Universidade de Aveiro Departamento de Engenharia Mecânica
2021

Inês Vilela dos Santos

Towards a predictive maintenance methodology of hydraulic pumps

Rumo a uma metodologia de manutenção preditiva de bombas hidráulicas

Universidade de Aveiro Departamento de Engenharia Mecânica
2021

Inês Vilela dos Santos

Towards a predictive maintenance methodology of hydraulic pumps

Rumo a uma metodologia de manutenção preditiva de bombas hidráulicas

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Mecânica, realizada sob orientação científica de António Gil D'Orey de Andrade Campos, Professor Auxiliar com Agregação do Departamento de Engenharia Mecânica da Universidade de Aveiro, e de Bernardete Coelho dos Santos, Investigadora Doutorada da Universidade da Bretanha do Sul, França.

Este trabalho teve o apoio financeiro dos projetos UIDB/00481/2020 e UIDP/00481/2020 - FCT - Fundação para Ciência e Tecnologia; e CENTRO-01-0145-FEDER-022083 - Programa Operacional Regional do Centro (Centro2020), no âmbito do Acordo de Parceria Portugal 2020, através do Fundo Europeu de Desenvolvimento Regional.

o júri / the jury

presidente / president

Prof. Doutor Nelson Amadeu Dias Martins
Professor Associado da Universidade de Aveiro

vogais / committee

Doutor Bruno Alexandre Abreu da Silva
Diretor Executivo da *Scubic*

Prof. Doutor António Gil D'Orey de Andrade Campos
Professor Auxiliar com Agregação da Universidade de Aveiro (orientador)

agradecimentos / acknowledgements

Quero agradecer ao meu orientador Gil Campos pela sua disponibilidade e orientação e pelo apoio e ajuda que foram uma parte determinante do sucesso desta dissertação de mestrado. Quero também agradecer à minha co-orientadora Bernardete Coelho pelo grande apoio dado em várias fases críticas do desenvolvimento deste trabalho. Um agradecimento muito especial ao Francisco, que foi incansável em incentivar-me e apoiar-me constantemente ao longo desta jornada. Uma vez que a conclusão deste trabalho de dissertação coincide com a conclusão do meu percurso académico, não posso deixar de agradecer a todos os que fizeram parte do mesmo, em especial às minhas amigas, pelas incontáveis horas de estudo e também pelas excelentes memórias que fomos colecionando ao longo destes cinco anos. Por último mas não menos importante, quero agradecer aos meus pais, a quem devo tudo o que sou, por todo o carinho e compreensão. À Babi, por estar sempre comigo e por sempre tentar ajudar-me a ser a melhor versão de mim própria.

keywords

Hydraulic Pumps, Maintenance, Fault Detection and Classification, Remaining Useful Life Estimation

abstract

Hydraulic pumps, essential elements in water supply systems, are mainly responsible for the high energy consumption associated with these systems. It is, therefore, relevant to keep the pumps running in their best possible conditions in order to avoid further consumption and costs, and also to anticipate possible pump failures. The best strategy to anticipate the occurrence of failures is to implement preventive and predictive maintenance plans, instead of corrective maintenance that is still widely applied. Thus, with the goal of developing a predictive maintenance methodology applied to hydraulic pumps, this dissertation aims to explore and investigate the applicability of two techniques that can be integrated into a maintenance plan: the detection and classification of failures and the estimation of the remaining useful life (RUL) of the pump. To implement the proposed tasks, simulated data and measured data from real systems were used, taken from online data repositories, with values recorded by sensors and with the identified condition of the system. The first technique allowed, through sensor data with the respectively identified faults, to train classification algorithms able to identify failures. In the first of the evaluated case studies, the best of the implemented algorithms identified the failures associated with the pump data with an accuracy of 82.9%, whereas, in the second of the evaluated case studies, the algorithm that presented the best performance obtained an accuracy of 94.6% in identifying the failure mode associated with the pump. The decision tree and ensemble trees algorithms proved to be the most suitable for the studied purpose. The second technique allowed to estimate RUL values from sensor data recorded from normal operation to system failure. Although the first RUL implemented case study was an engine, the second case study was a water pump. The methodology of the RUL model proved to be relevant because it managed, even with some deviations from the true values, to estimate acceptable values of RUL. An economic analysis was also carried out, highlighting the relevance of applying RUL estimation models in predictive maintenance methodologies for hydraulic pumps.

palavras-chave

Bombas Hidráulicas, Manutenção, Deteção e Classificação de Falhas, Estimativa do Tempo de Vida Útil Restante

resumo

As bombas hidráulicas, elementos essenciais nos sistemas de abastecimento de água, são os principais responsáveis pelos elevados consumos energéticos associados a estes sistemas. Torna-se, portanto, relevante manter as bombas a funcionar nas suas melhores condições possíveis de forma a evitar mais consumos e custos, e também antecipar possíveis falhas nas bombas. A melhor estratégia para antecipar o acontecimento de falhas passa pela implementação de planos de manutenção preventivos e preditivos, ao invés da manutenção corretiva que é ainda muito aplicada. Assim, com vista ao desenvolvimento de uma metodologia de manutenção preditiva aplicada às bombas hidráulicas, esta dissertação tem como objetivo a exploração e investigação da aplicabilidade de duas técnicas que podem ser integradas num plano de manutenção: a deteção e classificação de falhas e a estimativa do tempo de vida útil restante (RUL) de uma bomba. Para implementar as tarefas propostas utilizaram-se dados simulados e dados medidos a partir de sistemas reais, retirados de repositórios de dados online, com valores registados por sensores e com a condição do sistema identificada. A primeira técnica permitiu, através de dados de sensores com as respetivas falhas identificadas, treinar algoritmos de classificação capazes de identificar falhas. No primeiro dos casos de estudo avaliados, o melhor dos algoritmos implementados identificou as falhas associadas aos dados da bomba com uma classificação de desempenho de 82.9%, ao passo que, no segundo dos casos de estudo avaliados, o algoritmo que apresentou melhor desempenho obteve uma classificação de 94.6% na identificação do modo de falha associado à bomba. Os algoritmos de decision trees e ensemble trees demonstraram ser os mais indicados para o propósito estudado. A segunda técnica permitiu calcular previsões de valores do RUL a partir de dados de sensores registados desde uma operação normal até à falha do sistema. Apesar de o primeiro caso de estudo de implementação de RUL ter sido um motor, o segundo caso de estudo foi uma bomba de água. A metodologia do modelo de RUL demonstrou ser pertinente pois conseguiu, ainda que com alguns desvios em relação aos verdadeiros valores, estimar valores aceitáveis de RUL. Elaborou-se ainda uma análise económica que evidencia a relevância em aplicar modelos de cálculo de RUL em metodologias de manutenção preditiva de bombas hidráulicas.

Contents

Intentionally blank page.

List of Tables

Intentionally blank page.

List of Figures

Chapter 1

Introduction

1.1 Framework

Water is critical for social and economic development, food and energy production, and, therefore, essential for the welfare of all humankind. Since access to clean water constitutes the most important criterion for public health, it is the water supply industry that has the critical responsibility to guarantee its access to populations.

Water supply systems (WSS) transport water from sources to customers and also provide vital services to the industrial society. Efficiency and energetic sustainability has a major role in this sector: approximately 35% of the water supply expenses are related to energy costs, which translates into an annual worldwide expenditure of approximately 12 billion euros [1]. Thus, the continued increase in world population and energy prices lead the industries responsible for the supply to prioritize the efficient use of energy and water resources, which reveals to be a hard task. The complex configuration of the networks, the lack of flexibility in their operation, and the number of variables to be controlled result in nonlinear systems that are difficult to manage.

One of the most common devices in these water supply systems are the hydraulic pumps, often used to transport water from its source to storage tanks. Since these elements are the main responsible for the highest values of energy consumption [2], understanding how hydraulic pumps operate and how their efficiency can be increased is, therefore, an essential task.

1.2 Problem Identification

Hydraulic pumps are accompanied by performance curves, which are developed by the manufacturers themselves, seeking to indicate their characteristics and behavior under certain working conditions.

However, the information presented is only accurate for the first years of the device's life. Over time, the behavior and efficiency of the pumps begin to change in a way that is today relatively unknown and little controlled by the industry. These unforeseen shifts result, initially, in the malfunction of these elements, leading, later, to the appearance of damages that compromise their integrity. Additionally, a non predicted imminent malfunction or failure of a hydraulic pump inserted in a water supply network ends up compromising not only its individual operation but also the functioning of the entire system, which can become slower, behave in an undesirable way, or even be interrupted by a blockage of its pipelines. In the end, all these deviations converge to increasing costs.

Given this context, it is relevant to constantly monitor the behavior of a pump, with the intrinsic objective of anticipating failures, correcting them, and, thus, reducing costs and downtime.

1.3 Proposed Solution

Traditionally, pump maintenance is handled through a corrective approach: the problem is only corrected after it happens. Although this method does not require a monitoring effort, nevertheless it presents several disadvantages and particularly the possibility of compromising the entire hydraulic network, which, in turn, as previously mentioned, has major economic implications. Alternatively, in order to maintain a proper operation with the least possible associated costs, an ideal solution for a robust maintenance plan would be to combine the optimum operating conditions of a pump with real-time monitoring of its basic performance parameters, i.e., a predictive maintenance plan [1].

The presented work aims to explore and explain two techniques of different stages of the data treatment, which can be integrated into a possible predictive maintenance process for hydraulic pumps: fault detection and classification and estimation of the Remaining Useful Life (RUL) of a pump. Additionally, a straightforward economic analysis was defined and applied to appraise the financial viability of the studied techniques in comparison to traditional pump maintenance procedures.

To assess the effectiveness of the selected models, since it was not possible to collect real data from these types of devices in operation, data sets available in online repositories with different levels of detail were used.

1.4 Guidelines

In order to achieve the defined goals, this work was divided into the following five chapters:

- 1. Introduction: the theme of the dissertation is presented as well as a brief introduction of the work and its main objectives;
- 2. State-of-the-art: some concepts associated with pumps and their maintenance are explored, followed by the review of some predictive and preventive maintenance methodologies applied in the context of pumps and hydraulic systems;
- 3. Methodology: the reasoning used in the dissertation is presented, starting with an overview of the logical structure of the idealized predictive maintenance process, then followed by the presentation and explanation of the used techniques for fault detection and classification and RUL estimation and also for the economic analysis, ending later with a description of the tools applied in their implementation;
- 4. Results and Discussion: two different case studies are described for each of the considered techniques, as well as the results of their implementation and their respective discussion. Finally, the economic analysis is applied and evaluated to one of the sets of the obtained results for one of the case studies;
- 5. Conclusions: the conclusions and limitations of the present dissertation are recorded, as well as possible suggestions for future works.

[1] At the beginning, inserted in the context of predictive maintenance, this dissertation had as its final objective the creation of an expeditious methodology for assessing the quality of the status of a hydraulic pump compared to its initial operating state. For this purpose, it was intended to carry out local and virtual analyzes of pumps inserted in water supply systems. However, due to the current context of the Covid-19 pandemic, it was not possible to undertake the study of these devices and, consequently, the purpose of this master thesis was adapted.

Chapter 2

Literature Review

This chapter aims to present a review of concepts on topics related to hydraulic pumps and their principle of operation, the faults typically associated with these devices, and the main types of maintenance that can be applied to them. A literature review is also presented that focuses on the study of preventive and predictive maintenance methodologies applied to pumps and other hydraulic systems.

2.1 Concepts

2.1.1 Hydraulic Pumps

Working Principle

A pump is a mechanical device that is capable of moving fluid by converting mechanical power to hydraulic energy. It generates flow with enough power to overcome pressure induced by the load at the pump outlet. Pumps operate by creating low pressure at the inlet of the pump, which allows the fluid to be pushed from the reservoir, through the pump inlet, and into the pump, and then, by mechanical action, delivers the fluid to the outlet and into the hydraulic system.

Types of Pumps

Hydraulic pumps can be classified as hydrostatic or hydrodynamic, according to the method they use to move the fluid [3]. Hydrostatic pumps are positive displacement pumps and cause the fluid to move by trapping a fixed volume of the fluid and displacing it into the outlet. Positive displacement pumps produce the same flow at a certain speed despite the pressure at the outlet of the pump [4, 5]. Hydrodynamic pumps, or simply dynamic pumps, are more commonly used in industry than hydrostatic pumps. In hydrodynamic pumps, kinetic energy is given to the fluid by increasing the velocity of the flow. Kinetic energy is then converted into pressure as the flow exits the pump through the outlet [6]. There are various types of dynamic pumps and the most common are centrifugal pumps and axial-flow pumps. In centrifugal pumps, also referred to as radial-flow pumps, the fluid enters the pump along the axis and exits the pump perpendicularly to the axis. In general, centrifugal pumps can operate at higher pressures with lower flow rates [7]. In axial-flow pumps, the direction of the flow exiting the pump is the same as the fluid entering the pump. These types of pumps operate at higher flow rates and lower pressures.

Pump operation basics

The basic principle of operation of a pump is to convert mechanical energy to pressure. While operating, the impeller rotates and accelerates the fluid and then, the velocity of the fluid is

converted to pressure energy. The purpose of the pump is to transport and lift the fluid to a higher level. The volume of the fluid that must be moved is the flow rate and the height that the fluid must be lifted is the head [7]. These concepts - flow rate and head - are two of the main variables that must be taken into account to understand the basics of the operation of a pump. Other variables are inherent to the process, such as speed, pressure, power, and efficiency.

The efficiency of a pump is given by the ratio of the power delivered to the pump and the power delivered by the pump. The Brake Horse Power is the pump input that is the actual power delivered to the pump shaft. The Hydraulic or Water Horse Power is the pump output, which is the liquid power delivered by the pump. The pump input power is, naturally, greater than the pump output power due to the losses incurred in the pump [8].

The Best Efficiency Point (BEP) of the pump is an important parameter to determine if the pump is being correctly operated. At the BEP, the flow rate is the one that gives the pump its highest efficiency. Almost all the pumps do not operate at their BEP most of the time since the variables of the process are inconstant and unpredictable. Nevertheless, a pump that is properly designed can perform with a flow near its highest efficiency. A good practice is to maintain a flow between 80 to 110 percent of BEP, allowing to increase efficiency and decrease the risk of failures. However, most pumps do not achieve that range of efficiency [9].

Typical Failures in Hydraulic Pumps

An inefficient operation, such as a pump operating far from its BEP, can cause problems in the pump, like parts being deflected and wear excessively. In fact, many problems can arise within pumps, resulting in a decrease in the flow and consequently, interrupting the transport of the fluid.

McKee et al. classified the problems that affect water pumps as hydraulic failures, mechanical failures, and other types of failure [10]. Hydraulic failures result from modifications in the pressure of the system due to changes in the parameters flow rate, temperature, and velocity of the fluid flow. The most common hydraulic failures come from cavitation, but other failures can occur, such as radial thrust, axial thrust, pressure pulsations, and suction and discharge recirculation. Cavitation is the formation of vapor bubbles in the fluid when the pressure of the fluid is lower than its vapor pressure. The cavitation bubbles collapse in areas of higher pressure causing damage to the pump surfaces. This phenomenon result from a reduction in suction pressure, an increase in suction temperature, or an increase in the flow rate, and its damage is accompanied by the symptoms: erosion, noise, vibrations with characteristically high amplitude and low frequency, and a reduction in pumping efficiency [10]. High radial thrust and axial thrust can cause bearing, shaft, and seal damage [10]. Pressure pulsations are characterized as fluctuations in the basic pressure or head being developed by the pump and can cause instability of pump controls, high levels of noise, and vibration in suction and discharge piping [6, 10, 11]. Recirculation is a reversal flow that can occur near the inlet of the impeller - suction recirculation - or near the outlet of the impeller - discharge recirculation -, and it usually appears when the pump is operating at low flows [6,10]. The main mechanical failures are: (i) seal failure, which is mostly caused by the pump running dry; (ii) bearing failure, usually caused by contamination in the bearing oil or high heat resulting from an overload on the bearing; (iii) lubrication failure; (iv) fatigue; and (v) excessive vibrations, caused by unbalanced moving parts in the pump system, movement of the pipelines, and interactions between the fluid and its particles [10]. Some other types of failures that can arise in pumps are: (vi) erosion and corrosion, which are caused by structural problems; (vii) excessive power consumption, which can result from several problems and with numerous causes; and (viii) blockages, which occur due to materials or objects located either at the impeller inlet or outlet or between the impeller vanes [10].

2.1.2 Maintenance

The main purpose of maintenance is to provide the required capacity for production without losses, and consequently increasing productivity and safety and decreasing costs [8]. The technical

definition of maintenance implicates functional checks, replacing or repairing a device into a state in which it can perform its required function, and adjusting parameters in the system [12]. Since any machine or equipment is susceptible to tear and wear, maintenance has the important task of keeping the devices as long as possible in their original condition.

Maintenance programs can lead to savings, ensuring a greater pump capacity and a higher efficiency (Figure 2.1), resulting in a decrease of energy consumption [8, 13].

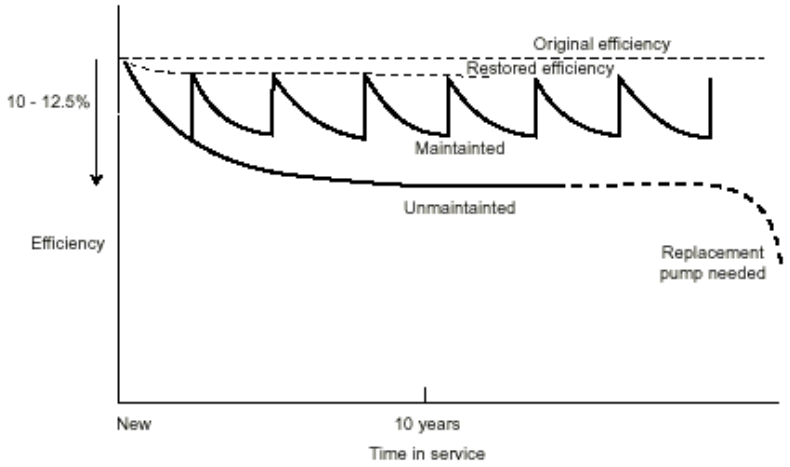


Figure 2.1: Progress of pump efficiency over time with and without maintenance plans [14].

Maintenance activities can be performed before or after failures occur, which is the major difference between the distinct maintenance categories. Maintenance exercises can be classified into three main categories: corrective, preventive, and predictive maintenance [14, 15].

Corrective maintenance, also referred as reactive maintenance or breakdown maintenance, is a run-to-failure type of maintenance. As its designation suggests, actions of maintenance are not performed until the device has failed [8]. This is a very common type of maintenance in industry and, generally, it is the least cost-effective option [16]. Nevertheless, breakdown maintenance can sometimes be cost-effective, depending on the cost penalty of unexpected failure [8].

Preventive maintenance is a scheduled type of maintenance that consists of a routine checkup of the equipment, such as calibration and analysis [17]. In preventive maintenance programs, appropriate actions are taken to prevent failures once every specific period of time [16]. Nevertheless, eventually, it may occur a breakdown in between two maintenance service periods, leading to unplanned downtime [14].

Using a predictive maintenance plan the condition of in-service devices can be determined using data collected from the device. The data is collected as a result of measurements, analysis, and parameters related to the state and condition of the system, that are compared with optimum performance parameters [18]. With the condition of devices determined, it can be predicted when maintenance should be performed [19]. Predictive maintenance was developed to complement preventive maintenance, preventing a part of the device to be switched by its lifetime [18].

In some of the literature, predictive maintenance is referred to as condition-based maintenance, since the condition and some parameters of the machine are read to elaborate a maintenance plan [20,21]. Condition or predictive maintenance can be applied as continuous monitoring, where the condition of the pump is continuously read and certain parameters are automatically adjusted as needed [16].

Some authors suggest that the best strategy is not either preventive maintenance or predictive maintenance, it is a combination of the two types of maintenance, properly applied to the system [8, 17].

2.2 Preventive and Predictive Maintenance Methodologies of Pumps and Hydraulic Systems

Various academic works based on different techniques and approaches and with models of different natures have been developed in the field of maintenance research and development applied to hydraulic pumps. The following is the result of a research that aimed to focus on the main techniques used to build predictive maintenance methodologies and, also, some preventive maintenance methodologies, not only for hydraulic pumps but also for other hydraulic systems.

Some authors suggest the Failure Mode Effect Analysis (FMEA) as a tool to establish predictive and preventive maintenance methodologies. FMEA is a process analysis tool that consists of a step-by-step approach for determining the possible failures on a process, a product, or a service. To develop a successful analysis all significant failure modes for each element in the system must be included [22]. PV Senthiil et al. explored the FMEA as a resource to evaluate a hydraulic press unit and to find its more susceptible to failure components [19]. They developed a predictive maintenance model using a service life prediction approach. The FMEA is used as a tool to identify the most critical components which fail more frequently. Based on the information of the critical components and their failures, it is possible to determine the condition of the equipment in service and to predict when maintenance should be performed. This approach with predictive maintenance techniques allows saving more costs than compared with routine or time-based preventive maintenance, once the maintenance tasks are performed only when it is necessary [19]. E. Lisowski and J. Fabis established the identification of potential failures and damages of a gear pump using an analysis based on the use of the FMEA matrix [23]. This quality method is used for identifying potential failures and foreseeing their effects. Preventing actions can be done before failures occur, based on the methodology developed by predicting the potential failures.

The evaluation of pumping systems performance is necessary to implement preventive and predictive maintenance actions. Efficiency tests are a tool to evaluate their performance. Leite et al. analyzed and compared the two main tests methods used in pumping systems: the conventional and the thermodynamic methods [13]. They identified the possible advantages and constraints on implementing those tests on a WSS.

Many authors use vibration analysis and thermography as tools towards predictive maintenance [9, 18]. Navega and Junior discuss three of the most common analysis performed in predictive maintenance of hydraulic pumps [18]. As stated by the authors, vibration analysis is extremely effective in finding particular problems within pumps, such as misalignment and unbalance, cavitation, and bearing defects. Additionally, thermography is a suitable analysis in maintenance programs to detect if components are operating at proper temperatures. Also, the authors include particle analysis in a pump predictive maintenance to verify the existence of sediments in undesirable areas of the pump [18]. Kernan states that pump vibration analysis is essential to every pump performance monitoring program [9]. Also, the author refers that the vibration level of a pump is related to where the pump is operating in relation to its BEP - higher vibrations refers to a pump operating far from its BEP.

Shen et al. proposed a method for gear fatigue life prediction based on virtual simulation technology [24]. The method is applied to the analysis of an aviation gear pump considering the influence of the internal flow field of the gear pump. The flow field of the gear pump and the movement of the gears are simulated using the Ansys software. According to the simulation results, the contact stress of the gear is calculated. The fatigue life of the gear is predicted using the nominal stress approach and miner cumulative damage principle. The proposed method is demonstrated to be efficient in the studied application, hence, it has a great influence on improving the life and reliability of the gear pump [24].

Rivera et al. focused on condition monitoring towards a predictive maintenance system of a hydraulic pump: it was developed an iterative and hierarchical methodology for the implementation of a predictive maintenance system in an industrial machine that remains in production [25]. The approach to the problem consisted of problem evaluation, gathering of expert knowledge,

data access, exploration of the data, and knowledge mapping to the data. The authors implemented different analytical methods on a running system and derived their results into health indicators, which have an important value in terms of anomaly and fault detection. There were considered six different methods for the implementation of the predictive maintenance system. The methods were studied and analyzed to understand the importance of the resulting health indicators. The results of indicators that are not relevant to the process are not taken into account in the next iteration of the methodology. The implementation of a state machine as a quality control system is discussed as well as the implementation of two signal models, a physical pressure signal model and a vibration model. The derived health indicators are able to detect anomalies in the acquired data [25].

Parrondo et al. developed a predictive maintenance system for a centrifugal pump based on signal monitoring of the fluid-dynamic operating conditions of the pump [26]. The proposed methodology consisted of an experimental study of the dynamic response of the pump under a variety of operating and abnormal conditions, in terms of acceleration and pressure signals. Then, a numeral succession for each signal and each of the studied anomalies are obtained. The numeral successions indicate the relative variations observed in the power spectrum at different frequencies. Finally, the most adequate signals for diagnosis are selected by comparing their sensibility to detecting operating condition changes and their capacity to identify each possible anomaly [26].

Bansal et al. rejected the need for constant monitoring of system parameters towards a predictive maintenance approach [27]. A real-time predictive maintenance system for machine systems based upon a neural network method was developed. The system accurately predicts the dynamic behavior of the machine based upon the interpretation of the motor current signature. A neural network is used to learn the non-linear function between system parameters and motion current system. This method avoids the need for costly measurement of system parameters. Additionally, an alternative approach for developing the system is explored. Simulation models are used to generate the data to train the neural network. The usage of neural network approach for condition monitoring system is validated with an overall correct classification percentage of 97,5% [27].

