FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Optimizing O&M plans for flexible hydropower systems

Xavier Tarrio Fernandes

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Prof. Luís Guímarães

Second Supervisor: Prof. Armando Leitão

July 06, 2020

Resumo

Nos dias em que vivemos, consideramos a eletricidade como um bem adquirido, e é inconcebível imaginar um mundo sem ela. Eletricidade esta que é proveniente maioritariamente de centrais de produção não renováveis. Tendo em conta o estado de mudança do nosso planeta, existe uma pressão política e social extra para a redução da utilização de energia proveniente de fontes não-renováveis, colocando um peso acrescido às centrais de energias renováveis, pois estas terão de ser competitivas o suficiente para responder às necessidades da rede elétrica. E é aqui onde se encontram as maiores dificuldades das energias limpas. Algumas delas, dependem diretamente do estado climatérico, como é o caso das energias eólicas (vento), solares (luz solar) e até mesmo ondomotriz (ondas e marés), não conseguindo responder durante todo o ano às condições impostas pela rede elétrica. Assim começaram a se estudadas energias cuja produção e estados fossem mais facilmente controladas, como é o caso das energias hídricas e de biomassa. Estas energias conseguem compensar momentos em que a procura energética é mais elevada e combater as exigências, no entanto, um problema comum a todas as energias renováveis, é a falta rapidez com que respondem a picos de procura energética que não conseguem ser previstos. E é aqui que entram projetos como o XFLEX.

O XFLEX é um projeto europeu que é constituído por várias instituições desde universidades, centros de investigação a empresas produtoras de turbinas, em que um dos principais objetivos, é o combate ao atraso na reação de resposta das energias renováveis, mais especificamente nas centrais hidroelétricas. Posto isto, o projeto foi dividido em 13 *workpackages* tendo cada um deles tarefas especificas que levarão à concretização do objetivo principal.

Esta dissertação enquadra-se no *workpackage* 3, em que o objetivo é a criação dum Supervisor Inteligente da central, na qual integra uma basta monitorização, modelação e controlo de todos os subsistemas envolventes de forma a conseguir avaliar a melhor decisão a ser tomada em cada situação para aumentar a produtividade, fiabilidade e flexibilidade da central. O objetivo que foi proposto nesta dissertação foi o desenvolvimento de uma metodologia para o cálculo de um índice de saúde. Índice este que representa o estado de saúde de todo o sistema de forma a auxiliar na tomada de decisão das estratégias de manutenção.

Assim sendo, numa fase inicial um estudo de fiabilidade é apresentado, motivado pelo facto de que os ativos duma central são sujeitos a inúmeros fatores de degradação. Consequentemente o trabalho desenvolvido foi um modelo de deteção de falhas e de avaliação do estado de saúde do mesmo. Toda esta análise foi desenvolvida apenas com acesso a dados sensoriais.

Uma vez que era pretendido o desenvolvimento de um único numero que representasse o estado de saúde do sistema, algoritmos de aglomeração e fusão foram desenvolvidos, permitindo aglomerar a informação dos indices de saúde dos diversos sub-sistemas. Algoritmos de otimização também foram utilizados para melhorar os resultados obtidos, para que um índice de saúde represente o estado de saúde do sistema o mais preciso possível para auxiliar nas tomadas de decisão das políticas de manutenção.

Por fim, os resultados realizados demonstram a performance do modelo desenvolvido, estando

acompanhados duma discussão e de possíveis melhorias a fazer no futuro, de forma a melhorar o desempenho do mesmo.

Abstract

Nowadays we consider electricity as guaranteed, and it is inconceivable to imagine a world without it. Electricity can be provided or produced by different energy source types, for instance, renewable and non-renewable energy production plants. Taking into account the unpredictable behavior of our planet, there is an extra political and social pressure to reduce the use of energy that comes from non-renewable sources, which place a burden on renewable energy plants as they must be able to respond to all needs of the power grid. And this is where the greatest difficulties of clean energies are found. Some of them depend directly on weather such as wind (wind), solar (sunlight), and even ondomotriz (waves and tides), failing to respond throughout the year to the demand defined by the power grid. Therefore, energies that are easier to control its production and predict its states have started to be studied with the hope to solve this seasonal problem. Hydropower and biomass are the most common energies that fit these requirements being able to compensate moments that energy demand is higher. However there is a problem common to all renewable energies that is the speedless that they respond to peaks in energy demand that can not be predicted. And this is where projects like XFLEX come in.

XFLEX is an European project that institutions like universities, research centers and turbine producing companies are part of. The main objectives are to combat the delay in the response of renewable energy, more specifically in hydroelectric power plants. The project is divided into thirteen (13) workpackages with different targets but with a common goal.

This dissertation is part of workpackage three (3), which aims to create an Intelligent Supervisor of the hydropower plant, where upon it integrates monitoring, modeling and controlling all the surrounding subsystems to be able to evaluate the best decision to be taken in each situation to increase the plant's productivity, reliability and flexibility. The objective proposed for this dissertation was the development of a methodology to build a composite health index. This index represents the health status of the entire system to better manage the maintenance policies while having the maximum performance, safety with the minimum possible costs.

Therefore, in an initial phase of the dissertation, a reliability study is presented, motivated by the fact that the assets of a plant are subject to numerous factors of degradation. Consequently the work developed is a model of fault detection and health status assessment. This entire analysis was developed only with access to sensory data.

Once the objective was to develop a single health representative number, agglomeration and combination algorithms were developed, allowing to fuse different sub-system information. Optimization algorithms were also used to improve obtained results. So health index represents the health status of the system as accurately as possible to assist in maintenance policy decision making.

Finally, achieved results demonstrate the outputs of the developed model followed by a discussion and possible improvements to be made in the future in order to improve performance.

Acknowledgements

First of all, I would like to thank both of my supervisors, Professor Luís Guimarães e Professor Armando Leitão, for the support and advice during these last months. Their professional, patience and positive attitude, along side their knowledge and tips strongly motivated me to execute the present work.

Next, my deepest gratitude goes to my family, my mom, dad, brother and borrowed sister, for being my unconditional support.

Furthermore, I would like to thank to all my friends, especially the ones that I made in this five year journey. YWNWA!

Lastly, despite all the headaches, the heated arguments and the zero sleep nights, I am in debt with FEUP Students' Union. Among other things, AEFEUP brought to my life amazing people and experiences. I am really grateful for that.

My sincere gratitude to all,

Xavier Tarrio Fernandes

“A pessimist sees the difficulty in every opportunity;
an optimist sees the opportunity in every difficulty.”

Winston Churchill

Contents

List of Figures

List of Tables

Acronyms and Symbols

ABM	Age Based Maintenance
ANN	Artificial Neural Networks
ARMA	Auto-regressive Moving Average
CBM	Condition Based Maintenance
CFD	Computational Fluid dynamics
CHI	Composite Health Index
CT	Critical Threshold
EDF	Eletricticite de France
EDP	Eletricidade de Portugal
EPFL	Ecole Polytechnique Federale de Lausanne
EPS	Eletric Power Systems
FD	Factor Divider
FEA	Finite Element Analysis
FMEA	Failure Modes and Effects Analysis
FPR	False Positive Rate
FT	Failure Threshold
HMM	Hidden Markov model
HPP	Hydropower Plants
HSMM	Hidden semi-Markov model
INESC TEC	Instituto de Engenharia de Sistemas e computadores, Tecnologias e Ciencia
KPI	Key Performance Indicatores
MC	Markov Chains
MM	Match Matrix
NLM	Non-Linear Models
PM	Predictive Maintenance
R&I	Research & Innovation
RCM	Reliability Centered Maintenance
ROC	Receiver Operating Characteristic
RUL	Remaining Useful Life
SMC	Sequential Monte Carlo
SPPS	Smart power plant supervisor
ST	Supervised Training
TBM	Time Based Maintenance
TPR	True Positive Rate
UM	Unsupervised Model
UT	Unsupervised Training
WP	Work packages
WT	Warning Threshold
XFLEX	Hydropower extending power system flexibility

Chapter 1

Introduction

The word reliability is at the tip of the tong of every engineer nowadays. This phenomenon start right after World War II while a group of engineers was developing a missile,and despite they had all the resources needed and high quality components, they could not make it right on the first attempts. A mathematician was called as a consultant, with the mission of analyzing the missile system. He quickly made a theorem called "product probability law of series components". It says that the reliability of a system is equal to the product of the reliabilities of the individual components which make up the system [56].

In this chapter it will be discuss the importance and the impact that good maintenance strategy has in a business. The main motivation that that led to the study of this subject, as well as the main objectives of this dissertation. A brief introduction of the integral project will also be made.

1.1 Context & Motivation

Asset reliability has an important role in every industry, this happens because we reach a moment where maintenance costs can represent, depending on the industry, between 15-70% of the production costs. As a result, this subject has had a tremendous growth in every type of industry decision making, according to [64].

These maintenance policies need to be well planned and followed strictly, and the reason is that capital intensive businesses require a large amount of investment in fixed assets, such as property, plant and equipment. Since they are so dependable on their equipment, their main focus is on increasing their availability and maintainability. These assets must work at/or below a maximum acceptable failure rate, as the system ages, this failure rate tent to increase [32], not only because of the use, but also because of the environmental condition, and sometimes because of the bad maintenance policies or the lack of it, which make the failure rate increase even faster.

Having into consideration everything that has been said before, it is imperative to have a well-thought maintenance strategy as well as an operation program, to increase as long as possible the asset's remaining useful life (RUL) as well as to decrease the costs and time of maintenance. Therefore, constant monitoring of the assets' degradation behavior and the understanding of the factor that is related to it is needed. These types of analysis are made to prevent potential faults, by predicting them or just reacting to them.

According to [18], the most important renewable source of electrical energy is the Hydropower, but it is not optimized to the fullest potential. The pressure to generate clean energy is increasing at a fast rate. This pressure is created by the society in general, but also, and largely, by the Environmental legislation such as the Kyoto Protocol. Hydro-power importance can be noticed in the results published by [38], which indicates that 86.31% of renewable energy is generated by a hydropower plant. However, hydropower plants are subject to environmental conditions, which can be very rigid, promoting the aging of components. This can lead to constructional, operational, and maintenance problems that could reduce significantly the plant electricity capacity. With a bad maintenance strategy, these conditions can lead to two scenarios but with the same output. The first one is the complete shut down of the infrastructure and the other one is reactive maintenance. Which both result in huge income losses and material costs.

Tacking into account the aforementioned context, in order to ensure availability and reliability of the equipment's, an effective maintenance plan is mandatory. The maintenance paradigm is changing, nowadays maintenance policies are described on the companies' budget with a plan to follow. This new approach has been replacing a simple policy named reactive maintenance, which is performed upon unexpected failures.

1.1.1 XFLEX Project

Hydropower extending power system flexibility, also known as XFLEX, is a European financed project, coordinated by *Ecole Polytechnique Federale de Lausanne* (EPFL) with nineteen participants, including INESC TEC, which the main objectives of it are:

- increase the potential of hydroelectric technologies in providing flexibility to electric power system, while achieving an improved average annual efficiency of the machinery, providing high availability of hydroeletric power and maximizing their performance;

- demonstrate the system integration methodology of hydroelectric technology solutions such as fixed and variable speed, pump power regulation, battery hybridization advanced monitoring and digitalization, and to draw the road-map for the deployment of the system integration to all kinds of European hydroelectric power plants, run of river, storage and pumped storage of all sizes; being existing, up-rated or new.

In order to achieve these goals, the project is organized into twelve work packages(WP). Each one has a clear focus on the tasks and contributions to tackling the overall project objectives. Following this organization, the methodology to be followed in the project consists of three different groups of activities.

- **Innovation Activities** - This activity is composed by the WP2 and WP3, which the R&I actions are developed focusing on the identification and selection of the advanced hydroelectric machinery technology portfolio required by the energy system (WP2), and on the enhancement of hydroelectric technological solutions consisting in a general methodology to fully exploit the hydropower potential, depending on the power plant conditions and characteristics (WP3).

- **Demonstration Activities** - The technological solutions will be implemented into demonstrators through these activities. They will be carried out in an iterative way, feedback about performance will be provided by utilities and OEMs to RTDs and technology providers to improve and adjust certain technological features. The developed solutions will be demonstrated in 7 plants to representatively cover the whole EU hydro fleet.

- **Deployment Activities** - In parallel, taking advantage of the hydroelectric flexible technology portfolio and of successfully implemented demonstrators, several activities will be performed with the aim of paving the way to XFLEX replication. WP11 build on top of previous WPs and analyses the benefits and impacts (social, environmental) of the flexible technologies and provide guidelines for an optimal industrial deployment.

This dissertation fits in the WP3 - Smart Power Plant Supervision: Integration Methodology of hydroelectric technology solutions, which the objective is to develop a methodology, called smart power plant supervisor (SPPS) that integrates advanced monitoring, modelling, control and communication with all HPP functional levels to increase the availability, flexibility and lifespan of the unit. However the whole completion of this project demands four years, which is more time than I had to conclude this dissertation, therefore this work is dedicated to develop methodology of a model capable of creating a composite health index that represent the overall system's health.

1.1.2 Approach and Goals

As was mentioned in the Section 1.1.1, this dissertation was comprised in work package 3, which aims to increase the availability and lifespan of the hydro units by continuous monitoring, however, continuous monitoring without any treatment can be unperceptive. Therefore, a methodology to transform the raw data from the sensors reading into a single value that can represent the health state of a complex system was suggested, which end up being the aim of this dissertation. In other words, the purpose of this work was to build a methodology capable of agglomerate different degradation data, of different components into one single value, which was called health index (HI), capable of predicting failures and describing the health condition of the system.

Due to the fact that the HI need to represent different components, from different parts of the system, it was necessary to deconstruct the HI system in four phases. Firstly, a brief data processing was used to identify which variables/components were useful or not for the representations of the different sub-systems, followed by the second phase, which consisted in the creation of a degradation curve for each measurement, that allowed to obtain an individual HI. After that, owing to the fact that a power station is a complex system, it was needed to agglomerate the diferents HI's of the sub-systems into one, making what was called a composite HI. A composite HI, according to [28], is a number that provides a basis for assessing the overall health situation of an asset. They are constructed to identify and predict failure modes of the overall system and its subsystems, and consequently, it is used to develop a representation of the asset degradation. And finally, the fourth phase was an evaluation of the model. Firstly, the model was trained to find which parameters were the best. Following by the test into data that the model never had contact. Finally, in order to analyze the accuracy and precision of the developed model, the results were organized into confusion matrices.

1.1.3 Dissertation outline

The following document will be structured as follows. In Chapter 2, a literature review approaching the main concepts that will be used in the following dissertation. Chapter 3 describes the problem in which this dissertation will be worked on, as well as the motivation that led to the creation of this project. The methodologies used to create the Health Index are described with high detail in chapter 4, which is followed by Chapter 5 where all the computational test and results are described. Lastly, a reflection about the results and the work that could be developed in the future are presented in Chapter 6.

Chapter 2

Literature Review

Nowadays, the permanent follow up of the health systems and respective subsystems is a major aspect. Therefore, monitoring data should be the most accurate and simple as possible to be able to be applied in multiple situations, such as maintenance plans and failure predictions to minimize short- and long-term costs [37]. This type of system's approach is becoming more frequent and is known as Condition-based Maintenance (CBM).

In this chapter will be reviewed some basic and optimized maintenance models. To better understanding we should acknowledge some concepts and techniques, represented further ahead in examples of maintenance policy and it proposes.

As the actual project is based on a CBM strategy, it is necessary to identify what kind of data will be collected and how it will be converted into reliable data. Here, some reliability models were tested.

To summarise, it will also be reviewed various health index and techniques capable of building one.

2.1 Reliability Concepts

When the word "reliability" appears in a conversation, it is automatically associated to the function of an equipment or system, in other words, if the equipment or system performs the task correctly for which they were made for. If that happens, we consider it reliable, if not it is considered as an unreliable equipment, based on [43].

However, the term "reliability" encompasses other terms as we will see in the following section:

Reliability can be define, according to [44] and [36], as:

" the probability of a device or system performing its function adequately for the period of time intended under the operating conditions intended."

Which means that it expresses the number of failures that happen on the equipment. This probability depends on the equipment's operation time. It is known as Reliability Curve or Survival function. It is expressed according to the following equation, as stated in [43]:

$$S(t) = P(T > t) \tag{2.1}$$

This equation tends for 0. Which means, theoretically, at the end, the equipment will fail, as we can see in Figure 2.1.

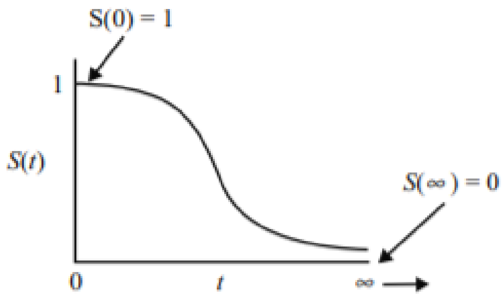


Figure 2.1: Theoretical survival curve, source [35]

S(t) is the probability of an equipment or system having a failure. *T* is a non-negative continuous random variable that represents the amount of time for a random equipment breakdown. And *t* is the actual time.

This curve represents the standard operational lifetime of equipment, however, it is possible to along the equipment's operation time. And that is possible through maintenance policies, which are going to be presented in the following section. In order to have a better understanding of the discussed matter, it is important to be aware of some concepts, as:

Failure - According to [4], a failure is the termination of the ability of an asset to perform its required function;

Fault - A fault is a state which the asset can perform his required function. The main difference between fault and failure is that, a fault is normally a failure consequence. A failure is an event while the fault is a state, according to [5].

Remaining Useful Life (RUL) - As the name indicates, Remaining Useful Life is the time left that the asset, according to condition analysis, have before having any type of failure. It is an unknown value, and normally is calculated through information that can be obtained from condition and monitoring [58].

Asset Health Index - it is a score which represents and reflects the asset condition and asset performance in terms of the asset's role [27].

2.2 Maintenance Policies

Over time, maintenance was forced to change in multiple ways. Its own concept, the variety of techniques and approaches have been evolving into more complex and competitive policies. Maintenance's exponential evolution and adaptation are a necessity created by the demanding and competitive business that we live today.[48]. These maintenance conceps are chronologically analyzed in Figure 2.2.

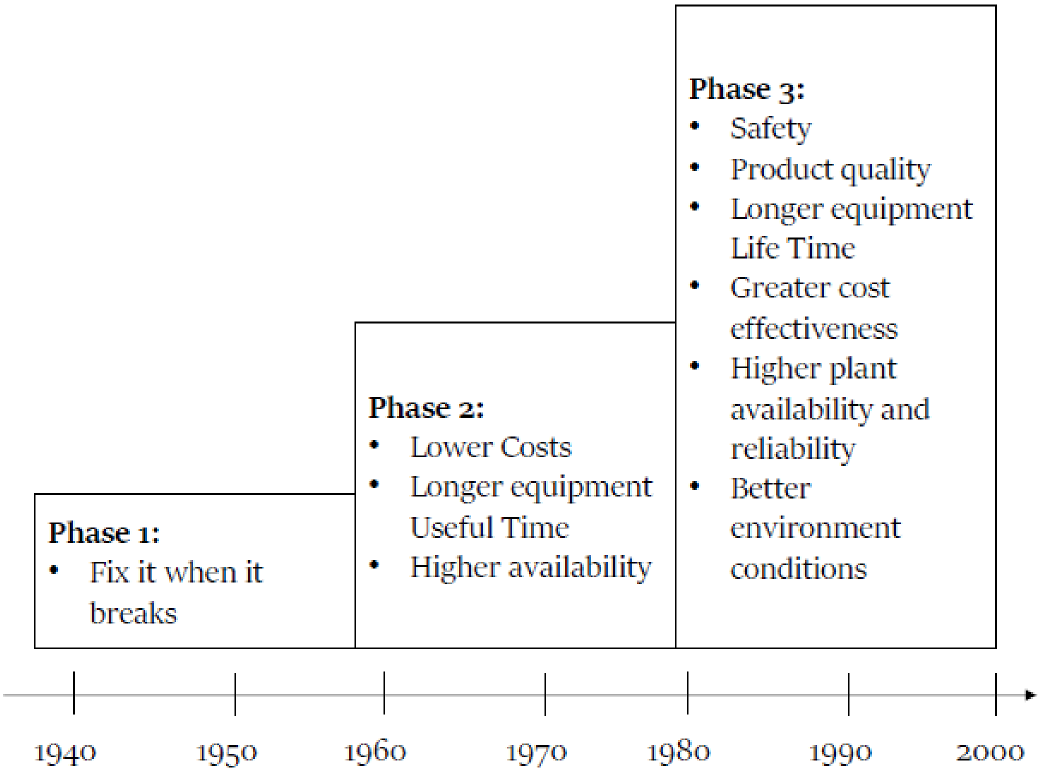


Figure 2.2: Evolution of Maintenance Expectations [48]

Historical breakthroughs were the main fuel to allow mind-changing toward maintenance. At the beginning, the amount of time wasted with equipment under maintenance was not valued, so failure prevention was never a priority. Although, World War II began and brought an opening rethink maintenance. The lack of resources leads to an economic rupture that was translated into expensive pieces. The priority was now equipment durability. Today is possible to assume that maintenance is faster, more flexible, and precise at the same time motivated by the *just in time* business models. [48].

All this development created a huge range of maintenance policies specifically thought to each business. There is no better or worst policy but one can be more suitable than others when we look into our business [63].

The next section will focus on some major maintenance policies and where they can be better applied. However, in order to understand the concept of maintenance policies, it is necessary to be comfortable with the concept of maintenance. According to [5], maintenance is defined as:

"Combination of all technical, administrative, and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function."

As we can see through its definition, maintenance is not only performed when some equipment has a breakdown, it is also performed to prevent all the consequences that result from the failure of equipment.

According to [47] the main focus of maintenance is to:

- Achieve optimum availability in order to increase utilization capacity;
- Apply maintenance policies in order to extend the useful time and, consequently, reduce costs.

So, to achieve these objectives there are several maintenance philosophies. Those philosophies can be categorized into two groups, based on the times that the maintenance is done. In Figure 2.3 these categories and these subcategories are represented.

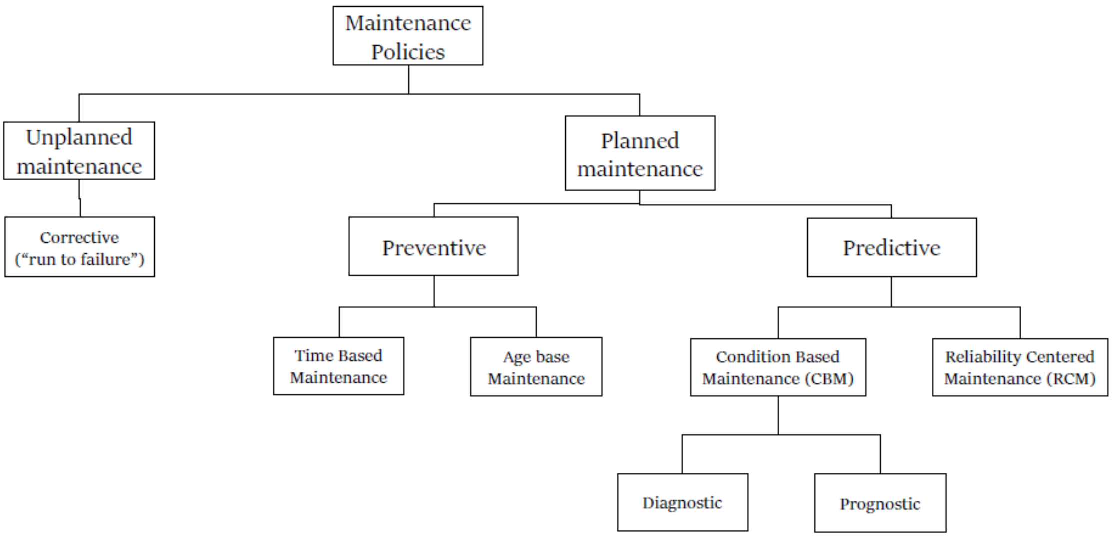


Figure 2.3: Taxonomy of maintenance philosophies

2.2.1 Unplanned Maintenance

Unplanned maintenance, also know as reactive maintenance, only happens when the equipment actually breakdown. There is not a specific maintenance strategy to solve the problem. Normally, when a failures occurs and this type of policy is implemented, it is associated with high incomes lost, related to the time that the asset is not producing plus the maintenance costs [63]. However, it is still used in specific business models. This approach is ideal for low-priority equipment, without which the companies production can continue running steadily.

