
25. Berjab, N.; Le, H.H.; Yokota, H. A Spatiotemporal and Multivariate Attribute Correlation Extraction Scheme for Detecting Abnormal Nodes in WSN. *IEEE Access* **2021**, *9*, 135266–135284. [CrossRef]
26. Lee, S.; Chai, J. An enhanced prediction model for the on-line monitoring of the sensors using the Gaussian process regression. *J. Mech. Sci. Technol.* **2019**, *33*, 2249–2257. [CrossRef]
27. Fu, X.; Wang, Y.; Li, W.; Yang, Y.; Postolache, O. Lightweight Fault Detection Strategy for Wireless Sensor Networks Based on Trend Correlation. *IEEE Access* **2021**, *9*, 9073–9083. [CrossRef]
28. Li, L.; Liu, G.; Zhang, L.; Li, Q. Sensor fault detection with generalized likelihood ratio and correlation coef fi cient for bridge SHM. *J. Sound Vib.* **2019**, *442*, 445–458. [CrossRef]
29. Rajesh, G.; Chaturvedi, A. Correlation analysis and statistical characterization of heterogeneous sensor data in environmental sensor networks. *Comput. Netw.* **2019**, *164*, 106902. [CrossRef]
30. Karmakar, G.C.; Das, R.; Kamruzzaman, J. IoT Sensor Numerical Data Trust Model Using Temporal Correlation. *IEEE Internet Things J.* **2020**, *7*, 2573–2581. [CrossRef]
31. Biswas, P.; Samanta, T. A Method for Fault Detection in Wireless Sensor Network Based on Pearson ' s Correlation Coefficient and Support. *Wirel. Pers. Commun.* **2022**, *123*, 2649–2664. [CrossRef]
32. Hoffmann, M. *PEANO—A Tool for On-Line Calibration Monitoring PEANO—A Tool for On-Line Calibration Monitoring*; OECD Halden Reactor Project—Institute for Energy Technology: Halden, Norway, 2014.
33. Coble, J.; Ramuhalli, P.; Meyer, R.; Hashemian, H.; Shumaker, B.; Cummins, D. Calibration Monitoring for Sensor Calibration Interval Extension: Identifying technical gaps. In Proceedings of the Future of Instrumentation International Workshop (FIIW) Proceedings, Gatlinburg, TN, USA, 8–9 October 2012; pp. 1–4. [CrossRef]
34. Kleber, N.; Dietlein, C.R.; Chisum, J.D. Cooperative Cross-Correlation Algorithm to Optimize Linearity of Fused RF Sensors. *IEEE Sens. J.* **2020**, *20*, 3766–3776. [CrossRef]
35. Chen, H.; Hsu, J.y.; Hsieh, J.Y.; Hsu, H.y.; Chang, C.h.; Lin, Y.J. Predictive maintenance of abnormal wind turbine events by using machine learning based on condition monitoring for anomaly detection. *J. Mech. Sci. Technol.* **2021**, *35*, 5323–5333. [CrossRef]
36. Alqahtani, A.; Ali, M.; Xie, X. Deep Time-Series Clustering: A Review. *Electronics* **2021**, *10*, 3001. [CrossRef]
37. Qiu, K.; Qin, T.; Pan, J.; Liu, S.; Shen, S. Real-Time Temporal and Rotational Calibration of Heterogeneous Sensors Using Motion Correlation Analysis. *IEEE Trans. Robot.* **2021**, *37*, 587–602. [CrossRef]
38. Lu, W.; Wen, R.; Teng, J.; Li, X.; Li, C. Data correlation analysis for optimal sensor placement using a bond energy algorithm. *Measurement* **2016**, *91*, 509–518. [CrossRef]
39. Dash, L.; Pattanayak, B.K.; Mishra, S.K.; Sahoo, K.S.; Jhanjhi, N.Z.; Baz, M.; Masud, M. A Data Aggregation Approach Exploiting Spatial and Temporal Correlation among Sensor Data in Wireless Sensor Networks. *Electronics* **2022**, *11*, 989. [CrossRef]