More authors applied neural networks as tools to develop predictive maintenance methodologies that include the task of fault detection [28,29]. Farokhzad et al. applied the usage of a neural network specifically to a centrifugal water pump [28]. An artificial neural network (ANN) based model with a back-propagation learning algorithm and multi-layer perceptron neural network is designed and developed for the purpose of fault detection. Simple statistical features were derived from vibration signals in the frequency domain by means of the Fast Fourier Transform method. The features are the input to the network model. The system classifies four different pump conditions with an overall accuracy of 100%. Zouari et al. developed a diagnosis system for centrifugal pumps [29]. The system is based on vibration measurements, signal pre-processing, and fault classification. Vibration signals of a centrifugal pump resulting from different induced faults were acquired. Then, using signal processing techniques, such as Principal Component Analysis (PCA) and Fischer Discriminant Analysis (FDA), a group of pertinent features is identified. Finally, using neural networks and neuro-fuzzy logic techniques, the detection and classification of the induced faults are performed. A global rate of system correct response of 96% is achieved on classifying sixteen types of anomalies [29].

Han et al. established a centrifugal pump backpropagation neural network model for the performance prediction of centrifugal pumps based on design and structure parameters [30]. A double hidden layer backpropagation (BP) neural network combined with the Levenberg-Marquardt algorithm was used. The proposed model achieves higher efficiency learning and better convergence accuracy compared with the traditional single hidden layer BP network.

Zhang et al. proposed a methodology for prognostics and health assessment on a double-suction centrifugal pump [31]. Time and frequency domain features are extracted from the pump vibration signals. Then, Principal Component Analysis (PCA) is applied to fuse features and to extract indicators (principal components). The methodology is concluded by calculating

a fault threshold for the pump using the BEP.

Calabrese et al. established a data-driven methodology that includes the steps of feature extraction, health indicators construction, anomaly detection, the building of degradation models, and RUL prediction [32]. The proposed method has several interesting characteristics: it does not require prior knowledge regarding faulty behavior, it does not rely on previously trained classification models, and it can be implemented online due to its low computational complexity.

Peng et al. reviewed some of the literature focusing on machine prognostics [33]. Prognostics addresses the use of automated methods to detect and analyze the degradation of systems performance and calculating the RUL towards a predictive maintenance plan. The main categories of prognostic models are explained as well as some of their advantages and disadvantages. For each type of model, some examples of methods, algorithms and techniques are described. Also, the application domains more suitable for each method and algorithm is provided. Furthermore, the advantages and disadvantages of each method and algorithm in terms of building the model, the data required for successful implementation, and the predicted accuracy of the model, are discussed [33]. The four categories of prognostic models are physical model, knowledge-based model, data-driven model and combination model. Physical model-based methodologies generally apply mathematical models that are directly related to physical processes that have consequences on the health of the components. Thus, such models require specific knowledge and relevant theory to the system under analysis. One of the advantages of the physical models is that these can be applied for different operating conditions without the need for recollecting new data. Although these can be applied to pumps and similar machines, physical models are hard to built and require specific knowledge. An alternative to physical models is knowledge-based models. Two typical examples of knowledge-based methodologies are expert system (ES) and fuzzy logic (FL). Essentially, an ES is a computer system that is programmed to exhibit the knowledge and decision-making ability of a human expert. ES is designed to solve particular domain problems and is represented in the form of rules. Nevertheless, the conversion of domain knowledge to rules is hard to perform. FL models are capable of representing, interpreting, and utilizing data and information that are vague, ambiguous, or imprecise. This methodology contrasts with conventional Boolean logic, where truth values of variables may only be the integer values 0 or 1. In FL, the truth values of variables may be any real number between 0 and 1. Both of these knowledge-based models are more suitable to be applied to manufacturing systems and power distribution equipment. FL models are widely used in fault detection and are usually incorporated with other techniques such as ES, Kalman filter, and ANN.

Data-driven methods are classified into two categories: statistical methods and artificial intelligence methods. Essentially, the data-driven artificial intelligence approaches for prognosis are ANNs and their variants. These neural networks can establish a complex regression function between a group of network inputs and outputs through a network training procedure. The classes of ANN more interesting to be applied in machine prognostics are polynomial neural networks, dynamic wavelet neural networks, self-organizing features map neural networks, and multilayer perceptron neural networks. These approaches are of great interest for prognostics because of their potential to enhance the speed of the process and to model analytically complex systems. ANN in prognostics is mainly used as a nonlinear function approximator to predict system failure features and trends. The data-driven statistical methods reviewed by Peng et al. [33] are state-space models, such as Bayesian networks, hidden Markov models and hidden semi-Markov models, hazard rate (HR) and proportional HR, and gray model. A Bayesian network is a probabilistic graph model that represents a set of variables and their dependencies through a directed acyclic graph. A common domain application in prognostics of Bayesian models is bearings fault forecasting. To successfully implement such models, a lot of historical state transition and fatal data are needed. A Markov model is a stochastic model used to model systems that change their state randomly. Markov models assume the Markov property, that is, assume that future states depend entirely on the current state and not on past states. Hidden Markov and semi-Markov models are types of Markov models that are commonly applied in the field of prognostics of pumps. Furthermore, they can be used to diagnose the health status of

machine components. HR and proportional hazard rate (PHR) are useful indicators in lifetime analysis and are used in many applications in the prognostics of RUL. PHR model differs from HR because it allows estimating the effects of different covariates influencing the RUL of a system. HR and PHR models have been successfully applied to the prognostics of pumps and turbines. A disadvantage is that the application of HR is restricted by the assumption that components are good as new after repair. Gray system theory is able to effectively deal with incomplete data for system analysis. The Gray model (1, 1) is a time series forecasting model that has been explored in the analysis of electric systems. Nevertheless, it is a newly introduced method in prognostics that requires more research.

Since in the real-world of prognostic processes the trends of the parameters are diversified and complex to be predicted by a single method, a combination method is a more appropriate approach. For example, the application of an ANN is often incorporated with knowledge-based techniques such as ES and FL. A successful combination model usually combines two or more techniques, theories, and algorithms to model the system [33].

Figures 2.2 and 2.3 provide an overview of the structures of a selection of methodologies that were previously covered. It is possible to identify similar blocks between the different methodologies: inputs, outputs, and methods.

The mentioned works differ in methods used, systems analyzed, the purpose of the study, however, they are quite similar in what concerns the structure of the methodologies on which they are based. In fact, it is possible to perceive a generic structure of a predictive maintenance methodology from the tables in the previously mentioned figures. Figure 2.4 presents a scheme of a generic predictive maintenance methodology that will serve as a guide for the construction of the methodology adopted in this work.

Article	Inputs	Outputs	Model/Method
Predictive maintenance model development using life prediction methodology [Senthiil *et al.* 2014]	List of components of hydraulic press, their respective functions and failures Critical components rod seal and piston seal: • geometry and material; • applied mechanical loads; • critical limit for the maximum deformation. Critical component oil: • fluid properties (viscosity, acid number, water, oil cleanliness, zinc (additive), viscosity index, flash point); • caution and critical limit for viscosity.	Predicted life cycle of the critical components	FMEA to identify the critical components ANSYS software simulation to determine the life cycles until the critical limit of deformation for critical components rod seal and piston seal Systematic method of oil analysis to predict the life service of the critical component oil based on the viscosity deviation analysis
Prediction of potential failures in hydraulic gear pumps [Lisowski and Fabiś 2010]	List of gear pump components and single parts, their respective functions and potential failures	Probability of potential failures for analysed functions Classification in levels of significance of potential failures	FMEA Similarity method
Development of a predictive maintenance system for a centrifugal pump [Parrondo *et al.* 1998]	Settings of operating and abnormal conditions: • variation of the flow rate comparing to the BEP flow rate • variation of the suction head • obstacle inserted at the inlet of the pump Pump performance parameters: • pressure; • flow rate; • rotational speed of the motor; • torque Measured vibration and pressure signals: • acceleration at the pump casing in 3 orthogonal directions (axial, transverse, vertical); • pressure at the inlet and at the outlet of the pump.	Correlation between measured signals and anomalies	Analysis of acceleration and pressure spectra Derivation of numeral successions indicating the relative deviations in the power spectra at different frequencies

Figure 2.2: Overview of a selection of the addressed methodologies (A).

Article	Inputs	Outputs	Model/Method
Towards a predictive maintenance system of a hydraulic pump [Rivera *et al.* 2018]	Production environment: • OEE (overall equipment effectiveness) of the plant; • environmental temperature. Machine environment: • machine signals; • oil quality; • weekly operation diagrams. Critical component (pump): • maintenance reports; • sensor data (vibration, pressure).	Health indicators: • Deviation from the ratio of the number of events of the machine; • Error distribution of the volume flow per second and, consequently, of the pressure; • Matching score of the vibrational fingerprint for each event (state) of the pump.	Different analytical methods: • Empirical signal evaluation; • State machine; • Physical model; • Operation spreadsheets; • Oil quality control; • Vibration model.
A real-time predictive maintenance system for machine systems [Bansal *et al.* 2004]	Motor current signature Training data (generated from a simulation model)	System parameters (load, torque, backlash) to classify different motor loads to detect abnormal electrical conditions	Neural network algorithm
Artificial neural network-based classification of faults in centrifugal water pump [Farokhzad *et al.* 2012]	Pump conditions: • healthy pump; • faulty impeller; • faulty seal; • cavitation. Measured vibration signals	Identification of the condition of the pump (healthy pump, faulty impeller, faulty seal, cavitation)	Statistical analysis to extract features from the vibration signals Artificial neural network model with a back-propagation learning algorithm and multi-layer perceptron neural network

Figure 2.3: Overview of a selection of the addressed methodologies (B).

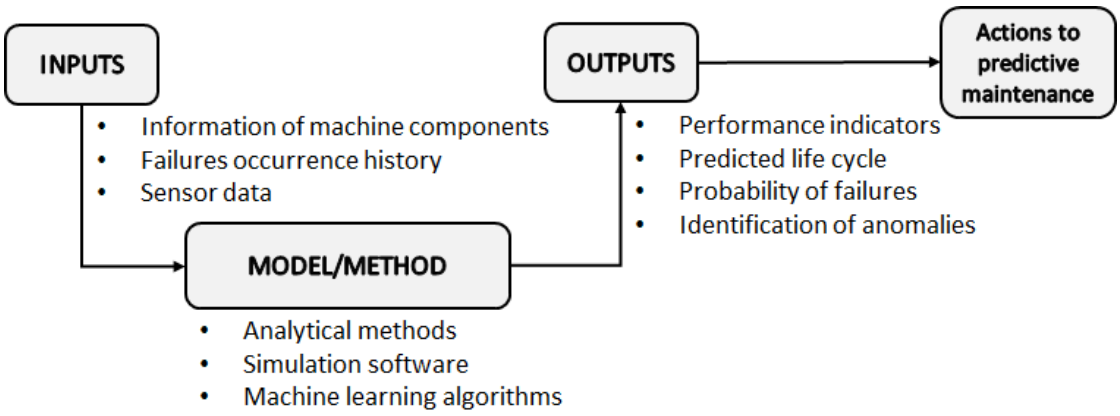


Figure 2.4: Generic predictive maintenance methodology.

Chapter 3

Methodology

In this chapter, the methodology used to develop the work is thoroughly discussed. Firstly, an overview of the logical structure of the idealized predictive maintenance process is explained, then followed by the description of the used techniques for fault detection and classification, RUL estimation, and economic analysis. Finally, a description of the tools applied in the implementation of the methodology is performed.

3.1 Predictive Maintenance Methodology Overview

In the previous chapter, with the review of the different approaches to the problem of predictive maintenance of hydraulic systems, it was possible to list, organize chronologically, and outline - in Figure 2.4 - the different common points of a generic process of predictive maintenance applied to pumps of the same nature. It was from this identified variety of inputs, scientific methods, and outputs that the final predictive maintenance methodology used in the scope of this dissertation was modeled, which is presented in Figure 3.1. and organized in three main blocks: Inputs, Tasks, and Outputs. Subsequently, a more detailed explanation of each of the blocks of the selected methodology is presented.

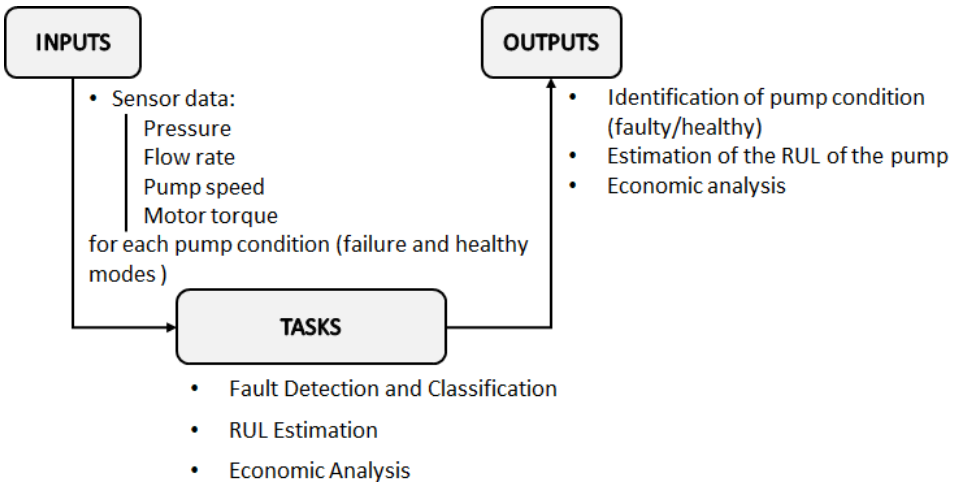


Figure 3.1: Adopted methodology.

The first stage to forecast a possible failure of a given hydraulic pump is to collect data related to its operating conditions - measurements, over a time span, of pressure, flow rate, engine torque, or even its speed (in the case of a rotational pump of variable speed) - which will later be used

as input to train models and classify any state of a similar pump from a new data set. This process is achieved through data sensors, strategically applied to the mechanism. To complete this step, it is also necessary to identify the corresponding pump condition to the collected data sensor, i.e., acknowledge if the obtained data characterizes a healthy pump - healthy mode - or a pump in a state of failure - failure mode.

Thereafter, the next stage of the selected methodology focuses on the treatment of the collected data, materialized through the fulfillment of three distinct tasks: (1) fault detection and classification, where classifiers are trained to identify different pump failures; (2) Remaining Useful Life (RUL) estimation, achieved by training models capable of estimating the remaining lifetime of the machine; and (3) an economic analysis, in order to assess the feasibility of maintenance.

Finally, having the previously presented set of tasks fulfilled, it is possible to label the conditions of any new set of sensor data of a similar pump through the achieved outputs: the pump condition (if it is healthy or at fault, and if so, which type of failure) and its respective RUL.

In the following sections, a further description of the three constituting tasks of the developed methodology is provided.

3.1.1 Fault Detection and Classification

The Fault Detection and Classification task is essentially about developing classification models, which are in their turn algorithms that are able to predict a target class based on a training data set. In classification problems, the learning is supervised since every sample in the data set is labeled with the corresponding class.

The adopted sub-methodology for this task consists of the following steps:

a) Data Processing

Data analysis is the first step to build a model for condition monitoring and predictive maintenance. In order to successfully design classification algorithms, it is necessary to organize and analyze a considerable amount of data and to track the condition the data represent.

Firstly, data were acquired and imported, and then prepared into a preferable configuration to handle the data. Frequently, it is necessary to clean and transform the data allowing the extraction of condition indicators. Specifying, data pre-processing can include:

- Outlier, and missing value and offset removal;
- Noise reduction, such as filtering or smoothing;
- Transformations between time and frequency domain;
- More advanced signal processing such as short-time Fourier transforms and transformations to the order domain.

This step typically results in a cleaned and transformed signal, able to allow further analysis to condense the signal information into condition indicators.

b) Feature Extraction

By definition, a feature is a distinctive characteristic or attribute of an object. Therefore, a condition indicator is here defined as a data feature that is useful to distinguish different behaviors as the system degrades or to discriminate between different operational modes. Features are the elements that will enable the classification of the different faults in the model training step.

As stated in the previous step, transformations between time and frequency domains are executed to be possible to obtain condition indicators. Simple statistical features can be derived from time-domain signals, such as mean, standard deviation, root-mean-square, and shape factor. High order moments of signals, particularly skewness and kurtosis, and impulsive metrics, such as peak value, impulse factor, crest factor, and clearance factor, can also be obtained from

time signals. Additionally, signal processing metrics are also extracted from signals in the time-domain. Such metrics include signal to noise ratio (SNR), total harmonic distortion (THD), and signal to noise and distortion ratio (SINAD). The features that are extracted from frequency-domain signals are power bandwidth, peak amplitudes, peak frequencies, natural frequencies, and damping factors. The mentioned features are illustrated in Figure 3.2.

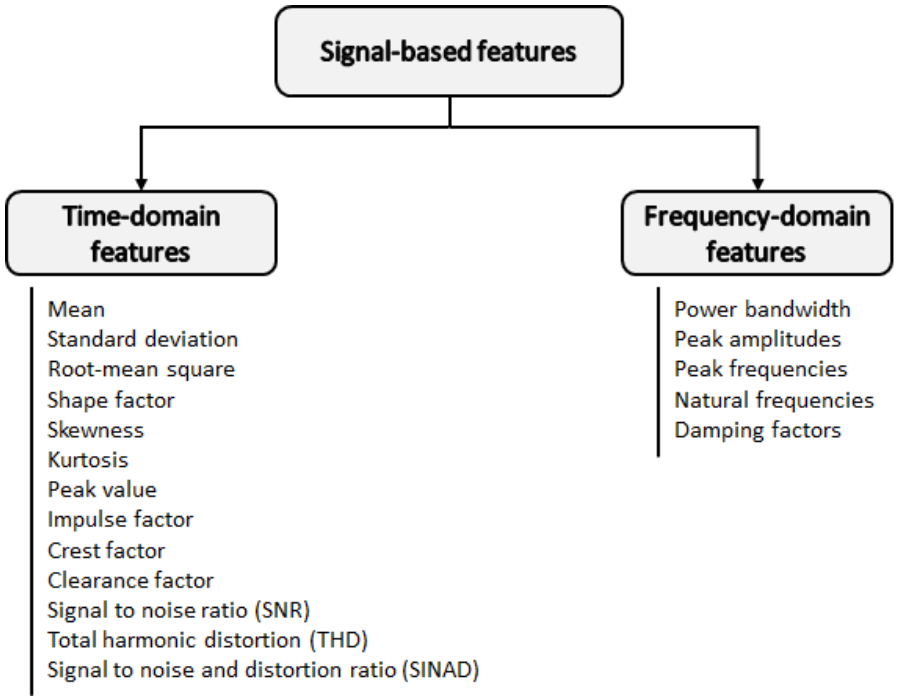


Figure 3.2: Signal-based features.

Although a considerable amount of features can be derived from data, not all of them are suitable to distinguish between the different fault types. So, in order to obtain relevant features to the intended purpose, it is necessary to rank them and select those with better classification. To achieve this purpose, it is possible to use classification ranking methods that allow to score and rank features by importance, enabling to assess how effectively each feature separates data with different condition labels. The ranking method chosen from the ones available in the literature and used software libraries, mentioned later in section 3.2, was the one-way analysis of variance, also known as one-way ANOVA.

Analysis of variance (ANOVA) is a statistical method used to develop and support explanations concerning the observed data. The variance of a feature measures how much impact that feature has on the response variable: if the variance is low, it suggests that there is a low impact of the feature on the response; otherwise, if the variance is high, there is a high impact of the feature on the response [34]. The ANOVA assigns sample variance to data groups and determines whether the variation arises within or among different groups. Variations around group and variation of group means around an overall mean are used to characterize samples. In case that variations within groups are limited comparing to variations between groups, a difference in group means may be inferred [35]. In essence, ANOVA tests for the difference in the group means by partitioning the total variation in the data into two components, variation between groups and variation within groups [34]. One-way ANOVA compares the variance in group means within a sample whilst considering one independent variable [36].

Following the ranking classification according to the one-way ANOVA method, the features are assigned with an ANOVA metric. This metric scores the capability of the features to distinguish different condition labels.

c) Model Training

In the model training step, different models are trained with the extracted features. Six different types of models are available in the selected software, in which are included decision trees, discriminant analysis, support vector machines, nearest neighbors, naive Bayes, and ensemble classification. Each type of these models has several subtypes that may differ in the inherent parameters to each model, or even in its structure. Subsequently, a more detailed explanation of each mentioned family of models is given.

As the name suggests, the decision tree learning process uses a decision tree, i.e., a decision support tool based on the structure of a tree as a predictive model. This structure is composed of multiple internal nodes that split into branches and end in terminal nodes, named leaves. Each internal node, branch, and leaf represents a test on an attribute, the outcome of the test, and the target class label, respectively [35, 37].

Discriminant analysis is a method used for data reduction, pattern recognition, and classification. It works by developing linear combinations of independent variables (predictors) that will discriminate between the different categories of the dependent variables. These linear combinations are functions - named discriminant functions [35, 38].

Support vector machines (SVM) are a group of related supervised learning methods that can be applied for regression analysis and classification. To classify data, an SVM finds the best hyperplane that separates data points from the two classes. This hyperplane is the one that is positioned with the maximum distance to the data points. In its turn, support vectors are the data points with the minimum distance to the hyperplane. Although SVM are generally applied to binary classification, they can also be used in multiclass classification. In such cases, the method consists of reducing the multiclass classification problem into multiple binary classification subproblems, each one with an associated SVM learner [35, 39].

The nearest neighbors (KNN) algorithm is a non-parametric method that is usually applied in regression and classification problems. The working principle of nearest neighbors classifiers is to find a set of the closest training samples - the nearest neighbors - to a data point and, from those samples, predict the label for that point. Thereafter, the classification of the data point is determined by the majority of labels of their nearest neighbors. In order to compute the distance, different metrics can be used [35, 40].

Naive Bayes classifiers are a family of classification algorithms based on Bayes's Theorem. Such theorem determines the probability of an event occurring given the probability of another event that has already occurred. The fundamental assumption of Naive Bayes classifiers is that each feature delivers an independent and equal contribution to the outcome [35].

Ensemble classifiers use multiple learning models to build one high-quality ensemble model with better predictive performance than each learning algorithm alone. The goal of ensemble classifiers is to reduce bias and variance. Two different principles to build an ensemble model are possible: the bagging method, which consists of building multiple models independently and average their predictions, resulting in a generally better-combined model since the variance is reduced; and the boosting method, which builds models sequentially, where each one of them reduces the bias of the combined model, with the inherent goal of combining multiple models to create a powerful ensemble [35, 41].

After training the different models, it is necessary to compare them with each other by assessing their performances. A simple way to assess the performance of a classifier is to compute its accuracy score. When using the cross-validation technique to divide the data into a training and validation set, the accuracy score corresponds to the one of all observations on validation data compared against the training data.

To assess how the classifier performs on each class, a useful tool to take advantage of is the confusion matrix. This matrix consists of a table layout that allows visualization of the performance of a classifier. Typically, the rows of the matrix represent the true class and the columns represent the predicted class. An abstract example of a simple confusion matrix is illustrated in Figure 3.3 [42], where P represents the positive class and N represents the negative class.

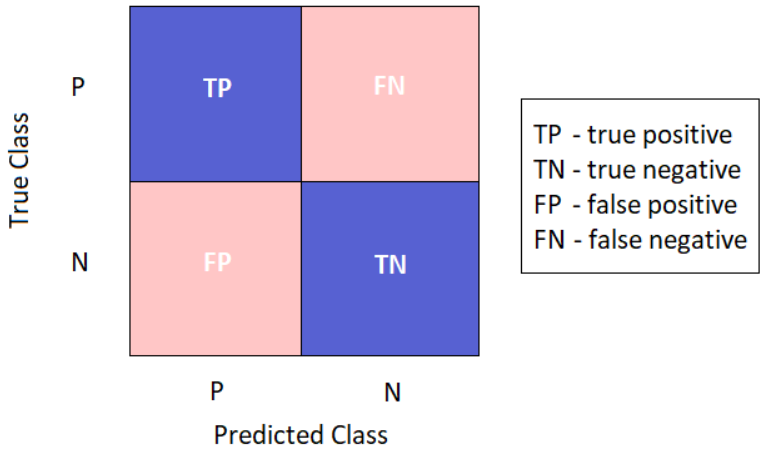


Figure 3.3: Example of a simple confusion matrix (adapted) [35].

The correct predictions are located in the diagonal of the table, i.e., the diagonal cells of the matrix show where the true class and the predicted class match. The confusion matrix can be computed in various forms: as the number of observations per predicted class versus the number of observations per true class; as the true positive rates (TPR) (Equation 3.1) - the proportion of correctly classified observations per true class - versus the false negative rates (FNR) (Equation 3.2) - the proportion of incorrectly classified observations per true class; and also as the positive predicted values (PPV) (Equation 3.3) - the proportion of correctly classified observations per predicted class - versus the false discovery rates (FDR) (Equation 3.4) - the proportion of incorrectly classified observations per predicted class. There is another value named as the false positive rate (FPR) (Equation 3.5), used in the metric receiver operating characteristic (ROC), that represents the proportion of the number of incorrectly classified observations and the number of observations of all classes except the predicted class.