Despite being the most primitive maintenance policy, it can bring some advantages, for instance, lower short-term costs, minimal planning required, and simpler process. Nevertheless like every maintenance policy has disadvantages which are, unpredictability, paused operation, higher long-term costs. This analysis between the bad and the goods has extreme importance in the decision of which maintenance policy should be implemented in the business [30] [9].

2.2.1.1 Corrective Maintenance

"Corrective maintenance is defined as the activity carried out after a failure has occurred and is intended to restore an item to a state in which it can perform its required function." , according to [37].

Basically corrective maintenance is known for their "fix it when it breaks" policy. Despite being the oldest type of maintenance and being still used in many specific businesses, this maintenance strategy lost some power. This started happening since the equipment complexity started to rise and the production deadlines started to be a lot more demanding [37].

2.2.2 Planed Maintenance

As the proper name indicates, planed maintenance is the one that is thought and planned usually before the failure. Although there are still different approaches withind this classification [37]. Scheduled maintenance can be divided into two categories as it is possible to see in Figure 2.3.

2.2.2.1 Preventive Maintenance

After proving, in some business, periodic maintenance improved system yield and income, preventive maintenance policy started to be adopted instead of waiting for failure. This policy usually uses a factor, such as time or age to do maintenance to prevent failure [42]. Since the maintenance intervals are scheduled, between this time, it could be needed to do corrective maintenance, which is non planned maintenance.

Time Based Maintenance

In this type of maintenance, the decisions are made based on failure time analyses. In other words, the expected lifetime, T, of equipment is estimated based on failure time data or used-based data. It is assumed that the failure behavior of equipment is predictable, based on failure rate trends, known as bathtub curves, as shown in Figure 2.4, according to [43].

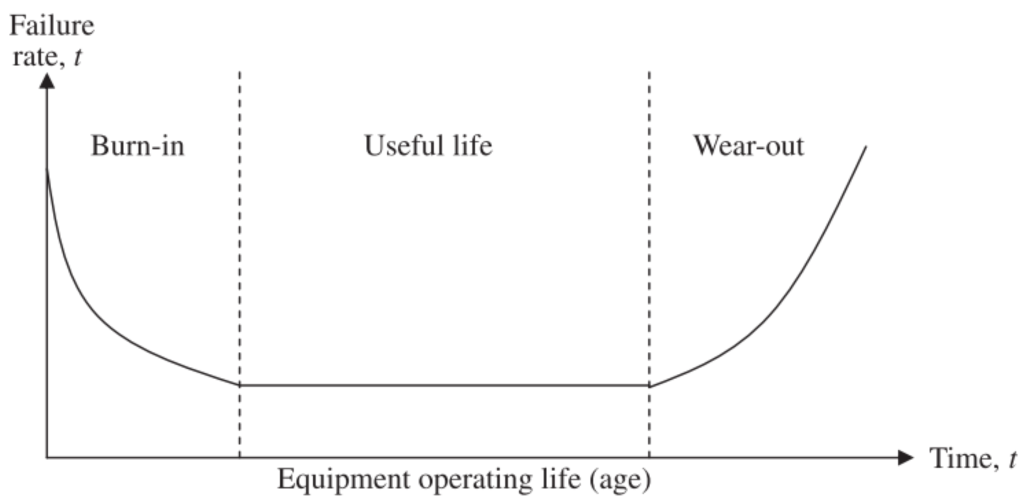


Figure 2.4: Bathtub curve [8]

Failure rate trends can be divided into three phases: burn-in, useful life, and wear-out. As we can see in Figure 2.4, at the early stages the failure rate is decreasing, this happens because it is when all the components are adapting to one another. Followed by a near-constant failure rate, also known as useful-life, and finally, at the end of their life cycles (wear-out), equipment experience increasing failure rates. [8]

The most challenging task is to determinant the intervals in which maintenance will be performed. Generally, this interval is defined according to the recommendations of the equipment manufacturer and the field experience of the technicians.

Age Based Maintenance

This strategy is an adaptation of the TBM. While TBM approaches starts to do periodic evaluations since the beginning of the equipment's life, age based maintenance only starts after the equipment reaches a certain age. Which is defined, usually, by the manufacturer. By deferring the early stage maintenance, it costs are reduced.

2.2.2.2 Predictive Maintenance (PM)

A predictive maintenance policy is made to be applied before failure but instead of doing it in periodic times, the system is studied in order to confirm if there is a necessity to do maintenance or not. Therefore, it can save money, time and human resources because the equipment is under maintenance only when is needed.

To conclude, schedule maintenance in predictive policy is always done after a health system study in order to reduce expenses and at the same time improve the equipment yield [42]. That being said, it is possible to classify two types of predictive maintenance, that are presented below.

Condition Based Maintenance

According to [31] condition based maintenance (CBM) can be define as:

"A maintenance program that recommends maintenance actions based on the information collected through condition monitoring."

The condition of a system is quantified by parameters that are continuously monitored and are system or application-specific. In CBM, we need to be aware of a concept called the actual state. The actual state of a system represents it actual condition, which is indicated by data collected from several sensors or other devices that serve to monitor [52] and gives us data that can be used to create a health index, which will be explained in a future section, of the operating conditions.

This health index is obtained through data analyzing and processing, and has great importance when the matter to be dealt with is maintenance decision. There are two main categories representing the techniques used to support maintenance decision-making: diagnosis and prognostics.

According to [60], diagnostic is the process of determining the health status and the equipment deterioration using data delivered by the condition-monitoring system. In other words, it is the data processing and analysis part. Based on [59], the diagnostic process can be divided into 3 phases.

- Fault detection - Which is responsible for detecting and reporting any abnormal operating condition;
- Fault isolation - Determining which component (equipment, system) is falling or has failed;
- Fault identification - Estimating the nature and extent of the fault.

Diagnostic has heavy importance when it comes to developing a successful prognostic because an adequate prognostic method begins with a robust diagnostics since the uncertainties of the estimated system conditions affect any future predictions. [26]

Several different definitions have been proposed to define "prognostics". Those are shown in Table A.1. Collectively, these definitions state that prognostic is the prediction of the health state of a system or sub-system, and besides that, it can have different formats. It can be represented by the time that the asset has until the next failure, by the probability of failing, the conditions of the system, or even by reliability percentage. Summing up, it is a representation of the system's state.

Similarly to diagnostic, prognostic can also be divided into phases, two to be precise. Which are:

- Confidence Interval Estimation - Estimating the confidence interval associated with the RUL prediction.

- Remaining Useful Life Prediction - It has to be able to identify the lead time to failure;

Founded on [59], the confidence value is necessary because of the intrinsic uncertainty associated with the degradation process, ambiguity in predicting the future conditions of the equipment and errors associated with both the diagnostic and prognostic models applied. Confidence intervals have greater importance in prognostic than in the diagnostic process because, the diagnostic can be more deterministic, in theory, than prognostic.

There are several techniques to make prognostics, and they will be presented in Section 2.4. However, there is one that outshines nowadays, which is the prediction of the Remaining Useful Life (RUL), but depending on the available data, not always it is possible to calculate RUL, so it is necessary to predict another value that represents the health system state, which will be explained in the following sections.

To summarize it, the flow of the relationship between diagnostic processes and prognostic are represented in Figure 2.5.

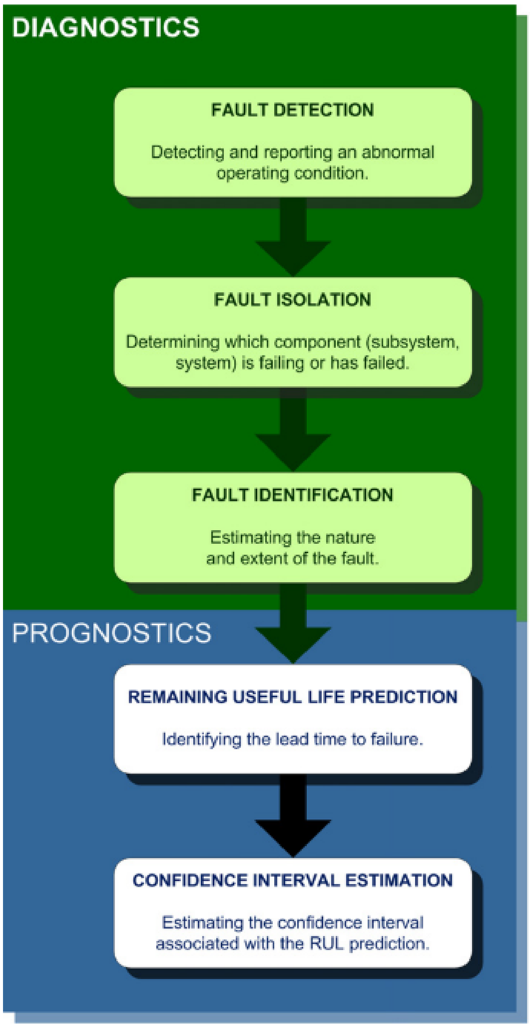


Figure 2.5: Diagnostic and Prognostic, based on [59]

Reliability - Centered maintenance:

As stated in [33], reliability centered maintenance (RCM) can be defined as:

"A method for developing and selecting maintenance design alternatives, based on safety, operational, and economic criteria. RCM employs a system perspective in its analyses of system functions, failures of the functions, and prevention of these failures."

At first, this type of model was implemented in the aviation industry because there is an high standard quality related to the risk of catastrophe. The need was to be able to watch all system to avoid any unnecessary maintenance. In the end, this policy created a balance in cost-effectiveness, safety, and asset availability during maintenance time [49]. In order to be able to achieve these objectives, an asset classification was made, depending on the importance that each one had in the system. If it happened to exist one low-risk and high-risk asset warning, the high-risk is prioritized because it has a bigger impact on the global system. On the other hand, a low-risk asset could reach failure status [14].

It is possible to divide the RCM into tasks, one of them is to analyze and categorize failure modes based on the effects of the failure on the system and the other is to assess the impact of maintenance schedules on reliability.

The failure analysis consists of the identifications of every failure mode and after that, categorization is done based on the consequences of each failure. In the end, failure modes and effects analysis (FMEA) is obtained.

This type of policy helped in developing a sustained knowledge about the failure modes and their impact on the system. However, it has some fragilities, this policy assumes that the system operates under nominal conditions. This means that they do not have into account the effects on the degradation and the influence on the system introduced by the load. This load depends on the production, which indirectly depends on the demand patterns. Operating in high loads, normally, is not harmful, in the short run, however in a long run could lead to lower availability and higher maintenance cost because of the degradation effects that the asset is undergoing. RCM deals with short to medium-term operational issues related to the equipment itself, therefore their objectives are a bit different from the ones defined as long term maintenance strategy [49].

In conclusion, through the years maintenance policies are changing and adapting to the new market always with one thing on their mind, improve the equipment availability while reducing its cost. This ratio is represented in Figure 2.6. It is possible to visualize what were the main concerns throughout the years.

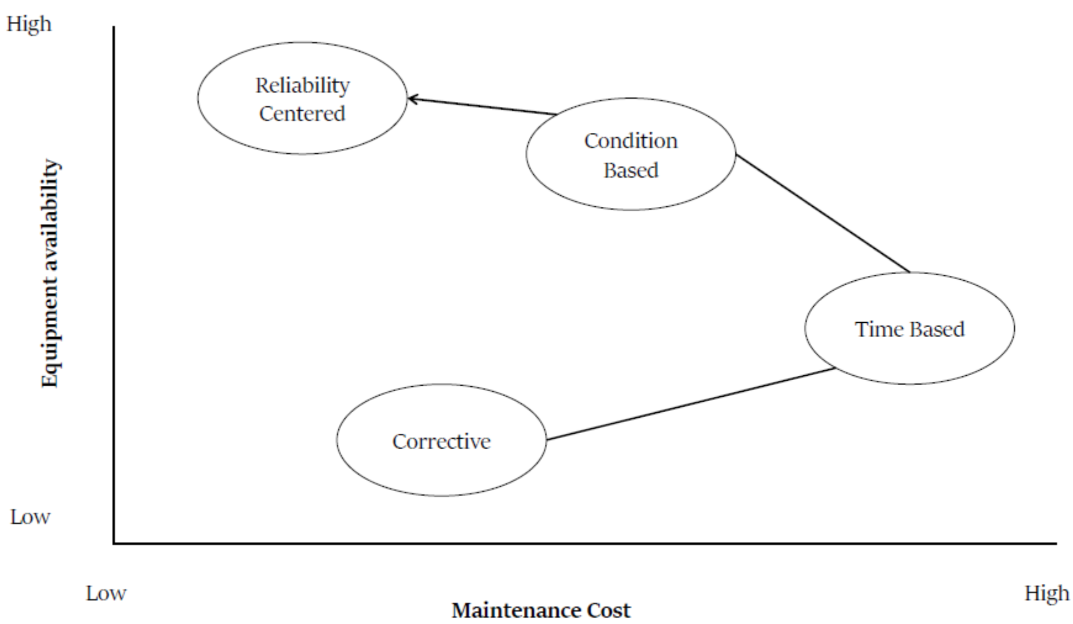


Figure 2.6: Maintenance Ratios Evolution [14]

2.3 Reliability Models

According to [49], reliability models are built with the purpose of "predict, estimate and optimize the survival or performance of equipment, the impact of the unreliability, and actions to mitigate that impact".

The most used modeling technique in the prognostic field is the calculation/prediction of Remaining Useful Life (RUL), which was already mentioned in the section before.

"Remaining Useful Life is the length of time that a machine, supposedly, has before it requires maintenance or replacement."

RUL is one of the most important tasks of the prognostic program. And because of that, many times prognostic is considered the same as RUL. More and more, prognostic models are being used in order to reduce the maintenance cost, improving the maintenance strategy, reducing the unplanned stoppages and improving operating efficiency.

There are several techniques for RUL's prediction, techniques that depending on the type of data, and the amount that is available is divided into models, in order to simplify the understanding. These different models are represented in Figure 2.7, and each model is detailed in the following sub-sections.

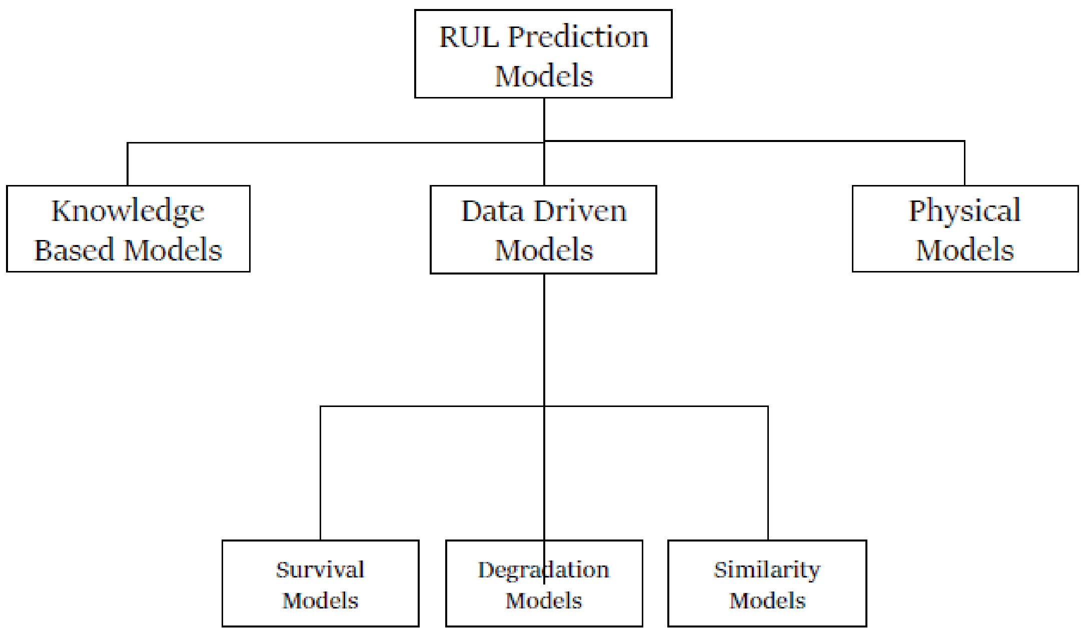


Figure 2.7: Prognostic Models [50]

2.3.1 Knowledge-based models

Knowledge-based models are designed with experience collected from working systems. So is required to have a qualified person with all the know-how to module a set of rules that can describe the system from start to finish.

The real definition of these models, according to [59], is:

" these assess the similarity between an observed situation and a data bank of previously defined failures and deduce the life expectancy from previous events."

This model is in constant improvement because its architecture becomes more and more detailed thanks to the increase of system understandings that is acquired over the time.

Fuzzy logic is the most common technique to make prognostic inside this model. It uses restricted rules to module the system to assure an overall check-up. Once this is an empiric system, it becomes very difficult to implement in complex systems because the rule set can be very hard to describe all the system even with a qualified person.

To analyze with more detail this prognostic technique, consult 2.4 where can be found more prognostic techniques.

2.3.2 Data-Driven Models

As the name indicates, data-driven models use data to model the system. Allow to collect information about the system in real-time and predict future health status. Bring some advantages like not having to rely specifically on a system expert, and also not have to be too much aware of the physical model of the system. Nevertheless, the amount of data collected is huge and there is a need to process the raw data. According to [45], the three pre-processing data methods are moving average (MA), weighted moving average (WMA), and exponential smoothing.

All data gathered could become an issue if it could not be collected properly or for some other reason not usable. To overstep this, there are data-driven models categories to apply according to the amount or the kind of data available [58].

2.3.2.1 Survival Models

Survival analyses is a statistical method that modules and analyze data during a specific period. The initial point is scheduled, but the finishing moment can be adjusted due to a special event, which in this project is an equipment failure. Once the failure is identified, it is possible to create a variable of interest which is the time since the beginning of the study until the failure [3]. In case that is not possible to identify the failure before ending the study and that failure happens, that information is called censored data. It's very important to analyze this data which is called right censuring [39], represented in Figure 2.8.

The data that is used to develop survival models can be extracted directly from a specific system in study or through a similar system. In both situations, reliability models are chosen not only by failure data but also by stress data. This model uses a covariant or explanatory variable

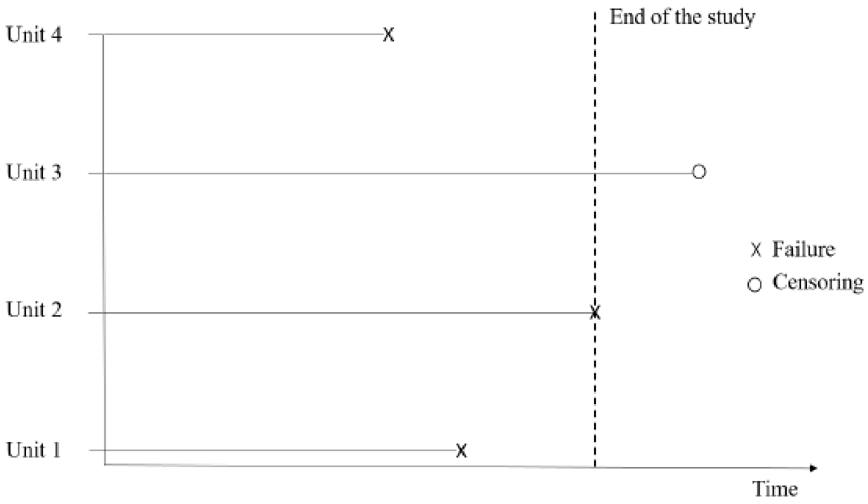


Figure 2.8: Failure and Censoring Time , source [51]

that is environment information which has a significant influence on the model because it relates straight to the material degradation curve [13]. The aim is to create a survival model that allows us to measure the remaining useful life, RUL.

Figure 2.9 represents a survival function obtained from the training data, with covariates [3].

In order to obtain a similar curve that is represented in Figure 2.9 it is necessary to use parametric, semi-parametric and non-parametric methods, such as hazard and survival functions. In a parametric method the probability distributions is already known, which means that, in order to estimate the model parameters, assumptions on failure or survival times need to be done.

On the other hand, non-parametric methods do not assume the "shape" of the data, and the estimations of the suitable form of the model along with the parameters is needed. [3]

There is a hybrid method called the semi-parametric method. They make assumptions on the type of relations between explanatory variables and survival times.

In the Table 2.1 the most common methods are presented.

Table 2.1: Survival Methods, based on [55]

Parametric distributions	**Non-Parametric**	**Semi-Parametric**
Weibull	Kaplan-Meier estimator	Cox's proportional hazard model
Gamma		
Exponencial and log-logistic		

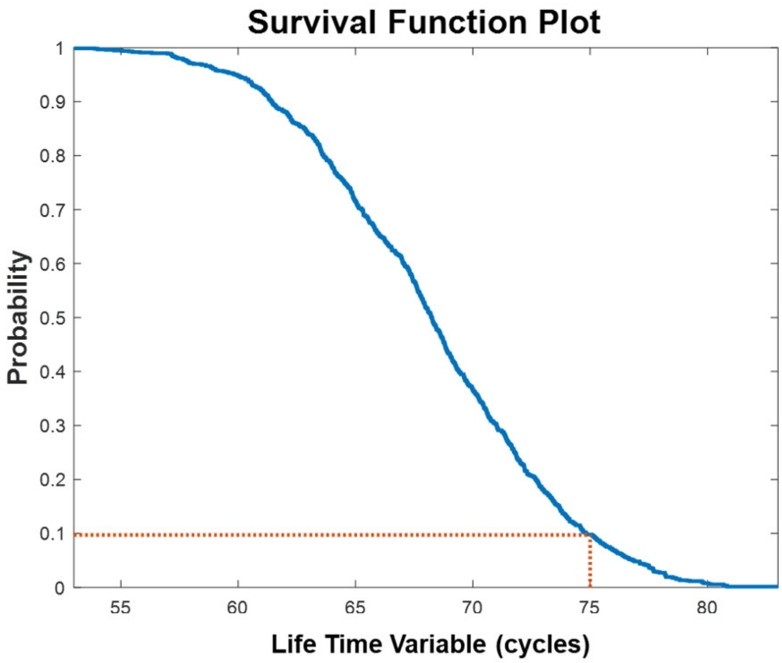


Figure 2.9: Survival Function plot, source [3]

2.3.2.2 Degradation Models

According to [29], the degradation models use degradation data that are acquired through continuous monitoring. They are used because lifetime data can be, sometimes, difficult to obtain.

Degradation is the cumulative change of a subject's performance characteristic over time, as stated in [29]. This accumulative changes will eventually achieve a threshold resulting in the equipment failure. This threshold normally is defined by experts [58]. For applying the degradation models it is necessary to have data and the failure threshold value associated with the condition of degradation parameter. The degradation parameter is a measure that is obtained by measuring directly the equipment and is connected with the health of the equipment.

The forecasting of the time that the condition parameter will cross the threshold is the way that RUL is predicted.

There are two types of condition parameters. Direct and Indirect Condition Monitoring (CM) data.

- **Direct CM**- "Data which can describe the under lying state of the system directly so that the prediction of the RUL is actually the prediction of the CM data reaching the threshold"

- **Indirect CM**- "Data which can only indirectly or partially indicate the underlying state of the system so failure event data may needed in additional to CM data for an RUL estimation purpose." According to [58].

The choice of the best degradation model for equipment is based on the models that were used in similar equipment evaluated by the same condition parameter. Once we have the degradation model, it is possible to predict the RUL, as we can see in Figure 2.10.