40. Zhang, X.; Wang, S.; Li, W.; Lu, X. Heterogeneous sensors-based feature optimisation and deep learning for tool wear prediction. *Int. J. Adv. Manuf. Technol.* **2021**, *114*, 2651–2675. [CrossRef]

41. Kendall, B. A New Measure of Rank Correlation. *Biometrika* **1938**, *30*, 81–93. [CrossRef]
42. Awasthi, S.; Travieso-González, C.; Sanyal, G.; Singh, D. *Artificial Intelligence for a Sustainable Industry 4.0*; Springer: Cham, Switzerland, 2021; pp. 1–311. [CrossRef]
43. Azar, K.; Hajiakhondi-meybodi, Z.; Naderkhani, F. Semi-supervised clustering-based method for fault diagnosis and prognosis: A case study. *Reliab. Eng. Syst. Saf.* **2022**, *222*, 108405. [CrossRef]
44. Martins, A.; Fonseca, I.; Farinha, J.; Reis, J.; Cardoso, A. Maintenance prediction through sensing using hidden markov models—A case study. *Appl. Sci.* **2021**, *11*, 7685. [CrossRef]
45. Schiff, A.J. Power Plant Surveillance and Diagnostics; In *Power Systems*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 1–384. [CrossRef]
46. Rafiee, H.; Aminizadeh, M.; Hosseini, E.M.; Aghasafari, H.; Mohammadi, A. A Cluster Analysis on the Energy Use Indicators and Carbon Footprint of Irrigated Wheat Cropping Systems. *Sustainability* **2022**, *14*, 4014. [CrossRef]
47. Shen, X.; Lin, X.; Zhang, K. *Encyclopedia of Wireless Networks*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 1–1566. [CrossRef]
48. Peng, K.; Tan, J.; Zhang, G. A Method of Curve Reconstruction Based on Point Cloud Clustering and PCA. *Symmetry* **2022**, *14*, 726. [CrossRef]
49. Stapathy, S.; Bhateja, V.; Favorskaya, M.; Adilakshmi, T. Smart Innovation, Systems and Technologies. In *Proceedings of the Fourth International Conference on Smart Computing and Informatics*; Satapathy, S.C., Bhateja, V., Favorskaya, M.N., Adilakshm, T., Eds.; Springer: Singapore, 2021; Voume 1. [CrossRef]
50. Yousefi-Darani, A.; Babor, M.; Paquet-Durand, O.; Hitzmann, B. Model-based calibration of a gas sensor array for on-line monitoring of ethanol concentration in Saccharomyces cerevisiae batch cultivation. *Biosyst. Eng.* **2020**, *198*, 198–209. [CrossRef]
51. Zhang, Y.; Wijeratne, L.O.H.; Talebi, S.; Lary, D. Machine Learning for Light Sensor Calibration. *Sensors* **2021**, *21*, 6259. [CrossRef]
52. Thrun, M.C.; Ultsch, A. Using Projection-Based Clustering to Find Distance- and Density-Based Clusters in High-Dimensional Data. *J. Classif.* **2021**, *38*, 280–312. [CrossRef]
53. Abdullah, D.; Susilo, S.; Ahmar, A.; Rusli, R.; Hidayat, R. The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data. *Qual. Quant.* **2021**, *56*, 1283–1291. [CrossRef] [PubMed]
54. Ali, M.; Fasaee, K.; Monghasemi, S.; Reza, M.; Shafiee, M.E.; Zechman, E.; Hashempour, P. A K-Sensor correlation-based evolutionary optimization algorithm to cluster contamination events and place sensors in water distribution systems. *J. Clean. Prod.* **2021**, *319*, 128763. [CrossRef]
55. Zhao, X.; Nie, F.; Wang, R.; Li, X. Neurocomputing Improving projected fuzzy K-means clustering via robust learning. *Neurocomputing* **2022**, *491*, 34–43. [CrossRef]