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{3.1}$$

$$FNR = \frac{FN}{P} = \frac{FN}{TP + FN} \tag{3.2}$$

$$PPV = \frac{TP}{TP + FP} \tag{3.3}$$

$$FDR = \frac{FP}{TP + FP} \tag{3.4}$$

$$FPR = \frac{FP}{N} = \frac{FP}{TN + FP} \tag{3.5}$$

A receiver operating characteristic (ROC) curve is another useful method to visualize the performance of a classifier since it shows the true and false positive rates for each class. An example of a ROC curve is presented in Figure 3.4. By the example given in the mentioned figure, an FPR of 0.04 indicates that the classifier assigns 4% of the observations incorrectly to the positive class, while a TPR of 0.47 indicates that the classifier assigns 47% of the observations correctly to the positive class. The area under the curve (AUC) is a measure of the overall quality of the classifier for the positive class. A large AUC value indicates a good classifier performance, having the perfect classification with the value of 1. In the case of the example of Figure 3.4, the AUC has the value 0.92.

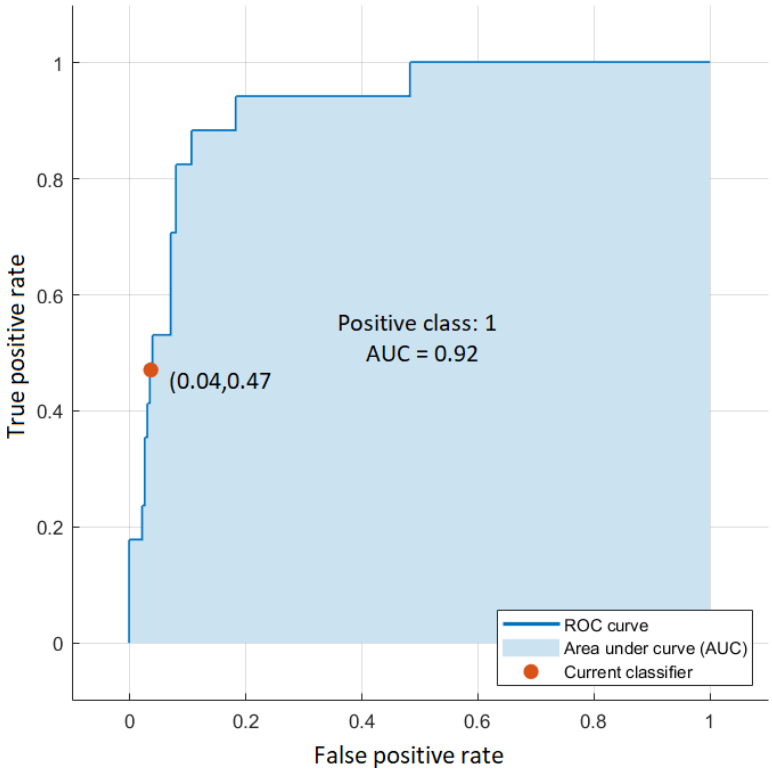


Figure 3.4: Example of a ROC curve [36].

3.1.2 Remaining Useful Life Estimation

Following the detection and classification of a machine condition, another important task towards a predictive maintenance methodology is the prediction of the remaining useful life (RUL) of the system. By definition, the RUL of a machine is the expected lifetime remaining before the machine reaches failure and requires maintenance or replacement.

The type of system data available is important for RUL estimation since it is a determining factor to be taken into account when choosing the model type for computation. Furthermore, system data can be a run-to-failure history of the machine, a known threshold value of a condition indicator that indicates failure or lifetime data. In the context of the present work, the needed data type for the adopted RUL estimation method is the run-to-failure history with sensor measurements.

The adopted sub-methodology for the task of estimating the RUL of a pump consists of the following steps:

a) Data Processing

Firstly, the data measurements are imported and, subsequently, the data set is divided into a training and a validation set, for later assessing the model performance.

Generally, sensor measurement data have associated noise and, hence, a noise reduction and data smoothing technique is applied to the sensor data. In this case, a moving average was used. Furthermore, smoothing time series data helps reveal underlying trends in the data. The smoothing is performed with a centered moving average with a window size in accordance with the length of the data. Smoothing is performed both on training and validation data sets.

The following approach of data pre-processing is data normalization, The data is normalized by computing the mean and standard deviation of each sensor for the entire training data set. The approach used to normalize the data is standardization, which transforms data to have a

resulting mean equal to zero and a resulting standard deviation equal to one. The standardization technique is also referred to as z-score and is computed using:

$$z_i = \frac{(x_i - \mu)}{\sigma}, \tag{3.6}$$

where x is the raw value, μ is the sample mean and σ is the sample standard deviation. The same process of data standardization is applied to the validation data set using the mean and standard deviation of the training data set.

b) Sensor Trendability and Sensor Fusion

The data set has measurements from several sensors but not all of them are suitable to incorporate the RUL model. Since only a group of the sensors possibly contributes to the fault development, those appropriate sensors for estimating the remaining lifetime until failure must be identified and selected. This section of the model is referred to as trendability analysis, where the sensors are selected based on their trends. A linear degradation model is estimated for each sensor of the training data set. The linear degradation model used follows [43]:

$$S(t) = \phi + \theta(t)t + \epsilon(t), \tag{3.7}$$

where ϕ is the model intercept and is a constant, $\theta(t)$ is the model slope and is modeled as a random variable with a normal distribution, $\epsilon(t)$ is the model additive noise and is modeled as a normal distribution with zero mean, and $S(t)$ represents the sensor signal [35]. The mean value of the parameter θ determines the trendability of the sensor. For the training data set, one slope per sensor is computed. The slopes of the signals are ranked. The sensors measurements with higher slope values are the ones more suitable to be used in RUL estimator models. A group of T sensors with higher slopes are selected to construct a health indicator. The selected sensors are also referred to as the trended sensors. The number T of sensors to be selected is determined by an iterative process. Several values of T are computed and the best number of sensors is decided in accordance with the RUL estimation results.

To build an RUL model, one single indicator is needed. Hence, a fusion of the selected sensors into a single health indicator is performed. The run-to-failure data is assumed to start at a healthy condition and to be degrading until it reaches failure. The health condition of the machine at the start is assigned a value of one and is assumed to be linearly degrading until the end, which is assigned a value of zero. This characterizes the theoretical behavior of the health condition. To approximate the theoretical behavior to real behavior, the linearly degrading health condition is adjusted with the trended sensors.

A linear regression model fits the health condition with the trended sensors as regressors. A multivariable regression is defined as:

$$h = w_0 + (\sum_{z=1}^{T} w_z S_z) + \epsilon, \tag{3.8}$$

where w_0 is the intercept term and it is a constant, the other w parameters are the regression coefficients or also referred to as weights of each of the variables S and ϵ is the error term [35]. In this analysis, the health condition is the h variable and the trended sensors are the S variables. As a result of adjusting the linear regression of the condition indicators with the trended sensors, each sensor has a corresponding weight. The weights are used to construct the real health indicator adjusted to the sensor data. The health indicator is constructed by multiplying the trended sensors' measurements with their associated weights. The linear regression and, therefore, the sensor weights, are computed concerning the sensors' training data. The same process of sensor fusion into a health indicator is applied to the validation data set. The trended sensors to be considered in the validation data are the trended sensors selected from the training data and the weights to be used to fuse the trended sensors are the already computed weights of the training data.

c) Train a RUL Estimator Model

To estimate the RUL, a residual comparison based similarity model is used [35]. Different deterministic models can be used for residual generation such as regression models (linear, exponential, polynomial) and auto-regressive moving average (ARMA) models [44]. The type of model to be selected is data-dependent: if a linear degradation pattern can be clearly distinguished in the data, a linear model is suitable to directly fit the data; the same applies with other degradation patterns and corresponding models. Furthermore, if no appropriate pattern is observed, ARMA models can be used [44, 45].

A distance measure is used to assess the similarity between the health indicators' behavior. Various distance functions can be computed. A simple definition of distance using the 1-norm of the residual is used. The distance between data i and j is calculated as:

$$d(i,j) = \|h_j - \hat{h}_j, i\|, \tag{3.9}$$

where h_j is the health indicator of data sample j and $\hat{h}_j, i$ is the estimated health indicator of data j using the model identified in data i.

The similarity level is evaluated using the similarity score defined as [35]:

$$score(i,j) = e^{-d(i,j)^2} \tag{3.10}$$

For RUL estimation of a given validation data sample, the model finds the nearest k training data samples and fit a probability distribution based on the k data samples. The number of nearest data samples k is chosen according to the length of the dataset. The model estimates RUL as the difference between the median of the probability distribution and the current lifetime value of the data sample to be estimated.

3.1.3 Economic Analysis

After the presentation and description of the tasks that constitute the predictive maintenance methodology proposed in this dissertation, the question arises of the potentiality of the benefit in implementing such actions. Thus, the need arose to outline a simple economic analysis that contemplates generalized costs and/or savings with the application of the methodology. More specifically, this economic analysis aims to assess the feasibility of applying a predictive maintenance model that includes RUL estimation.

The reasoning used in this analysis is based on the comparison of the generic costs associated with the application of the knowledge extracted from the RUL estimation model with the generic costs of not applying maintenance activities until failure. Although in reality in the operation of pumps some preventive maintenance tasks are performed even without a prediction of when the failure will occur, these are not taken into account in this analysis[1].

As the name implies, in the elaboration of the generic economic analysis, generic RUL values were used and it was assumed that the model will estimate the RUL with a certain deviation from the real value. The prediction of the RUL model will most certainly not coincide with the real value, thus, it was made the conservative assumption that the RUL model will estimate values between a 20% lower and higher deviation from the real values.

Figure 3.5 shows the evolution of the true RUL and the estimated RUL. The line in Figure 3.5 that illustrates the evolution of the true generic RUL starts at the value 100, which coincides with the start of the pump operation (time = 0), and decreases linearly until reaching the value 0, coinciding with the time of the pump failure (time = 100) - red line in the figure. The two blue dashed lines in Figure 3.5 illustrate the lower and upper limits of the range of predicted values as estimates of the RUL model. The lower limit line has a negative deviation of 20% with the

[1]It was not possible to collect values from a real scenario of a pump in operation, as well as it was not possible to have access to concrete values of costs associated with the operation of the pump and of consequent failures and costs of the arrangements. Although it may call into question the adequacy of the assumptions and considerations decided in the present work, it was considered pertinent to demonstrate this economic analysis.

true RUL line and the upper limit line has a positive deviation of 20% with the true RUL line. The gray shaded area corresponds to the range of possible predictions of RUL values, according to the assumed estimate.

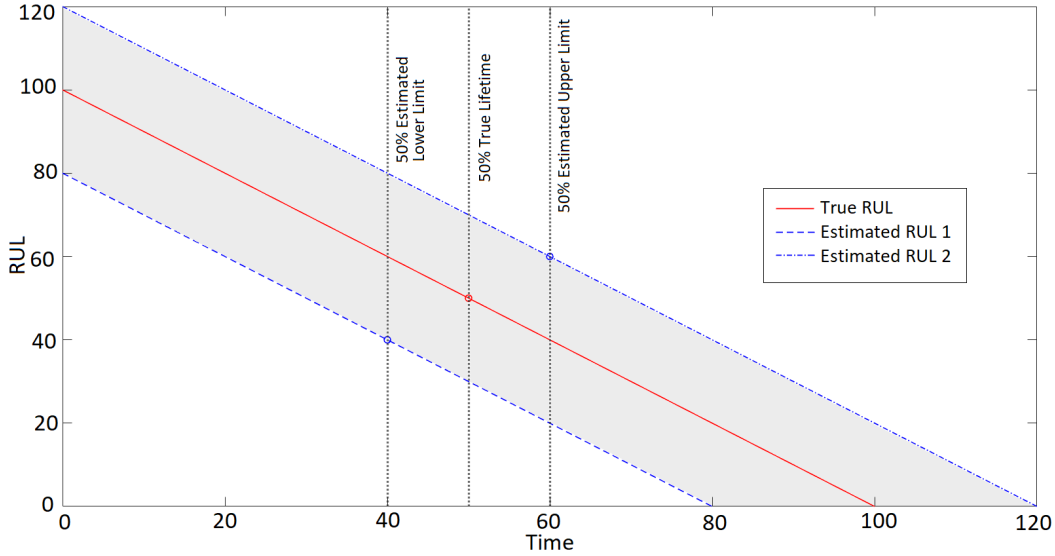


Figure 3.5: Generic RUL: Estimated RUL 1 - 20% lower deviation, Estimated RUL 2 - 20% higher deviation.

It was defined that the moment to make decisions and perform maintenance actions to prevent the failure predicted by the model is half the life of the pump, which corresponds to the moment when the time that the pump has already operated equals the remaining time until failure.

The chart in Figure 3.5 shows three vertical dashed lines whose intersection with the respective RUL line indicates the 50% RUL point. Each of the three lines corresponds to 50% of the RUL of the lower and upper limits, and 50% of the true RUL.

Having defined 50% of RUL as the moment for applying maintenance actions, the associated costs were compared. Three scenarios were outlined: (i) one where the pump operates normally until the moment of failure is reached and, later, that failure is repaired and the pump is put back into operation, and (ii) the other two scenarios correspond to situations in which the pump operates normally up to the moment of 50% RUL, where maintenance activities are applied to prevent failure and, thus, prolonging the life of the pump beyond the estimated. The difference between the last two scenarios is that in one of them the 50% of RUL corresponds to the lower prediction limit of the estimated RUL - generic value 40 - and in the other corresponds to the upper prediction limit of the estimated RUL - generic value 60. These three scenarios are illustrated in Figure 3.6. The red lines represent the generic cost curve associated with the operation of the pump without predictive maintenance applied - first scenario. The dark green and light green lines reflect the generic cost curves associated with the operation of the pump with maintenance actions applied at the time of 50% of estimated RUL, lower and higher, respectively, resulting from the predictions obtained with the RUL estimation model. The dark green and light green lines reflect the generic cost curves associated with the operation of the pump with maintenance actions applied at the time of 50% of estimated RUL, lower and higher, respectively, resulting from the predictions obtained with the RUL estimation model. The three lines start at zero time and with a cost associated with the acquisition cost (Ca) of the pump. The cumulative cost increases over time and the slope α associated with this increase corresponds to the pump's operating costs, which, fundamentally, are the costs related to energy consumption. The slope α depends on the pump's power, the energy tariff, and also on the pump's operating time. Strictly speaking, the slope α should increase incrementally over time due to the inherent decrease in

pump efficiency over the time of operation [8].

Given an acquisition cost (Ca), the cost of major failure (Cf) will be $Cf \approx 0.3Ca$, and the cost of maintenance actions (Cm) will be $Cm \approx 0.05Ca$. These proportions result from assumptions and considerations, and that, although the variables involved depend on several factors such as the type of pump and the nature of the failure and can assume wide ranges of values, they seem reasonable considering the literature [6, 46, 47].

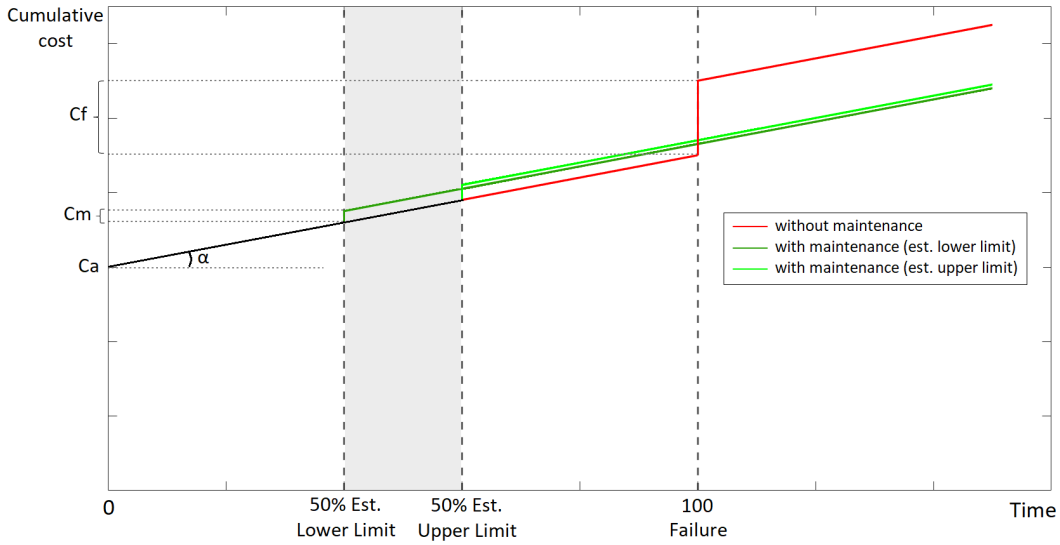


Figure 3.6: Cost comparison between the three scenarios.

Since maintenance costs increase as the time to failure decreases [48], the cost of maintenance actions associated with the upper limit of the RUL estimate is slightly higher than the cost associated with the lower limit of RUL.

The costs considered in this analysis refer to generic failures and do not contemplate the associated cost of a specific failure neither their necessary preventive actions to avoid it. The advantage of anticipating the failure event may be greater or lesser depending on the nature of the failure. Nevertheless, with the elaborated analysis, it is noticeable the saving of time that allows the planning of maintenance actions and the reduction of the associated costs, thanks to the RUL prediction of the pump.

3.2 Implementation

All parts constituting the methodology previously described were implemented in Matlab®. The used methodology in this dissertation for RUL Estimation was inspired in several methods described in the Predictive Maintenance Toolbox™ documentation of Matlab [49]. Moreover, two apps from the software were used for the graphic and interactive implementation of the steps of the Fault Detection and Classification task explained in 3.1.1. Both data processing and feature extraction steps were implemented in the Diagnostic Feature Designer app [35]. This app allowed to compute and explore the data signals, the transformations between time and frequency domains, and the feature computation and ranking, all using a multi-function graphical interface. The model training step was implemented using the Classification Learner app [35], which allowed to test several types of classifiers, and, also, enabled to automate the parameters optimization and to compute and visualize interactively the results.

Chapter 4

Results and Discussion

In this chapter, two case studies are presented and described for each of the techniques explained in the previous chapter. Following the description of each case study, the results and the respective discussion are also presented. Finally, the economic analysis defined in the previous chapter is applied to one of the sets of results obtained in one of the case studies.

4.1 Case studies - Fault Detection and Classification

Given the impossibility of collecting experimental data, a search in online repositories aiming to find hydraulic pumps data sets was performed. Data sets were selected from Predictive Maintenance Toolbox examples of Matlab [49] due to the nature of the data, which come from hydraulic pumps, and because of the suitability of the data sets on implementing the adopted methodology. Two data sets retrieved from examples of the referred toolbox documentation were used, indicated next as Case Study 1 and Case Study 2.

4.1.1 Case Study 1

The first case study to be presented consists of a set of data from a triplex reciprocating pump which was collected from a Simulink model. The pump data contains 240 flow and pressure measurements for different fault conditions. Additionally, each measurement contains 1201 observations, which are organized in a time interval from 0 seconds to 1.2 seconds, meaning that each observation was recorded every 0.001 seconds. Despite not influencing the following analysis, it is important to note that these values do not seem to reflect real-time values: it is not usual to record, in a pump, considerable variations in flow or pressure every 0.001 seconds. Such can be explained by the fact that the data set comes from a pump simulation model. An attempt was made to ascertain the actual periodicity of the data, but such information was not available at the source of the data.

The simulation model that originated the data was configured to include three types of failures: leaking pump cylinder, blocked pump inlet, and increased pump bearing friction. It also includes different degrees of severity for each fault. The data covers conditions with no faults, all faults, and combinations of one or two faults, so, there are eight fault modes. The fault modes and corresponding code used to identify the condition of the data are presented in Table 4.1.

The pump specific speed is 950 rpm and its nominal pressure and nominal flow are 7 bar and 38 lpm (2.28 m^3/h), respectively.

The following steps concern data processing and feature extraction, which, as already mentioned, were implemented using the capabilities of the software of the Diagnostic Feature Designer app from Matlab.

Table 4.1: Fault codes used to label the data in case study 1.

Fault mode	**Label**
No fault (healthy pump)	0
Leaking cylinder	1
Blocked inlet	10
Leaking cylinder and blocked inlet	11
Bearing friction	100
Bearing friction and leaking cylinder	101
Bearing friction and blocked inlet	110
All faults (leaking, blocked and bearing)	111

Figures 4.1 and 4.2 present the flow curves and pressure signals, respectively. In each of the figures, the upper subfigure shows several samples of curves concerning all fault modes, and, for clearer observation of the behavior of a curve, the lower subfigure shows one of those samples isolated, which refer to healthy mode. The different fault modes are not distinguishable by simply observing the signals, hence, time-domain features were extracted from flow and pressure signals. For both signals, all the thirteen time-domain features listed in Figure 3.2 were extracted.

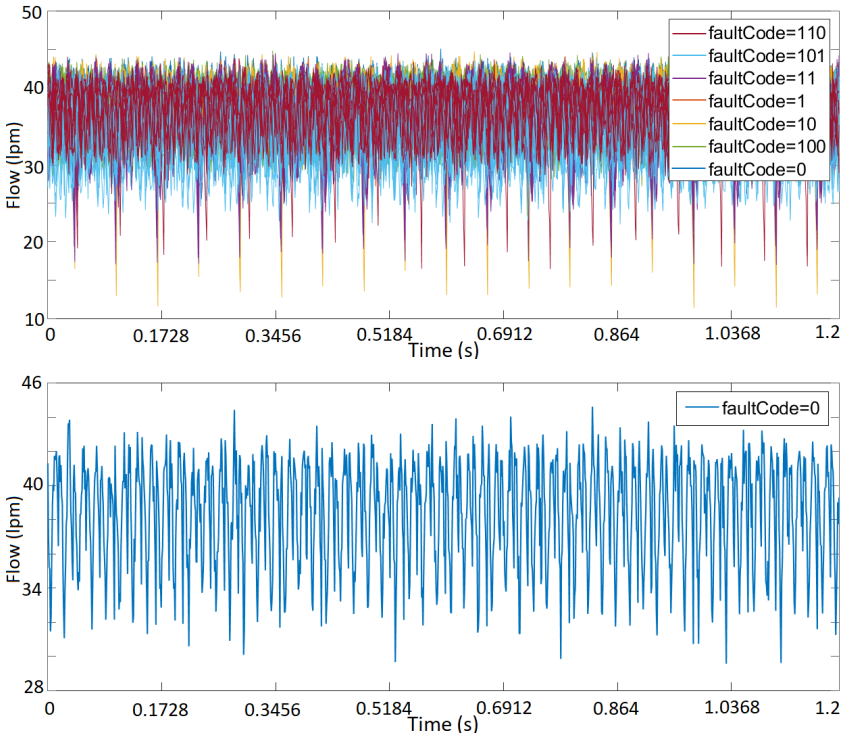


Figure 4.1: Flow signal trace - case study 1.

After extracting the time-domain features, the frequency-domain features were obtained. For this purpose, the frequency spectra of both signals were computed using an auto-regressive model (which was selected from the models available in the software) [35]. Figures 4.3 and 4.4 present the power spectrum of both signals in the frequency-domain, from which all the frequency-domain features listed in Figure 3.2 were extracted. Each of the frequency-domain features - except the power bandwidth - is computed for each spectral peak. In order to obtain relevant frequency-domain features, it is necessary to select a frequency band that contains well-defined

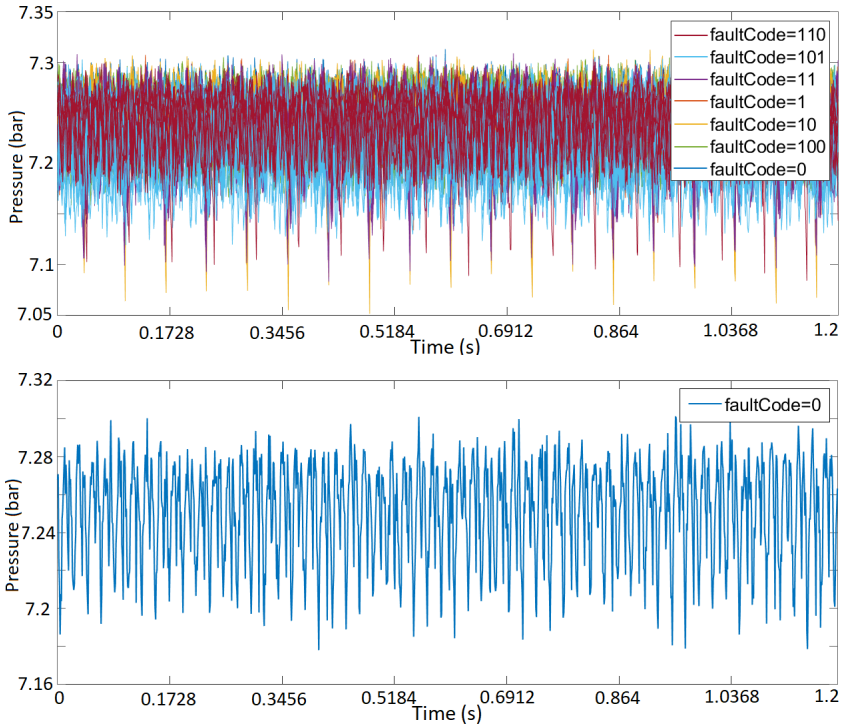


Figure 4.2: Pressure signal trace - case study 1.

and distinguishable peaks. For both signals, it was defined a frequency band from 25 to 275 Hz containing the first five peaks.