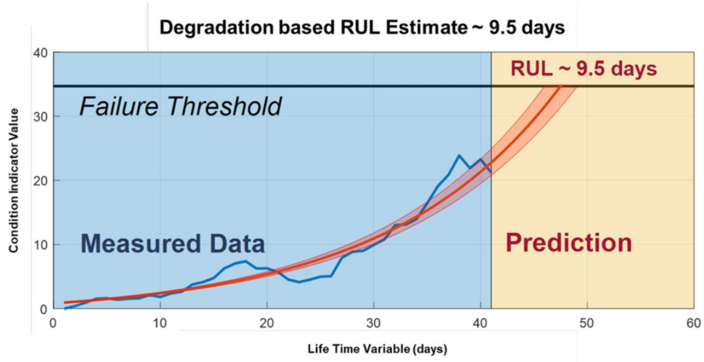


Figure 2.10: Degradation based RUL prediction, based on [3]

It is possible to use a model to estimate the time distribution required for degradation signal to reach the predefined threshold [23].

As it is possible observe, in Figure 2.10, the degradation curve has an upward trend, which means that over the years there are an accumulative degradation, that is understandable having into account all the stress and environmental factors. Based on [25], [41], [23], [54], and [61] the accumulative degradation of a measurement is described by an exponential curve. This assumption came after an exhaustive study of the behavior of sensor data over a long period, which was done by [41]. With the objective of represent the data in a function several fitting curves were tested, concluding that the exponential were the best fit, which are represented in Figure 2.11. It was assumed based on the scatter plots that the sensor data $s_{i,j,k}$ satisfies:

$$s_{i,j,k} = \phi_k + \theta_{i,k} e^{\alpha_{i,k} * j + \beta_{i,k} * j^2 + \tau_{i,k}(j) - \frac{\sigma_k^2}{2}} \tag{2.2}$$

where ϕ_k is a constant deterministic parameter for sensor $k, \beta_{i,k}, \alpha_{i,k}$ and $\theta_{i,k}$ are random variables, and $\tau_{i,k}(j)$ is the normally distributed random error term with mean 0 and variance σ_k^2.

Lastly, there are various degradation models and it is possible to divide them into two separate classes. The discrete degradation models, which any Markovian Models include, and the continuous degradation models. Degradation Path Model, Gamma process, Wiener process are the most commonly used.

2.3.2.3 Similarly Models

These models are the most precise models when the objective is to predict RUL. It is used run-to-failure historical data is available. Which means that we have a complete history from similar equipment with similar conditions. That included data from a healthy state, degradation, and failure.

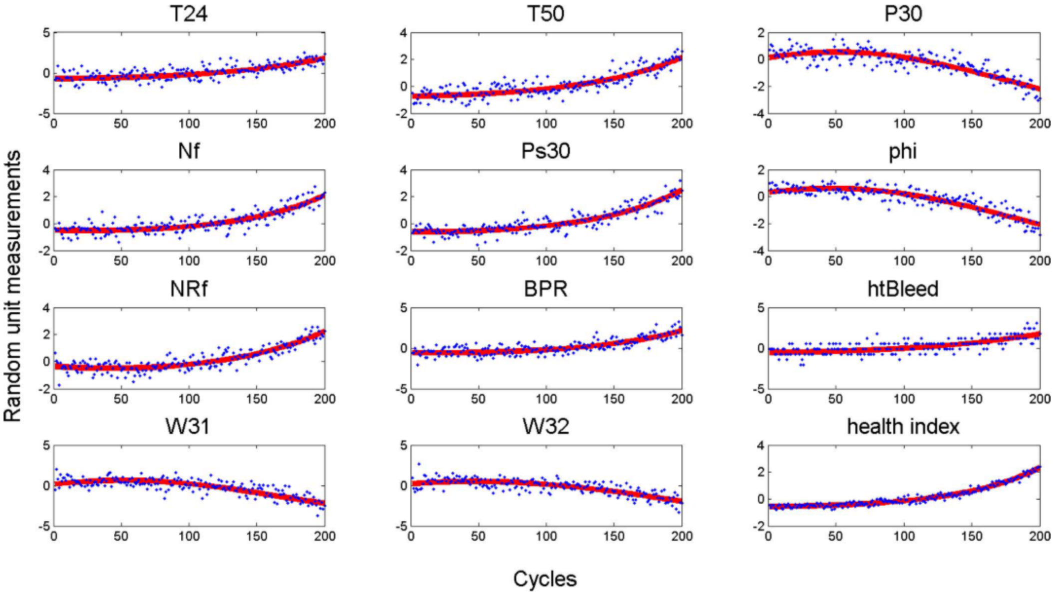


Figure 2.11: Degradation signals representation, Source [41]

Basically, the data that we are monitoring from our equipment is confronted and compared with many different degradation profiles built from the complete available run-to-failure data from various similar equipment, in order to find the closest match. [3] The time failures of the other models (red points on Figure 2.12) will be fit into a probability distribution. The RUL estimations are the difference between the average lifetime of the training equipment to the current life position of the test unit, as we can see in Figure 2.12.

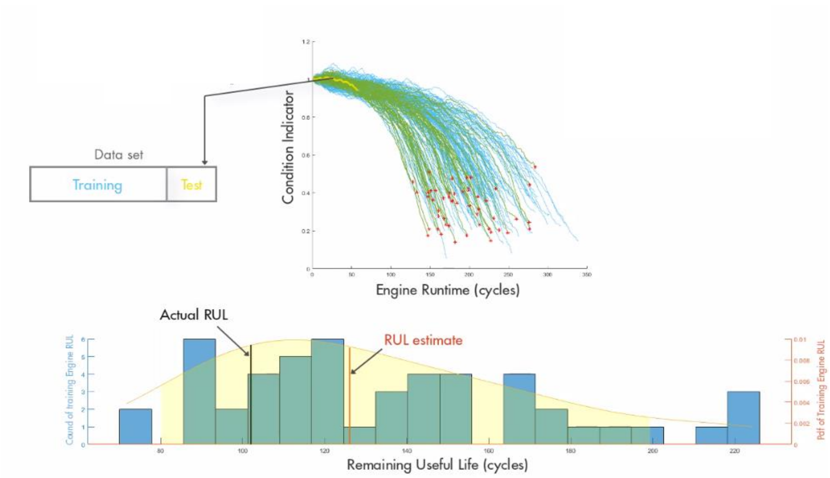


Figure 2.12: Failure Time distribution, source [3]

The most common distributions used in this models are: Weibull,exponential and normal distribution.

2.3.3 Physical Models

Also known as physics, failure or behavioural models quantitatively characterise the behaviour of a failure mode with the help of physical laws. In order to develop a physical model, a deep understanding of the system behaviour in response to stress is needed. The RUL prediction in this type of model is calculated through solving a deterministic or several deterministic equations from extensive empirical data. This data came from engineering knowledge and through the acquisition from specific laboratory or field experimentation.

Failure models can be described using a series of dynamic ordinary or partial differential equations, that can be solved with Lagrangian or Hamilton dynamics. It also can be described through a state-space method [59].

Once a physical model is completed, the real data from sensors are compared with the output from the model. Differences between reality and the model are called residuals. High residuals indicate that a fault has occurred, small residuals occur under normal conditions.

The main advantages of these models are the ability to directly incorporate an existing understanding of the physical mechanisms of failure that, in many cases, were already intensively studied and tested empirically. When they are sufficiently complete, behaviour models tend to outperform other types of models.

However, this model have their issues and disadvantages. The main problem faces is that the behaviour of the system must be derivable from first principles, this may not be possible due to an imperfect understanding of how the failure mechanism behaves under operating conditions. It is needed a considerable accuracy and reliability on multi-variable data, that is rarely available in order to assign the parameters for all aspects.

2.4 Asset Health Index

Asset Health Index or Health Index is a very useful tool to represent a simple or complex condition system. This technique combines various data into a single numeric value [14] [27].

As was mentioned in Section 2.2.2.2, there are several techniques to demonstrate the system's condition, being the RUL the most used technique. Health index techniques that are going to be presented in the flowing section are prognostics techniques that depending on the available data give us a representation of the health of the system.

Having into account what was said before, constructing a composite Health Index (CHI) has two major steps. The first one is the development of the HI, which is predicting the asset health status, and the second one is the agglomeration of various HI's to create a more complex one, that can represent a bigger system. This aggregated or composite health index is used because they usually can measure multidimensional concepts that cannot be captured by a single index [22].

Despite having many techniques to obtain and health's system representations, every prognostic result has similar objectives, are presented in Figure 2.13.

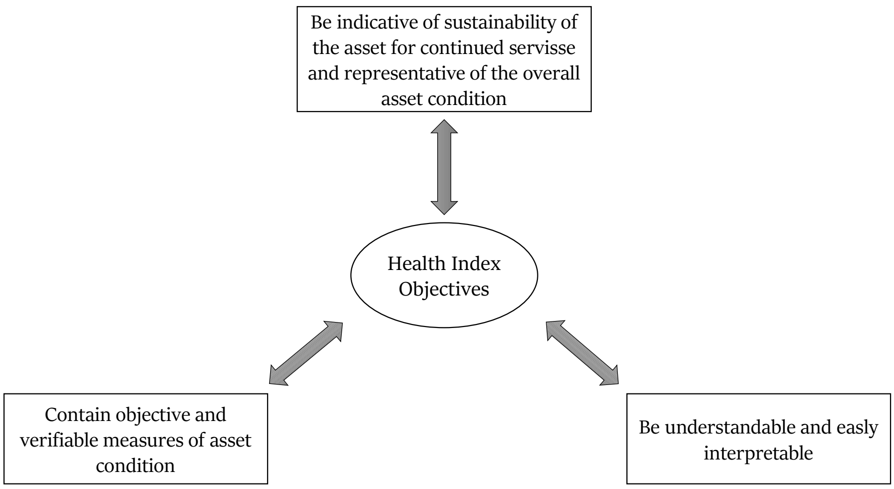


Figure 2.13: Health Index main objectives [14]

This subsection will be divided into two parts, the first one will explain the different prognostic techniques that are used to predict an HI. In other words, how to transform the input data, that is provided by sensors, into an HI number, and the other, different techniques that are used to make composite HI, which means techniques that are used to add various HI into one without losing data sensitivity.

2.4.1 Health Index construction techniques

As was described in Section 2.4, a health index is a single numeric value that represents the overall health system. Usually, this value is between 0-1 and conjugate all the data that is relevant for making a condition evaluation. The development of a model that can represent a system's health needs to follow some steps in order to obtain the best result. The first step is data acquisition, which is one of the most important. A sensor acquires data and sends to the computer that is responsible for processing it. However, before processing, it is needed to verify if the data itself has some anomalies. The data suffers a small inspection with the objective of founding sensor failures, reducing false alarms. After acquiring data, a pre-processing is made. Pre-processing routines, usually, include data compression, filtering, signal averaging, and Fast Fourier Transforms [42]. Subsequently, the condition indicator is defined or extracted. These are values that after analyzing the faulty data stand-out. In the cases that we do not have reported any fault, it is possible to identify the condition indicator by analyzing the efficiency of the system and compare which values are associated with the worst results [42]. Lastly, a prognostic technique to predict the condition's health of the system is used, and there is a panoply of techniques, which some are described below.

2.4.1.1 Match Matrix (MM)

It is a technique used to do prediction on a far horizon. It has been used in non-stationary multivariate time series and non-linear dynamical systems. It is usually described as an enhanced Auto-regressive Moving Average model (ARMA). This type of technique needs historical data to obtain the best results [40].

The foundation of this method is not too complex. In a first state, it is created a matrix that represents the paste maintenance cycles, with p features vectors, that represent these cycles. After that, a similar matrix is created, but this time with the current maintenance cycles, with q feature vectors.

The degradation pattern of the system is compared by multiplying the before made matrixes [40].

$$Q * P = S \tag{2.3}$$

Each element of the S matrix represents the similarity between features vectors of the past and current runs, which form the Match Matrix (MM). Thereafter, a linear parametric prediction technique, such as ARMA modeling, is used to compute the MM and predicting the HI.

This technique is not the most suitable for problems with large datasets, because of the computational power needed to make individual comparison of each element [34].

2.4.1.2 Particle Filtering

This technique is normally referred to as a Sequential Monte Carlo (SMC) method that uses a statistical method called Bayesian inference. Firstly, this technique was proposed for recursive Bayesian filters [24]. These filters represent the future distribution of the variables in a particle system. Once new data arrives, the system adapts and evolves recursively.

A particle system, as the name indicates, is a system composed of particles, which contain information of unknown parameters, that are estimated and updated as a form of the probability density function (PDF) in the Bayesian process using observations. The particle moves from states that are more likely to happen or that can represent better the system. This movement and likelihood are represented in Figure 2.14.

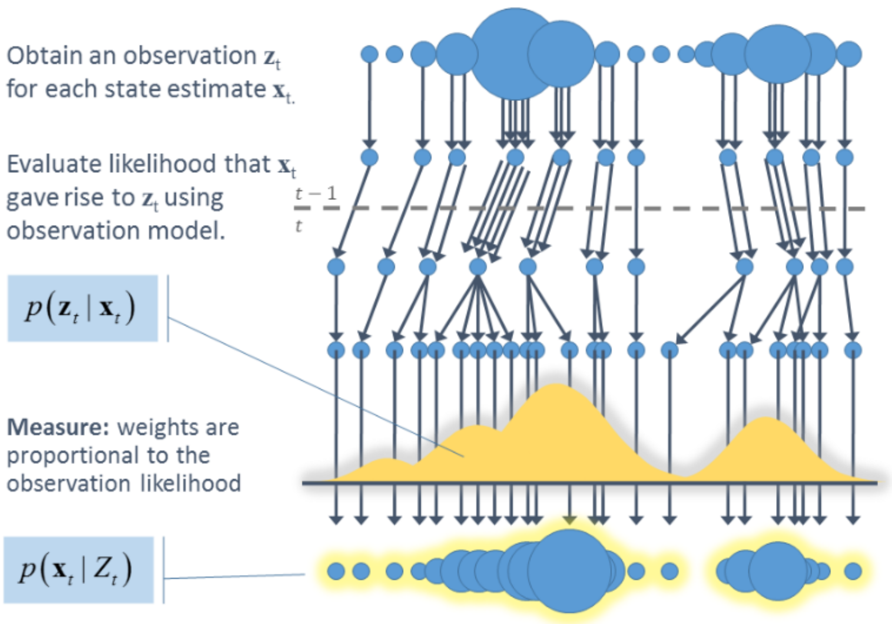


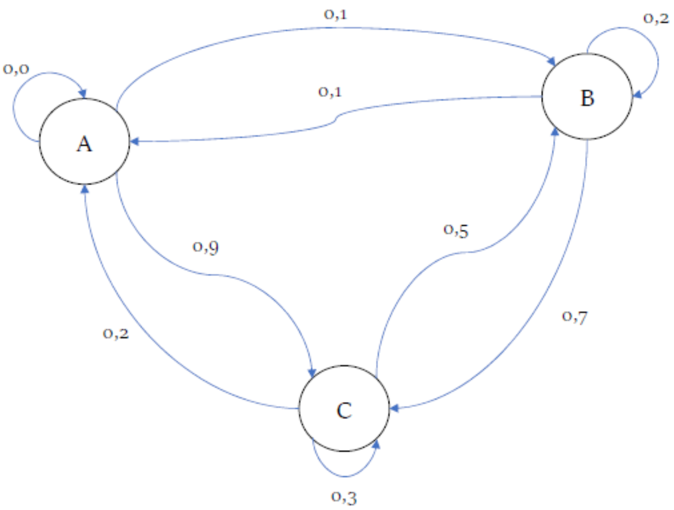
Figure 2.14: Particle Filtering technique [2]

This technique is capable of dealing with non-linearity, non-standard posterior distribution, and multivariate [10]. However, the vulnerability of the model's filter can make it collapse for a single point, which is not ideal.

Particle filtering has been applied for forecasting and prognosis extensively, and it is achieving remarkable results. However, the computational power needed to perform this type of model is really demanding, leading to some studies to reduce that demand [34].

2.4.1.3 Hidden Markov model (HMM) and Hidden Semi-Markov model (HSMM)

Hidden Markov model (HMM) is a statistical approach based on the principle of Markov chains (MC) for modeling signals that have a finite number of states. According to [15], MC is a sequence of events that each one depends only on the previous state. In other words, the MC basis is the same as the state machine model used in robotics to model the different actions and their interactions. But instead of having conditions in the transitions, they have probabilities, as we can see in Figure 2.15.



$$P = \begin{bmatrix} 0,0 & 0,1 & 0,9 \\ 0,1 & 0,2 & 0,7 \\ 0,2 & 0,5 & 0,3 \end{bmatrix}$$

Figure 2.15: Example of a Markov Chain

In the example above, probabilities between states are fixed, which means it is a constant. However, depending on the complexity of the problem, they can be variable, with a function associated.

It is possible to map all the features of a system with an MC if the average-state-change is efficient and if the number of states is sufficient. This technique is suitable for non-linear and non-stationary systems [34].

Hidden semi-Markov model (HSMM) is a specific and modified type of HMM which considers temporal structures [16] [17]. HSMM is an improvement of the HMM and is not bounded by the Markov chain assumption making one of the reasons that it analysis powerfully real-life problems [34].

These types of models started to be used in the prediction universe recently, being good on predicting the RUL. One thing that can be a problem for this model, is that they require an immense computational power, besides, they need a pre-defined threshold [16].

2.4.1.4 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) are known as the non-linear information processors. This technique is widely used in data-driven prognostic calculations.

The neuron is the basis of an ANN, which interacts with another one through a numerically weighted connection. A group of neurons organized in a specific way is called a layer, and an ANN is built by many layers, as we can see in Figure 2.16

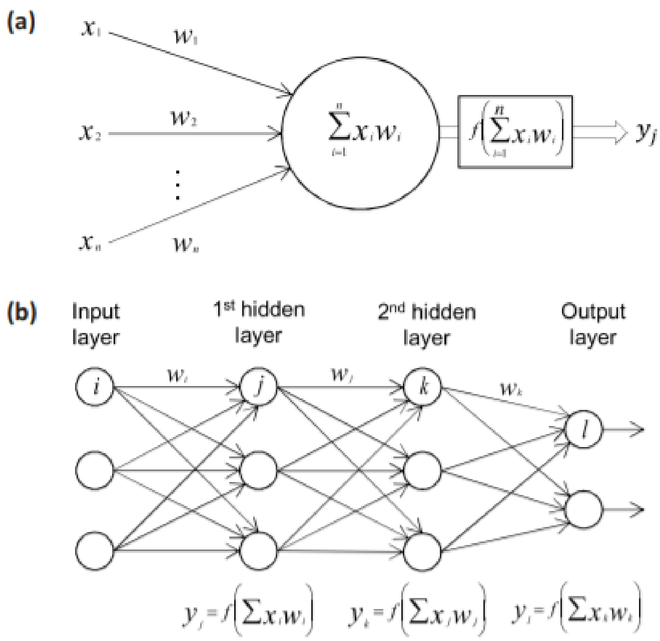


Figure 2.16: Block of deep Neural Network, from [62]

Each input (x_i) has a associated weight (w_i). The sum of all weighted inputs $\sum(x * w)$ is used in a non-linear function, called activation function f, to transform the pre-activation level of the neuron in an output level, (y_i). There is also a bias, usually used in conjunction with the activation function, but to simplify it was omitted. After that, the result which is the output of a neuron is the input for the next layer's neuron [62].

In order to obtain the best results, it is necessary to train the network and there are two types of training. The supervised training (ST) and unsupervised training (UT). ST, according to [57], is based on training a dataset where we already have the correct classification made. On the other hand, UT is a training where the network has to learn and organize information without providing an error signal to evaluate the potential solution.

Another classification, is the NN itself, they can be a feed-forward network or a dynamic network. A feed-forward network represents a network where the set of outputs does not affect the response to the next set of inputs. On the contrary, in dynamic networks, the inputs depend on the output of the previous node and to the previous iterations of the network itself. Having into consideration all that was said before, there are different types of ANN, and these types are mainly because they use different activation functions.

This model can work with non-linearity and time-varying dynamics, but it has some negative points. It is highly dependent on well-trained data and has no transparency on how decisions are reached in the network. And last but not least, the optimization of the results, the standard methods to the optimal solution is yet to be established.

2.4.1.5 Fuzzy Logic

Fuzzy logic is usually used in the control industry, however, it can be used in other areas, for instance, to make a prognostic (HI). This type of model was created to work with uncertainties that systems had. Back in the days, the logic used was binary logic, which only has two values, zero (0) or one (1). This type of logic created a very restrict system, and all the values that were in the between were adulterated. Because of that need to understand what was happening in the "dark" area, the fuzzy logic was developed [11].

The fuzzy logic is based on fuzzification, which means that the inputs need to be converted into fuzzy representation thereafter, those modified inputs can be compared against the fuzzy set rules. These rules have a restricted format, which is represented below by Equation (2.4).

$$\textbf{IF } x \textit{ is } A \textbf{ AND/OR } y \textit{ is } E \textbf{ THEN } z \textit{ is } U \tag{2.4}$$

It is easier to understand the rules, and the fuzzy logic by analyzing Figure 2.17

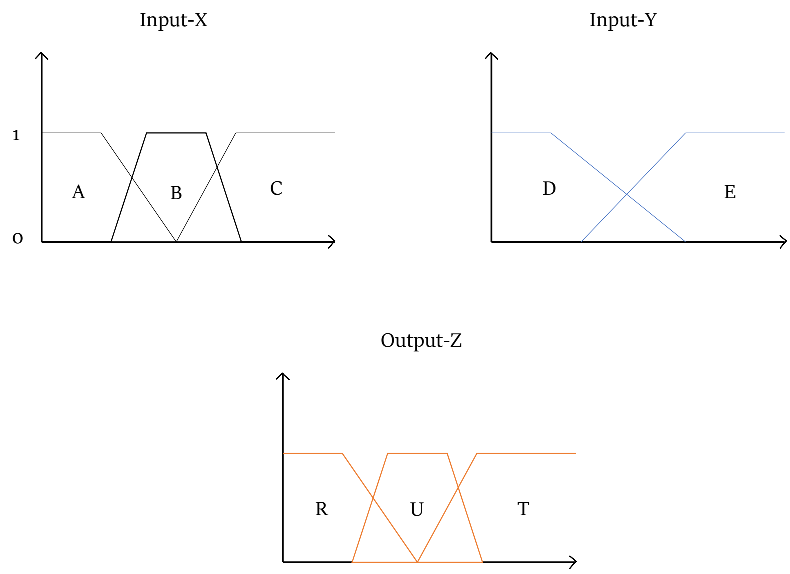


Figure 2.17: Fuzzy Memberships

Each data input has one set of memberships (classification) which is the representation of the numeric input data into linguist data. Once transformed, they are compared with rules. These rules are handily made by an expert, which has to have a big understanding of the system in order to modulate all the possibilities in a set of rules.

In short, memberships are used to fuzzificate all datasets, which result in the linguistic variables. These variables are mapped into rules in one single output that represents the Health Index [59].

A table which gathers all the advantages and disadvantages of these techniques that were talked above can be found in Table A.2.

2.4.2 Agglomeration Techniques

Nowadays, the need for information to help to make decisions is a must, but not every information is helpful nor reliable, which is the first problem in constructing a composite HI. There are two distinct sides when the subject is composite HI. On one hand, there is a group of people that support that a composite HI can give miss leading information about the overall health system situation. Also uphold that the development of the aggregation equation requires a lot of assumptions and arbitrary decisions, which is not the ideal when the health systems depend on it [22].

On the other hand, proponents of composite HI, defend that the straightforwardness that it brings in decision making is a real advantage. Having only one number to analyze and discuss make the overall view of the health system more clear. They also agree that the main problem that exists in the construction of a composite HI is the lack of good and appropriate information [22].

Despite some nonconformities about their reliability of describing the system's health condition, composite HI are being more and more adopted and studied. Therefore, there are different techniques to agglomerate several HI into one. Some of these techniques are presented below.

2.4.2.1 Adding and Technique

As the name implies, this technique relies on the sum of all individual component's health indexes. This can be misleading, as we can see in the following example:

Several components from a group of generators are being study, and the results are represented in Table 2.2.