56. Wang, J.; Wang, K.; Jia, R.; Chen, X. Research on Load Clustering Based on Singular Value Decomposition and K-means Clustering Algorithm. In Proceedings of the 2020 Asia Energy and Electrical Engineering Symposium (AEEES), Chengdu, China, 28–31 May 2020; pp. 831–835. [CrossRef]
57. Fritz, M.; Behringer, M.; Tschechlov, D.; Schwarz, H. Efficient exploratory clustering analyses in large-scale exploration processes. *VLDB J.* **2021**, *31*, 711–732. [CrossRef]
58. Ezugwu, A.E.; Ikotun, A.M.; Oyelade, O.O.; Abualigah, L. A comprehensive survey ofclustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Eng. Appl. Artif. Intell.* **2022**, *110*, 104743. [CrossRef]
59. Wickramasinghe, A.; Muthukumarana, S.; Loewen, D.; Schaubroeck, M. Temperature clusters in commercial buildings using k -means and time series clustering. *Energy Inform.* **2022**, *5*, 1. [CrossRef] [PubMed]
60. Borlea, I.d.; Precup, R.e.; Borlea, A.b. Improvement of K-means Cluster Quality by Post Processing Resulted Clusters. *Int. Conf. Inf. Technol. Quant. Manag.* **2022**, *199*, 63–70. [CrossRef]
61. Separation, I.C.; Li, K.; Member, S.; Cao, X.; Ge, X.; Member, S.; Wang, F.; Member, S.; Lu, X.; Member, S.; et al. Meta-Heuristic Optimization-Based Two-Stage Considering Intra-Cluster Compactness and Inter-Cluster Separation. *IEEE Trans. Ind. Appl.* **2020**, *56*, 3375–3384. [CrossRef]
62. Zong, P.; Jiang, J.; Qin, J. Study of High-Dimensional Data Analysis based on Clustering Algorithm. In Proceedings of the 2020 15th International Conference on Computer Science Education (ICCSE), Delft, The Netherlands, 18–22 August 2020; pp. 638–641. [CrossRef]
63. Rodrigues, J.; Martins, A.; Mendes, M.; Farinha, T.; Mateus, R.; Cardoso, A.J. Automatic Risk Assessment for an Industrial Asset Using. *Energies* **2022**, *15*, 9387. [CrossRef]
64. Feng, Y.; Xu, W.; Zhang, Z.; Wang, F. Continuous Hidden Markov Model Based Spectrum Sensing with Estimated SNR for Cognitive UAV Networks. *Sensors* **2022**, *22*, 2620. [CrossRef]
65. Rabiner, L.R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE* **1989**, *77*, 257–286. [CrossRef]
66. Rabiner, L.R.; Juang, B.H. An Introduction to Hidden Markov Models. *IEEE ASSP Mag.* **1986**, *3*, 4–16. [CrossRef]
67. Martins, A.; Fonseca, I.; Torres, F.J.; Reis, J.; Cardoso, A.J.M. Prediction Maintenance based on Vibration Analysis and Deep Learning—A case study of a drying press supported on Hidden Markov Model. *SSRN* **2022**. [CrossRef]
68. Boechat, A.A.; Moreno, U.F.; Haramura, D. On-line calibration monitoring system based on data-driven model for oil well sensors. *IFAC Proc. Vol.* **2012**, *45*, 269–274. [CrossRef]

69. Saucedo-Dorantes, J.J.; Jaen-cuellar, A.Y.; Delgado-prieto, M.; Romero-troncoso, R.D.J.; Osornio-rios, R.A. Condition monitoring strategy based on an optimized selection of high-dimensional set of hybrid features to diagnose and detect multiple and combined faults in an induction motor. *Measurement* **2021**, *178*, 109404. [CrossRef]
70. Dhamande, L.S.; Chaudhari, M.B. Compound gear-bearing fault feature extraction using statistical features based on time-frequency method. *Measurement* **2018**, *125*, 63–77. [CrossRef]