In Figures 4.3 and 4.4, the upper subfigures present several samples of curves concerning all fault modes, and the lower subfigures present one sample of a curve of healthy mode with the five peaks marked with yellow points, as well as the frequency band (25-275 Hz) highlighted in a red shaded area.

After the extraction of time-domain and frequency-domain features, the most relevant ones for the purpose of classifying the fault modes were selected. Of the methods available in the software, the ANOVA method was used to classify the features, which resulted in a feature ranking with 68 features with ANOVA metrics ranging from 0 to 132. Among the features that obtained the highest ANOVA metrics, the following stand out: root-mean-square of flow and pressure signals, first natural frequencies of flow and pressure power spectra, first peak amplitudes of flow and pressure power spectra, and mean of flow and pressure signals. A total of 49 features with ANOVA metric greater than 10 were selected to train the classification algorithms. As previously mentioned, the algorithms available in the Classification Learner app were used to implement them with the selected features.

The approach followed for this step was: train all available models without further improving the parameters associated with each algorithm, analyze the models and select the ones that stand out for the best performances, and finally, optimize the parameters intrinsic to the algorithms that obtained the best results. Initially, all models were trained with predefined parameters. Table 4.2 presents the results of a selection of algorithms for an average of 10 analyzes. The algorithms based on the method of discriminant analysis were not able to produce acceptable results because some of the classes have singular covariance matrices for the values of the features, which caused the algorithms to fail. For several algorithms, the results obtained are not satisfactory, as is the case of the SVM algorithms and the KNN algorithms, which obtained accuracy values below or close to 50%, and the Naive Bayes algorithms, which obtained accuracy values near 67%. The

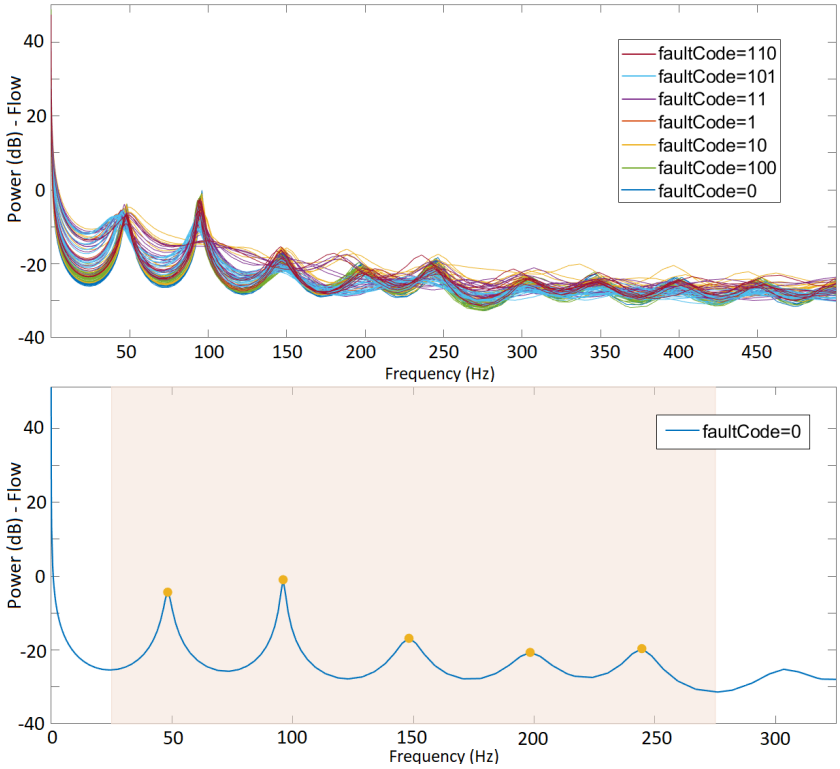


Figure 4.3: Flow power spectrum - case study 1.

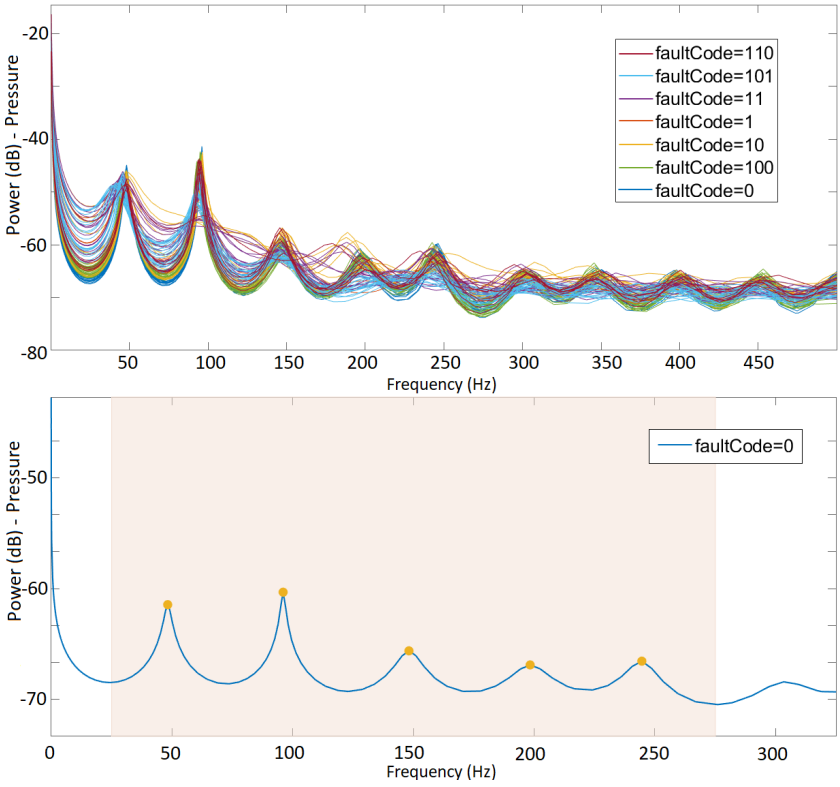


Figure 4.4: Pressure power spectrum - case study 1.

best results were obtained for models of the types of ensemble classifiers and decision trees, with values above 73%. The justification for these results may lie in the fact that the evaluation and classification structure of these algorithms is the most suitable for the data of the case study under analysis. If the data were related to variables other than those in the data set under analysis - flow and pressure - or included data from other variables in addition to these, the quality of the results obtained for the various types of algorithms would be different. Moreover, the results would also be different if other groups of features had been used to train the algorithms, however, since the case study under analysis presents data for only two variables, it was not sought to further reduce the number of features used for the implementation of the algorithms as the results could decrease in quality due to the reduction of the amount of data.

Using the capabilities of the software, the parameters chosen for the algorithms that obtained the best results were optimized using hyperparameter optimization [35]. This technique is used to automate the selection of the parameter values. For the given model types, the software tries different combinations of parameter values and returns a model with the optimized parameters, i.e., parameter values that minimize the model classification error.

Table 4.2: Results of a selection of algorithms for an average of 10 analyzes - case study 1.

Model	**Accuracy**	**Model**	**Accuracy**
Fine Decision Tree	78.3%	Medium Decision Tree	76.3%
Linear Discriminant	Fail	Quadratic Discriminant	Fail
Gaussian Naive Bayes	67.8%	Kernel Naive Bayes	66.2%
Linear SVM	50.1%	Quadratic SVM	52.6%
Fine Gaussian SVM	30.3%	Medium Gaussian SVM	49.7%
Fine KNN	48.7%	Medium KNN	45.8%
Cubic KNN	46.2%	Weighted KNN	48.5%
Ensemble Subspace KNN	73.4%	Ensemble Subspace Discriminant	77.1%
Ensemble Boosted Trees	80.1%	Ensemble Bagged Trees	80.8%

The best results of the hyperparameter optimization were obtained for an ensemble of decision trees with the ensemble method bagging, having reached an accuracy of 82.9%. In Figure 4.5 the results of the hyperparameter optimization for the ensemble classifier are illustrated. The best results for the parameters were an ensemble with 24 learners of the type of decision trees, a maximum number of splits of 54 with an associated minimum classification error of 0.163. Moreover, such results were achieved using only 44 features as predictors of the model - the best 44 features with ANOVA metrics greater than 15. The features that were not included in this group are: damping factors of peaks 2, 3, 4, and 5 of flow and pressure spectral signals, natural frequencies of peaks 2, 3, 4, and 5 of flow and pressure spectral signals, peak frequencies of peaks 3, 4 and 5 of flow and pressure signals, and kurtosis of flow and pressure time-domain signals. Therefore, in general, for frequency-domain features, only the features that concern the first peak are needed to perform fault classification.

As mentioned, the best results obtained with this model were for an overall accuracy of 82.9%. Regarding the performance of the algorithm in each class, such results are presented in the confusion matrices in Figures 4.6 and 4.7. In Figure 4.6, the values of the confusion matrix refer to the number of observations for each predicted class and true class, whereas in Figure 4.7, the values of the confusion matrix refer to TPR (true positive rate) and FNR (false negative rate) for each true class.

Analyzing the two figures for, as an example, label 0 (healthy mode) it can be concluded that:

- 44 observations were assigned to class 0, 42 of which were correctly attributed and 2 of these observations were attributed incorrectly, because in reality they correspond to class 10 (Figure 4.6);

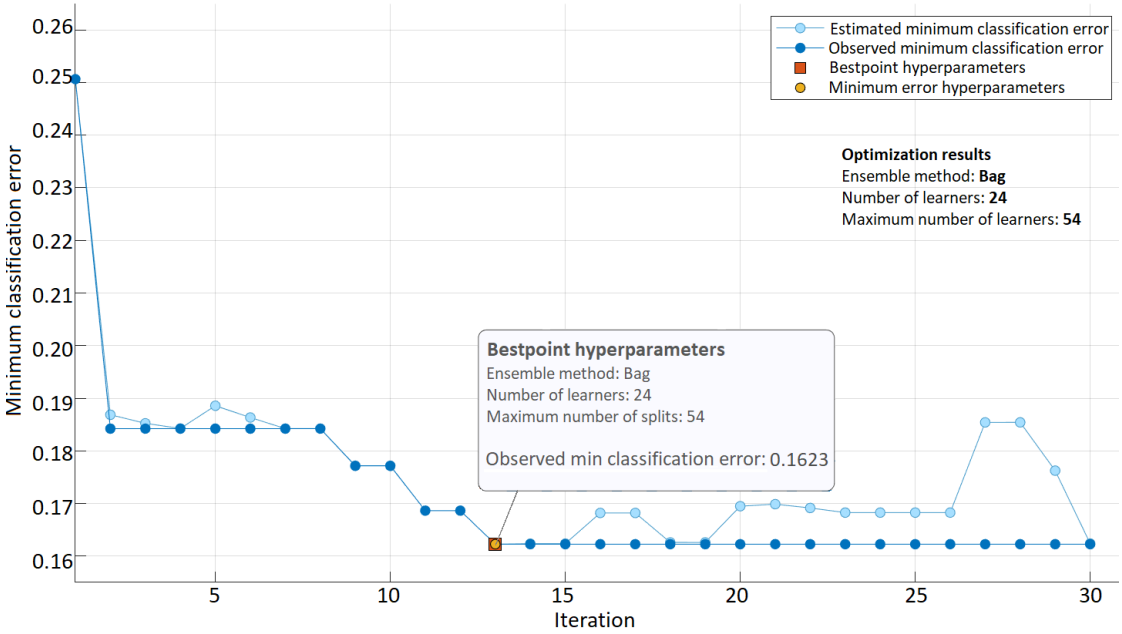


Figure 4.5: Minimum classification error plot for the ensemble classifier.

- No measurements relative to class 0 were assigned to another class, that is, all measurements corresponding to class 0 were correctly assigned to that class (the algorithm/model correctly classifies 100% of healthy mode observations) (Figure 4.6);
- The 2 observations are a false positive of class 0 and, therefore, a false negative of class 10, which corresponds to 4.9% of the class 10 observations to be assigned to class 0 (Figure 4.7);
- The 42 observations are a true positive of class 0, therefore, the TPR of class 0 is 100%, $TPR = \frac{42}{42} = 1$ (Figure 4.7);

Regarding label 11 (leaking cylinder and blocked inlet), for example, it can be concluded that:

- There are a total of 17 measurements belonging to class 11, of which only 9 of these measurements were correctly identified, and of the remaining 8, 4 of them were incorrectly assigned to class 1, 1 was incorrectly assigned to class 110, and 3 were incorrectly assigned to class 111 (Figure 4.6);
- Several measurements were incorrectly classified as belonging to class 11 and, in reality, 3 of these measurements correspond to class 1, 1 corresponds to class 10, 1 belongs to class 101, 1 belongs to class 110, and 3 correspond to class 111 (Figure 4.6);
- The 9 observations that belong to class 11 and were correctly assigned to this same class constitute the true positives of the class, therefore, the TPR of class 11 is 52.9%, $TPR = \frac{9}{4+9+1+3} = \frac{9}{17} = 0.529$ (Figure 4.7);
- The remaining observations that belong to class 11 and were incorrectly assigned to other classes make a total of 8 observations incorrectly classified, so, the total FNR associated with class 11 is 47.1%, $FNR = \frac{8}{17} = 0.471$ (Figure 4.7);
- The observations that were incorrectly attributed to class 11 constitute the FNR of the other classes to which they truly belong.

From the confusion matrix with the number of observations (Figure 4.6), the obtained accuracy can be confirmed: the quotient between the sum of correctly assigned observations (diagonal entries) and the total number of observations, reflects the model's accuracy, which is 82.9% ($\frac{42+24+37+9+35+20+21+11}{42+27+41+17+40+24+32+17} = \frac{199}{240} = 0.829$).

Figure 4.8 shows the ROC curves for classes 0 and 11, which results were previously explained. Taking class 0 as the positive class, the FPR of 0.01 indicates that observations that were incorrectly assigned to class 0 correspond to 1% of all observations that were not assigned to such class, and the TPR of 1 indicates that all observations that correspond to class 0 were correctly classified. The AUC for this class is 1, which means that the classifier is perfect at distinguishing the healthy mode. Regarding class 11 as the positive class, the FPR is 0.04, which indicates that the observations that were incorrectly classified as belonging to class 11 correspond to 4% of all observations that were not attributed to this class, and the TPR of 0.53 indicates that only 53% of class 11 observations were correctly classified. The AUC for this class is 0.93, which, despite being a high value, was the lowest obtained AUC value among all classes.

In general, the model satisfactorily identified classes 0, 1, 10, and 100, which correspond to the observations regarding the healthy operation of the pump and observations regarding only each of the three failures, and presented more difficulties in correctly classifying the observations regarding two or three simultaneous failures. Better accuracy results could have been achieved if there were more sets of measurements for failure modes. In fact, the classes in which the algorithm had the worst performance are those with the least amount of observations associated: the classes referring to two or three simultaneous failures have between 17 and 27 observations, whereas the healthy and one failure classes have between 32 and 42 observations. However, a global average of 30 observations for each failure mode does not represent a considerable amount of data, so it would be beneficial to have access to a data set with more measurements to obtain better classification results.

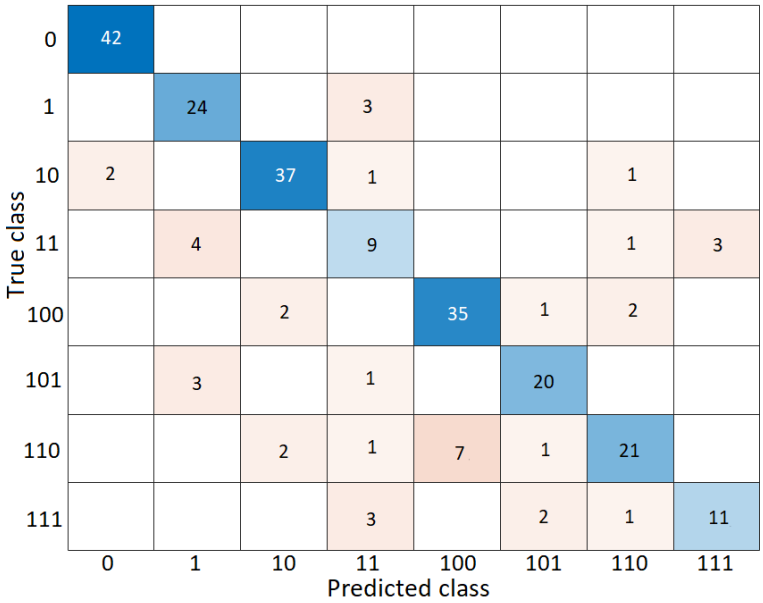


Figure 4.6: Confusion matrix with the number of observations per predicted class and true class - optimized ensemble.

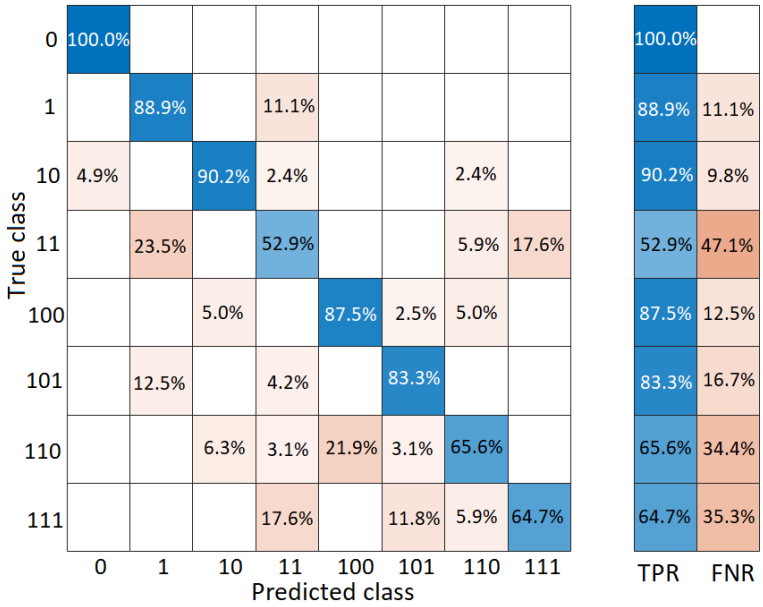


Figure 4.7: Confusion matrix with TPR and FNR - optimized ensemble.

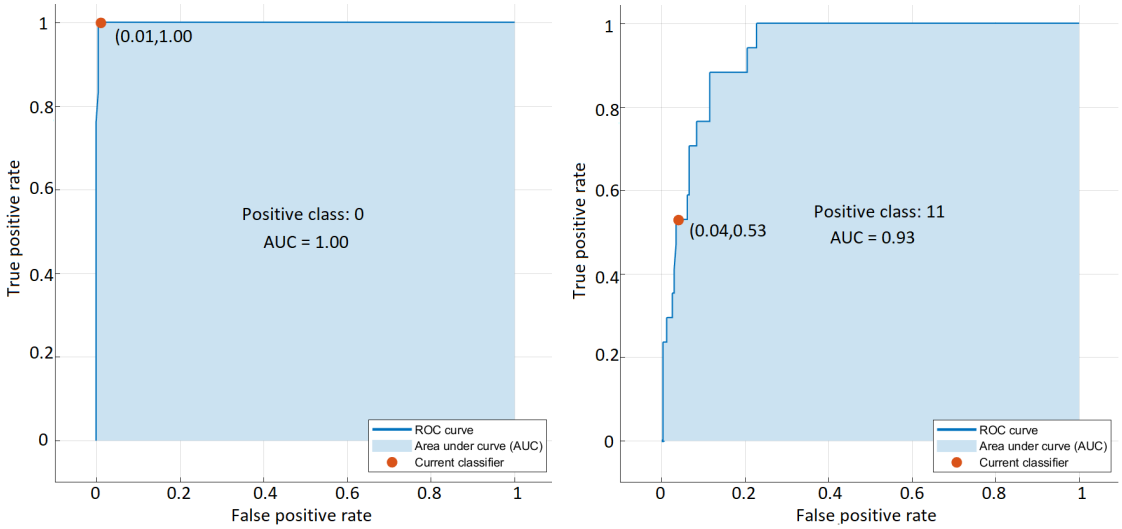


Figure 4.8: ROC curves - optimized ensemble.

4.1.2 Case Study 2

The second case study to be presented consists of faulty and healthy data from a centrifugal pump. The set of data was retrieved from Matlab documentation and is based on a centrifugal pump model from an analysis presented in [50]. The data set contains 500 measurements of flow rate, head, speed, and torque for each of the six fault modes and healthy mode, making a total of 3500 measurements. The types of faults covered in the data set and the corresponding label used to identify them are presented in Table 4.3. For each measurement, there are 201 observations organized in the data file with a time interval from 180 seconds to 200 seconds, meaning that each observation was recorded every 0.1 seconds. As in the previous case study, also for this data set, the time records may not refer to real recording times and, instead, may refer to simulation times. Additionally, an attempt was also made to find out the real periodicity of the data but such information was not available at the source of the data.

Table 4.3: Labels used to classify the types of faults and healthy operation of data in case study 2.

Type of fault	**Label**
No fault (healthy pump)	0
Wear at clearance gap	1
Small deposits at impeller outlet	2
Deposits at impeller inlet	3
Abrasive wear at impeller outlet	4
Broken blade	5
Cavitation	6

The following steps, implemented using the capabilities of the Diagnostic Feature Designer app, concern data processing, and feature extraction.

Figure 4.9 shows the flow, head, speed, and torque signals curves. An example curve for each of the seven failure modes is illustrated in the figure. For the case of flow and head signals, although with difficulty, differences in the curves of different fault modes can be perceived - curves associated with distinct fault modes occupy slightly different ranges of values. For example, fault mode 2 - small deposits at impeller outlet - and fault mode 3 - deposits at impeller inlet - have flow rates that are globally lower than fault mode 0 - healthy pump - which makes sense, since obstructions to the passage of fluid between the impeller vanes lead to a decrease in the flow rate [10]. Regarding the speed and torque curves, differences between the different failure modes are not distinguishable.

For each of the four signals, all thirteen time-domain features listed in Figure 3.2 were extracted, making a total of 52 features related to the time-domain. Then, the power spectrum of all signals was computed for the subsequent extraction of frequency-domain features. Figure 4.10 shows the spectral domains of each signal, with an example curve of each fault mode being illustrated. From these signals, all the frequency-domain features listed in Figure 3.2 were extracted. The first five signal peaks were used in a frequency range of 0.3-3.2 Hz to compute the features referring to the frequency-domain.

Combining time-domain features with frequency-domain features, a total of 136 features were obtained. Using the ANOVA method, the features were classified and obtained ANOVA metrics between 0 and 266. The features that obtained the highest positions in the ranking were root-mean-square, mean value, peak value, shape factor, clearance factor, impulse factor, and crest factor of torque, head, and flow signals.