Table 2.2: Composite HI - Adding Technique, based on [27]

Factor	**Trf1**	**Trf2**	**Trf3**
DGN Main Tank Score	5	0	0
Dieletric Score	2	10	0
Thermal Score	3	0	0
Oil Score	1	0	3
Mechanical Score	1	10	0
DGA LTC Tank Score	7	0	10
Operational Score	3	10	0
Subject Matter Expert Score	8	0	10
Design/manufacturer Score	3	3	10
SUM	33	33	33

The score represents the value that was given to each component, according to the data that was previously analyzed.

According to the example, it is possible to see that each transformer has the same score, theoretically, means that all transformers need an intervention with a similar sense of urgency. However, it is clear that while transformer number 1 is suffering from aging degradation, transformer number 2 and 3 has serious and specific issues, which are represented by the scores with value 10. Those scores mean that the condition of the component brings imminent danger and failure. This is the problem of the simple addition system because can hide problems. Similarly, the average technique has the same problem. Making impossible to understand the real state of the system's health.

2.4.2.2 Logarithmic Technique

In order to fight the main problem of adding and averaging technique, which was the lack of sensitivity in the data variance, the logarithmic technique was created. Starting by changing the scale, depending on the data range that it is considered can go to 1-100 or 1-1000. Higher than that it starts to be difficult to interpret the meaning of the HI. A score range is defined, which means that the individual HI needs to be classified depending on their evaluation. This can be understood by the example below, based on [27].

Taking the same example that was given in Section 2.4.2.1, a dielectric's health condition is being analyzed, and depends on the individual HI represented on Table 2.2. Now, these scores are going to be subjected to a transformation having into consideration the log scale condition codes, Table 2.3.

Table 2.3: Log Scale Dieletric Condition Codes, based on [27]

Code	Description
1	No known problems
3	Slightly unusual dissolved gas signature
10	Possible arcing/sparking or partial discharge fault
30	Severe arcing/sparking or partial discharge fault
100	Very severe arcing/sparking or partial discharge fault, transformer at high risk of failure

After the transformation, the adding technique is made. Once the values are in a logarithmic scale their variance are more likely to be noticed in the final result, as we can see in Table 2.4

It is possible now to notice that Trf2 and Trf3 need an urgent maintenance intervention, while the Trf1 do not.

2.4.2.3 Weighted Average

The Weighted average is the most used technique, based on [12], when we are talking about composite HI.

Table 2.4: Logarithmic Condition Codes

Factor	**Trf1**	**Trf2**	**Trf3**
DGN Main Tank Score	10	1	1
Dieletric Score	3	100	1
Thermal Score	3	1	1
Oil Score	3	1	3
Mechanical Score	3	100	1
DGA LTC Tank Score	30	1	100
Operational Score	3	100	1
Subject Matter Expert Score	30	1	100
Design/manufacturer Score	3	3	100
SUM	88	308	308

Firstly, the identification of the key indicators is a task with high importance (health indexes for a single component that is part of the equipment in the study). Consequently, it is necessary to determined how important they are to the overall health of the equipment, which means the relative importance. This identification is based on the failure modes analysis, which is the analysis of the most common reasons for failures, as well as failure trends and correlations across datasets. Then, the definition of evaluation calculations is decided. According to [12], there is no standard way of calculating assets HI. Each organization will play different values on the various factors involved. Within an organization, multiple stakeholders must come together to help decide what is important and how is calculated.

This method is very used, however, it still has some large gaps. The first is the location of weights to the test is also based on the experience, which means that it differs from one expert to another. The second disadvantage is the numerical thresholds between the score can not be precisely found, creating a fuzzy region, that statement can be found in [7].

2.4.2.4 Maximum and Multiple Increment

Based on [6], the maximum and multiple increment (MMI) technique is a method to merge information, typically numeric without losing many data features.

Using MMI technique ensures that the aggregated score is driven by the strongest individual reading.This technique uses a threshold-based dynamic, called, critical threshold (CT) . It has normally a high value and represents the limit to consider that the system entered in an emergency state. This threshold is also responsible to indicate which of the two sub-aggregation algorithms will be used.

Firstly each reading,M_c, where c represent the component that the measurment M is part of, is compared with the CT. If $M_c < CT$, means there is no problematic situation happening or beginning to happen and it is used the Steady Algorithm. In this algorithm the resulting aggregated score is determined only by the lowest and the maximum of the individual reading. However, if $M_c > CT$, means that the system has entered in the danger zone. Some readings have an abnormal values, which can lead to an abnormal situation. In that case, the sub-aggregation algorithm used is the Unsteady Algorithm. In this algorithm the resulting score will be determined by having into account the highest value of the readings, and a given number of the next highest readings, depending on their values, but can not be higher than the maximum number of combined factor, which is a calibration parameter. Which represent the number of reading that should have impact on the final result.

Chapter 3

Problem Description

In this chapter, the problem that led to this thesis will be presented. In the first section, it will be done a brief description of itself, as well as, what led to the seeking of a solution. Furthermore, it will be presented a nomenclature with some general concepts that will be used throughout the document. In the second section, it will be presented the way that the problem was approached as well as the main encountered difficulties.

3.1 Contextualization and Nomenclature

This technological explosion that we are living brought with it easier ways to deal with problems, faster ways to finish our tasks and so many other advantages. However, the cost of that was the exponential increase in the system's complexity. Systems with more relations within, more inputs, outputs, variables, and less manpower operating them. For that reason, the constant monitoring of the systems gains a crucial role. On a complex system, like an HPP, it is possible to have hundreds of sensors sending readings, in real-time, which in a few minutes transforms in gigabytes of data, that makes it almost impossible to understand and analyse on naked eye and that is exactly the foundation of the problem that will be dealt in this dissertation. The need of having a easier value to analyse, that represents data that is presented in all the sensors readings, is increasing. On decision making is an enormous benefit, since, firstly, it is not necessary to be an expert in the system to understand the meaning of that value, and secondly, it is a user-friendly health indicator.

Therefore, this dissertation aims to develop a methodology that can build an interpretable HI value, that can accurately represent the system's health. Before proceeding, it is necessary to understand some concepts that will be used throughout the document, which is described in Table 3.1:

It is also possible to see the relationship between these concepts in Figure 3.3.

Table 3.1: System Nomenclature

Name	Lable	Definition
Measurement	M_c	Measurement or Condition Indicator of the component c, are the readings of the sensors that can give information about the health status of the component that is in.
Component	C_{csS}	Equipment that is monitored by measurements. A group of components that work with the same purpose are considered a sub-system. C_{csS}, mean the component c of the subsystem sS
Sub-system	sS	A group of component
System	S	Hydropower plant and all their interactions

3.2 Problem approach

As was said in Section 1.1.1, the objective of WP3 is to create an SPPS. The SPPS is a system integration methodology of hydroelectric technology solutions based on advanced control and deep knowledge of the HPP units by integrating the operating time-series statistics, the model tests, the numerical simulations, and the field test results in the unit baseline database to generate a so-called multidimensional hilchart and state estimator constantly updated by advanced monitoring of the HPP unit health indexes. This will enable the increase of the HPPs' potential. These interactions are represented in Figure 3.1.

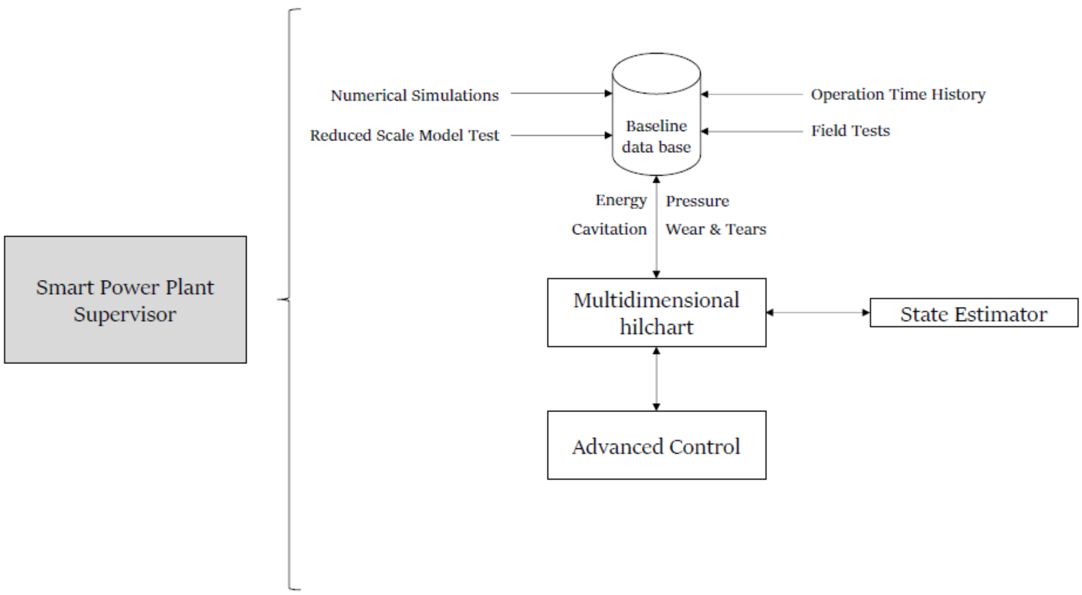


Figure 3.1: SPPS data flow

The current section is addressed to, firstly describe the main features of the provided data and secondly, to define the main goals settled for global HI problem, that can be, achieved by applying the methodology introduced in chapter 4.

3.2.1 Data specification

The data that was going to be used in the construction of the global HI had three sources. There were going to be data that represented the operation time statistics of the power plant. This data will be collected from real HPP that is being monitored. In which it is possible to extract some features from HPP and understand the different states that the turbine has, along with the highest and lowest-demand intervals.

Physical test data will also be available. These tests are being done on a reduced scale HPP. This methodology will be defined to test the extension of the operating range of the hydroelectric unit and establish its limits in respect of the wear and tear, and residual lifetime. This includes the definition of the measurements to be performed, in other words, the thresholds of the system.

Finally, data from numerical simulations will be also a source of data to analyse. Specific software will be used in order to do these simulations. That software will simulate the hydropower plant and the EPS under extended operating conditions to identify critical conditions. Tests that simulate a rupture from fatigue will be done, extending the operating range to extreme limits. Despite having three types of data sources, these tests are all connected and that connection can be seen in Figure 3.2, which represents the system's data flow.

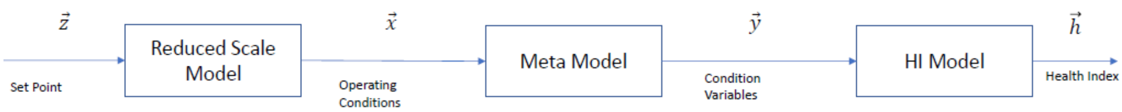


Figure 3.2: System's data flow

However, due to bureaucratic delays and, mainly, because of COVID-19, it was not possible to have access to the data described below. And in order to solve that lack of data, a wind power plant dataset, shared by EDP Open Data [1], was used to develop the methodology described in Section 4.

3.2.2 Asset Health state prediction

The hydropower is gaining more and more importance in the European Network of Transmission System Operators for Electricity, which already counts with a total capacity of 237.5 GW of hydroelectric power [21]. A hydroelectric power complex system is constituted by many components that need to work perfectly together in order to obtain the best results. However, not every HPP are equal. There are three different types of HPP, depending on the capacity that their reservoir has and the type of turbine that they use [18]. Starting with the largest, reservoir or storage type. This HPP has the biggest reservoir of the three existing types and confines the water behind the dam to enable the flow regulation throughout the year, depending on the size of the reservoir. Usually these plants are used as a based load plant but in extreme situations they can be also peak load plants [20]. They use either Francis or Pelton turbines [18]. Run-of-River (RoR) also called a diversion facility. They are usually used to respond to the base demands of the grid. It is not required, in some specific cases, a dam. They use a penstock to channel a portion of the river through the powerhouse. The Kaplan turbine is the perfect fit for this type of HPP [18]. Lastly, the pumped-Storage facility. This type of HPP is called pumped because when there are profitable times they can pump water from a lower reservoir to the upper storage. A profitable time is when the electricity that will be used to pump has a lower price, other times that the pumping is necessary to make is when a dry season is predicted and it is needed to have water in the upper storage. They use Francis turbine types because it is the only type that can rotate in both ways[20] [18], in other words, can work normally and can also pump the water to upper reservoirs.

As was said before, an HPP is a complex system with many components. In Figure 3.3, the subsystems, and components that need to be monitored in the XFLEX project are represented.

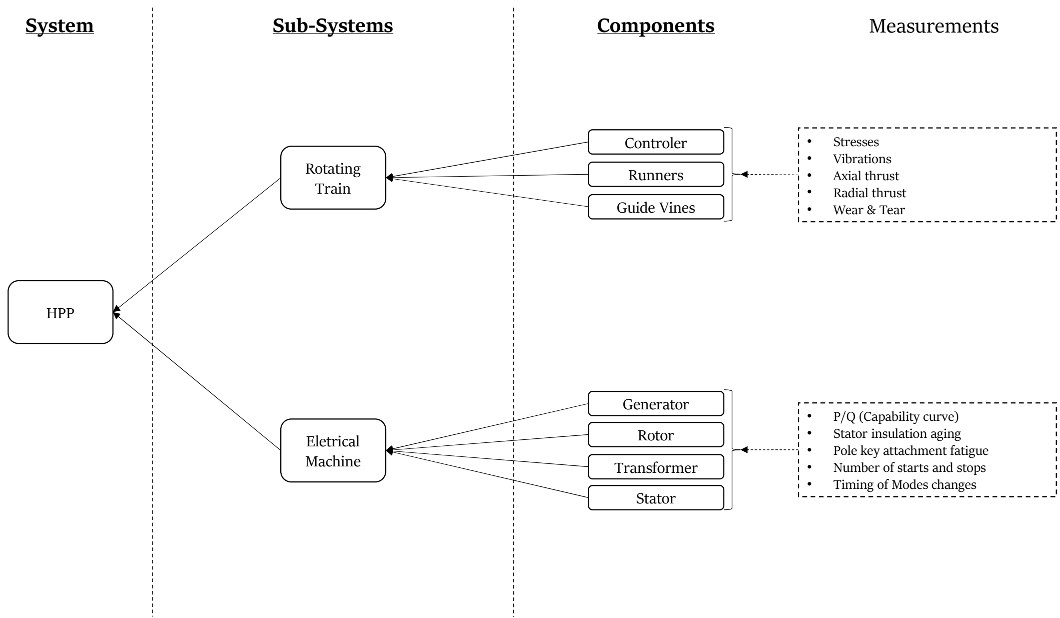


Figure 3.3: Hydropower station studied architecture

Having the architecture of the system in mind, the objective was to build a methodology that could lead us from measurement data to systems HI. In other words, a function or a group of functions that has as inputs the readings of the sensors and as outputs the composite HI.

Starting from sensor readings, these are the most primitive and raw set of data of the system. As it is possible to see in Figure 3.3, many measurements can be the condition indicators of a component. Therefore, it is important to understand which of them can represent the best health condition of the component in study. Once the components have their representative measurements it is time to transform these into degradation data. This means that the raw data is passed through a degradation model function, in order to be transformed into a condition indicator. This transformation is represented by the following equation.

$$h_{ic} = f(\vec{M_c}) \tag{3.1}$$

where h_{ic} represents the condition indicator of a component, i represent the measurement that is being studied, c the respective component and M_c represents a specific sensor reading. In case the component only has one measurement to represent his health status, the h_{ic} correspond also to the component health index, otherwise, it is necessary to agglomerate all the h_{ic} from the respective measurements to represent the component health index.

$$\vec{M_c} = (M1_c, M2_c, M3_c, ..., Mn_c) \tag{3.2}$$

where n, represent the number of measurements taken from the component c. Anew, a sub-system is a group of components, which means that their individual HI needs to be aggregated in order to obtain the sub-system HI. Nonetheless, a question arose, which was, if the components had the same impact on the sub-systems. Are there any component more important to the functioning of the overall system? Intending to evict this problem, the aggregation model, Section 4.2, had in consideration the worst case. Then, the degradation data obtained in (3.1) passes through an aggregation function, which is output is the sub-system HI.

$$h_{subsystem} = f(h_{M_{1c}}, h_{M_{2c}}, ..., h_{M_{nc}}) \tag{3.3}$$

Finally similarly to the mentioned example, the system is composed by several subsystems. Therefore, the overall system HI is calculated by the aggregation of the constituents sub-systems HI.

$$h_{system} = f(h_{subsystem_1}, h_{subsystem_2}, h_{subsystem_3}, ..., h_{subsystem_n}) \tag{3.4}$$

To sum up, the following flowchart represents the steps from the raw data to a composite HI.

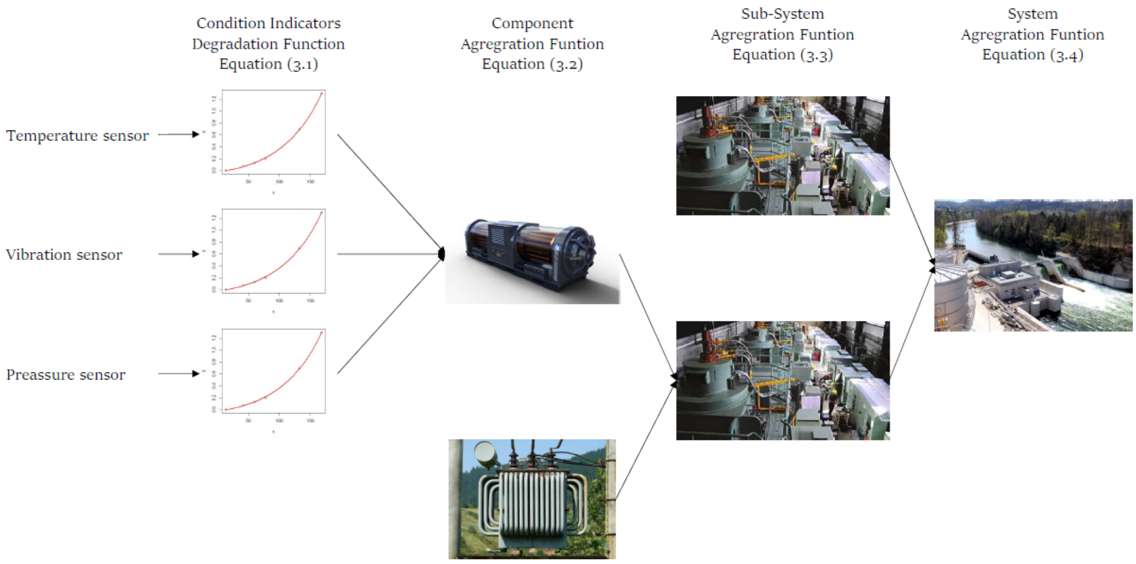


Figure 3.4: Systems Data Flow

Chapter 4

Reliability model and agglomeration methodologies

In this chapter, the methodologies that were developed in this dissertation are presented. In that regard, it is possible to divide into degradation modeling and agglomeration technique. The degradation model is applied to all measurements of each component, in order to obtain their reliability curve and understand their health state, which will influence the overall health state of the system. Firstly, it was needed to separate the different components into different groups, that we called sub-systems. They were clustered according to the area that they worked on (hydraulic machinery, rotating train, and electrical machinery), as it is possible to see in Figure 3.3. The model that was created took into account all the different sub-systems but only the electrical machinery sub-system was tested. The available data is composed by readings of the sensors and the failure report. Having this, it is possible to analyze the behavior of the system before the failure happens and try to find some patterns or features that can indicate future failures.

Figure 4.1 represents the phases that were needed to be conducted in order to obtain a model capable of representing the health status of the component, as well as, predicting future failures. For that an initial data processing was made to clean the noisy data, next the degradation curve was drawn based on that it. Lastly, the failure prediction phase is where we define and compare the HI reliability, as well as the probability of a failure associated with an HI. This reliability is tested by changing some parameters and analyze the results, and these different tests, with the objective of finding the best results, are represented by the feedback arrow. All these details will be explained in Section 4.1. Once we have all the cumulative degradation represented and associated with an HI, it is needed to fuse them in only one number. In Section 4.2, we describe the steps required to conclude that stage and therefore to obtain the composite score. The overall system architecture is represented in Figure 4.2.

To summarize, the desired output of these methodologies is a composite HI that ranges between 0-1, and whose returned value forms the basis to predict future failures.

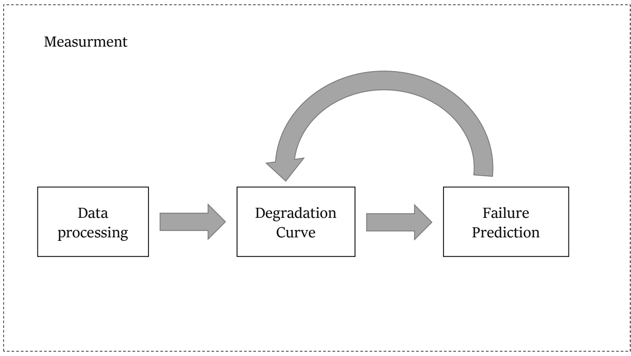


Figure 4.1: HI Single Component Model

4.1 Degradation Model

In this section it will be described the methodology that allows us to determine the HI of a single measurement. A measurement is a value that is obtained from the sensors that are monitoring the system. A component is, usually, monitored by different sensors that analyzes different parts of the component. Some system conditions can be directly reflected by reading the sensors. In this type of situation, the detection of failure states becomes possible through the analysis of the behaviour of the acquired data. However, there are conditions that are not directly detectable by the analysis of a simple measurement. This was one of the reasons that led to the creation of a number that could represent several states.

4.1.1 Data Processing

The data-driven models, as the name indicates, depend fully on data but not every data is usable nor relevant. So there is the need of doing a pre-processing which includes selecting the relevant data for the model construction.

The data that fed the algorithm consists of raw sensor data and a failure report. This type of data allows to understand how the system works before, after, and during a failure. However, some pre-processing techniques were first required to manipulate the data more easily. As mentioned above, the sub-system that had bigger emphases was the Electrical Machinery, especially the generator.

The data that we had was organized in a way that each component, from each equipment, only had associated the sensors that represented their health state. Focusing on one component at the time, we performed a correlation test on the measurements to better understand how the system

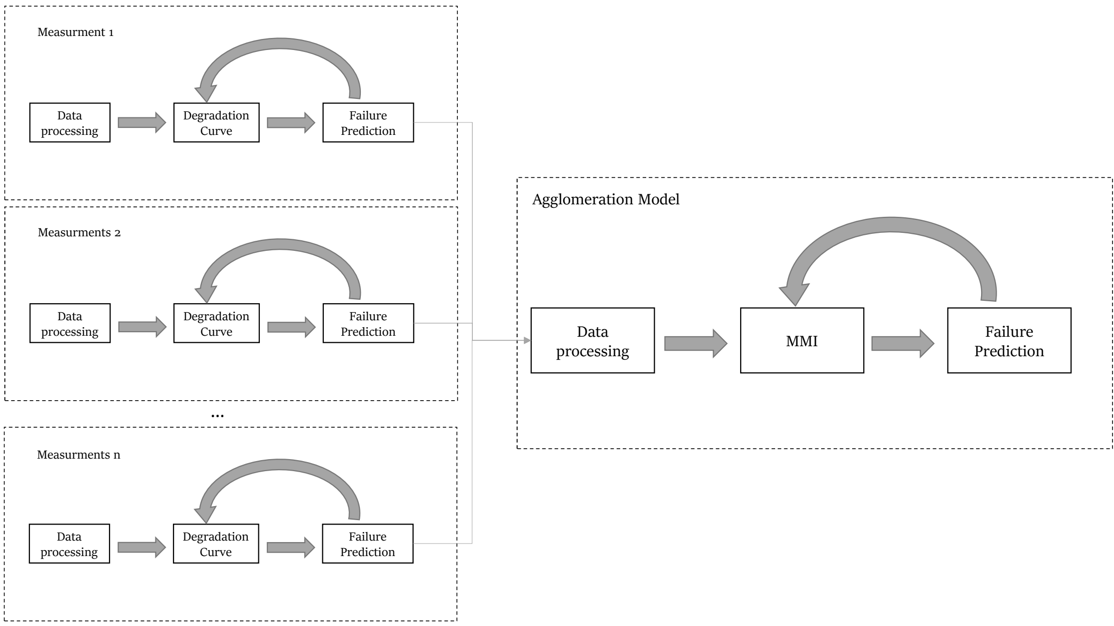


Figure 4.2: HI Model architecture

was interconnected and especially, to identify redundant variables, in other words, that represent the exact same characteristic. For instance, maximum temperatures and average temperatures of the same component. In case of their existence, those are deleted to reduce the probability of over-fitting the model. Once the sensor data is cleaned, the dataset from each equipment becomes composed by the time series data and the reading for that specific time.