Three groups of features were selected to train the classification algorithms. The first group of features used to train the algorithms was the one that contains features with ANOVA metrics greater than 1, which corresponds to the 65 features with the highest rating. All models available in the app were evaluated and Table 4.4 presents the results of accuracy for a selection of models. The results of most of the algorithms presented in the mentioned table are not satisfactory, and

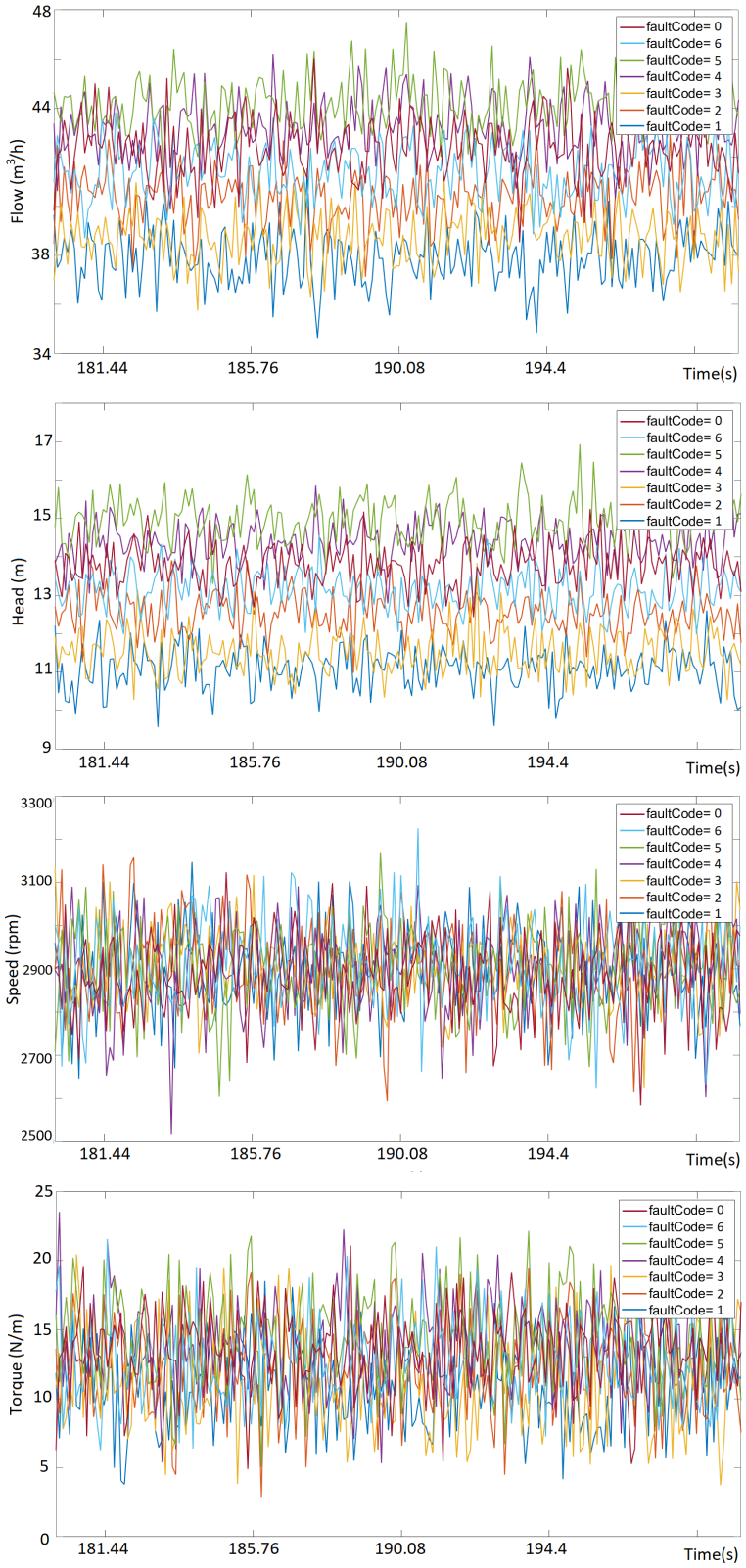


Figure 4.9: Signal trace - case study 2.

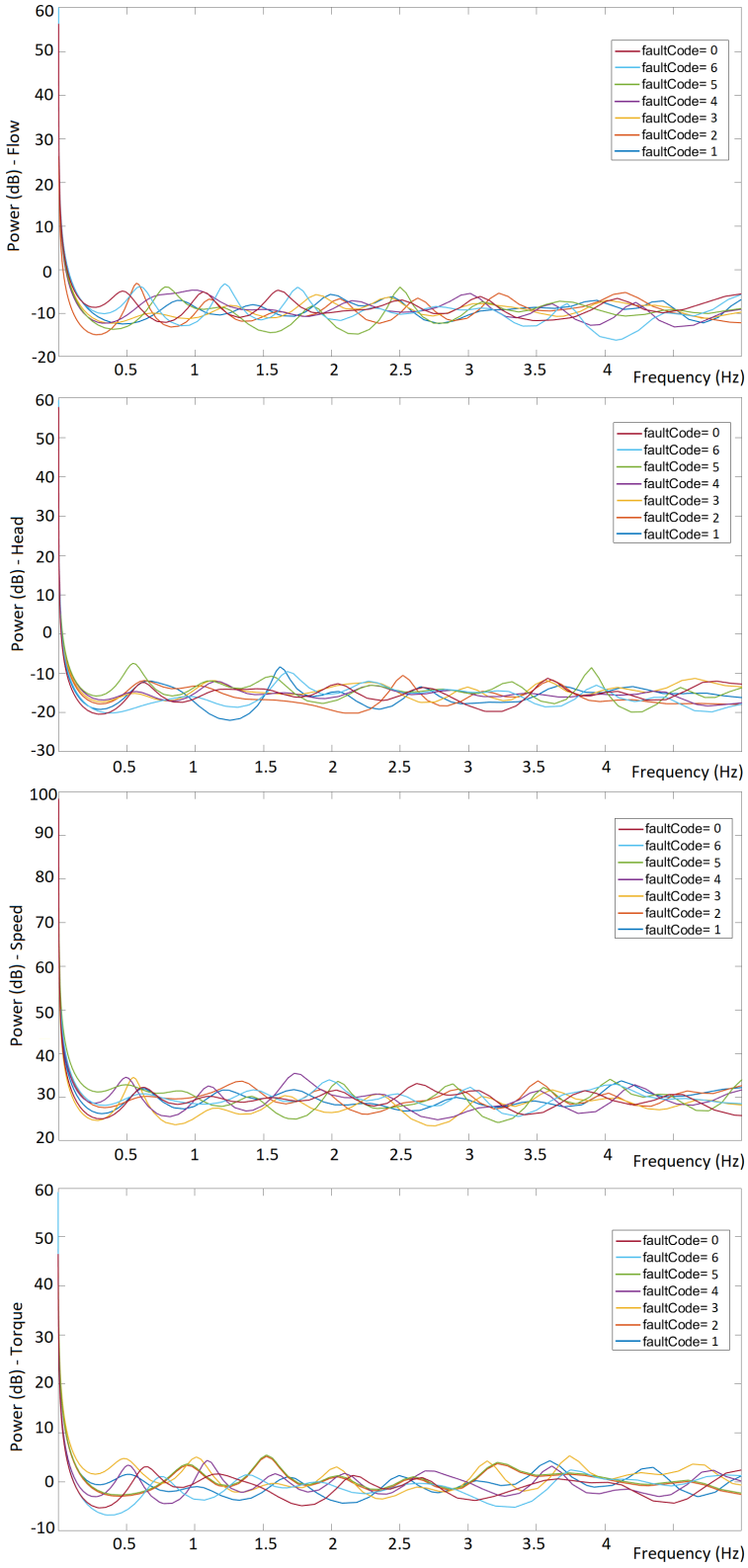


Figure 4.10: Power spectrum - case study 2.

even some of them fail and do not show results. Some algorithms may fail because the features used have invalid values or because they have constant values for the same response class. These situations do not invalidate the performance of all algorithms as some are robust to the presence of values with the mentioned characteristics. The Fine Decision Tree algorithm obtained the best results for the first set of features evaluated, with an accuracy of 91.6%, followed by the Ensemble Bagged Trees with 89.7% and Ensemble Boosted Trees with 72.1%. The remaining algorithms did not obtain satisfactory results, and all, except the Medium Decision Tree, which obtained an accuracy of 62.1%, obtained accuracy values below 22%. The difference in performance values between Fine Decision Tree and Medium Decision Tree can be explained by the fact that the Fine Tree algorithm has a more complex structure with more splits or branches, which can be beneficial for the successful implementation of the analyzed data.

Table 4.4: Results of a selection of algorithms trained with first group of features for an average of 10 analyzes - case study 2.

Model	**Accuracy**	**Model**	**Accuracy**
Fine Decision Tree	91.6%	Medium Decision Tree	62.1%
Linear Discriminant	Fail	Quadratic Discriminant	Fail
Gaussian Naive Bayes	21.1%	Kernel Naive Bayes	21.3%
Linear SVM	14.2%	Quadratic SVM	14.3%
Fine Gaussian SVM	14.2%	Medium Gaussian SVM	14.3%
Fine KNN	14.2%	Medium KNN	14.1%
Cubic KNN	14.2%	Weighted KNN	14.3%
Ensemble Subspace KNN	15%	Ensemble Subspace Discriminant	17.5%
Ensemble Boosted Trees	72.1%	Ensemble Bagged Trees	89.7%

The second group of features used to train the algorithms was the one that contains features with ANOVA metrics greater than 2, which corresponds to the 21 features with the highest ranking metric. All models available in the app were evaluated and the results for a selection of these models are shown in Table 4.5. Contrarily to the results presented in Table 4.4, the results of a large part of the algorithms are quite satisfactory. For the present group of features, none of the algorithms failed and the Fine Decision Tree, Linear and Quadratic Discriminant, Linear and Quadratic SVM, and Ensemble of Bagged Trees algorithms all had accuracy results above 90%. As with the first set of features, the algorithm that obtained the best rating for the second set of features was the Fine Decision Tree algorithm, with an accuracy of 92.4%. This group of features, which corresponds to an even smaller selection of the features with the highest ANOVA metrics, proves to be interesting because, by using any of the available algorithms, good performance can be achieved.

Table 4.5: Results of a selection of algorithms trained with second group of features for an average of 10 analyzes - case study 2.

Model	**Accuracy**	**Model**	**Accuracy**
Fine Decision Tree	92.4%	Medium Decision Tree	62.2%
Linear Discriminant	90.7%	Quadratic Discriminant	92.2%
Gaussian Naive Bayes	24.6%	Kernel Naive Bayes	44.3%
Linear SVM	90.1%	Quadratic SVM	91.4%
Fine Gaussian SVM	74.7%	Medium Gaussian SVM	79.1%
Fine KNN	67.1%	Medium KNN	65.1%
Cubic KNN	62.4%	Weighted KNN	67.9%
Ensemble Subspace KNN	89.7%	Ensemble Subspace Discriminant	89.5%
Ensemble Boosted Trees	71.9%	Ensemble Bagged Trees	91.8%

The third and last group of features evaluated was the one that contains all time-domain and frequency-domain features referring to head and flow rate. This group of features was defined because they only concern the head and flow rate signals, which are the quantities most commonly measured in pump analysis. All models available on the app were evaluated and the results related to a selection of these models are described in Table 4.6. The results are generally similar to those obtained for the first group of features: most of the results are not satisfactory and for the discriminant algorithms no results were obtained - there was a failure in the implementation of the algorithms. However, it should be noted that the Fine Decision Tree algorithm not only obtained the best accuracy for this third group of features, but it was the algorithm that obtained the best accuracy in relation to all groups of features - 92.7%. In this context, the parameters of the Fine Decision Tree algorithm were optimized using hyperparameter optimization, having reached an accuracy of 94.6%. Figure 4.11 illustrates the results obtained in the hyperparameter optimization of the decision tree classifier.

Table 4.6: Results of a selection of algorithms trained with third group of features for an average of 10 analyzes - case study 2.

Model	**Accuracy**	**Model**	**Accuracy**
Fine Decision Tree	92.7%	Medium Decision Tree	44%
Linear Discriminant	Fail	Quadratic Discriminant	Fail
Gaussian Naive Bayes	21.2%	Kernel Naive Bayes	22.2%
Linear SVM	14.3%	Quadratic SVM	14.3%
Fine Gaussian SVM	14.2%	Medium Gaussian SVM	14.3%
Fine KNN	14.3%	Medium KNN	14.2%
Cubic KNN	14.3%	Weighted KNN	14.3%
Ensemble Subspace KNN	16.2%	Ensemble Subspace Discriminant	15.3%
Ensemble Boosted Trees	68.1%	Ensemble Bagged Trees	78.3%

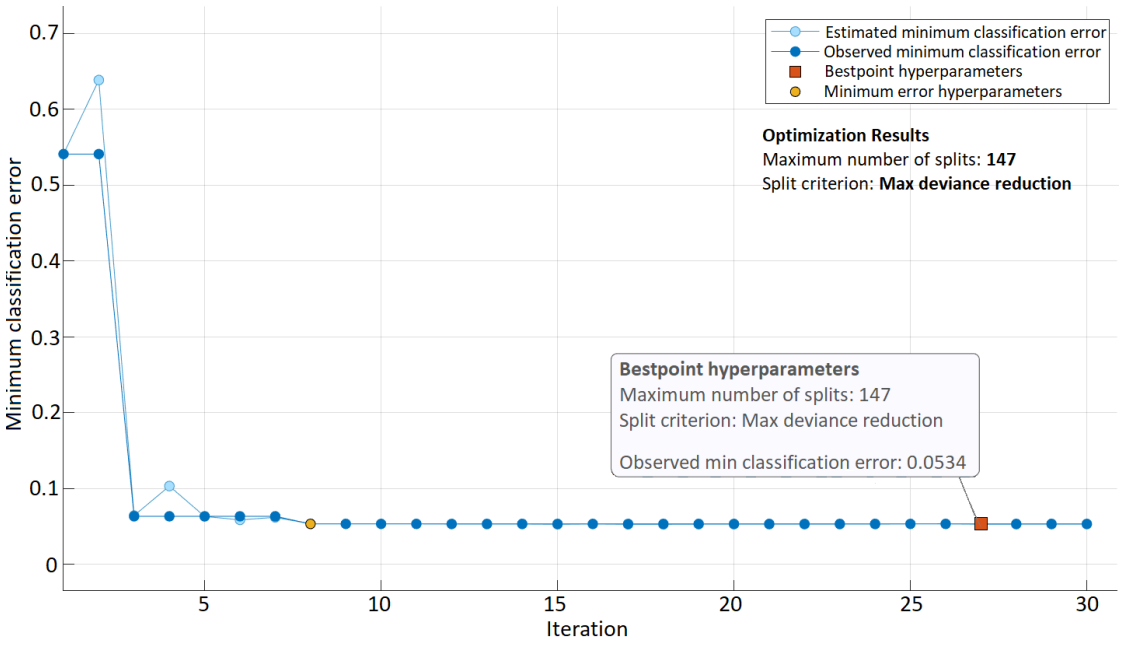


Figure 4.11: Minimum classification error plot for the decision tree classifier - case study 2.

Regarding the performance of the optimized decision tree algorithm in each class, such results

are presented in the confusion matrices in Figures 4.12 and 4.13.

In Figure 4.12, the values of the confusion matrix refer to the number of observations for each predicted and true class, whereas in Figure 4.13, the values of the confusion matrix refer to TPR and FNR for each true class. Note that each of the seven classes has 500 associated measurements. From the two figures, for label 0 (healthy mode), for example, it can be concluded that:

- 505 observations were attributed to class 0, from which 499 were correctly attributed and the other 6 observations were incorrectly attributed, because actually, 3 of these observations belong to class 2, 2 belong to class 4, and 1 belongs to class 6 (Figure 4.12);
- Only 1 measurement from the 500 measurements that correspond to label 0 was incorrectly classified, and this measurement was attributed to class 4 (Figure 4.12);
- From the observations belonging to class 0, 99.8% were correctly classified, corresponding to the TPR of class 0, $TPR = \frac{499}{500} = 0.998$ (Figure 4.13);
- The 6 observations that were attributed to class 0 and not belong to such class constitute the FNR of the classes that they actually belong to (Figure 4.13);

Regarding label 6 (cavitation), for example, it can be concluded that:

- 520 observations were identified as belonging to class 6, however, only 440 of these observations truly belong to class 6, while of the others, 78 observations refer to class 2 and 2 observations refer to class 3 (Figure 4.12);
- From the 60 observations which true class is class 6, 1 was attributed by the classifier to class 0, 56 were attributed to class 2, and 3 were attributed to class 3 (Figure 4.12);
- The 440 correctly identified observations of class 6 constitute the TPR of this class, which is 88%, $TPR = \frac{440}{500} = 0.88$ (Figure 4.13);
- The remaining 60 observations that belong to class 6 and were attributed to other classes constitute the FNR of class 6, making a total of 12%, $TPR = \frac{1+56+3}{500} = 0.12$ (Figure 4.13);
- The observations that were incorrectly attributed to class 6 constitute the FNR of the other classes to which they truly belong.

From the confusion matrix with the number of observations (Figure 4.12), the obtained accuracy can be confirmed: the quotient between the sum of correctly assigned observations (diagonal entries) and the total number of observations, 3500, reflects the model's accuracy, which is 94.6% ($\frac{499+500+418+495+480+481+440}{3500} = 0.946$).

Furthermore, from Figures 4.12 and 4.13 it can be concluded that: 78 observations in which the true class is class 2 were identified as belonging to class 6 and, 56 observations in which the true class is class 6 were identified as belonging to class 2. Therefore, the classifier reveals some difficulty in distinguishing these two fault modes (small deposits at impeller outlet vs. cavitation); 18 observations in which true class is class 4 were attributed to class 5 and, 19 observations in which the true class is class 5 were attributed to class 4. Although it represents only 3.6% and 3.8% of the observations of class 4 attributed to class 5 and observations of class 5 attributed to class 4, respectively, it reveals some difficulty of the classifier in distinguishing these two fault modes (abrasive wear at impeller outlet vs. broken blade).

The ROC curves for classes 0 and 6 are presented in Figure 4.14. Taking class 0 as the positive class, the TPR of 1 indicates that all observations that correspond to class 0 were correctly classified, and the FPR of 0.01 indicates that observations that were incorrectly assigned to class 0 correspond to 1% of all observations that were not assigned to such class. The AUC value for class 0 is 1, which means that the classifier is perfect at distinguishing the healthy pump mode. Considering class 6 as the positive class, the TPR of 0.88 indicates that 88% of class 6

observations were correctly classified, and the FPR is 0.03, which indicates that the observations that were incorrectly classified as belonging to class 6 correspond to 3% of all observations that were not attributed to this class. The AUC for this class is 0.95, which, despite being a high value, was the lowest AUC value obtained among all classes, indicating the high suitability of the algorithm in classifying all of the seven different fault modes.

In general, the model classifies the observations of all failure modes very well. This is probably due not only to the considerable amount of data - 500 observations for each failure mode - but also to the pertinence and suitability of the data and the careful analysis and implementation of all steps related to the task in question. However, due to the large volume of data, the time taken to implement the steps of the task under analysis was much higher than for the previous case study.

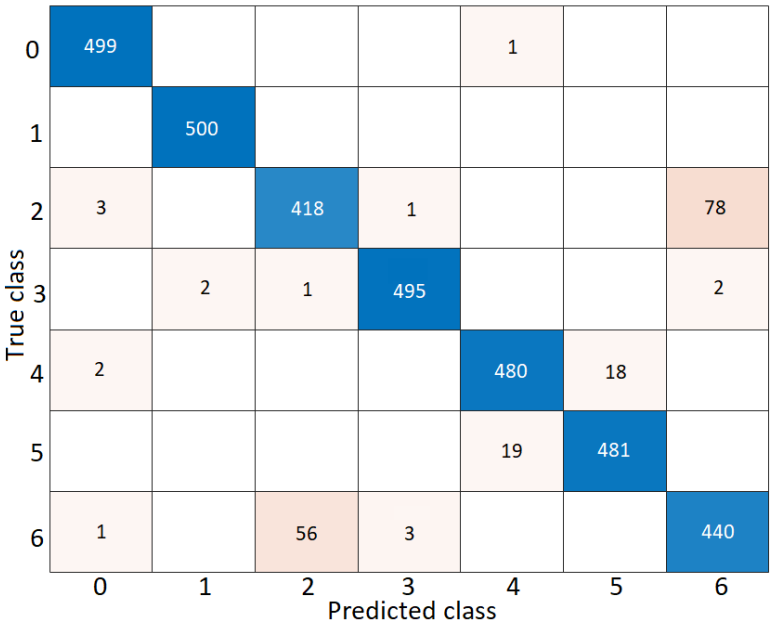


Figure 4.12: Confusion matrix with the number of observations per predicted class and true class - optimized decision tree.

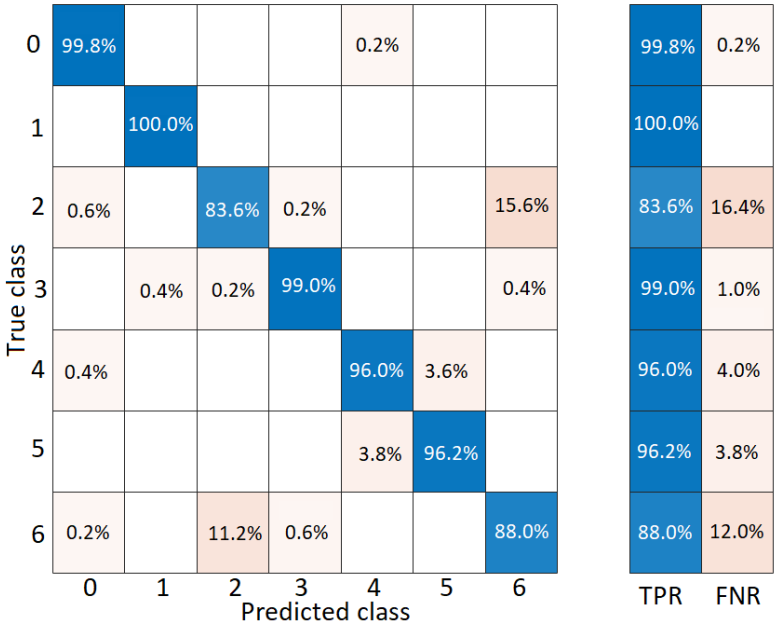


Figure 4.13: Confusion matrix with TPR and FNR - optimized decision tree.

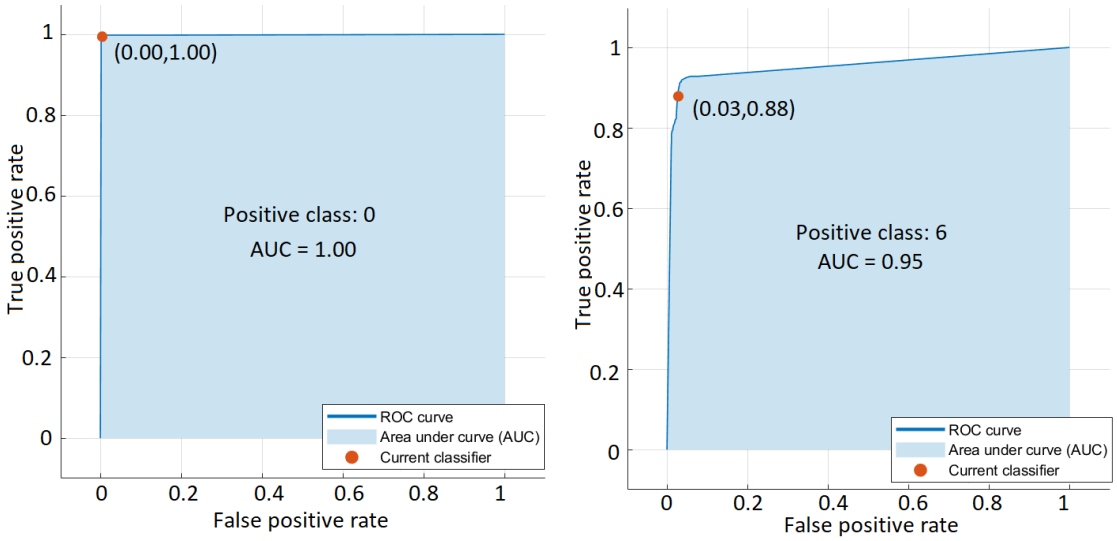


Figure 4.14: ROC curves - optimized decision tree.

4.2 Case studies - Remaining Useful Life Estimation

4.2.1 Nasa Case Study

The first case study used to implement the methodology of RUL Estimation is a data set retrieved from the Prognostics Data Repository of Nasa [51] (author of the data set [52]). The repository file contains several data sets. For this case study, the set used is the first training data set named "train_FD001" in the repository file. The data set has simulated data of turbofan engines and it consists of one hundred multivariate time series. The authors of the data stated the time series data can be considered to be from a fleet of engines of the same type, and that each engine starts with a different degree of initial wear that is unknown but can be considered to be in a normal condition, i.e., the engine is not considered to be faulty [52]. The time series end at system failure, thus, the data set covers the lifetime span of engines. The lifetime is measured in units of engine operational cycles. Regarding the set of data used in the dissertation work, the faults covered in the engine data concern the same fault type.

The data set was provided in a text file with twenty-six columns of values. Each row is a measurement of data of a single operational cycle and each column is a different variable. The first column is the engine id; the second column is the time in operational cycles; the third, fourth, and fifth columns are operational settings imposed on the engines; the other columns are measurements of the different sensors. The operational settings do not have an influence on the data of the selected set for this case study, hence, these were not taken into account in the analysis.

Data were organized in a text file and so, the first step was to convert and organize the data in an array of tables. The array is composed of one hundred multivariate time series tables with data concerning time in operational cycles and sensor data from twenty-one different sensors. Data were randomly partitioned for model validation: 80% of the time series were used to train the model and 20% of the time series were used for validating the model, which resulted in two data sets, a training data set with 80 time series and a validation data set with 20 time series.

In Figure 4.15 a sample of 10 time series of training data set of sensors 1-4 raw data is illustrated. The raw training data of sensors 5-21 is illustrated in Figure A.1.

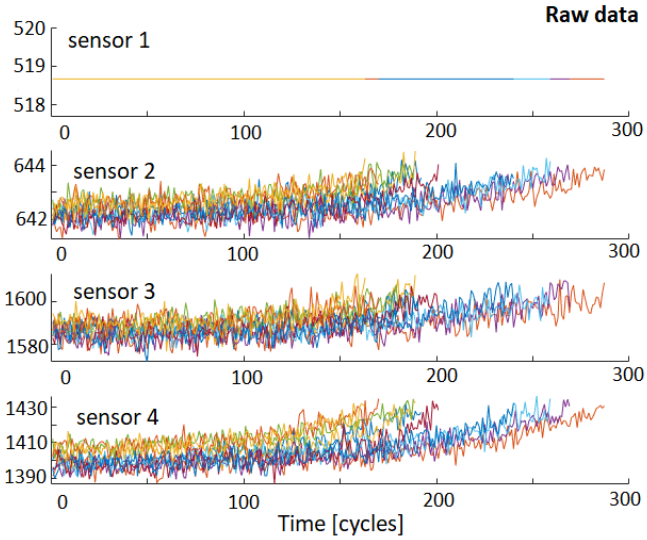


Figure 4.15: Sample of 10 time series of raw training data set of sensors 1-4.

For data noise reduction, a moving average smoothing was performed on raw data with a window size=10 (Figure 4.16: smoothed data of sensors 1-4, Figure A.2: smoothed data of sensors 5-21). Degradation trends became more clear for some sensors after data smoothing.

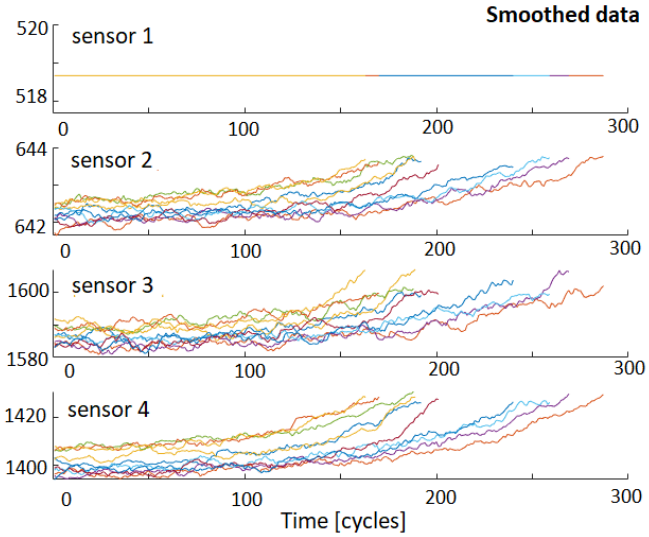


Figure 4.16: Sample of 10 time series of smoothed training data set of sensors 1-4.

The same process of data smoothing was performed both in training and validation data sets. Since different sensor measurements have very different data ranges, data normalization was performed (Figure 4.17: normalized data of sensors 1-4, Figure A.3: normalized data of sensors 5-21). Training data were normalized using a z-score described in Equation 3.6. The statistics computed from training data were used to normalize the validation data. Some sensors have a standard deviation very close to zero. These sensor data are not relevant since a nearly constant signal is not useful for RUL prediction.

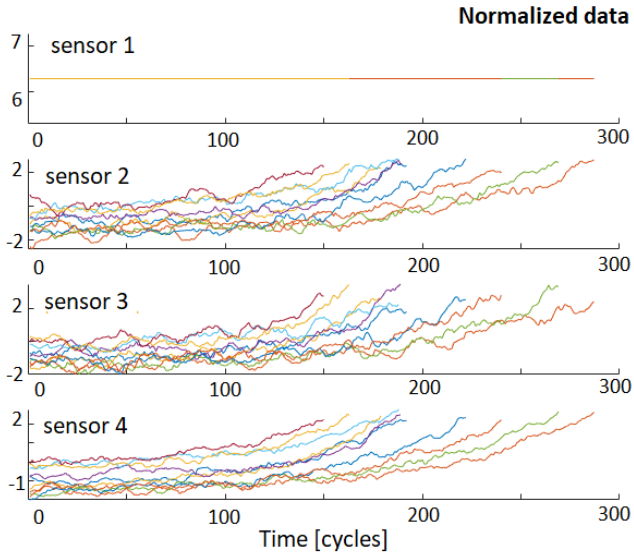
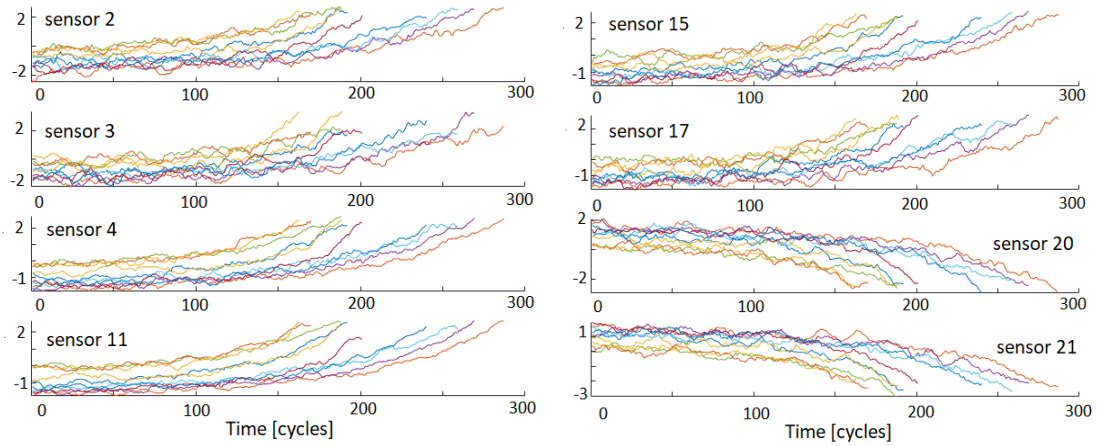


Figure 4.17: Sample of 10 time series of normalized training data set of sensors 1-4.

To construct health indicators from sensor data, the techniques mentioned in 3.1.2 in the step sensor trendability and sensor fusion were used. A linear degradation model described in Equation 3.7 was estimated for each sensor of training data and the slopes of the signals were ranked. The eight sensors with the higher mean values of slope parameter θ were selected: sensors

2, 3, 4, 11, 15, 17, 20, and 21. Several values for the number of selected sensors were computed as an iterative process. A model with eight selected sensors was the one that produced the best estimation results. Data from the eight selected sensors are present in Figure 4.18. Some sensors show a positive trend and others show a negative trend.



(a) (b)

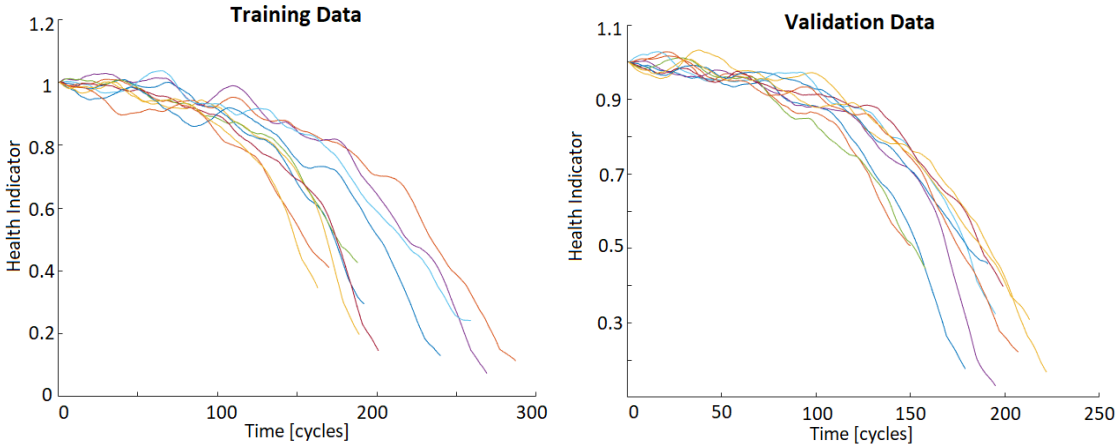
Figure 4.18: Sample of 10 time series of selected sensors from trendability analysis: sensors (a) 2, 3, 4, 11, (b) 15, 17, 20, 21.

A column for health condition was created for each time series data. The theoretical health condition starts with the value one and reaches the end of the data lifetime with the value zero. A linear regression model fitted the health condition with the most trended sensors as regressors. The Equation 3.8 assumes the form

$$h = w_0 + w_2S_2 + w_3S_3 + w_4S_4 + w_11S_11 + w_15S_15 + w_17S_17 + w_20S_20 + w_21S_21 + \epsilon \quad (4.1)$$

The weight parameters w were computed and, for each time series, a single health indicator was constructed by multiplying the sensor measurements S with their associated weights w. The health indicators of validation data were computed using the estimated weights of training data. A sample of health indicators of the training and the validation data are illustrated in Figure 4.19. Both on training data and validation data, the health indicators assume a notorious degradation trajectory until the end of the lifetime (failure).

A residual similarity model was built using data from the training set. In the configuration used, the model fitted the health indicators data with a second-order polynomial and found the nearest twenty data samples in training data to estimate the RUL. To evaluate the model, samples of validation data until 30%, 50%, and 70% of their lifetime were used and their RUL was predicted. The partition of validation data sets in the 3 breakpoints was executed to simulate the real context of a predictive maintenance plan with an RUL estimation methodology of pump systems, where new data are being collected in real-time and new predictions are regularly updated. In the following figures, the results of the RUL estimator model are presented. The k nearest neighbor plot illustrates the trajectory of health indicator of the validation data sample being evaluated until the breakpoint together with the twenty nearest health indicators of the training data set (Figures 4.20a, 4.21a and 4.22a). The RUL estimation plot illustrates the true RUL value and the RUL value estimated by the model at each breakpoint. Furthermore, a probability density function along with a 90% confidence interval on the estimation based on its nearest neighbors is plotted (Figures 4.20b, 4.21b and 4.22b). The estimation gets better as more percentage of lifetime data becomes available: at the 30% breakpoint, the model estimated the RUL at 156 and the true RUL was 129, which means the model overestimated the RUL by



(a) (b)

Figure 4.19: Sample of 10 time series of health indicators of: (a) training data and (b) validation data.

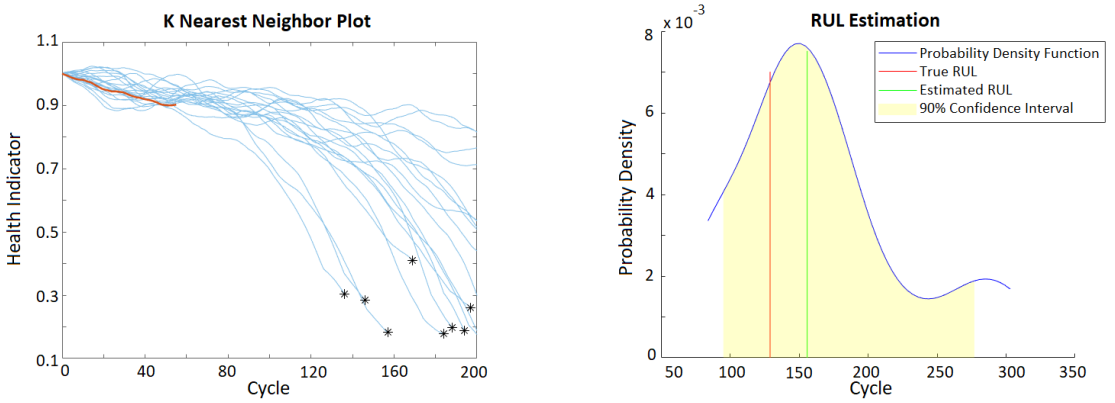
27 cycles; at the 50% breakpoint, the model estimated the RUL at 105 and the true RUL was 92, which means the model overestimated the RUL by 13 cycles; at the 70% breakpoint, the model estimated the RUL at 59 and the true RUL was 55, which means the model overestimated the RUL by 4 cycles.

RUL estimation was performed for the whole validation data set (all twenty time series). The prediction error was computed as the difference between the true RUL and the estimated RUL. Three different ways to interpret graphically the prediction error of the model for the entire validation data set are presented in the following figures.

In Figure 4.23 the histograms of the error, together with its probability distribution for each breakpoint, are presented. The area of each bar is the relative number of observations. The probability density estimates (plotted as the red lines in Figure 4.23) are based on a normal kernel function [35]. The probability distribution of the prediction error gets more concentrated around zero as the percentage of validation data used in the estimation gets higher.

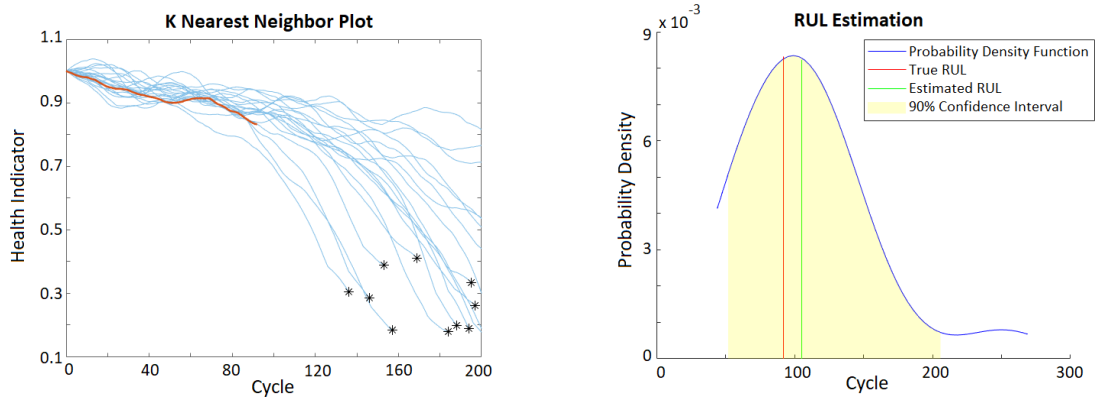
The median, the 25th and 75th percentiles, and outliers values of the prediction error of each breakpoint are illustrated in box plots in Figure A.4. On each box, the red central mark indicates the median of the prediction error, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Outlier values are plotted with the red symbol "+".

The mean value and the standard deviation of the prediction error are illustrated in Figure 4.24. Although the breakpoint at 70% of the validation data set has a mean prediction error higher than the mean of the error at 30%, the standard deviation becomes inferior for predictions more close to the end of the lifetime.



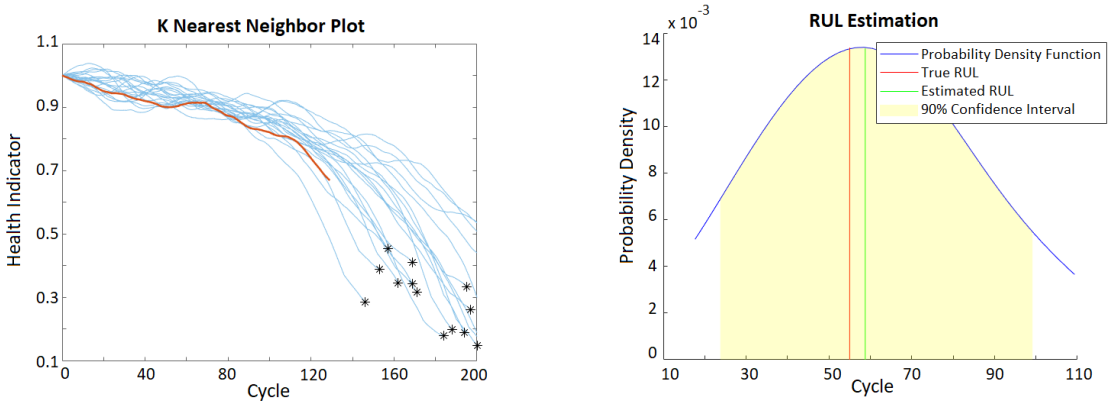
(a) (b)

Figure 4.20: RUL estimation for 30% breakpoint of a sample of validation data: (a) nearest neighbor plot, (b) probability density function.



(a) (b)

Figure 4.21: RUL estimation for 50% breakpoint of a sample of validation data: (a) nearest neighbor plot, (b) probability density function.



(a) (b)

Figure 4.22: RUL estimation for 70 % breakpoint of a sample of validation data: (a) nearest neighbor plot, (b) probability density function.

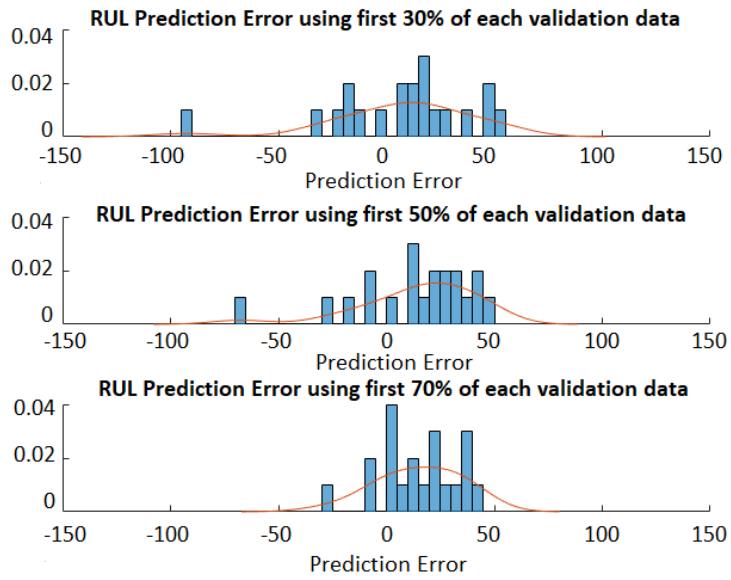


Figure 4.23: Histogram of the error and probability distribution for each breakpoint of validation data set.

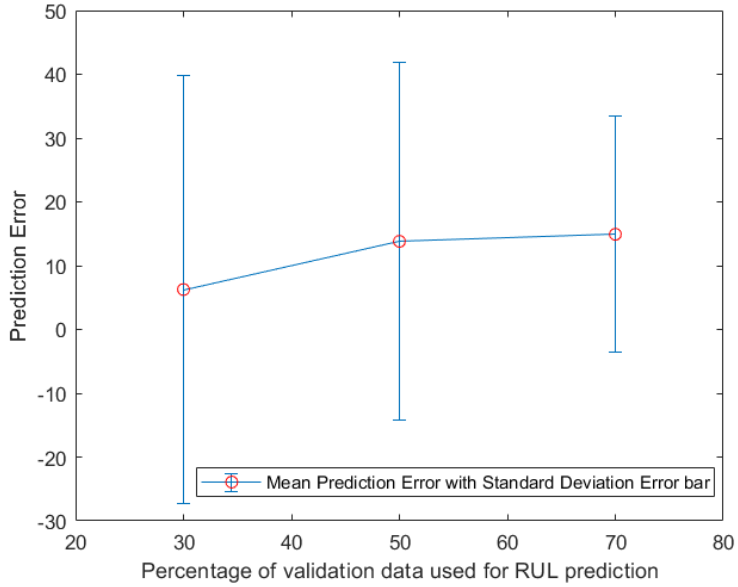


Figure 4.24: Mean prediction error and standard deviation error bar for each breakpoint of validation data set.

4.2.2 Kaggle Case Study

The second case study of the methodology of RUL Estimation is a data set retrieved from Kaggle, an online data repository. The data set was provided by a team that maintains water pumps [53]. The author of the data reported that, during the time span of the data set, 7 system failures occurred and the causes of the failures are unknown.

The set is related to a real water pump and contains information from 52 different sensors. No information concerning what the different sensors measure is provided. Data were collected every minute during five months, resulting in 220320 observations. A column containing the output label that describes the machine status is also present in the data set. Normal conditions, broken conditions, and recovering conditions are the possible machine status covered in the data set. The sensor measurement values are all raw values, subsequently, there are a considerable amount of missing values in the data set.

First, an exploration of the data was done for better interpretation and to possibly formulate a hypothesis about the pump measurements.

Exploratory Data Analysis

The original data set file was stored in a .csv format. The data set was converted into a table, re-arranged, and organized in a mat-file. Table 4.7 presents some segments of data illustrating how the converted data set structure is organized.

Table 4.7: Segments of data to illustrate the structure of the whole data set.

Count	Timestamp	Sensor 1	...	Sensor 52	Machine Status
0	01-Apr-2018 00:00:00	2.4654		201.3889	normal
1	01-Apr-2018 00:01:00	2.4654		201.3889	normal
2	01-Apr-2018 00:02:00	2.4447		203.7037	normal
...					normal
17155	12-Apr-2018 21:55:00	0		324.6528	broken
17156	12-Apr-2018 21:56:00	0		341.7245	recovering
...					recovering
18099	13-Apr-2018 13:39:00	0.3060		38.7731	recovering
18100	13-Apr-2018 13:40:00	0.3060		38.7731	normal
...					
220319	31-Aug-2018 23:59:00	2.3965		234.0856	normal

The measurements were collected every minute from 1st of April, 2018 until 31st of August, 2018. That resulted in a total of 220320 observations, each observation with values of the 52 different sensors and the label of the pump status. The measurements collecting was initiated with the pump in a normal operating condition. Numerous observations were registered in a normal condition until a failure occurred. A single observation is registered with a broken condition label, indicating that the failure occurred in that specific minute. The following observations concern a recovering condition. Then, multiple observations are registered with recovering condition until the pump is in a normal condition again. Throughout the measurement collecting, 7 system failures were registered. From the total 220320 observations, 205836 concern a normal pump operation, 7 observations are related to the broken condition and 14477 refer to recovering condition of the pump (Figure 4.25). Thus, normal operation represents 93.43% of the data set, recovering condition represents 6.57% and broken condition is only 0.003% of the data set.

In order to visualize how long it took for each recovering state to reach a normal operation, the number of observations for each recovering state was counted. The same process was applied to each normal state. The duration, in hours, of each normal state and each recovering state are illustrated in bar charts in Figures 4.26a and 4.26b, respectively.

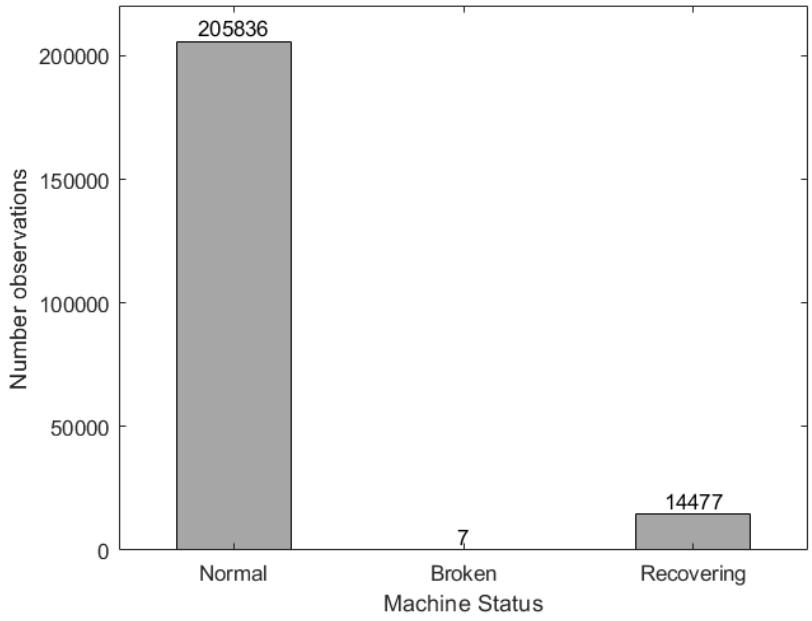
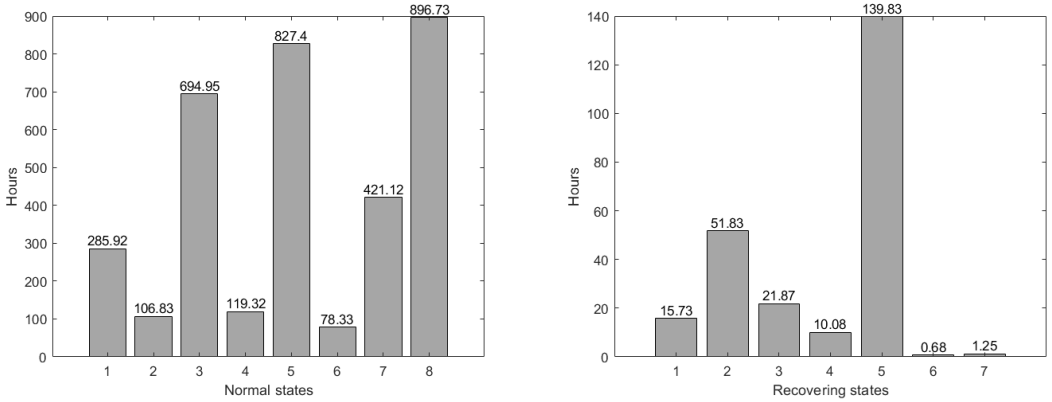


Figure 4.25: Number of observations of each machine status in the data set.



(a) (b)

Figure 4.26: Time [hours] in each normal state (a) and each recovering state (b).