Since the event that we are studying is the failure of the system, and since we had a report of failures available, it was important to understand if there is any common characteristic between the reported failures: if the failures occurred during the launch, or during the steady phase, of working. Most of the failures occurred in a steady-state, therefore the data that represented the system shut down and the beginning of the launch were eliminated. These data were identified taking into account the rotation speed and the energy production that was being sent to the grid at that very moment. We also resorted to other criteria to clean the data, based on environmental and physical characteristics of the system. Taking the example of the turbine, the manufacturers of this equipment wrote in their technical sheet some conditions that need to be strictly followed in order to preserve the health of the equipment. In a hydro turbine is the amount of water that flows in it while in a wind turbine is the cut-in wind speed. Bearing this in mind, every data that did not respect those conditions meant that the equipment was initializing the process, stopped, or finalizing. In this case, this type of data was irrelevant for the analysis of the event of interest, a failure. To sum all the criteria up, Figure 4.3, describes all methods that were used to clean the given data.

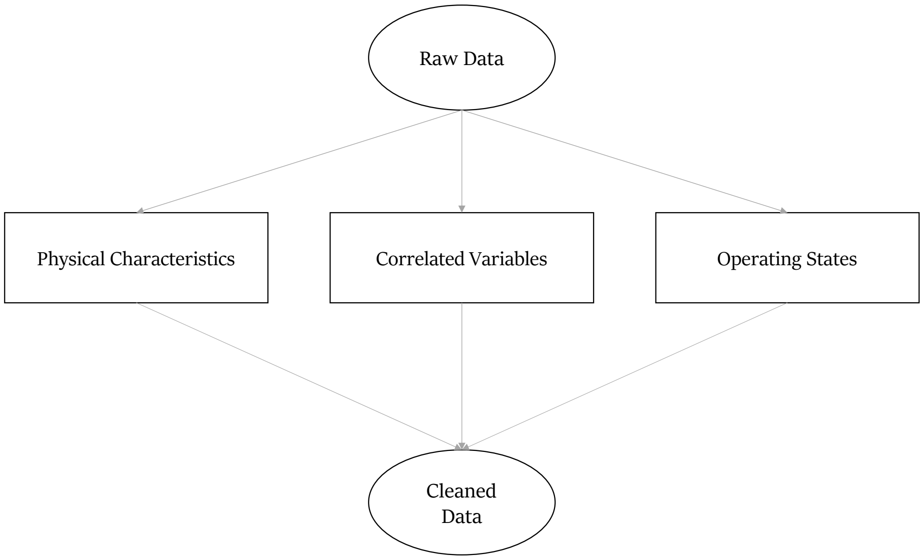


Figure 4.3: Cleaning Data Criteria

4.1.2 Degradation Curve

After cleaning the data, it was possible to construct a degradation model, representing the state of each measurement, i.e. describing the cumulative degradation of each measurement. It can also be seen as the amount of stress that each equipment is subjected to in that specific time.

In the section below, it will be explained every step that leads to the degradation and similarity models building, as well as the difficulties that were encountered in the development of an unsupervised model (UP) and finally the transformation of measurement into a HI, by fitting and training data. The shape of the curve that would transform the raw data into a degradation indicator was based on degradation models, mentioned in Section 2.3.2.2. We decided to apply it, because a considerable part of our methodology was based on accumulative stress and how we could represent and predict it. However, even with the available data from constant monitoring, it was not possible to use a degradation model to obtain the HI. This is a consequence of the degradation models being supervised.

A supervised model is a model that has information about the system beforehand. It can be based on similar systems, historical data, or even empirical data. This information makes it possible to develop conditions indicators, which can be specific sensor readings or groups of factors that represent the conditions of the system over time. Furthermore, we defined a condition threshold that once reached, the component is stopped and receive maintenance in the sub-system. The theoretical curve that can represent this behavior was represented in Section 2.3.2.2.

In the other hand, the model that was developed in this dissertation is an unsupervised model, which means that the only available data are the sensor readings and some manufacturer information about the equipment. As a result, we do not possess a well defined threshold nor conditions indicators to establish a limit in the degradation of components. Therefore, and based on the studied done by [41], presented in Section 2.3.2.2, it was assumed that the degradation of each M_c could be represented by an exponential shape function.

The exponential curve used in the development of the model is given by Equation (4.1) and represent the possibility of a failure to happen. As mentioned before, the model is not supervised so a run-to-failure analysis is not possible to do. Instead it is necessary to evaluate what the sensor reading data represent. Bearing in mind a component with cumulative degradation starts to present signs of malfunction, decrease its performance, and their working conditions (leakages, temperature, vibrations, etc.) are not the same. Therefore, a model that represented the performance deviation, as well as the variation of work condition, could represent the health state of the system. So the form of the curve which is represented by Equation (4.1) was based on the degradation curves aforementioned. However represents the state of the equipment if the reading of the sensor has a specific value.

$$y = ae^{bx} + c \tag{4.1}$$

Since the representation of a complete cycle of working, from installation to maintenance, was not available, it was not possible to draw a curve that would describe the relationship between the sensor readings with the HI. In that way some assumptions needed to be made in order to enable drawing such curve.

In order to understand how the sensor readings were organized, it were represented in a boxplot, such as the one in the Figure 4.4. This type of analysis provides a notion of how the values of the dataset under study are distributed. Boxplots are a standardized way of representing the data distribution summarized in five values [lower limit, first quartile (Q1), median, third quartile (Q3), and upper limit]. The median represents the value that is exactly in the middle of the dataset. The first and third quartiles respectively represent the value that is at 25 percent and 75 percent of the data. The upper and lower limits are the values that determine when a value is considered outlier or not. An outlier is an atypical value, in other words, it is a value that stands out from the rest of the series under analysis. It is possible to calculate this limit through the following equations:

$$LowerLimit = Q1 - 1.5 * IQR \tag{4.2}$$

$$UpperLimit = Q3 + 1.5 * IQR \tag{4.3}$$

$$IQR = Q3 - Q1 \tag{4.4}$$

where IQR, stands for interquartile range, and as it is possible to deduce through the name, the value represents the spreading level of the data, in torn of a point.

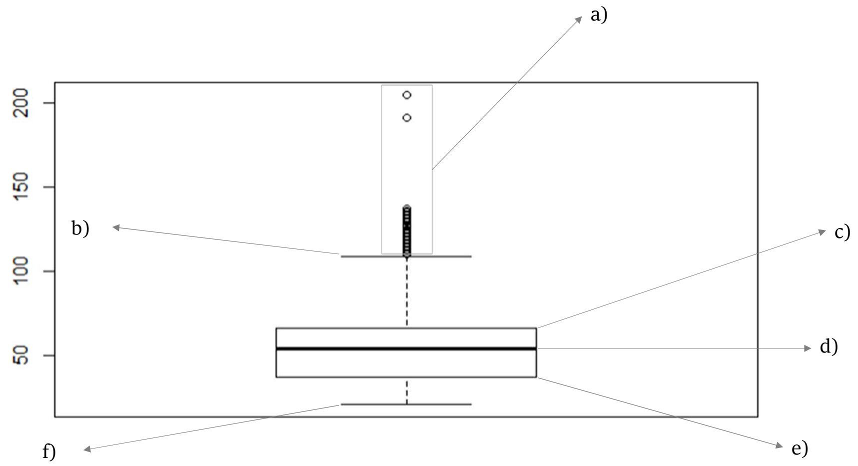


Figure 4.4: Extraction of features for developing the fitting curve. (a)-Outliers; (b)-Upper Limit; (c)- Third quartil; (d)- Median; (e)-First quartil; (f)- Lower limit

By analyzing the boxplot we can infer that there is a similarity between the outliers and the values that preceded a faulty behavior. Hence it was noticeable that some values were not desirable to reach. As a result, it was assumed that this group of reading would represent an imminent danger to the system. Knowing that, an HI was defined for them, which is represented by area A in figure 4.5. The B area represents the maintenance area, which means, that if the HI is on the interval of values, fairly urgent maintenance it is needed. The C zone represents the steady area of using, it presents some oscillations though not much pronounced. The D zone represents the equipment condition working on normal conditions and well-coordinated, and finally, the E zone incorporates the values when the machine is starting, which the sensor reading is not that high, with the exception of the starter, in that specific case the starting values need to be studied deeply.

Once the areas of HI are defined and there are some specific points that represent different areas of the dataset, it is necessary to do some assumptions to create a new dataset that can represent the prediction HI curve.

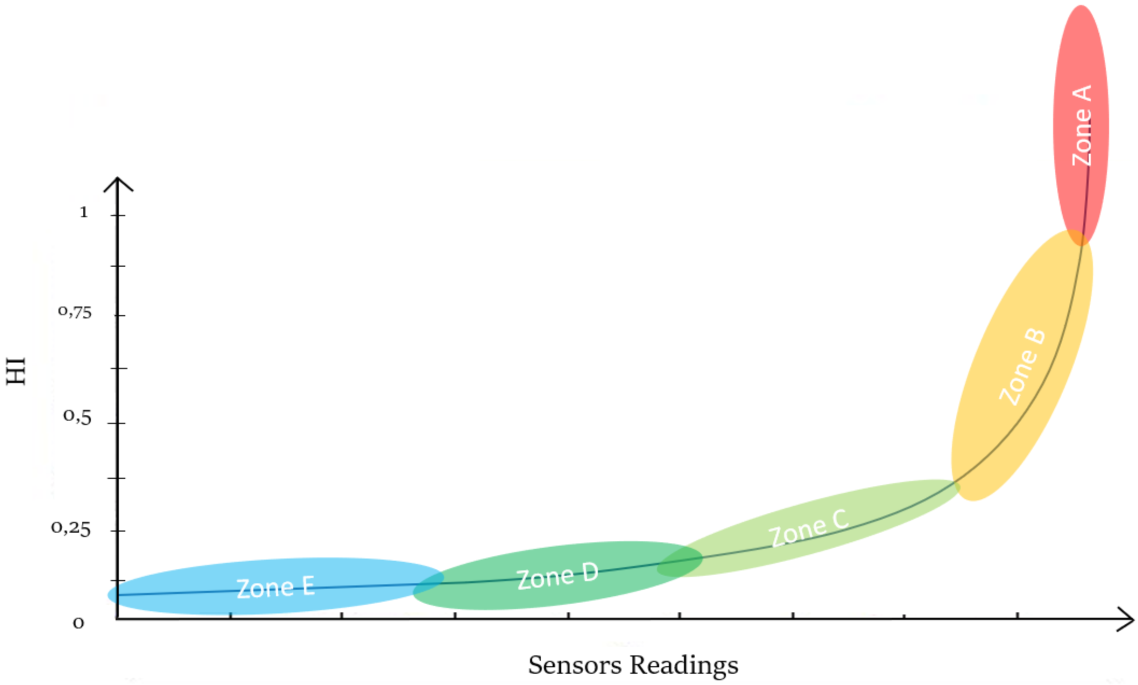


Figure 4.5: HI Area classification

Starting of with the E zone, as was mentioned before, is not a problematic zone for the system. Ergo, the ideal value to associate is the lower limit, which also represents the lower readings of the sensors. The intervals of HI that are associated with each zone are established and represented in Table 4.1.

The D zone and the C zone represent the operational zones, and these areas can be represented by the most common values in the dataset, which are the values that are between the first quartile and the third. Entering on the warning zone, B zone, the values are essencially distributed around the upper limit. Lastly, the danger zone in which there are not any left values in the boxplot to represent this area. Yet, it is an area that should never be reached, in that matter, it was assumed that the point that could represent this area is the sum of the Upper limit with the IQR, $A_{Zone}Point = UpperLimit + IQR$. Thereby, and having one point, at minimum, representing each specific zone, as it is possible to see in Figure 4.6, for instance, it is possible to fit an exponential curve like in

Table 4.1: Points assumption Zones

Assumption Points	**Boxplot points**	**HI area**	**HI score**
#1	LowerLimit	Zone E	0-0.049
#2	First Quartil	Zone D	0.05-0.24
#3	Median	Zone D	
#4	Third Quartil	Zone C	0.25-0.49
#5	UpperLimit	Zone B	0.5-0.89
#6	UpperLimit + IQR	Zone A	0.9-1

Figure 4.1 to represent the prediction HI curve.

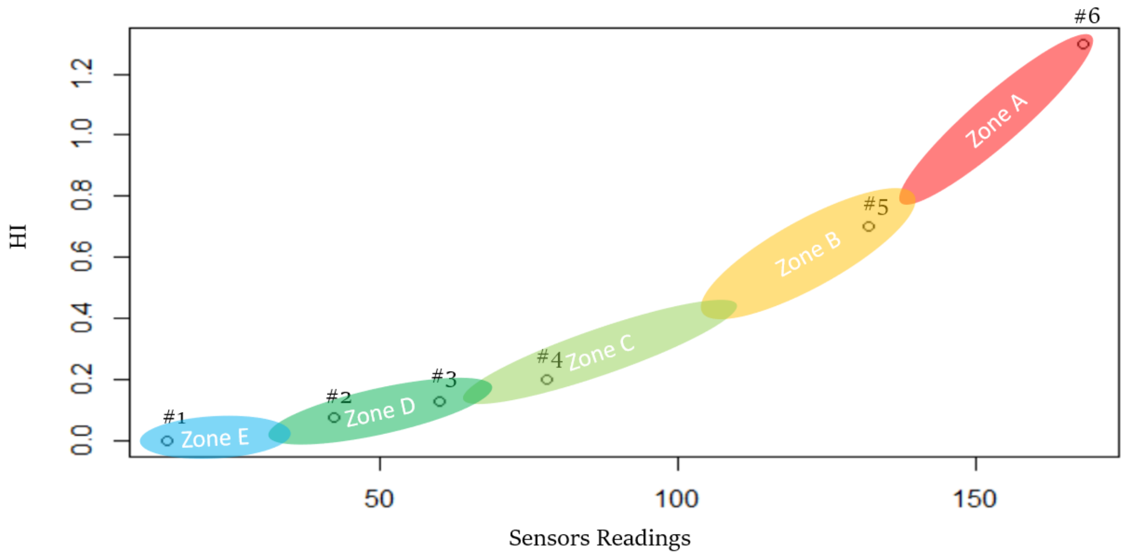


Figure 4.6: Assumptions points representation

4.1.2.1 Fitting Curve

Once we had the new dataset, where the y values are HI scores and the x values are the box plot points, it was possible to build a function that fits that data. It was tested different fitting curves, as we can see in Figure 4.7. Nevertheless, the exponential curve was the one presenting the best results.

The exponential curve is a non-linear function, which means that a non-linear model (NLM) should be used to fit the data. However, the NLM needs to be fed with an initial estimate to the iterative optimization routine, which is a hard task to be accurate. Despite being a non-linear function, it is still possible to use linearly squares to fit the parameter a, b, and c of the Equation (4.1).

Firstly, a new variable z is defined,

$$z = ln(y - c_0) \tag{4.5}$$

where,

$$c_0 = min(y) * 0.5 \tag{4.6}$$

According to [53], c_0 can be defined as half of the minimum value y. This assumption made to be able to calculate the linear parameters. This value will be used in the non-linear model as an

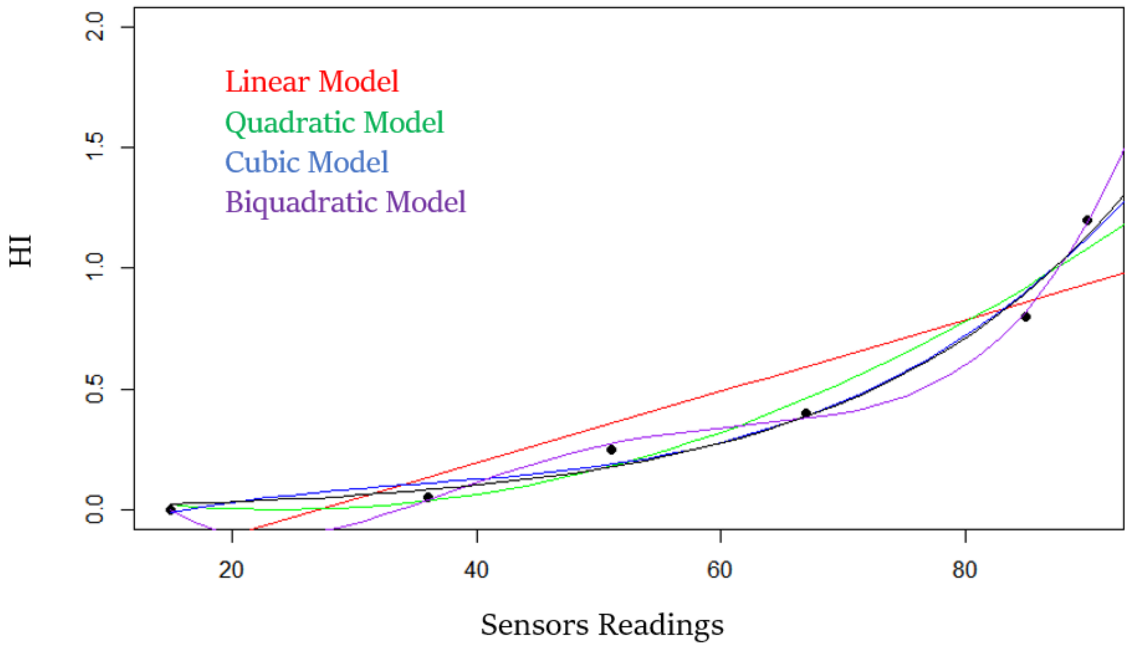


Figure 4.7: Linear regression test curves

initial estimate to calculate the parameter c of the Equation (4.1). Equation (4.5) can be written as:

$$\begin{aligned} z = ln(y - c_0) &= ln(ae^{bx}) \\ &= ln(a) + bx \end{aligned} \tag{4.7}$$

which is a linear equation, with slope b and the y-intercept is given by $ln(a)$. A linear model is fitted in the new variable z. The first coeficient,fc, represents the value of the y-intersection and the second coeficient, sc, the slope, which is represented by the Equations (4.8) and (4.9).

$$\begin{aligned} fc &= ln(a) \\ e^{fc} &= a \end{aligned} \tag{4.8}$$

$$sc = b \tag{4.9}$$

Once we calculate the parameters a and b, we have the initial estimates for the non-linear model fitting. As was mentioned before, the c parameter is initialized with the value of c_0. The NLM will modify the initial parameters that were given to them, with the objective of getting the best data representative curve. Afterwards, we get the optimized parameters, all the sensor readings are computed and mapped by this HI curve.

4.1.3 Fault prediction

The model explained above allowed the assignment of a number that represents the health status of a sensor reading. Despite that, the classification of the healthy state as a faulty state is not linear. Since this is not a supervised model, there is not a well define HI threshold that when it reaches, that value is automatically classified as a fault. Therefore, it were defined several values for a threshold, named Failure Threshold (FT), and tested in order to find the best fit. FD will be explained in more detail on Section 5.2. Bearing that in mind, the model that was created to predict and classify a state as failure can be divided into two phases, time-independent prediction and time-dependent prediction. They are considered in two different phases because one is an upgrade from the other.

4.1.3.1 Time independent prediction

Both of the models that are going to be presented are based on a degradation based model. This means that the threshold technique is used. Since this is a unsupervised models the Failure Threshold values that were tested were taken into account the HI that the system has in moments of failures.

This model is called time independent prediction because it does not take into consideration what were the previous state of the system. The state of the system is the predicted HI in a specific interval of time. In order to understand how it work's, it is necessary to understand how the fault identification is made.

In a first state, the data is divided into specific time intervals, dictated by a constant named step. The step indicates which are the size of data that we are going to consider once at a time. For instance, the system is monitored in intervals of 10 minutes during two years, this is a enormous quantity to process and manipulate. So a step is defined to divide the data into bigger groups, i.e, if the step is equal to one day, the raw data will be divided in groups of on day, which are easier to process. An example will be demonstrated for better understanding. The second state is to define another constant, that is called lag. Lag is the interval of time that the developed model need to identify the reported failure in order to be considerate as a good prediction. With the variation of lag, it is possible to study how reactive or predictive the developed model is.

Supposing that the failure report and the predictions were the following:

Table 4.2: Example of Failure Report

Asset	Component	Timestamp	Remarks
T6	Generator	20/12/2019 02:53:00	Temperature sensor failure
T6	Rotor	01/01/2020 06:37:00	High Temperature
T7	Hydraulic Group	12/06/2020 18:26:00	Error in pitch regulation

Table 4.3: Example of prediction data

Timestamp	HealthIndex	Fault Prediction	Failure
12/12/2019	0.2	FALSE	FALSE
13/12/2019	0.26	FALSE	FALSE
14/12/2019	0.35	FALSE	FALSE
15/12/2019	0.38	FALSE	FALSE
16/12/2019	0.48	FALSE	FALSE
17/12/2019	0.56	TRUE	FALSE
18/12/2019	0.60	TRUE	FALSE
19/12/2019	0.78	TRUE	FALSE
20/12/2019	0.3	FALSE	TRUE
21/12/2019	0.25	FALSE	FALSE

Tables 4.2 and 4.3 represent a hypothetical failure report and a hypothetical and very simplified prediction data. These are just an example to put in practice the concepts that were presented before. In the table 4.3 the failure threshold used was 0.5. In a first analysis we can see that the step is one day and the lag is 0, which means that the HI model has to identify only the failure in the day that it was registered. If the lag was for example two days the fault prediction would, already, be able to identify it. In the case that is represented by the tables, there is one failure that is not identified, because none of the values of the Fault Prediction and the Failure are correspondent.

The HI that is represented in each day, represent the maximum value that was registered in that same day. And that is the only criteria used in this model, they do not take into account the previous HI to do the classification of the fault. That is noticed on the 20^{th} of december, when it has a Health index of 0.3 but a failure is reported, which could lead to wrongly predictions. The dependence on the previous values is going to be described in the following section.

4.1.3.2 Time dependent prediction

A time-dependent prediction has into account the results of previous iterations, to have a better prediction. That being said, this type of prediction can be considered as an upgrade from the prediction model presented in the previous section.

This upgrade is based on the use of neural networks to be able to make a prediction of the failure state or not, taking into account the previous states. These model would start by choosing the best type of ANN that fit the pretended results. Which, according to [57], would be recurrent neural networks, more specifically, Long Short-Term Memory or Gated Recurrent Unit. The responsible model for the built of each measurement HI was unsupervised, however, once we got these values, it is possible to construct a neural network around them, considering a supervised system. This way, the neural network will learn from the HI, and from the fault classification that was made in Section 4.1.3.1, as in combination with the failures that were reported. This learning through training is also called error back-propagation algorithm. This algorithm trains the NN based on the input-output samples and finds error signals, which normally is the difference between the desired output and the obtained output. Based on this, it is possible to divide error back propagation learning into two steps:

- **Forward Pass**: The input vector, which in this case is the fault classification and the values of HI, propagates forward in the neuron network. And appears a signal output defined by: $y(n)\phi(v(n))$, where $v(n)$ is the induced local field of a neuron defined by $v(n) = \sum w(n)y(n)$, according to [57] and [46]. The output that is calculated is compared with the desired output, $d(n)$, and creates the error for that neuron, $e(n)$.

- **Backward Pass**: The error signal that was calculated before, is propagated backward through the network. Which calculates the local gradient for each neuron that allows the synaptic weights of the network to change in conformity with the delta rule, $\Delta w(n) = \alpha * (d(n) - y(t)) * x(n)$, where α is a constant called learning rate, and $x(n)$ the input.

This way, the inputs of the NN would be the HI, as well as their classification, as a failure or not. All the parameters that will be used are described on Table 4.4.