There are 8 normal operation states and only 7 recovering states because the measurements data concluded in the eighth normal pump state, i.e., after the seventh failure and the seventh recovering state occurred. The eighth normal state was the one that lasted longer, with almost 897 hours (37.3 days). Additionally, the normal state that follows the eighth one in terms of duration is the fifth normal state, with 827 hours, which is approximately 34.5 days. The duration of the sixth normal state is only 78 hours or 3.3 days, which makes this the normal state that has the lower duration. The fifth failure was the one that took the most time to recover and reach a normal operation. The pump was in the fifth recovering state for almost 140 hours, which is about 5.8 days. The failure that took less amount of time to recover was the sixth failure. The pump was in the sixth recovering state for 0.68 hours, which in minutes is only 41 minutes.

The following step in the data analysis was to check for missing values in the data set. First, the data for each of the 7 broken states was interpreted and the possible existence of missing values when the failures occurred was determined. The sensors with missing values at the time of a failure are sensors 51 and 52. Table 4.8 presents the data for each broken state for those sensors. The data concerning the other sensors are not displayed since no missing values were found. Evaluating the mentioned table, it can be concluded that for the fifth failure there is no value for sensor 52. Moreover, for the seventh failure, there is no value for sensor 51. No conclusion on whether these missing values will compromise the data on sensor 51 and sensor 52 can be performed at this point of the analysis. Further investigation was executed to verify such a hypothesis.

Table 4.8: Data of sensors 51, 52 for each of the 7 failures.

Count	**Timestamp**	**Sensor 51**	**Sensor 52**	**Status**
17155	12-Apr-2018 21:55:00	401.9097	324.6528	broken
24510	18-Apr-2018 00:30:00	177.6620	183.7384	broken
69318	19-May-2018 03:18:00	246.2384	257.5231	broken
77790	25-May-2018 00:30:00	220.1968	267.3611	broken
128040	28-Jun-2018 22:00:00	32.4074	NaN	broken
141131	08-Jul-2018 00:11:00	192.1296	174.7685	broken
166440	25-Jul-2018 14:00:00	NaN	205.7292	broken

As it was referred, there are a considerable amount of missing values in the data set. The percentage of missing values in each sensor was calculated, sorted in descending order, and organized into a table (Table 4.9).

It can be verified that some of the sensors have a high percentage of missing values. Sensor 16 has 100% of missing values, which means that no measurement was collected concerning this sensor. Thus, in the following analysis, sensor 16 was not considered. The sensors with a percentage of missing values higher than 1% were analyzed with more detail. Such sensors include sensor 51, sensor 52, sensor 1, sensor 8, sensor 9, sensor 7, and sensor 10.

In Figure 4.27 the data for the sensors with higher percentages of missing values throughout the five months are shown. The machine status associated with the time intervals is also present in the figure. The normal states are represented with green lines and the recovering states are represented with red lines. In the transition between these states, failure occurs, which is illustrated by broken states. For a clearer visualization, the sensor data was normalized to have a data range between 0 and 1.

There is no data for sensor 51 after the sixth failure until the end of the time interval of the measurements. That resulted in almost two months without values for sensor 51. The existence of this data gap could be explained by a possible failure of the sensor and an associated high cost to repair it. Another possible cause is that the team that manages the pump system concluded that the sensor was not providing relevant information to understand the behavior of the system and turned it off. Nevertheless, none of these assumptions can be confirmed.

Sensor 1 has a small period with missing values during the first recovering state and another

Table 4.9: Percentage of data missing values for each sensor sorted in descending order.

Sensors	**Missing values [%]**	**Sensors**	**Missing values [%]**	**Sensors**	**Missing values [%]**
Sensor16	100	Sensor40	0.0123	Sensor12	0.0086
Sensor51	34.9569	Sensor41	0.0123	Sensor13	0.0086
Sensor52	6.9821	Sensor42	0.0122	Sensor14	0.0086
Sensor1	4.6333	Sensor43	0.0123	Sensor20	0.0073
Sensor8	2.4741	Sensor44	0.0123	Sensor21	0.0073
Sensor9	2.3180	Sensor45	0.0123	Sensor22	0.0073
Sensor7	2.1778	Sensor46	0.0123	Sensor24	0.0073
Sensor10	2.0856	Sensor47	0.0123	Sensor25	0.0073
Sensor2	0.1675	Sensor48	0.0123	Sensor28	0.0073
Sensor31	0.1185	Sensor49	0.0123	Sensor29	0.0073
Sensor30	0.0327	Sensor50	0.0123	Sensor32	0.0073
Sensor33	0.0309	Sensor15	0.0095	Sensor34	0.0073
Sensor18	0.0209	Sensor27	0.0091	Sensor35	0.0073
Sensor19	0.0209	Sensor3	0.0086	Sensor36	0.0073
Sensor23	0.0186	Sensor4	0.0086	Sensor37	0.0073
Sensor26	0.0163	Sensor5	0.0086	Sensor38	0.0073
Sensor17	0.0141	Sensor6	0.0086		
Sensor39	0.0123	Sensor11	0.0086		

missing value period during the second recovering. Additionally, during the fifth recovering state, sensor 1 has a time interval of missing values. These three periods without data for sensor 1 results in a missing value percentage of 4.6% (Table 4.9). The same situation of missing values during the fifth recovering characterizes sensors 8, 9, 7, and 10. The percentage of missing values for these sensors is between 2% and 2.5% (Table 4.9). However, it is considered that the data coverage of the five sensors (sensors 1, 8, 9, 7, 10) is reliable. Since all of these sensors stopped the measurements collecting after the fifth broken state, this failure was possibly different from the other failures.

Concerning sensor 52, there is a time period with missing values in the data. That period begins before the fifth failure occurs and ends during the fifth recovering state. Sensor 51 was discarded for the following analysis since it has almost 35% of data with missing values, which includes the missing value in the seventh failure.

A better understanding of the behavior of sensor 52 during the fifth recovering state is derived from Figure 4.28. Data of this sensor are illustrated from June 19 until July 7. Sensor data stopped being recollected before the fifth failure, that is, during a normal operating state. The fifth failure occurred on July 28 at 22 p.m., in a time period where sensor 52 was not receiving data. The data gap ends on the first of July when the pump was in a recovering state. Thenceforth sensor 52 does not have more associated missing values. A total of 15355 observations of sensor 52 have an associated missing value. Although such problem could be fixed by different methods that fill missing data, that would result in more than 10.5 days of value estimation. It is considered that such approach is not the best one and, sensor 52 was discarded.

Three sensors were discarded from the analysis: sensor 16, sensor 51, and sensor 52. The next step was to calculate statistical measures such as the mean, minimum and maximum values, and standard deviation. The mentioned statistical measures concerning all sensors in the analysis are presented in Table 4.10.

Towards a more graphical interpretation of the statistical values, see Figure A.5 and Figure A.6. The mean value is represented with a red circle, the minimum and maximum values are represented as the bottom and top edges of the green bar, and the standard deviation is represented as the edges of the blue bar.

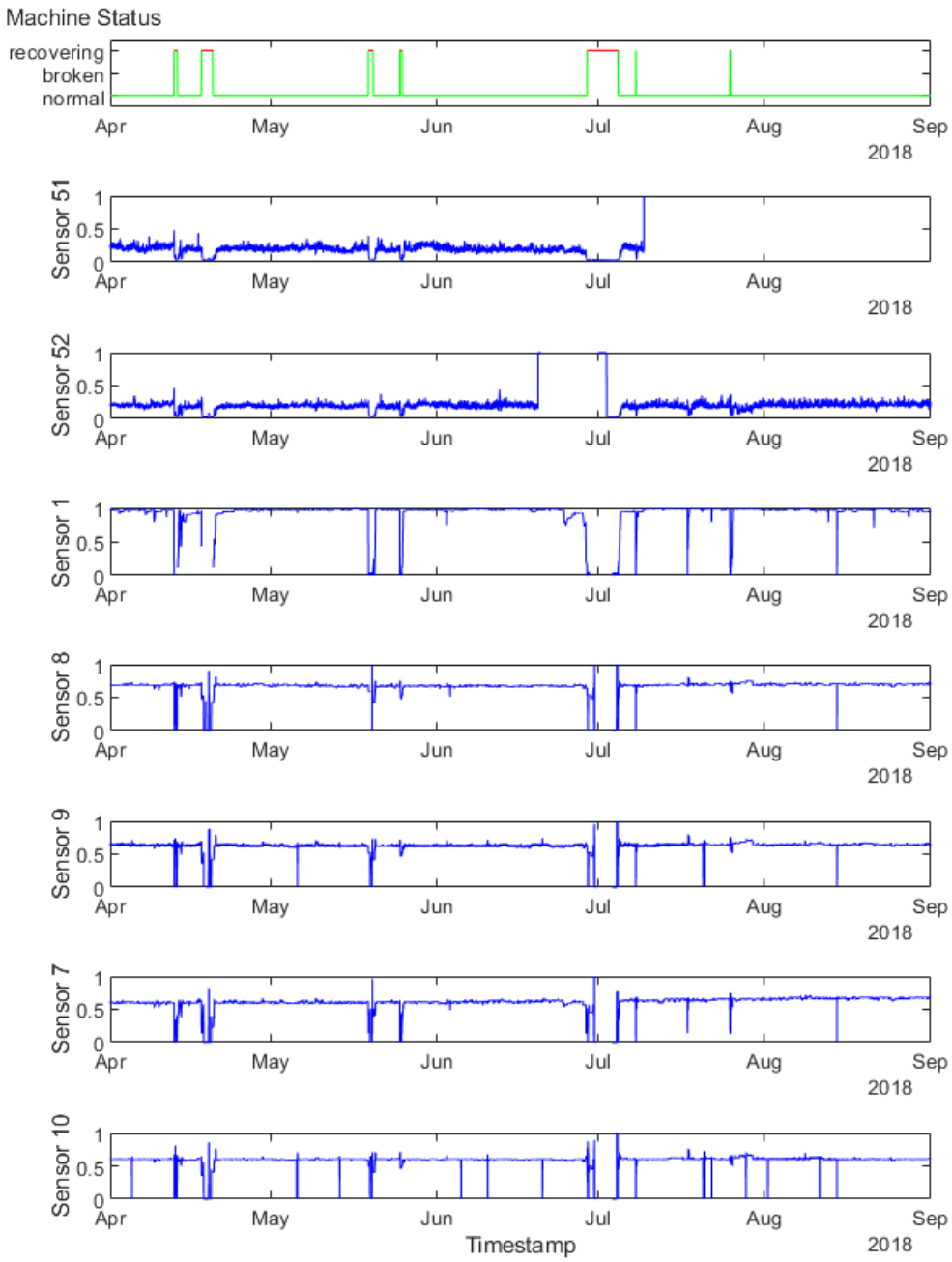


Figure 4.27: Data concerning sensors with high percentage of missing values.

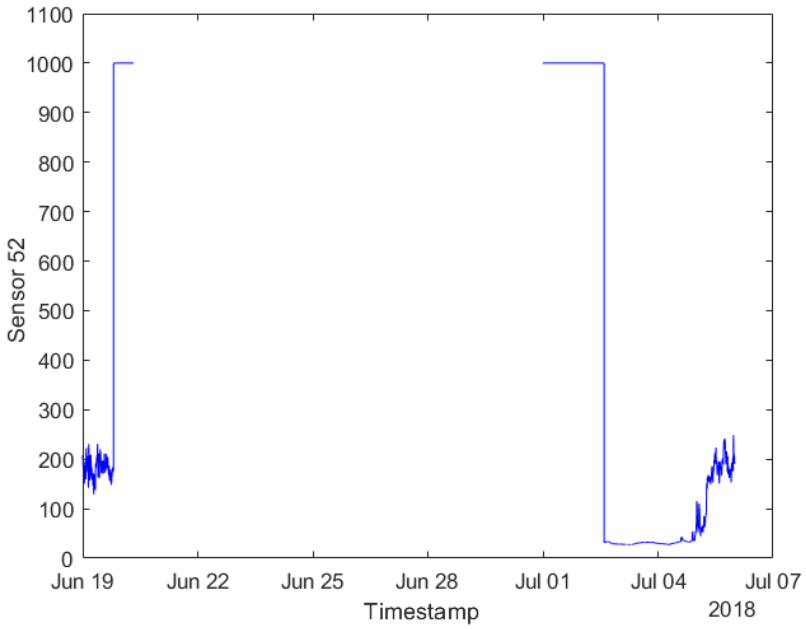


Figure 4.28: Data of sensor 52 before the fifth broken state and during the fifth recovering state.

From the analysis of the listed statistics, it can be assumed that all of the sensors are analog sensors since the information is represented in continuous values. Additionally, all sensors have positive values, most of the sensors have zero as the minimum value and some sensors have the maximum values expressed as whole numbers. The maximum values expressed as whole numbers can be a consequence of truncated measures at the end of the scale of an analog sensor.

Analyzing Figure A.5 and Figure A.6 one can see that, although for some sensors the minimum and/or maximum values far deviate from the mean value, the standard deviation has a low value, i. e., the majority of the data tend to be close to the mean. It is the case of sensors 39 until sensor 48: the maximum values for these sensors are much higher than the mean, however, since they have low standard deviations (in comparison with the maximum values), one can conclude that most of the data are concentrated around the mean value. Furthermore, sensors 1 and 19 contrast with all the other sensors since they have very low data ranges.

As it was referred, no information concerning what the sensors measure was provided by the author of the data set. Nevertheless, one can assume that, since the data is related to a water pump, some sensors can be related to pressure, flow rate, temperature, vibration, motor torque, motor current, and others.

In Figure 4.29, a heatmap based on the correlation coefficients between the different sensors is illustrated. The correlation coefficient is a measure of the linear dependence of two random variables. The values of the coefficients can range from -1 to 1: -1 represents a direct and negative correlation, 0 represents no correlation, and 1 represents a direct and positive correlation. The correlation heatmap indicates that sensors from sensor 15 until sensor 27 have high positive correlations with each other. Additionally, some other sensors have a moderate to a high positive correlation. Such is the case of sensors 7, 8, 9, and 10. There are other strongly correlated groups of sensors, however, those correlations are not as strong as the first one. To verify the correlation between sensors, the data for each sensor was plotted and analyzed.

Figures A.7, A.8, A.9, A.10, A.11, A.12 illustrate the behavior of each sensor throughout the whole time window of five months. Sensors 7, 8, 9, and 10 have very similar behaviors and their data range covers similar values. The same situation happens for sensors 11, 12, and 13. All sensors between sensor 15 and sensor 31 have identical behaviors, although some of those sensors

Table 4.10: Statistics for each sensor.

Sensors	**Mean**	**Min**	**Max**	**StD**	**Sensors**	**Mean**	**Min**	**Max**	**StD**
Sensor1	2.4	0	2.6	0.4	Sensor27	786.4	43.2	1214.4	246.7
Sensor2	47.6	0	56.7	3.3	Sensor28	501.5	0	2000	169.8
Sensor3	50.9	33.2	56.1	3.7	Sensor29	851.7	4.3	1841.2	313.1
Sensor4	43.8	31.6	48.2		Sensor30	576.2	0.6	1466.3	225.8
Sensor5	590.7	2.8	800	144.1	Sensor31	614.6	0	1600	195.7
Sensor6	73.4	0	99.9	17.3	Sensor32	863.3	23.9	1800	283.6
Sensor7	13.5	0.01	22.3	2.2	Sensor33	804.3	0.2	1839.2	260.6
Sensor8	15.8	0	23.6	2.2	Sensor34	486.4	6.5	1578.6	150.8
Sensor9	15.2	0.03	24.4	2.1	Sensor35	234.9	54.9	425.6	88.4
Sensor10	14.8	0	25	2.1	Sensor36	427.1	0	694.5	141.8
Sensor11	41.5	0	76.1	12.1	Sensor37	593.0	2.3	984.1	289.4
Sensor12	41.9	0	60	13.1	Sensor38	60.8	0	174.9	37.6
Sensor13	29.1	0	45	10.1	Sensor39	49.7	24.5	417.7	10.5
Sensor14	7.1	0	31.2	6.9	Sensor40	36.6	19.3	547.9	15.6
Sensor15	376.9	32.4	500	113.2	Sensor41	68.8	23.4	512.8	21.4
Sensor17	416.5	0	739.7	126.1	Sensor42	35.4	20.8	420.3	7.9
Sensor18	421.1	0	599.9	129.2	Sensor43	35.5	22.1	374.2	10.3
Sensor19	2.3	0	4.8	0.8	Sensor44	43.9	24.5	408.6	11.1
Sensor20	590.8	0	878.9	199.4	Sensor45	42.7	25.8	1000	11.6
Sensor21	360.8	0	448.9	101.9	Sensor46	43.1	26.3	320.3	12.9
Sensor22	796.2	95.5	1107.5	226.7	Sensor47	48.1	26.3	370.4	15.6
Sensor23	459.8	0	594.1	154.5	Sensor48	44.3	27.2	303.5	10.4
Sensor24	922.6	0	1227.6	291.8	Sensor49	150.9	26.3	561.6	82.3
Sensor25	556.2	0	1000	182.3	Sensor50	57.1	26.6	464.4	19.1
Sensor26	649.2	0	839.6	220.9					

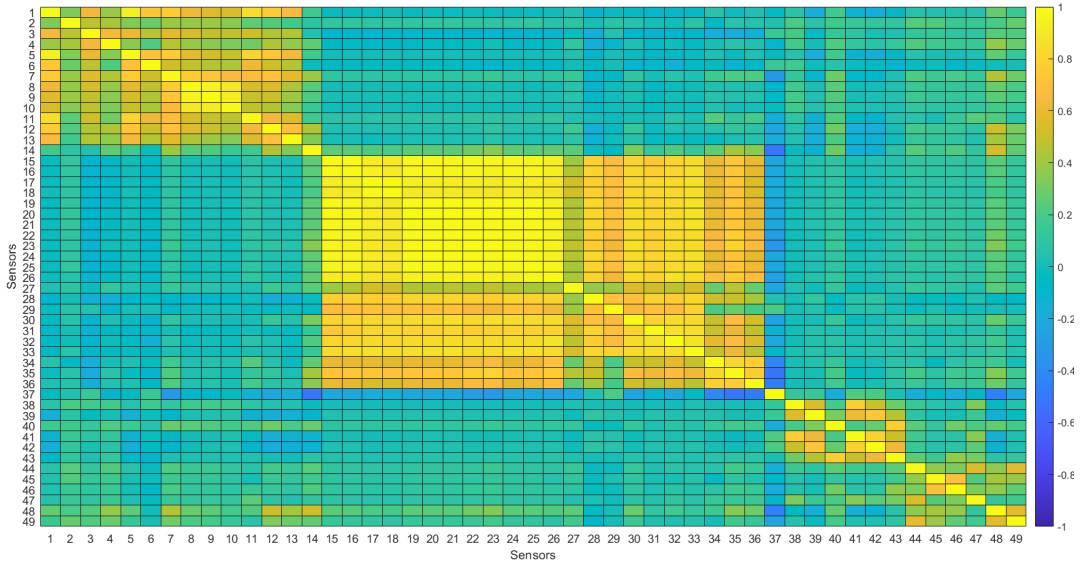


Figure 4.29: Correlation heatmap of the sensors.

have some variations throughout the data. Sensor 38 has a continuously diminishing trend and a lot of noise associated with the data in comparison with other sensors. Sensors 39, 40, 42, and 43 seem to share a similar trend, and also sensors 46, 47, and 48, but not so strong as the other group of sensors.

RUL Estimation

The case study consists of a data set with 7 different run to failure histories. The cause of each failure is not known and no strong assumptions on the nature of the failures can be made. Nevertheless, since the failures concern the same water pump, the assumption that the failures relate in some similar way was made. The data set was divided into 7 data sets. Each data set is a time series containing the data from the beginning of the normal state until the broken condition. For the purpose of RUL estimation, the data related to the recovering of the pump was not necessary. It was only needed data concerning the normal operation until the occurrence of failure. Data covering the eighth normal operation of the pump was discarded from this analysis since it is unknown when the eighth failure occurred, therefore, it is not relevant for RUL estimation. Each data set was organized with data of the sensors measurements and the corresponding time. No label related to the state of the pump is needed, the data starts at a healthy operation and ends with the occurrence of failure.

In Figure 4.30, the lifetime of each series is illustrated in minutes and in the equivalent time in hours and in days, for better understanding. Time series 3 and time series 5 have a significantly higher lifetime, comparing with the other time series. These two time series have almost double of the lifetime of the rest of the time series of the data set.

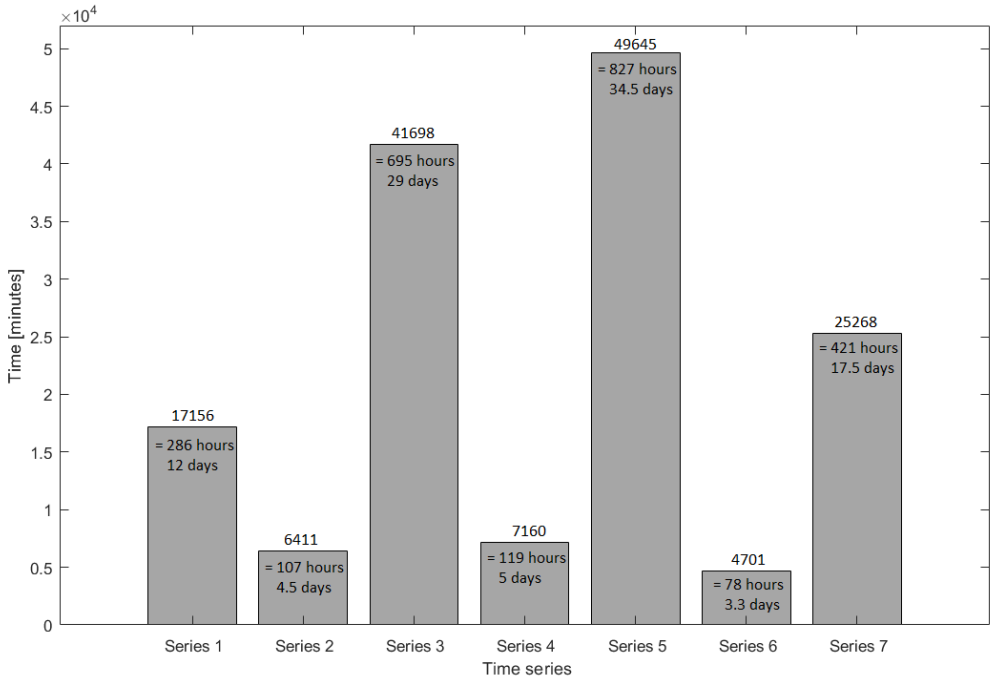


Figure 4.30: Lifetime of each time series.

Some methods to fill in missing data assume that the previous and next values of a given missing value are available. However, that is not the situation of the data in the analysis. There are several consecutive missing values in the data sets. To overcome this problem, a flexible and robust method to fill in missing data must be used, such as interpolation. Given the trend of the signals, a linear interpolation was used to fill in the missing values.

Global Models

The case study has a small data set dimension: only 7 data time series. Hence, for validating the model, a leave-one-out cross-validation partition of the data sets was performed. The leave-one-out cross-validation is a particular case of the k-fold cross-validation because the number of folds is equal to the number of time series, and is adequate for small data sets. The data set was split into a training data set, containing 6 of the 7 data sets, and a validation data set, containing the other data set. The cross-validation was repeated 7 times, with each time series used once as the validation data. A total of 7 models were trained. These models are referred to as "Global Models".

The results of the entire RUL estimation methodology for one of those models are presented. The results of the other models are mentioned and used to average the prediction error of the models. The following results refer to the Global Model 4, i.e., the model trained with the time series 1, 2, 3, 5, 6, and 7, and validated with the time series 4.

In Figure 4.31 signals of sensors 1, 2, 3, and 4 of training data are illustrated. A moving average smoothing with a window size = 500 was applied for data noise reduction. A sample of training data sensors after smoothing is in Figure 4.32. Training data was normalized using the z-score (Equation 3.8). Signals of a sample of sensors of normalized training data are in Figure 4.33.

Then, a trendability analysis of the sensors was performed to find the most adequate sensors to construct a health indicator. The techniques described in the methodology concerning sensor trendability and sensor fusion steps were applied. The selection of the number of trended sensors to be used in sensor fusion was an iterative process. The most adequate number of sensors to be selected is eight and these sensors are sensors 1, 2, 3, 5, 6, 7, 11, and 43 (Figure 4.34). As it can be seen in Figure 4.34, a given sensor has different trends among different time series.

A linear regression fitted a theoretical health condition with the previously selected sensors from the training data set, the weight parameters were computed and used to construct the health indicators of training data (Figure 4.35) and validation data (Figure 4.36). Health indicators do not demonstrate the expected linearly degrading behavior. Such situation occurs possibly due to the fact that sensors have different trends among the time series.