Table 4.4: Neural Networks Parameters

Parameters	**Values**	**Ranges**
Input	HI;HI classification	0-1;FALSE/TRUE
Output	HI classification	FALSE/TRUE
Loockback	Intervale of values to get based on to predict the future	0-N

where N represent the length of the dataset. This parameters are still on testing, and can suffer changes. There are more parameters that need to be define on the NN like the number of epochs, the batch size, learning rate. Despite having an important impact on the performance of the NN, their associations with predicting the next failure are not as direct as the *lockback*, which represents the interval of previous iterations used to predict the following one. It is possible to change this parameter, however, the higher the *loocback* value, the higher computational power is needed.

However, using a NN with the dataset available may lead to an issue. Has was mentioned, the NN would be trained based on failure events, though, the number of failure events compared with non-failure events is insignificant. Which, in machine learning language is called an unbalanced dataset. This makes it difficult to train the NN to detect patterns on the data. Therefore, techniques as under-sampling, which is the reduction of the abundant class, with the objective of balance will be used. Or use clusters in the abundant class, in other words, the abundant class would be clustered in r groups. Thereafter, the medoid, which is the center of each cluster, would be taken and finally, the model is trained with the rare class, which in this specific case is the failure events, and with the medoids only.

4.2 Agglomeration Model

Since the HPP is composed by many systems, and sub-systems, that in turn are represented by multipe measurments, there are many types. So, when the aim is to represent the health status with a single value, it is necessary to merge the data. As was mentioned in Section 2.4.2, the MMI technique was the method that this agglomeration model was based on. Throughout this section, the method that was implemented will be described.

According to Chapter 2, this model is driven by the strongest h_{ic}, which in health evaluations systems is important. However, it is also important to keep in mind the impact that the others h_{ic} can have on the system. This is because, once a component is failing on one element it is forcing others, in that way, their degradation can increase drastically. In conclusion, all the biggest h_{ic} has the biggest importance but we can not bleach the impact of the others.

That being said, once the individual HI are calculated, they are added to a dataset, a column for each individual HI. First things first, it is defined the CT, which was mentioned in Chapter 2, and the warning threshold (WT). As the name indicates, this threshold is in the verge between an maintenance warning intervention or not. It is after this threshold that the value starts to have negative impact on the final result of the aggregated HI.

After that, each HI from a specific time series is compared with the critical threshold. This comparison will determine which algorithm will be used. If there is any individual HI that is equal or bigger than the critical threshold, unsteady state algorithm is used otherwise it is used the steady state algorithm.

4.2.1 Unsteady State Algorithm

As was mentioned before, this algorithm is used when an individual HI is higher than the critical threshold. In that sense, it is needed to discover which one it is. For that, it is applied a function that gives us the greater value between the individual HI. This value is saved to have into consideration later, and it called Maximum Value (MV).

$$MV = max(\vec{h_{ic}}) \tag{4.10}$$

where N is the number of measurements that the system has.

After that, a variable called Factor Divider is calculated. FD is the number of all the particular HI that are bigger than the CT. This means, that represent the number of HI that need to be still considered as important as the greatest HI, and possibly dangerous for the system because they also represent a really bad equipment health condition. This variable has an inverse scale, which means that the more there are high HI, the least its value will be. The maximum value that can reach is N, and the least is $zero(0)$, meaning that every individual HI surpass the critical threshold.

$$FD = N - q \tag{4.11}$$

where q correspond to the number of h_{ic} that surpassed the CT. Once we got the FD, the α and β are calculated. Those variables quantify the impact that the rest of the h_{ic}, once again, except the biggest HI, has in the classification. α has a range between 0-1, and the higher the number, the least is the impact on the composite HI. When it comes to β, this relates α with the FD. when the value of α is low means that there are many $h_{ic} > CT$.

In order to be have impact on the final HI, the h_{ic} need to be bigger than the WT. That information it is kept on a vector, $\vec{u}_i$, which will be used afterwards. There after, α is calculated, Equation (4.12).

$$\alpha = ((h_{M_1c}) * u_1) - CT) + ((h_{M_2c}) * u_2) - CT) + ... + ((h_{M_Nc}) * u_N) - CT) \tag{4.12}$$

Thereafter, it is possible to calculate β, which is represented on Equation 4.13.

$$\beta = \frac{\alpha}{FD} \tag{4.13}$$

Finally, the composite HI, 4.14 is calculated by adding the β to the MV, which represent the maximum value of the individual HI, as was aforementioned.

$$Composite\_HI = MV + \beta \tag{4.14}$$

4.2.2 Steady State Algorithm

In case of not having any individual HI higher than the CT, Steady State Algorithm is the chosen one to agglomerate the data. In contrast with Unsteady State Algorithm, the minimum value of the individual HI is held, but this time on mV.

$$mV = min(\vec{h_{ic}}) \tag{4.15}$$

After that, the maximum value is assigned to the θ. And similarly, with Unsteady Algorithm, ϕ represents the impact that the others HI will have in the final composite HI.

$$\theta = max(\vec{h_{ic}}) \tag{4.16}$$

$$\phi = \frac{\theta - CT}{FD} \tag{4.17}$$

And similarly to the first algorithm, the composite HI was represented by the sum of mV with ϕ, however in this algorithm it is necessary to verify if the result is a positive number. Because of that, the sum of mV with ϕ is saved in an auxiliary HI variable, HI_{aux}. In case of being negative,

HI_{aux} is rounded to zero, Equation (4.19). Thereon, the mean of the $\vec{h_{ic}}$ is calculated, Equation (4.20), and summed to HI_{aux}, to create the composite HI, Equation (4.21).

$$HI_{aux} = mV + \phi \tag{4.18}$$

$$\begin{aligned} &if(HI_{aux} < 0) \\ &\implies HI_{aux} = 0 \end{aligned} \tag{4.19}$$

$$\overline{h_{ic}} = \frac{\sum_{i=1}^{n} \vec{h_{ic}}}{n} \tag{4.20}$$

$$Composite\_HI = \frac{\overline{h_{ic}} + HI_{aux}}{2} \tag{4.21}$$

Summarizing, the used agglomeration model had two different approaches depending on the values of the individual HI. Giving both at the end the value that represents the HI of the system, as is described in figure 4.8.

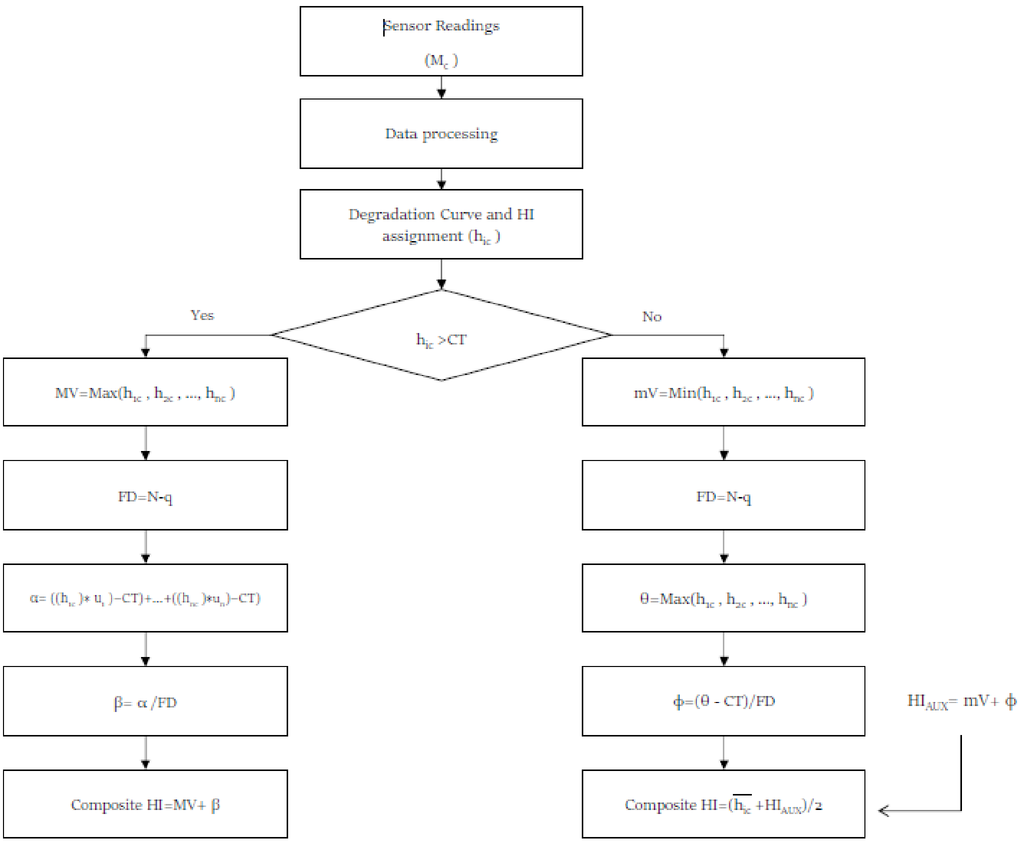


Figure 4.8: Agglomeration Methodology flowchart

Chapter 5

Computational Tests

To improve the model is necessary to understand which parts are not going to be able to achieve the intended result. In that way, model evaluation is very important to its own development. As a result, this chapter will be divided into four sections. The first Section will describe data that was used to develop the model, as well as the way that it was organized and obtained. In the second Section will be presented techniques that were used to evaluate the model. Thereafter, results of the model training will be detailed in Section 5.3. And finally, the model test results will be presented in Section 5.4.

All training and testing used time independent prediction because it was not possible to upgrade this model to a time dependent prediction.

5.1 Data Description

As mentioned in Section 3.2.1, the data that initially would be used to build the methodology above described, were not available, due to COVID-19. As a result, a similar dataset from EDP Open Data was used,[1]. Despite the used dataset belongs to a wind power plant, the architecture was very similar to the HPP, especially the sub-system of electrical machinery that will have a deeper analysis. The wind power plant architecture is presented in Figure 5.1, and as mentioned before, it is very similar to the one that was presented in Section 3.2.2.

The dataset available did not have all types of data that would be available from different tests, but only sensor readings were available in this dataset. There were taken eighty-one different measurements from five different turbines over the years 2016 and 2017. These readings were made every ten minutes, as it is possible to see in Figure 5.2.

Each sensor was assigned, by manufacturer, to a specific sub-system, represented in 5.1.

A failure report was also available, which contained the time that failure was reported, the turbine that failure occurred, in which component and what was the problem. In order to have a better system understanding, a wind turbine characteristics sheet was provided. In this paper, were described some values that system should not reach to function with the best yield possible and safer. Characteristics such as cut-in wind speed, cut-out wind speed, maximum rotor speed, were

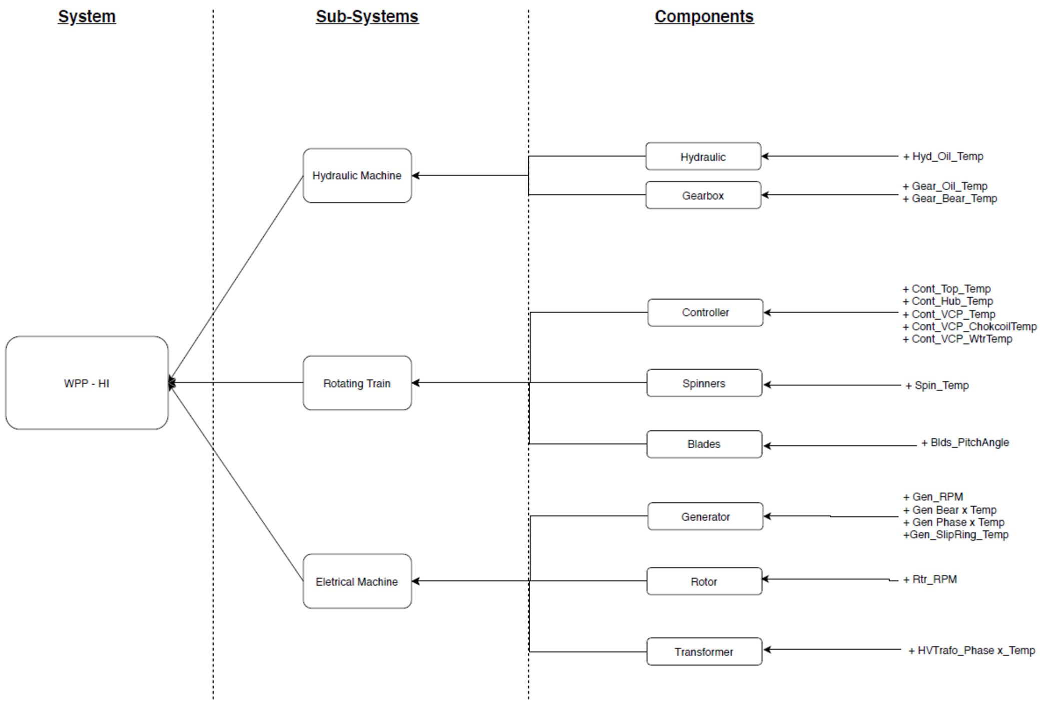


Figure 5.1: Wind-Power station architecture

some examples described. This information is very useful in the pre-processing because can give in formation about system boundaries.

In the next section will be presented an illustrative example of the construction of a composite HI to clarify some concepts mentioned before.

5.1.1 Illustrative Example

This section will demonstrate an example from raw data reception until composite HI building.

5.1.1.1 Data cleaning and pre-processing

Taking into account the enormous amount of data that was available, some cleaning and pre-processing techniques were used. In the first state, each sensor reading was divided into a new dataset based on their component. With that, it was possible to eliminate data that was not related to any interest component. Some of those are represented in Table 5.1. Were reduced from 81 variables to 34.

Turbine_ID	Timestamp	Gen_RPM_Max	Gen_RPM_Min	Gen_RPM_Avg	Gen_RPM_Std	Gen_Bear_Temp_Avg	Gen_Phase1_Temp_Avg	Gen_Phase2_Temp_Avg	Gen_Phase3_Temp_Avg	Hyd_Oil_Temp_Avg	Gear_Oil_Temp_Avg
T01	01/01/2016 00:00	1277,4	1226,1	1249	9	41	58	59	58	30	44
T01	01/01/2016 00:10	1268,3	93,3	999,7	435,9	41	58	59	59	30	44
T01	01/01/2016 00:20	1394,4	220,3	774	486,1	41	57	57	57	30	43
T01	01/01/2016 00:30	1306,6	1212,4	1257,1	17	40	56	57	57	30	44
T01	01/01/2016 00:40	1320,9	1224,1	1257,7	18	40	57	58	57	30	44
T01	01/01/2016 00:50	1462,6	1249,5	1328,5	48,3	40	58	59	58	30	45
T01	01/01/2016 01:00	1350	1228	1265,5	22,5	40	59	60	59	30	45
T01	01/01/2016 01:10	1335,2	1225,1	1268,7	18,6	40	59	60	60	30	45
T01	01/01/2016 01:20	1370,7	1230,8	1279,4	25,6	40	60	61	60	30	45
T01	01/01/2016 01:30	1375,7	1212,2	1272	33,5	41	61	61	61	30	46
T01	01/01/2016 01:40	1317,8	1211,7	1248,9	12	41	60	61	60	30	45
T01	01/01/2016 01:50	1287,3	1230,2	1251,3	9,7	41	60	61	61	30	45
T01	01/01/2016 02:00	1281,9	1218,7	1251,4	9,7	41	61	62	61	30	45

Figure 5.2: Excerpt of available data

Table 5.1: Unrelated Data

Descriptor	Description
Amb_WindDir_Abs_Avg [^{o}]	Average wind absolute direction
Blds_PitchAngle_Avg [^{o}]	Average angle
Cont_VCP_ChokcoilTemp_Avg [^{o}C]	Average temperature in the choke coils on the VCS-section
Grd_Prod_CosPhi_Avg	Average actual phase displacement
Grd_Prod_PsbleInd_Avg [kVAr]	Average possible inductive reactive power

Thereafter, those datasets were divided into two groups. The training data and the testing data. The training data was composed by data from four turbines (1,6,9,11), while the test data was from turbine 7. The data was distributed, based on the number of failures that existed. Turbine 6 and 7 were the ones with more failures, so the model will be trained with the turbine 6 data and tested on turbine 7 data. After organizing the data, it passes through a cleaning process based on the characteristics sheet given by the manufacturer [19]. This characteristics sheet has the interval of values that some specific measurements must have in order to have a healthy system. For instance, the system can not work at a certain wind speed and that speed was defined in the paper. Other criteria used to filter data was the power curve. This curve relates to the power (kW) that is produced and sent to the grid with the wind velocity. It is represented in Figure 5.3.

By analyzing the curve, it is possible to notice that production below $80kW$ are discarded, as a result, all data that were associated with production in that value order were eliminated. Lastly, it was made an inspection throughout all dataset in order to find any corrupted or incomplete data. Initially, the dataset had 80000 elements and was reduced to 55000.

Once the data is ready to be manipulated, it is created the degradation curve and consequently its HI, which will be presented in the next section.

5.1.1.2 Degradation Curve

As was mention before, the wind power plant sub-system that was more similar to the HPP was the Electrical Machinery, as a result, that was the one who got more attention, especially the generator component.

As it is possible to observe in Figure 5.1, the generator health status is represented by several measurements. That is the type of component that needs the fusion of all individual indexes to obtain is own HI. Based on the methodology described in Section 4.1.2, a boxplot graphic was

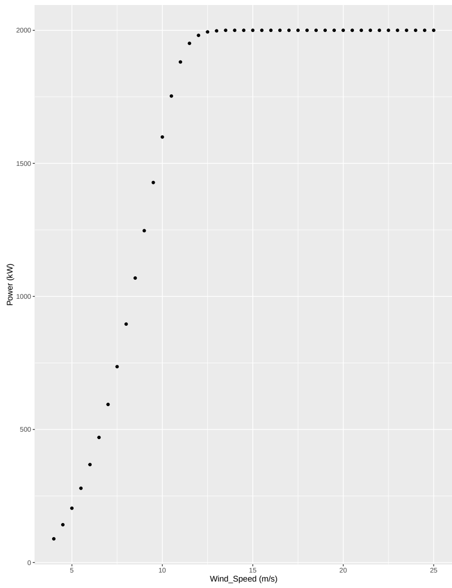


Figure 5.3: Power curve

applied to each generator measurements, in order to understand which were the boundaries. The Figure 5.4 represents that distribution.

As it is possible to see in Figure 5.4, the measurement phase is divided into three. Each one of them will be treated as a individual measurements and afterwards they will be aggregated using the method described in Section 4.2. After analysed the data distribution and taken the representative data points, described in Table 5.3, the HI assumptions were made, based on Table 4.1 and the areas represented in Figure 4.5. The results were:

Table 5.2: Generator boxplot's references points

Reference Points	**BearPhase_Temp (^{o}C)**	**Phase1_Temp (^{o}C)**	**Phase2_Temp (^{o}C)**	**Phase3_Temp (^{o}C)**	**SlipRing_Temp (^{o}C)**
Lower Limit	12	14	14	14	10
1^{o}Quartil (Q1)	35	42	42	42	24
Median	45	60	60	60	29
3^{o}Quartil (Q3)	58	77	76	76	35
Upper Limit	92	129	127	127	51
Outliers	115	164	161	161	62

Table 5.3: HI assumptions results

Reference Points	**HI Assumptions**
Lower Whisker	0.0002
1^{o}Quartil (Q1)	0.077
Median	0.13
3^{o}Quartil (Q3)	0.2
Upper Whisker	0.7
Outliers	1.3

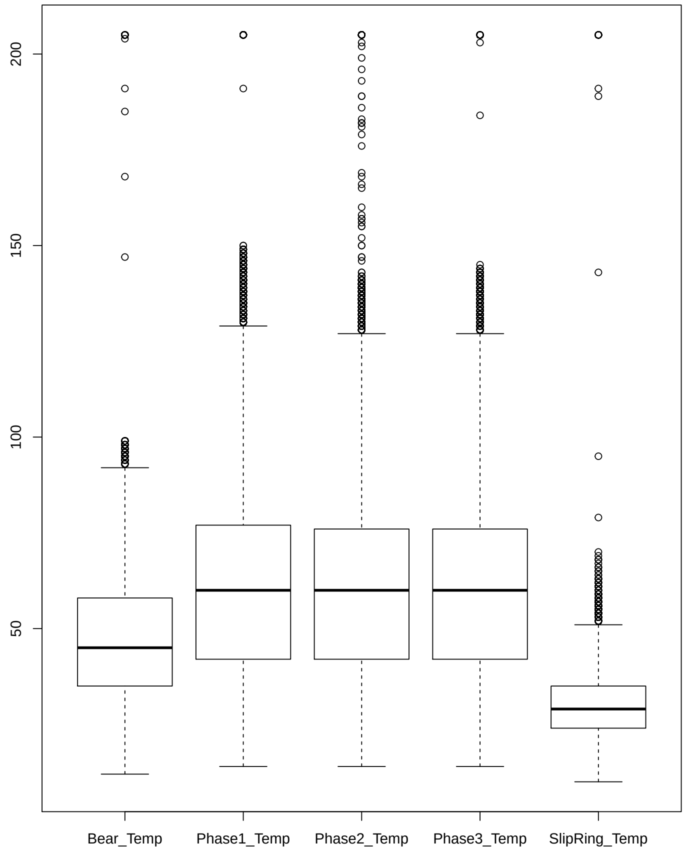


Figure 5.4: Generator data distribution

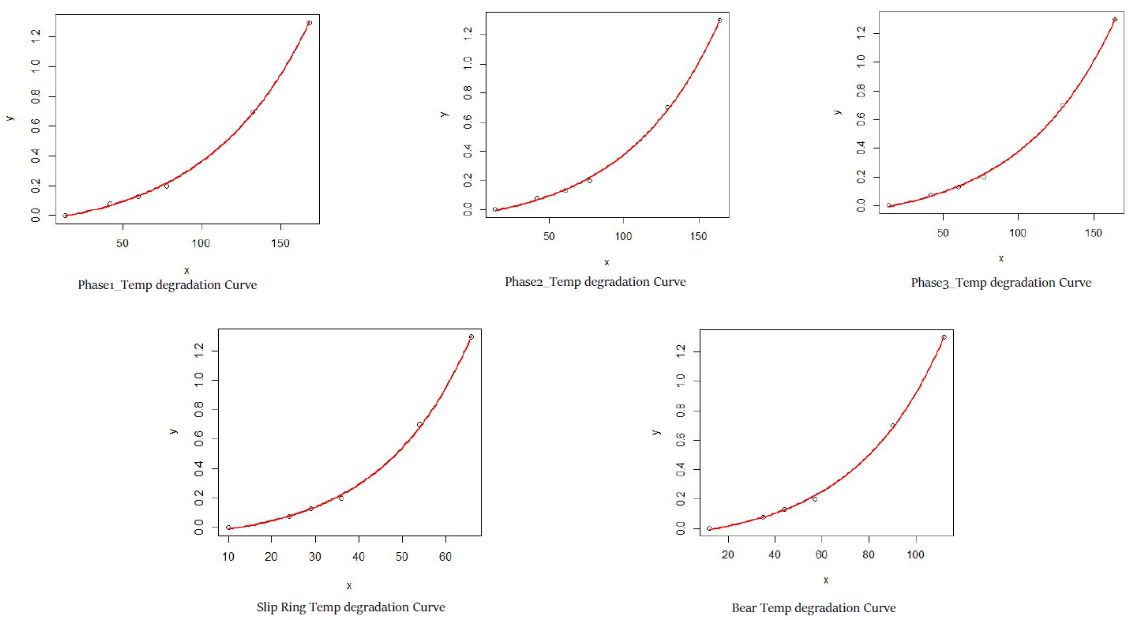


Figure 5.5: Generator's degradation curves

where the $y - axis$ represents the value of the HI and the $x - axis$ the respective measurement read by the sensor. The presented data correspond to turbine 6 analysis because there were more failures to train the model. Nonetheless, this methodology was also tested on the other turbines.

The HI assumptions were not random. To obtained the values that are presented in Table 5.3, a cycle of several values was made, and to encounter the best fit, the p-value of the fitting curve was the evaluation criteria. The algorithm of the cycle is presented in Figure 5.6.

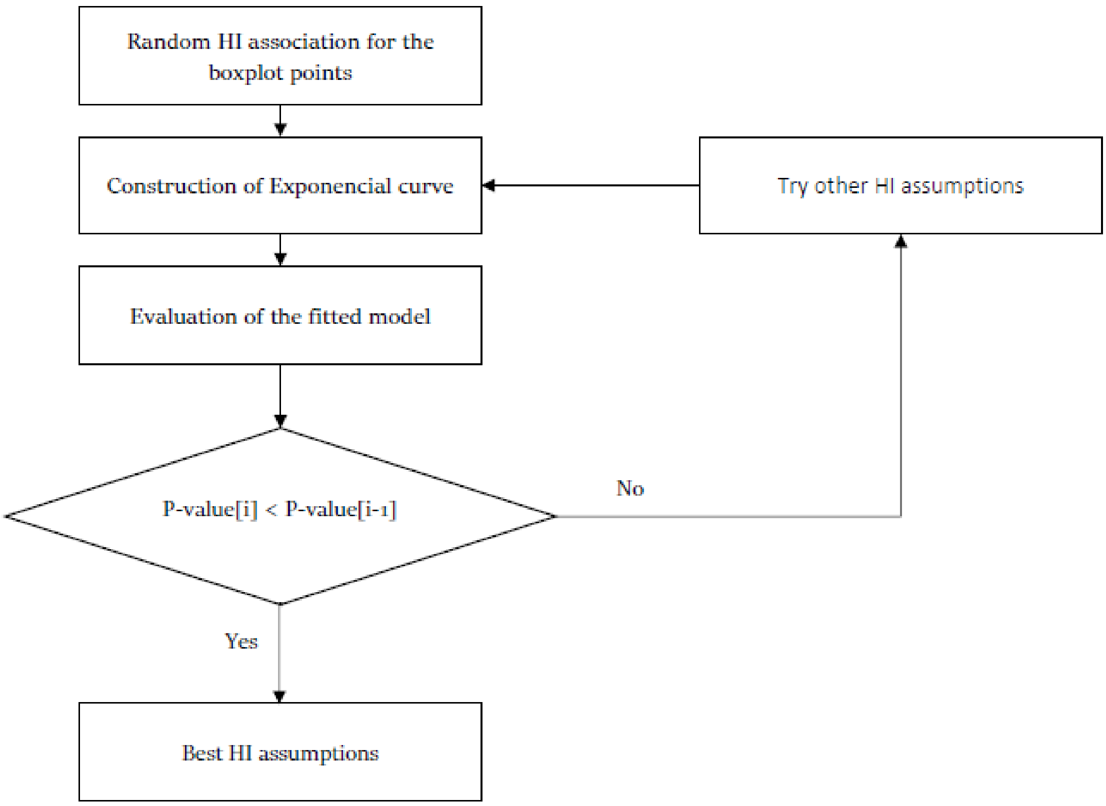


Figure 5.6: HI attribution points flowchart

At this point, all data necessary to represent the generator HI is available. It is only needed to aggregate that data into one single value. That aggregation will be presented in the next section.

5.1.1.3 Aggregation

When theoretical degradation curves for each measurement were defined, it is possible to calculate the respective HI. To do that, each reading needs to passe through the function that is represented by Equation (4.1), but this time with the parameter founded previously in the linear fitting.

As mentioned before, the phase temperature measurement is composed by three individual measurements, and to obtain the overall HI, it is needed to agglomerate each one. Based on the methodology described in Section 4.2, it was necessary to define some aggregation thresholds, which are the CT and the WT. In order to obtain the best combination of these two, several tests were made, which will be presented later. For the sake of the example, will be considered $CT = 0.8$ and $WT = 0.5$. By analyzing Table 5.4, it is possible to understand that we have two different situations.

Table 5.4: HI Phases-Turbine 6

Turbine ID	TimeStamp	Phase1_Temp	Phase2_Temp	Phase3_Temp	HI_Phase1	HI_Phase2	HI_Phase3
T06	2016-01-01 00:00:00	51	52	53	0.09811076	0.10194194	0.10676738
T06	2016-01-01 00:10:00	51	52	52	0.09811076	0.10194194	0.10292417
T06	2016-01-01 00:20:00	50	51	51	0.09451977	0.09816459	0.09914241
T06	2016-01-01 00:30:00	49	50	50	0.09098431	0.09444774	0.09542109
T06	2016-01-01 00:40:00	50	50	51	0.09451977	0.09444774	0.09914241
...							
T06	2016-08-22 15:00:00	136	138	122	0.7392656	0.8116141	0.5972773
T06	2016-08-22 15:10:00	136	138	122	0.7392656	0.8116141	0.5972773
T06	2016-08-22 15:20:00	136	138	122	0.7392656	0.8116141	0.5972773
T06	2016-08-22 15:30:00	136	139	122	0.7392656	0.8116141	0.5972773
T06	2016-08-22 15:40:00	136	139	122	0.7392656	0.8270044	0.5972773

Starting by considering the observation made in $2016-01-01$ at $00:00:00$, the CT is higher than every HI phase, which indicates that the agglomeration HI that should be used is the Steady State Algorithm. The obtained results are:

- $mV \quad = min(\vec{h_{ic}})$
  $= 0.09811076$
- $\theta \quad = max(\vec{h_{ic}})$
  $= 0.10676738$
- $\phi \quad = \dfrac{\theta - CT}{FD}$
  $= -0.2310775$
- $HI_{aux} \quad = mV + \phi$
  $= -0.13296674$

Since the HI_{aux} is negative, it is rounded to 0, and, as describes in 4.21, the composite HI is calculated by the mean between the individual HI and the HI_{aux}.

- $if(HI\_Phase1[i]) < 0$
  $\longrightarrow HI\_Phase1[i] = 0$
  $if(HI\_Phase2[i]) < 0$
  $\longrightarrow HI\_Phase2[i] = 0$
  $if(HI\_Phase3[i]) < 0$
  $\longrightarrow HI\_Phase3[i] = 0$

- $\overline{HI\_Phases} = \frac{\sum_{i=1}^{n} HI\_\vec{P}hases}{N}$
$= 0.09811076$

- $CompositeHI = \frac{\overline{HI\_Phases} + HI_{aux}}{2}$
$= 0.04905538$

On the other hand, the observation registered on $2016-08-22$ at $15:00:00$, has a value bigger than CT. This means that the aggregation algorithm that was chosen was the Unsteady State Algorithm. It is also possible to notice that both of the remaining HI have higher values than the warning threshold, which means that their value will have a considerable impact on the final HI.

- MV $= max(h_{\vec{Phase_{ic}}})$
$= 0.8116141$

- $FD = N - q$
$= 3 - 1$
$= 2$

Thereafter, to be able to calculate α, it was needed to identify the HI that was bigger than the warning threshold, which is the $\vec{u}_i$, presented in Section 4.2. Based on Equation (4.12):

- $\alpha = ((HI\_Phase1 * u_1) - CT) + ((HI\_Phase2 * u_2) - CT) +$
$((HI\_Phase3 * u_3) - CT) + ((FD + 1) * CT)$
$= 1.336543$

Finally it is possible to calculate $Var3$ and lastly, the $compositeHI$.

- $\beta = \frac{\alpha}{FD}$
$= 0.6682715$

- $Composite\_HI = MV + \beta$
$= 0.8116141 + 0.6682715$
$= 1.4798856$

As it is possible to see, the value of composite HI is bigger than 1, due to the importance that this aggregated approach gives to the biggest values. The biggest HI is 0.8116141, on the analyzed reading, which is a value that represents a big probability of failure. On top of that, all others were higher than warning thresholds, which leads to a drastic increase of the composite HI.

Finally, all the HI necessary to calculate the generator HI are calculated. The steps to construct the generator HI are similar to those that were described before, in that way, the final results are presented in Table 5.5. Figure 5.7 represents the relation between the individual HI and the final one, over a specific time interval.

Table 5.5: Composite HI Results - Turbine 6

Turbine ID	TimeStamp	HI_BearTemp_Temp	HI_SlipRing_Temp	HI_PhaseTemp	Composite HI
T06	2016-01-01 00:00:00	0.11459326	0.09744818	0.04905538	0.02452769
T06	2016-01-01 00:10:00	0.11459326	0.09744818	0.04905538	0.02452769
T06	2016-01-01 00:20:00	0.11459326	0.08793098	0.04725988	0.02362994
T06	2016-01-01 00:30:00	0.11459326	0.08793098	0.04549215	0.02274608
T06	2016-01-01 00:40:00	0.10876002	0.09744818	0.04725988	0.02362994
		...			
T06	2016-08-22 15:00:00	0.7950617	0.6814006	1.109576	1.400063
T06	2016-08-22 15:10:00	0.7715824	0.6814006	1.109576	1.388323
T06	2016-08-22 15:20:00	0.7487118	0.6814006	1.109576	1.376888
T06	2016-08-22 15:30:00	0.7487118	0.6814006	1.109576	1.376888
T06	2016-08-22 15:40:00	0.7264344	0.6814006	1.117271	1.371520

5.2 Model Evaluation

In this section, it will be presented the criteria used for considering an HI a failure, an example to explain the definition of lag and the method adopted to evaluate the performance of the developed model.

In the developed methodology, there are three variants thresholds. The critical threshold, the warning threshold and the failure thresholds. The first two control the impact and the importance of the individual HI on the aggregation method. Whereas the FT is responsible to classify an observation as a fault or not, based on their HI. Firstly, it is important to understand how the dataset is organized. As mentioned before in Section 4.1.3.1 the step indicates in which size of groups the data will be organized. In other words, data is initially organized in intervals of 10 minutes, the step controls that interval. If $step = 1\_day$, the data will be organized and divided in intervals of 1 day, and the HI that represents that interval will be the maximum value of the several observations in that group. Figure 5.8, represent that transformation.

After that, each HI that represents the interval defined by the step is compared to the FT, and if is higher, that observation is considered as a fault, else it is not. In order to obtain the most suitable FT, several values were tested and will be demonstrated in Section 5.3.

Another concept that was mentioned before but it will be explained with more detail is lag. According to Section 4.1.3.1, lag is the interval of time that the developed model needs to identify the reported failure to be considerate as a good prediction. When a failure happens, it is registered with a specific date. However, the process of degradation that leads to that date starts to happen before the report. In that way, it is defined by us the interval of time that we considered that the system started to degrade. It is at that time interval that our model needs to identify a failure to be considered as a good classification. This concept will be clarified further ahead with an example. There are 4 possible results when we are doing a state prediction, represented by Figure 5.9.

- True Positives (TP) - Represent the number of times that the HI model predicted correctly a failure state;
- True Negative (TN) - Represent the number of times that the HI model predicted correctly a non-failure state;

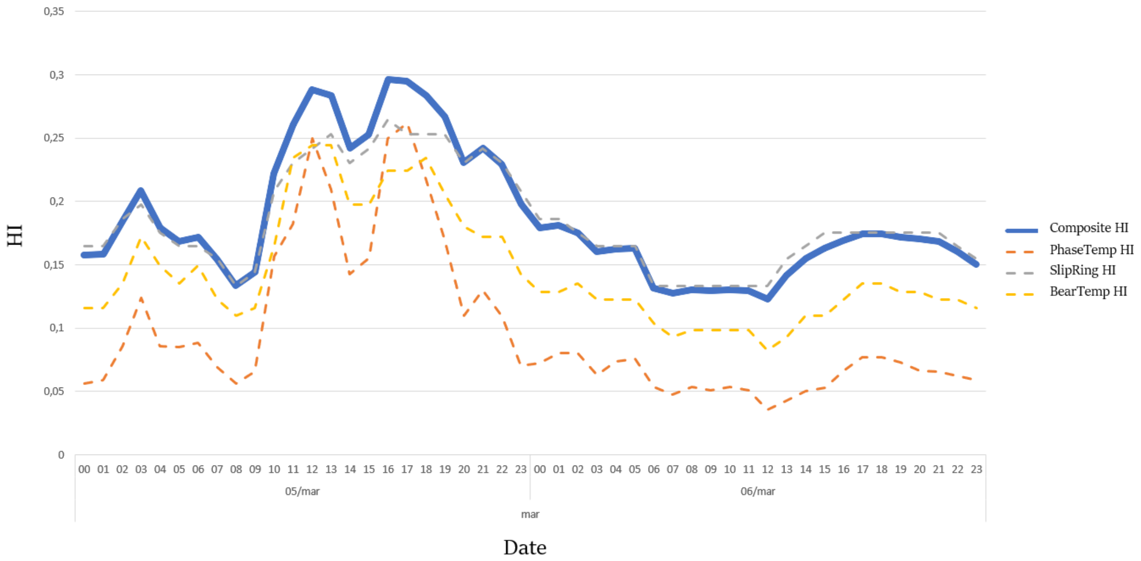


Figure 5.7: Composite HI evolution

- False Negative (FN) - Represent the number of times that the HI model predicted wrongly a non-failure state, in other words, it did not predict the failure;

- False Positive (FP) - Represent the number of times that the HI model predicted wrongly a failure state, in other words, predict that failure was happening when it was not;

In Figure 5.9 is represented two datasets that are compared in order to evaluate the model. On the top, it is the vector that has represented dates that the failures were reported. Once a failure is reported, observations before are also considered as in a failure state, according to the defined lag. On the bottom, it is presented HI's vector. The color classification has the intention to demonstrate which were the values considered as a fault or not. In other words, which were the values that $h_{ic} > FT$. The green color indicates that the h_{ic} is lower than FT and the red color higher. When the model predicts a good health condition and the system did not have any failure, it is considered a true negative. This is possible to observe on index 1 of the example. Following the same line of reasoning, when a failure is reported and the HI predicts that observation is an unsteady state, it is considered as a true negative. Described by index 13, 14, and 15. However, not every prediction is correct. On index 4, the developed model detected an unsteady state despite not being registered any failure. That is defined as a false positive. Nonetheless, that is not the most worrying misclassification. The most dangerous misclassification happens when a failure is reported and the model did not predict any bad health condition. This result is represented by indexes numbers 7 to 10 and it is called a false positive. That is the method used to evaluate the prediction model. Theoretically, each reported failure should only have one true positive associated with them, however as seen in Figure 5.9, that is not true. To understand the precision of

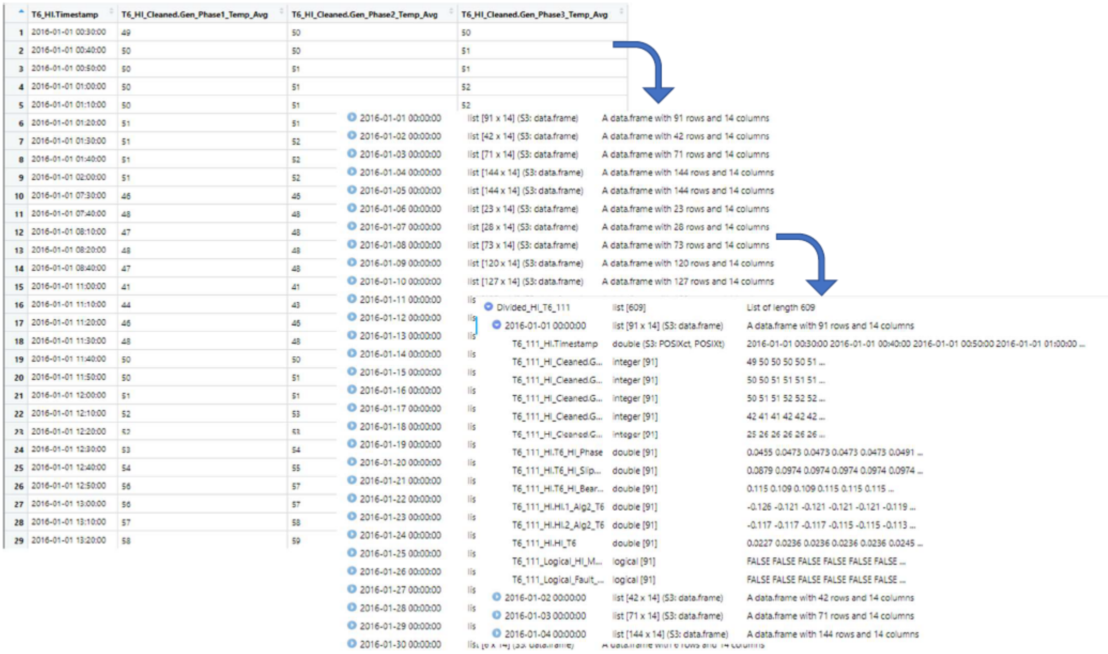


Figure 5.8: Data aggregation according to step - $Step = 1\ day$

the system, besides the before mentioned classification, the number of reported failures that were predicted is also used.

When observations number are low, it is possible to interpret the results, but when the dataset and the number of observations is high, like in this example, ends up being almost impossible. As a result, data is organized on confusion matrices, which is a table layout to represent this type of data, Figure 5.6. Each row represents the instances in a predictive failure and each column the real failure.

Table 5.6: Confusion Matrix Layout

		Actual Failure	
		F	**T**
Predictive Failure	**F**	TN	FN
	T	FP	TP

Once the data is organized on the confusion matrix layout, it is easier to evaluate the performance of the model. It is possible to obtain different types of measurements from the confusion matrix. Some are presented below:

- **Error Rate** - is the ratio between the incorrect prediction and all dataset.

$$ERR = \frac{FP + FN}{TP + TN + FP + FN} \tag{5.1}$$

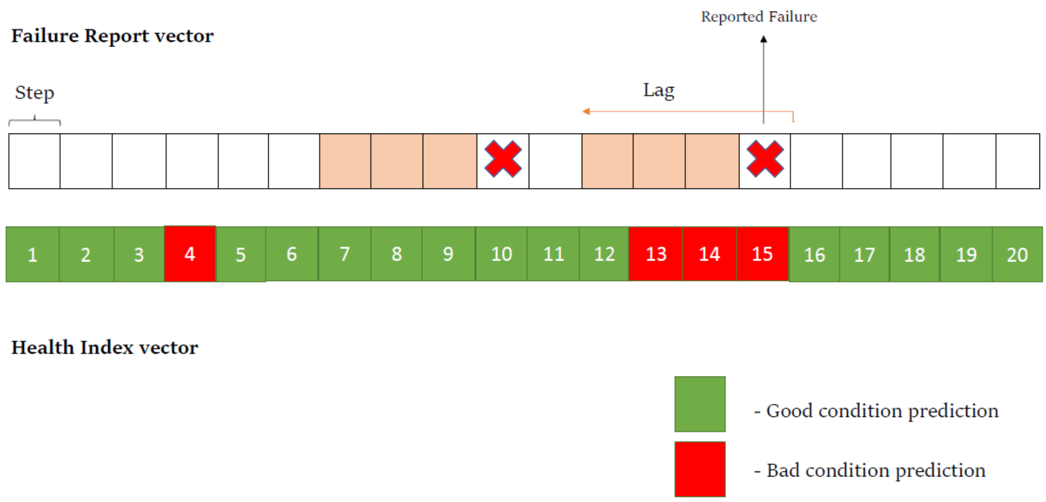


Figure 5.9: Fault Classification

- **Accuracy** - is calculated as the number of all correct predictions divided by all dataset.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.2}$$

- **Recall or True Positive Rate (TPR)** - is the ration between the correct positive predictions divided by all positive. It is also known as the probability of fault detection.

$$TPR = \frac{TP}{TP + FN} \tag{5.3}$$

- **False Positive Rate (FPR)** - is calculated as the number of incorrect positive predictions divided b the total number of negatives. It is also known as the probability of false alarms.

$$FPR = \frac{FP}{TN + FP} \tag{5.4}$$

- **Matthews correlation coefficient (MCC)**- This measurement is used to evaluate the quality of a binary variable. It evaluates the correlation coefficient between the predicted classification and the observed one. Returns a value between -1 and 1. Being one the perfect correlation, which means that every prediction was on point. And the opposite, -1, means that the prediction model was total of.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5.5}$$

There are still more ways to evaluate the performance of a prediction model. Due to an unbalanced dataset, as mentioned in Section 4.1.3.2, some results are not representative of the model performance. As a result, not every evaluative measurement was approached.

5.3 Model Tuning

This section will be presented with the results of the training. In order to optimize the model, a grid search test was applied, testing CT, WT and FT for different thresholds. The performance of the model depends on the values used on these variables, and, as a result, several values were needed to be tested. This values are presented in Table 5.7. The following results derived from the evaluation of the model on turbine 6 with a $lag = 7$ $days$ and a $step = 1$ day. The fault report is presented in Figure 5.10.

Table 5.7: Validation test data

	Tested Values				
Critical Threshold	0.6	0.7	0.8		
Warning Threshold	0.4	0.5	0.6		
Fault Threshold	0.46	0.5	0.53	0.56	0.6

	Turbine_ID	Component	Timestamp	GTM	Remarks
1	T06	GENERATOR	2016-07-10 19:50:00	00:00	Generator replaced
2	T06	GENERATOR	2016-07-24 17:00:00	00:00	Generator temperature sensor failure
3	T06	GENERATOR	2016-09-04 08:10:00	00:00	High temperature generator error
4	T06	GENERATOR	2016-10-02 17:10:00	00:00	Refrigeration system and temperature sensors in generator ...
5	T06	GENERATOR	2016-10-27 16:30:00	00:00	Generator replaced

Figure 5.10: Turbine 6 - Generator's failure report

Having into account all the possible value combination shown in Table 5.7, we obtained 45 confusion matrices to analyze, few of them are represented in Table 5.8.

Table 5.8: Example of Confusion Matrices : (a) Critical Threshold=0.6; Warning Threshold=0.4;Fault Threshold=0.46. (b) Critical Threshold=0.7; Warning Threshold=0.4;Fault Threshold=0.53. (c) Critical Threshold=0.8; Warning Threshold=0.5;Fault Threshold=0.56. (d) Critical Threshold=0.8; Warning Threshold=0.6;Fault Threshold=0.53.

	Fault_N	Fault_P
Predict_N	505	22
Predict_P	69	13

a)

	Fault_N	Fault_P
Predict_N	526	25
Predict_P	48	10

b)

	Fault_N	Fault_P
Predict_N	547	30
Predict_P	27	5

c)

	Fault_N	Fault_P
Predict_N	543	27
Predict_P	31	8

d)

To evaluate each result aforementioned techniques were used, Section 5.2. The following figures represent how the accuracy and the false positive rate fluctuate with different threshold combinations. This analysis is important to understand which is the best thresholds combination to apply on the test data. Figures 5.11, 5.13 and 5.15 describes the evolution of the accuracy with different thresholds, which can be compared with the Figures 5.12, 5.14 and 5.16 that the TPR evolution. On $x-axys$ values of the FT, and on $y-axis$ represents the Accuracy or the TPR, respectably, of the model. Each graph has three curves, each one is associated with a different CT. In each graph the WT is constant.

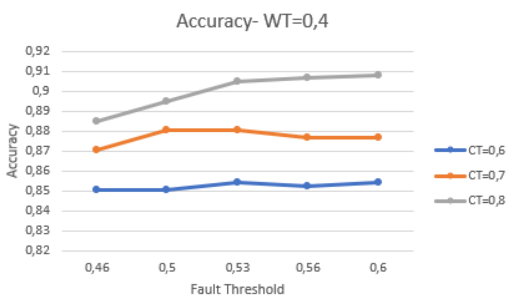


Figure 5.11: Accuracy evolution - $WT = 0.4$

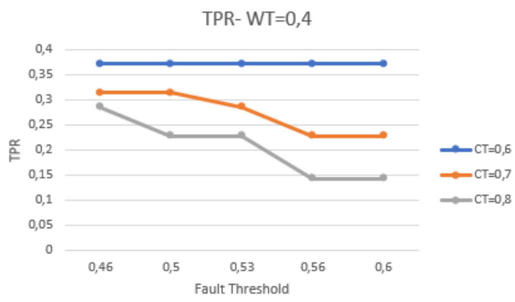


Figure 5.12: TPR evolution - $WT = 0.4$

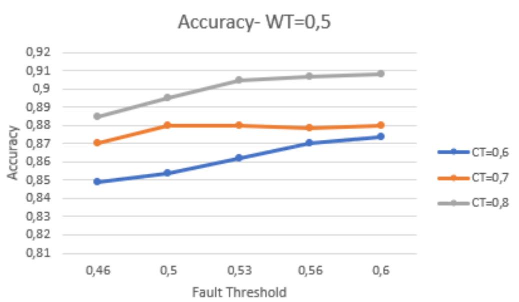


Figure 5.13: Accuracy evolution - $WT = 0.5$

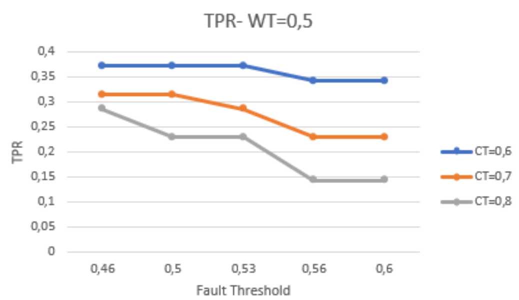


Figure 5.14: TPR evolution - $WT = 0.5$

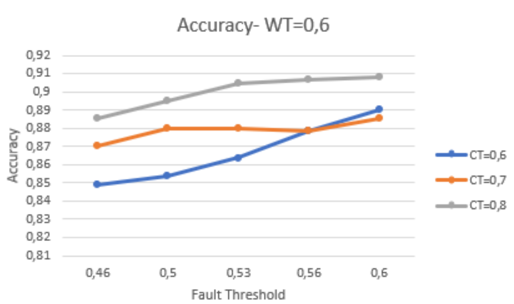


Figure 5.15: Accuracy evolution - $WT = 0.6$

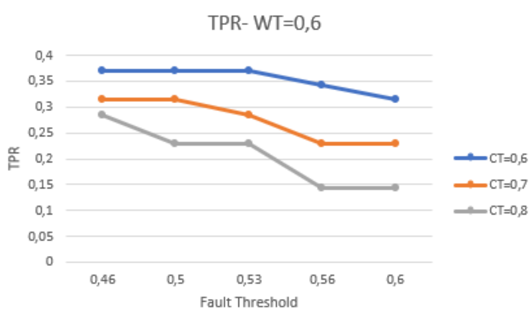


Figure 5.16: TPR evolution - $WT = 0.6$

As it is possible to conclude by graphics analysis, accuracy values are very satisfactory. However, these values are not representative of the model performance. The accuracy measurement do not have into account the imbalance of a dataset. Therefore, measurements like the false positive rate and the Matthew correlation coefficient are more suitable to make the analysis needed to find the best group of thresholds.

By analyzing the TPR graphics, it is more noticeable the impact that the threshold combination has on the number of good predictions. However, it is not clear which of the combination will give the best result. To do that, it is necessary to relate positive and negative ratings. Therefore, a ROC curve was applied to the matrices data, Figure 5.17. The ROC curve gives the relation between the probability of a correct prediction and the probability of doing a wrong prediction, false alarm.

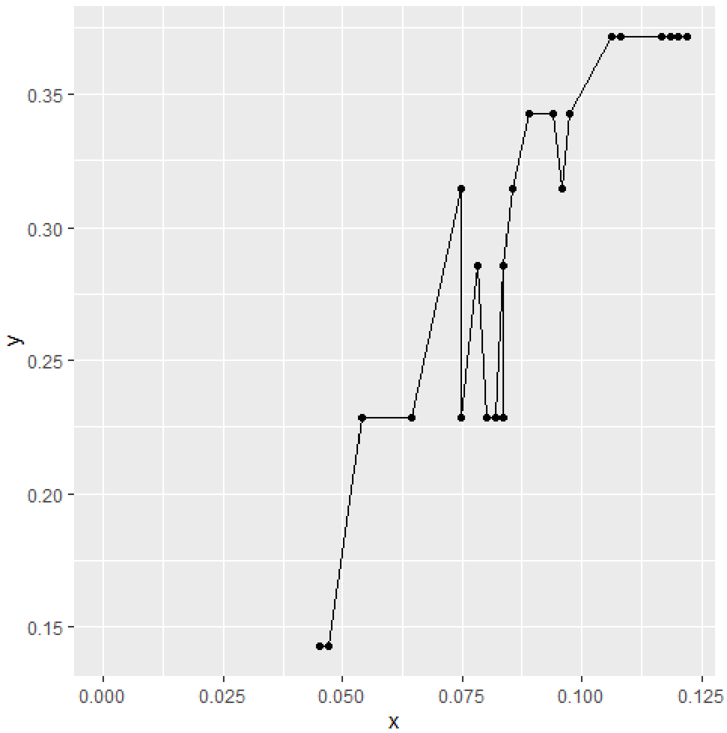


Figure 5.17: ROC curve - Turbine 6

After analysing the ROC curve, it was possible to have an idea of which combination were the most suitable for the model. To clarify this doubt, the MCC was used. The MCC variation curves are presented in Figure A.1,A.2 and A.3. The combination that had better results MCC score with $MCC = 0.2387406$, was:

- CT = 0.6
- WT = 0.6
- FT = 0.56

To conclude, several tests were made to obtain the best parameter for the model. Some of them were inconclusive due to the imbalance dataset. Nonetheless, others were capable of giving satisfactory results. The parameters that were calculated in this section will be tested on the turbine data and the results will be presented in the next section.

5.4 Turbine 7 - Model Test

All tests done before were intending to build a robust predictive model that could evaluate the health conditions of a system, without having any previous contact with the data. And that is the reason that turbine 7 was chosen to be the data test. The turbine 7 data was not used to construct the model, in that way, it will be used to test the model.

In figure 5.18 it is presented the generator failures that were registered and the ones that we want to identify.

	Turbine_ID	Component	Timestamp	GTM	Remarks
1	T07	GENERATOR	2017-08-20 05:50:00	00:00	Generator damaged
2	T07	GENERATOR_BEARING	2016-04-30 14:40:00	00:00	High temperature in generator bearing (replaced sensor)
3	T07	GENERATOR_BEARING	2017-08-20 06:00:00	00:00	Generator bearings damaged

Figure 5.18: Failure Report Turbine 7

The results that were obtained are represented in the following confusion matrix, table 5.9.

Table 5.9: Turbine 7- Test Results

	Fault_N	Fault_P
Predict_N	467	10
Predict_P	128	4

As was mentioned before, the confusion matrix is complemented by another information, which is the number of failures that were identified. In this case, all the failures were successfully predicted, at least with 2 days of difference. However, the model predicted many failures that did not occur, as it is represented by the number of false positives in the confusion matrix. This amount of false positives demonstrates that the threshold combination is not well suitable for this dataset. This means that the value of thresholds, mainly the FT could be lower. However, it is positive that the model was able to predict all the reported failures, as well as obtain low values of false negatives.

Bering that in mind, the results were satisfactory but the model still has to be improved on the precision level. Unfortunately, it was not possible to implement the technique presented in 4.1.3.2 which the main objective was to reduce these types of the wrong classification.

Chapter 6

Conclusions and Future work

Prediction methods have been gaining more and more popularity this last years in several applications, however the construction of an health index conducted on hydropower station is not very common. On this dissertation it is exploited a model that evaluate the healthy state of any complex system with hope to encourage studies similar to this.

In this dissertation were developed two methodologies that complement each another. The prediction failure model is based on the HI that were created previously on the system health evaluation model.

The system health evaluation model and the prediction failure model were developed within the scope of XFLEX project in order to be able to create a methodology that evaluate all dynamics of a hydropower plant.

The HI model was developed, as the name implies, to assess the system health status, with the aim of helping in the decision-making of which maintenance strategies should be implemented. Thus ensuring a healthy system preventing sudden and dangerous stops for the system to function.

Bearing in mind that the model was unsupervised, the representation of the health status of the system made by the developed HI was fairly accurate, taking into account that most of the registered failures were positively identified. This way, it is also possible to conclude that the degradation curves created were able to represent the state of decay of the various measures analyzed.

Since it would not be practical or perceptible to have several HI, an aggregation algorithm was used, which, taking into account the test results in the detection of failures, its is possible to said that corresponded to what was asked.

However, some false alarms were verified, based on the HI, that did not correspond to the reality of the state of the system. Therefore, and as a future work, it would be interesting to develop a more comprehensive model, in order to prevent false predictions. Techniques that were presented in Section 4.1.3.1 can be a solution to this problem. Another solution, could be the development of a model that takes into account all the dependencies and relationships of the components since complex systems such as hydroelectric station have high correlation between them.

Appendix A

In table A.1 is presented the different definitions of the concept prognostics.

Table A.1: Prognostic Definitions, source [59]

Author	Prognostics is...
ISO13381-1	An estimation of time to failure and risk for one or more existing and future failure modes
Engel	The capability to provide early detecting of the precursor and/or incipient fault condition of a component, and to have the technology and means to manage and predict the progression of this fault condition to component failure
Hess	Predictive diagnostics, which includes determining the remaining life or time span of proper operation of a component
Wu	The prediction of future health states and failure modes based on current health assessment, historical trends and projected usage loads on the equipment and/or process
Luo	Failure prognosis involves forecasting of system degradation based on observed system condition
Brotherton	The ability to assess the current health ofa part for a fixed time horizon or predict the time to failure
Katipamul	Address(ing) the use of automated methods to detect and diagnose degradation of physical system performance, anticipate future failures, and project the remaining life of physical systems in acceptable operating state before faults or unacceptable degradations of performance occur
Lewis	Prediction of when a failure may occur. To calculate the remaining useful life on an asset
Smith	The capability to provide early detection and isolation of precursor and/or incipient fault condition to a component or sub-element failure condition, and to have the technology and means to manage and predict the progression of this fault condition to component failure
Baruah	Prognostics builds upon the diagnostic assessment and are defined as the capability to predict the progression of this fault condition to component failure and estimate the remaining useful life (RUL)
Heng et al.	The forecast of an asset's remaining operational life, future condition, or risk to completion

The MCC variation is presented in the following curves.

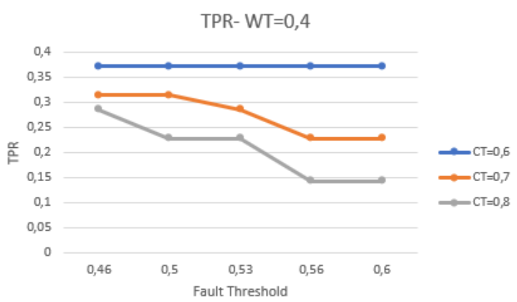


Figure A.1: MCC evolution - $WT = 0.4$

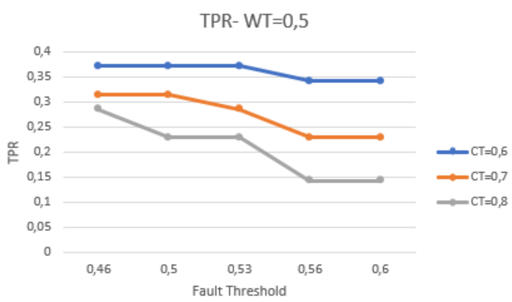


Figure A.2: MCC evolution - $WT = 0.5$

Table A.2: Advantages And Disadvantages of prognostic techniques

Approach	Advantages	Disadvantages
Macth Matrix	• Feasible for non-stationary processes • Can deal with high dimensional feature space • Can predict what kind of failure will happen when providing signatures corresponding to several failure modes	• Requires sufficient historical data from different operation cycles • Requires degradation data in the historical data • Can be computationally inefficient when the datasets are large
Particle filtering	• Provides non-linear projection, as well as multivariate and dynamic processes modelling • Applicable to non-linear system, non-linear and/or non-Gaussian noise • Higher accuracy than other existing filtering algorithms	• Requires significant resources including computation time and historical data to perform well in systems with a higher dimension or more particles • Requires a large amount of data to avoid degeneracy problem • Needs a threshold level
Hidden Markov model (HMM) and Hidden semi-Markov model (HSMM)	• Can perform fault and degradation diagnosis on non stationary signals and random dynamic systems • Can model different stages of degradation so failure trend does not need to be monotonic • Does not require specific knowledge of failure mechanism progression	• Requires large amount of training data, proportional to the number of hidden states, for accurate modelling • Prognosis projection relies on a failure threshold • Can only be an estimation due to the unobservable true state transition
Artificial neural network (ANN)	• Can model complex, multi-dimensional, unstable or non-linear systems • High capabilities in approximation, classification and noise-immunity through non-linear information processing • Remarkable ability in generalizing and deriving meaning from imprecise or complex data	• Requires data pre-processing to limit the number of data inputs and reduce model complexity • Trial and error to determine the most appropriate model can be time consuming • Operates as a “black box” without documentation of qualitative information of the model ; not clear about how decisions are reached in a trained network
Fuzzy Logic	• Enables condition classification based on histories • Can deal with incomplete, nosy or imprecise data and complex or unknown systems • Easy to interpret due to its expression in linguistic rules, resulting in a transparent model structure	• Does not provide indication of time to failure or probability of failure • Linguistic terms may compromise the accuracy of the model • Not feasible when membership functions are complicated to be determined

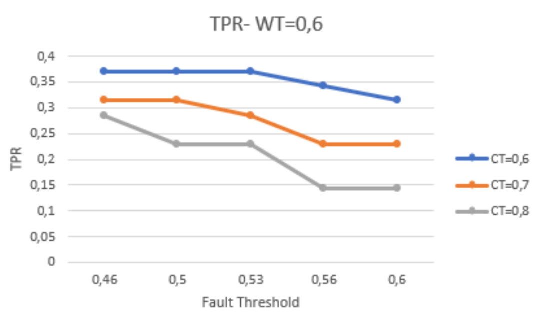


Figure A.3: MCC evolution - $WT = 0.6$

References

[1] Explore — EDPOpenData. https://opendata.edp.com/explore/?sort=modified{&}refine.keyword=Wind+Farm+1{&}refine.keyword=80. Accessed: 2020-01-30.

[2] Particle Filter : A hero in the world of Non-Linearity and Non-Gaussian. https://towardsdatascience.com/particle-filter-a-hero-in-the-world-of-non-linearity-and-non-gaussian-6d8947f4a3dc. Accessed: 2020-05-03.

[3] Three Ways to Estimate Remaining Useful Life for Predictive Maintenance - MATLAB & Simulink. https://www.mathworks.com/company/newsletters/articles/three-ways-to-estimate-remaining-useful-life-for-predictive-maintenance.html.

[4] ISO 14224:2006. Technical report, 2006.

[5] BS EN 13306:2017 Maintenance. Maintenance terminology, 2010.

[6] DNO Common Network Asset Inidices Health & Criticality- Version 1.1. Technical report, 2017.

[7] Ahmed E.B. Abu-Elanien, M. M.A. Salama, and M. Ibrahim. Calculation of a health index for oil-immersed transformers rated under 69 kV using fuzzy logic. *IEEE Transactions on Power Delivery*, 27(4):2029–2036, oct 2012.

[8] Rosmaini Ahmad and Shahrul Kamaruddin. An overview of time-based and condition-based maintenance in industrial application. *Computers and Industrial Engineering*, 63(1):135–149, 2012.

[9] Imad Alsyouf. Cost Effective Maintenance for Competitive Advantages. Technical report.

[10] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, feb 2002.

[11] Guanrong Chen, Trung Tat Pham, and NM Boustany. Introduction to Fuzzy Sets, Fuzzy Logic, and Fuzzy Control Systems. *Applied Mechanics Reviews*, 54(6):B102–B103, nov 2001.

[12] Delloite. Asset Health Indices - A utility industry necessity.

[13] Doctoral Dissertations, Jamie Baalis Coble, and Jamie Baalis. Trace: Tennessee Research and Creative Exchange Merging Data Sources to Predict Remaining Useful Life-An Automated Method to Identify Prognostic Parameters Recommended Citation. Technical report, 2010.

[14] Elizabeth da Silva Domingues. Improved Condition Monitoring of Composite Insulators. page 307, 2012.

[15] Ming Dong and David He. A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology. *Mechanical Systems and Signal Processing*, 21(5):2248–2266, jul 2007.

[16] Ming Dong and David He. A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology. *Mechanical Systems and Signal Processing*, 21(5):2248–2266, jul 2007.

[17] Ming Dong and David He. Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis. *European Journal of Operational Research*, 178(3):858–878, may 2007.

[18] Ugyen Dorji and Reza Ghomashchi. Hydro turbine failure mechanisms: An overview. 2014.

[19] EDP. Wind Turbine Characteristics. 2010.

[20] Dominique Egré and Joseph C. Milewski. The diversity of hydropower projects. *Energy Policy*, 30(14):1225–1230, nov 2002.

[21] ENTSOE. Power Statistics. https://www.entsoe.eu/data/power-stats/. Accessed: 2020-04-30.

[22] Diego Galar, Berges Peter, and Uday Kumar. The need for aggregated indicators in performance asset management. Technical Report 1, 2014.

[23] Nagi Gebraeel and Jing Pan. Prognostic degradation models for computing and updating residual life distributions in a time-varying environment. *IEEE Transactions on Reliability*, 57(4):539–550, 2008.

[24] Neil Gordon, D.J. Salmond, and A.F.M. Smith. *Novel approach to nonlinear and linear Bayesian state estimation*, volume 140. Institution of Electrical Engineers, apr 1993.

[25] A. Grall, C. Bérenguer, and L. Dieulle. A condition-based maintenance policy for stochastically deteriorating systems. *Reliability Engineering and System Safety*, 76(2):167–180, may 2002.

[26] Andrew Hess, Giulio Calvello, and Peter Frith. Challenges, issues, and lessons learned chasing the "Big P": Real predictive prognostics part 1. In *IEEE Aerospace Conference Proceedings*, volume 2005, 2005.

[27] R J Heywood and Dr T Mcgrail. Generating Asset Health Indices Which Are Useful and Auditable. Technical report.

[28] Thorhallur Hjartarson, Bruno Jesus, David T. Hughes, and R. Myles Godfrey. Development of Health Indices for Asset Condition Assessment. In *Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference*, volume 2, pages 541–544, 2003.

[29] Jiawen Hu and Piao Chen. Predictive Maintenance of Systems Subject to Hard Failure Based on Proportional Hazards Model. *Reliability Engineering and System Safety*, page 106707, 2019.

[30] Infraspeak. Corrective Maintenance: Definition, Advantages and Examples [2020]. https://blog.infraspeak.com/corrective-maintenance/. Accessed: 2020-04-22.

[31] Andrew K.S. Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance, oct 2006.

[32] V. Jayabalan and Dipak Chaudhuri. Cost Optimization of Maintenance Scheduling for a System with Assured Reliability. *IEEE Transactions on Reliability*, 41(1):21–25, 1992.

[33] Richard B. Jones. *Risk-Based Management*. 2011.

[34] Man Shan Kan, Andy C.C. Tan, and Joseph Mathew. A review on prognostic techniques for non-stationary and non-linear rotating systems. *Mechanical Systems and Signal Processing*, 62:1–20, 2015.

[35] David G Kleinbaum and Mitchel Klein. Statistics for Biology and Health Survival Analysis A Self-Learning Text Third Edition. Technical report.

[36] David G. Kleinbaum and Mitchel Klein. Introduction to Survival Analysis. pages 1–54. 2012.

[37] Ranganath Kothamasu, Samuel H. Huang, and William H. Verduin. System health monitoring and prognostics - A review of current paradigms and practices. *International Journal of Advanced Manufacturing Technology*, 28(9):1012–1024, apr 2006.

[38] A. Lejeune and S. L. Hui. Hydro power: A multi benefit solution for renewable energy. In *Comprehensive Renewable Energy*, volume 6, pages 15–47. Elsevier Ltd, jan 2012.

[39] Kwan-Moon Leung, Robert M. Elashoff, and Abdelmonem A. Afifi. Censoring Issues in Survival Analysis. *Annual Review of Public Health*, 18(1):83–104, may 1997.

[40] Jianbo Liu, Dragan Djurdjanovic, Jun Ni, Nicolas Casoetto, and Jay Lee. Similarity based method for manufacturing process performance prediction and diagnosis. *Computers in Industry*, 58(6):558–566, 2007.

[41] Kaibo Liu, Nagi Z. Gebraeel, and Jianjun Shi. A Data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis. *IEEE Transactions on Automation Science and Engineering*, 10(3):652–664, jul 2013.

[42] Canh Ly, Kwok Tom, Carl S. Byington, Romano Patrick, and George J. Vachtsevanos. Fault diagnosis and failure prognosis for engineering systems: A global perspective. In *2009 IEEE International Conference on Automation Science and Engineering, CASE 2009*, pages 108–115, 2009.

[43] F Maciel Barbosa. Introdução à Fiabilidade de Sistemas Elétricos de Energia. Technical report, 2013.

[44] A. R. Majeed and N. M. Sadiq. Availability and reliability evaluation of Dokan Hydro Power Station. In *2006 IEEE PES Transmission and Distribution Conference and Exposition: Latin America, TDC'06*, 2006.

[45] Luca Marinai and Riti Singh. Gas-path Diagnostics and Prognostics for Aero-engines Using Fuzzy Logic and Time Series Analysis. Technical report, 2004.

[46] Jae H. Min and Chulwoo Jeong. A binary classification method for bankruptcy prediction. *Expert Systems with Applications*, 36(3 PART 1):5256–5263, apr 2009.

[47] R Keith Mobley. *An Introduction to Predictive Maintenance (Second Edition)*. 2002.

[48] John Moubray. *Reliability-centered Maintenance - John Moubray* .

[49] D. N.P. Murthy, A. Atrens, and J. A. Eccleston. Strategic maintenance management. *Journal of Quality in Maintenance Engineering*, 8(4):287–305, 2002.

[50] C Okoh, R Roy, J Mehnen, and L Redding. Overview of Remaining Useful Life Prediction Techniques in Through-life Engineering Services. 2014.

[51] Alexandra Oliveira. Asset management optimization : Integrated Operations and Maintenance (O & M) Planning. 2019.

[52] Ying Peng, Ming Dong, and Ming Jian Zuo. Current status of machine prognostics in condition-based maintenance: A review. *International Journal of Advanced Manufacturing Technology*, 50(1-4):297–313, sep 2010.

[53] V Pereyra and G Scherer. Exponential Data Fitting and Its Applications, 2010.

[54] Aisong Qin, Qinghua Zhang, Qin Hu, Guoxi Sun, Jun He, and Shuiquan Lin. Remaining Useful Life Prediction for Rotating Machinery Based on Optimal Degradation Indicator. 2017.

[55] Ahmed Ragab, Mohamed Salah Ouali, Soumaya Yacout, and Hany Osman. Remaining useful life prediction using prognostic methodology based on logical analysis of data and Kaplan–Meier estimation. *Journal of Intelligent Manufacturing*, 27(5):943–958, oct 2016.

[56] Marvin Rausand and Arnljot Hyland, editors. *System Reliability Theory*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA, nov 1994.

[57] R Sathya and Annamma Abraham. The Science and Information Organization International Journal of Advanced Reasearch in Artificial Intelligence. Technical report.

[58] Xiao Sheng Si, Wenbin Wang, Chang Hua Hu, and Dong Hua Zhou. Remaining useful life estimation - A review on the statistical data driven approaches, aug 2011.

[59] J. Z. Sikorska, M. Hodkiewicz, and L. Ma. Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, 25(5):1803–1836, jul 2011.

[60] Mohammadreza Tahan, Elias Tsoutsanis, Masdi Muhammad, and Z. A. Abdul Karim. Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: A review, 2017.

[61] Jan M. van Noortwijk, Roger M. Cooke, and Matthijs Kok. A Bayesian failure model based on isotropic deterioration. *European Journal of Operational Research*, 82(2):270–282, apr 1995.

[62] Sandra Vieira, Walter H.L. Pinaya, and Andrea Mechelli. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications, mar 2017.

[63] Hongzhou Wang. A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139(3):469–489, jun 2002.

[64] Ming Yi You, Fang Liu, Wen Wang, and Guang Meng. Statistically planned and individually improved predictive maintenance management for continuously monitored degrading systems. *IEEE Transactions on Reliability*, 59(4):744–753, dec 2010.