A residual similarity model was built based on training data. In the selected configuration the model used a linear regression as the method for data fitting and residual generation, and the number of nearest data samples in training data to estimate the RUL was three. To evaluate the model prediction, the partition of the validation data in 3 breakpoints used and described in the previous case study 4.2.1 was also applied. The graphical illustration of RUL estimation results for time series 4 is in Figure 4.37 - 30% breakpoint -, Figure 4.38 - 50% breakpoint -, and Figure 4.39 - 70% breakpoint.

At 30% of time series 4, the model estimated the RUL at 3652 minutes and the true value of RUL was 5012. At 50% of time series 4, the model estimated the RUL at 2219 minutes and the true value of RUL was 3580. Finally, at 70% of the validation time series, the model estimated the RUL at 787 minutes and the true value was 2148 minutes. For the three breakpoints, the model underestimated the RUL value by 1361 minutes. That corresponds to 22.7 hours, which is less than one day. Such results are satisfactory since the lifetime of time series 4 is 119 hours, which corresponds to 5 days.

As it was previously referred, the validation process was repeated using each time series once as the validation data. That resulted in seven models with some differences in the selected sensors at trendability analysis and, consequently, different health indicators. The results of RUL estimation were very distinct among the seven global models. An overview of the RUL estimation results for each global model is listed in Table 4.11.

RUL estimation results were worst for global model 3 and global model 5. Inclusively, the estimation results of these models are negative values, which means the models are estimating that the failure has already occurred at the time of the three lifetime breakpoints. Such can be explained through the lifetime of time series 3 and time series 5: these are the time series with lifetime values significantly higher in comparison with the other series (see Figure 4.30). RUL

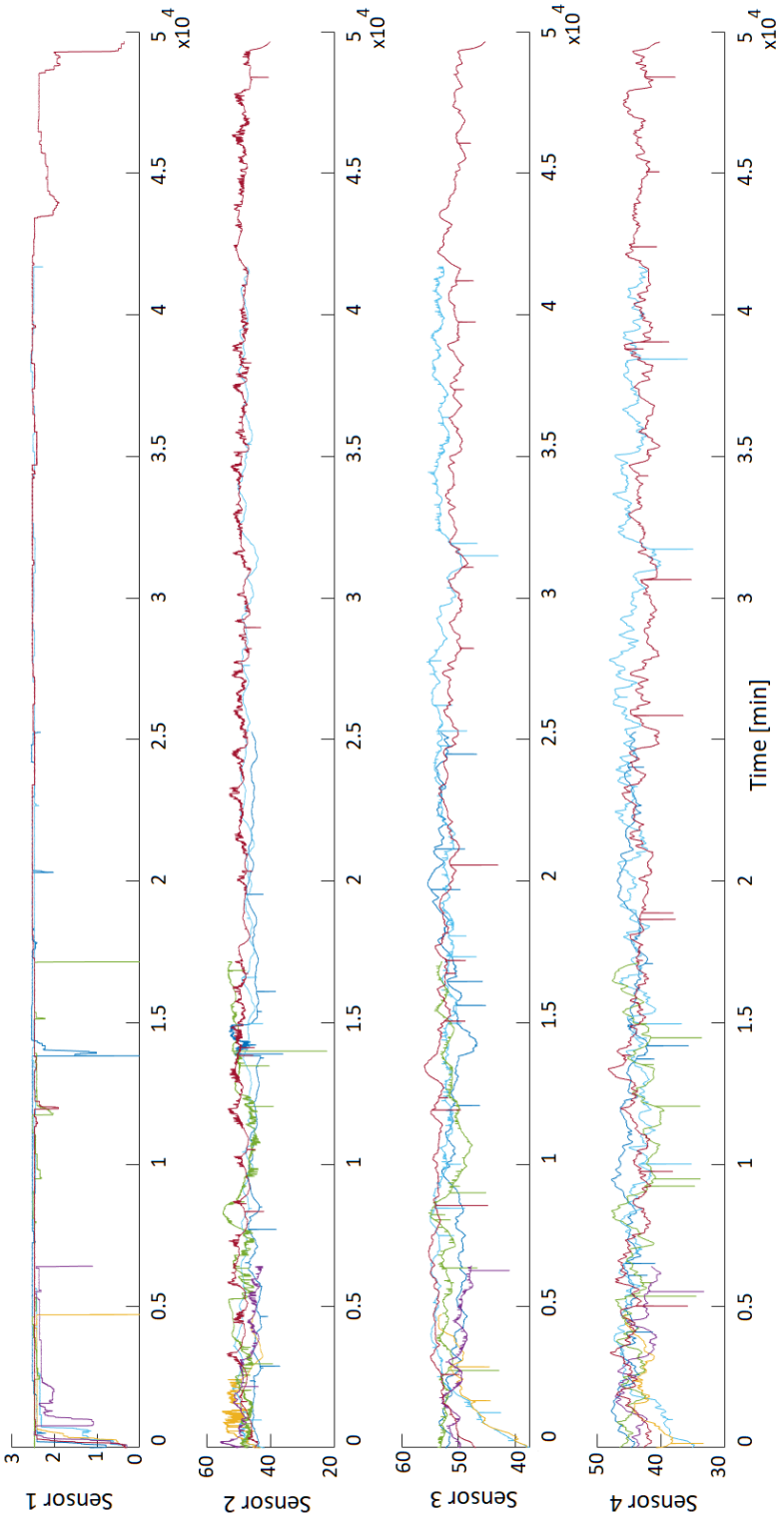


Figure 4.31: Sample of sensors of training data - raw signals.

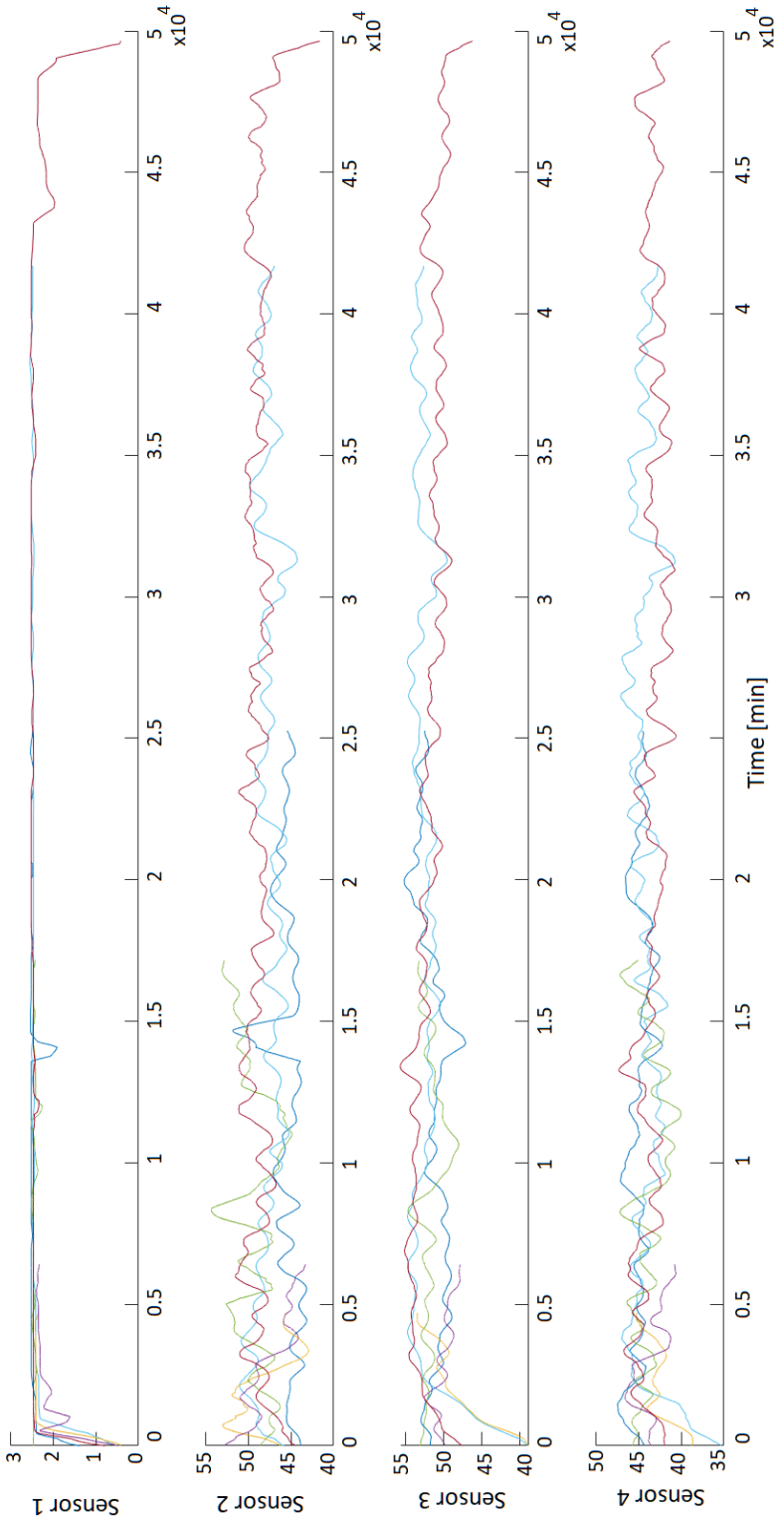


Figure 4.32: Sample of sensors of training data - smoothed signals.

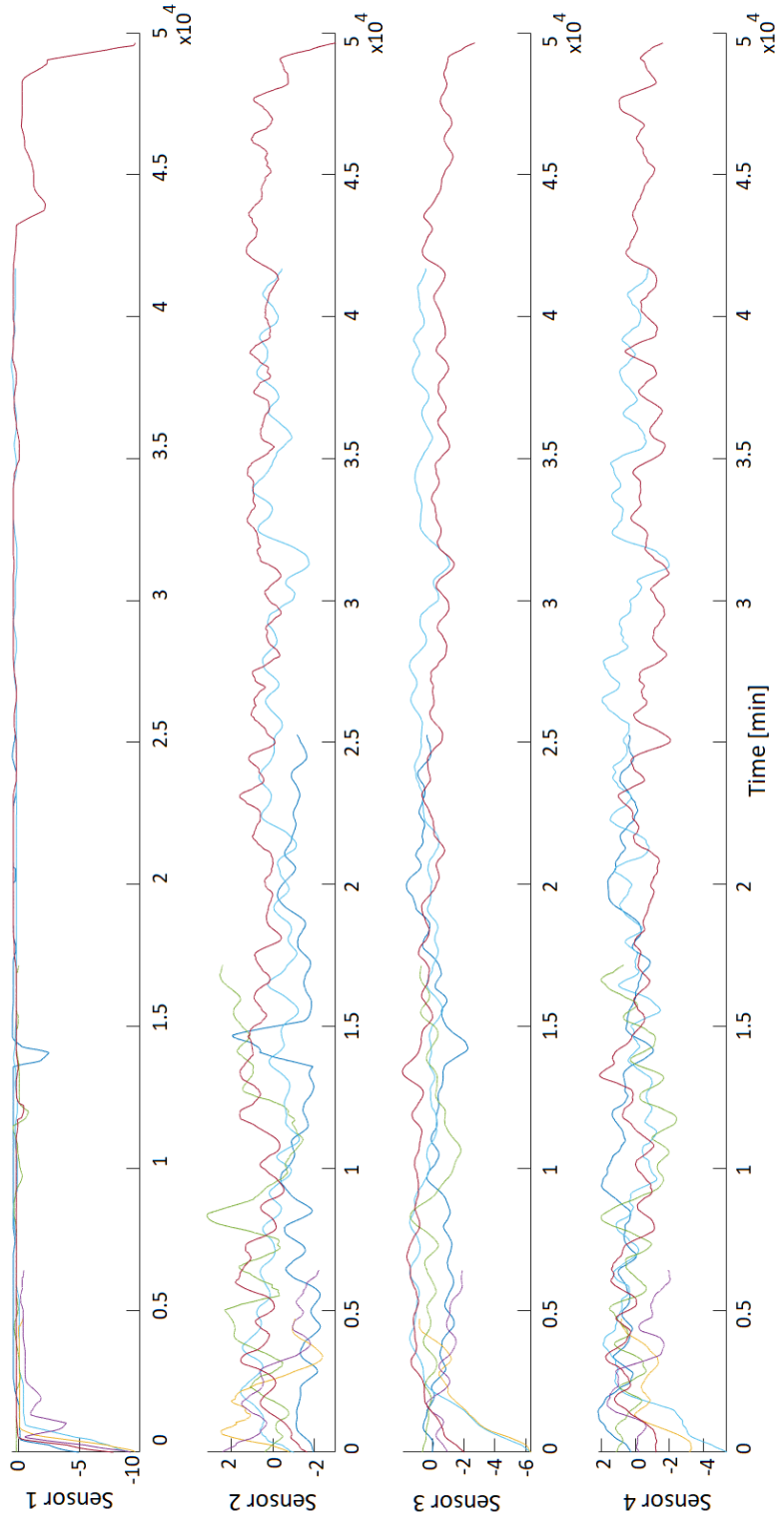


Figure 4.33: Sample of sensors of training data - normalized signals.

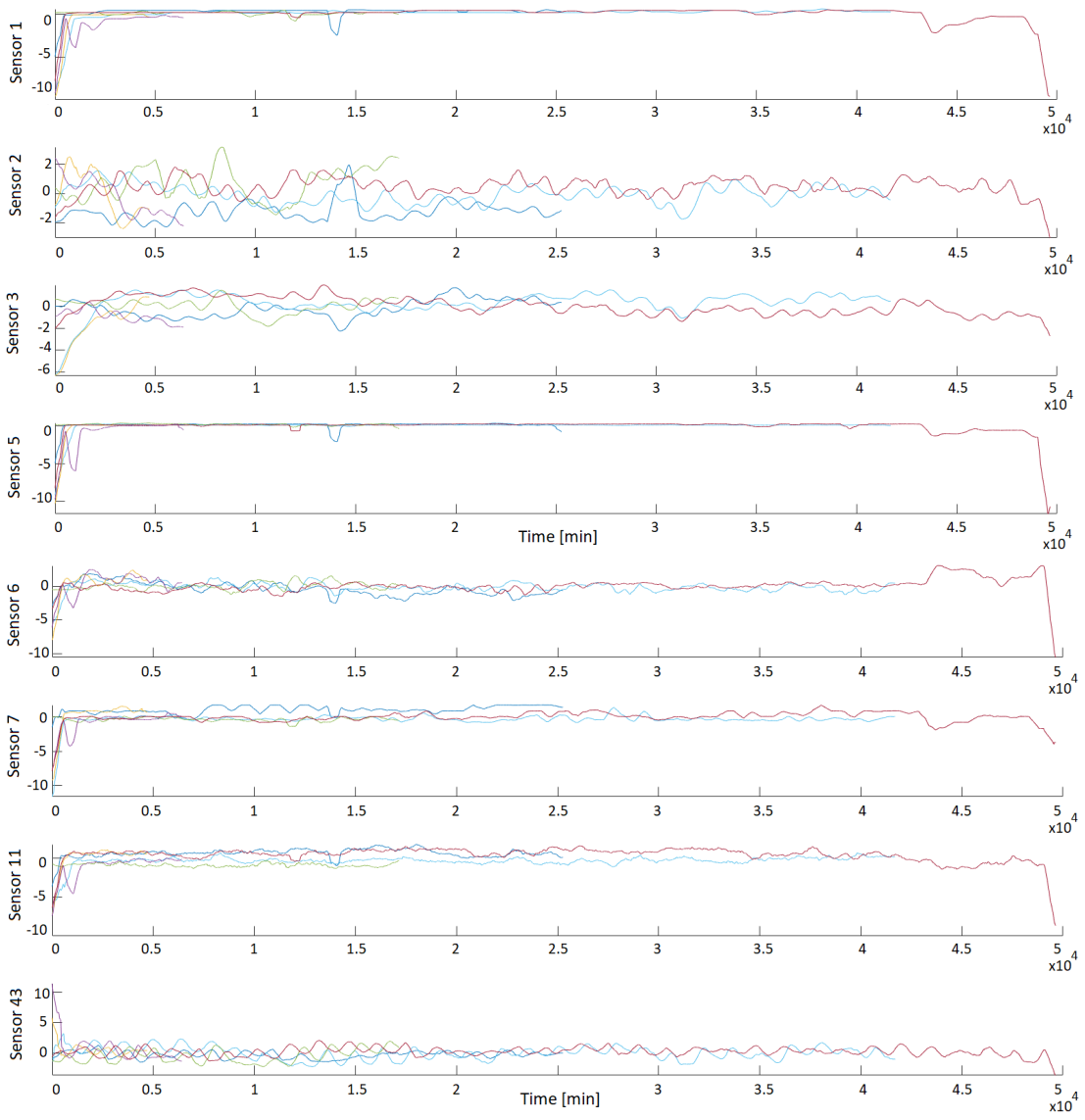


Figure 4.34: Training data signals of selected sensors from trendability analysis.

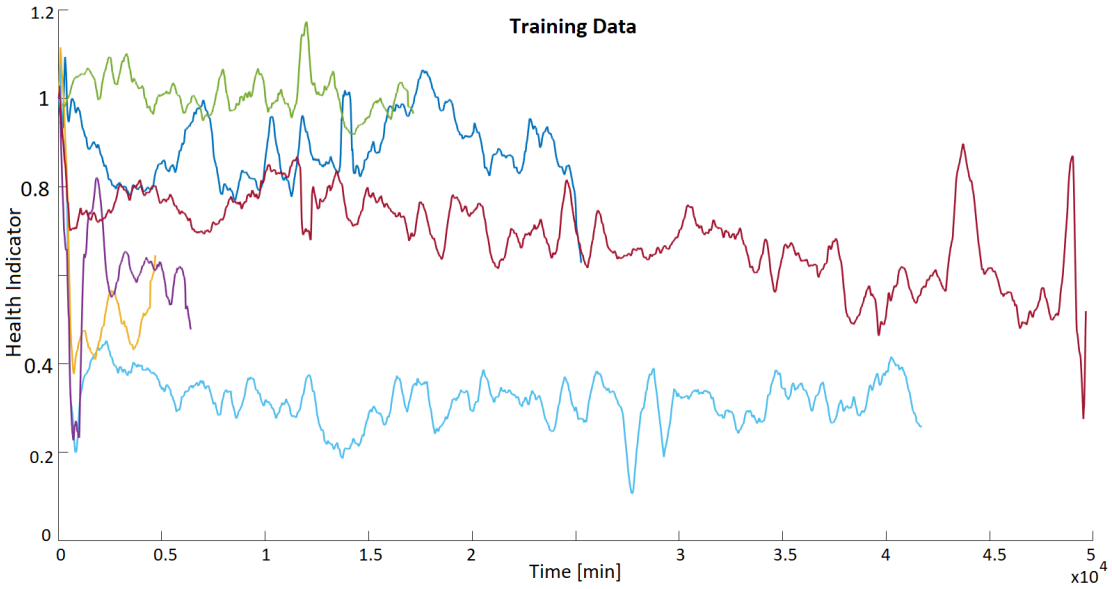


Figure 4.35: Health indicators of validation data.

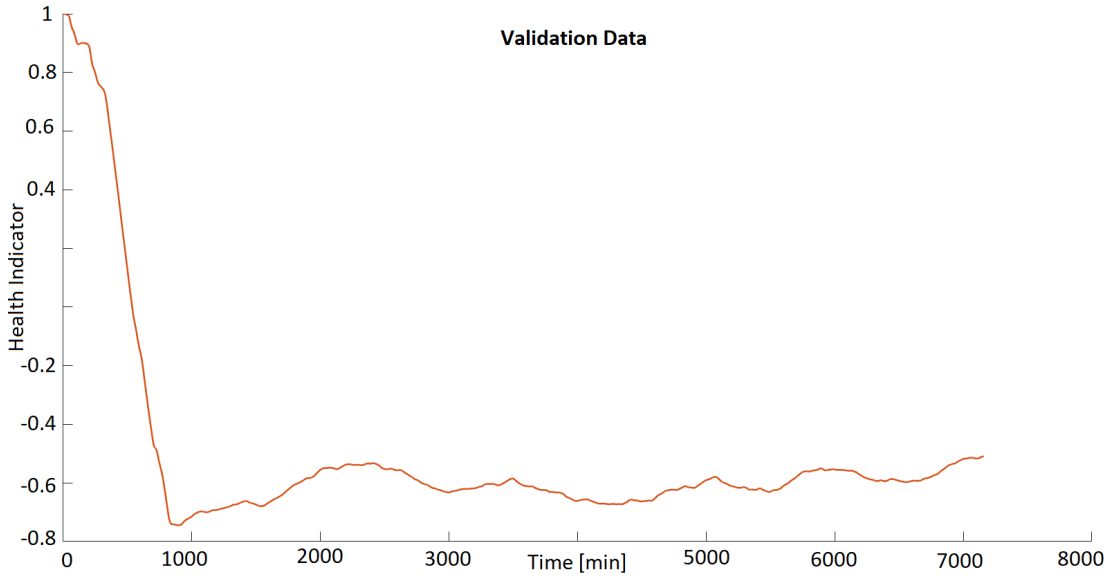


Figure 4.36: Health indicators of training data.

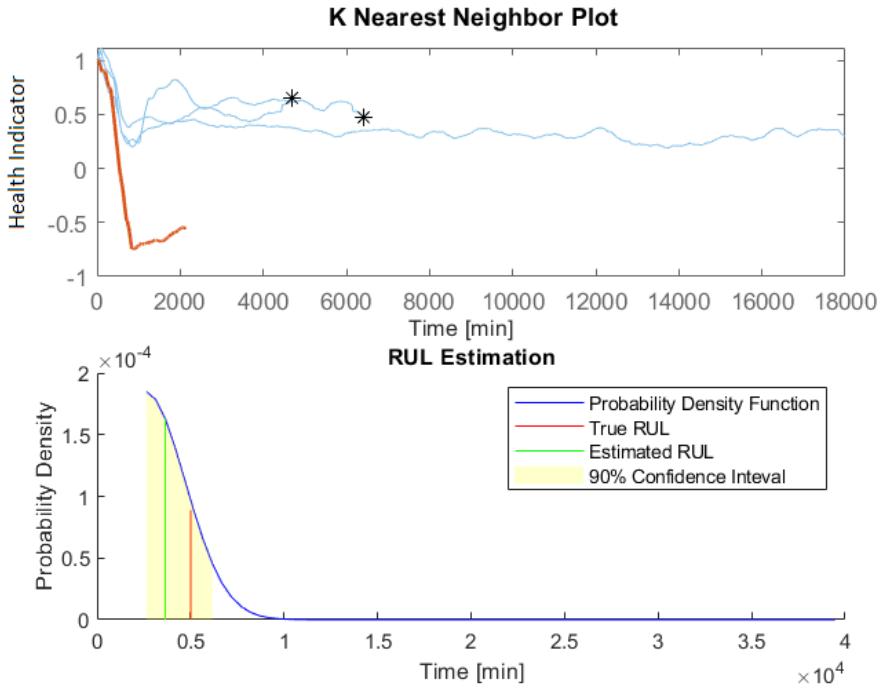


Figure 4.37: RUL estimation results for 30% of time series 4.

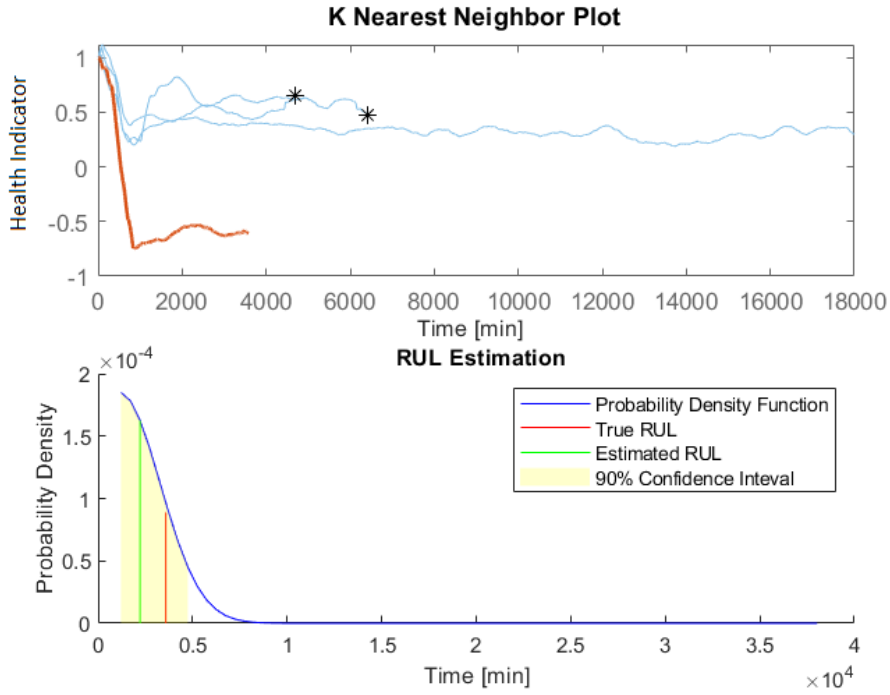


Figure 4.38: RUL estimation results for 50% of time series 4.

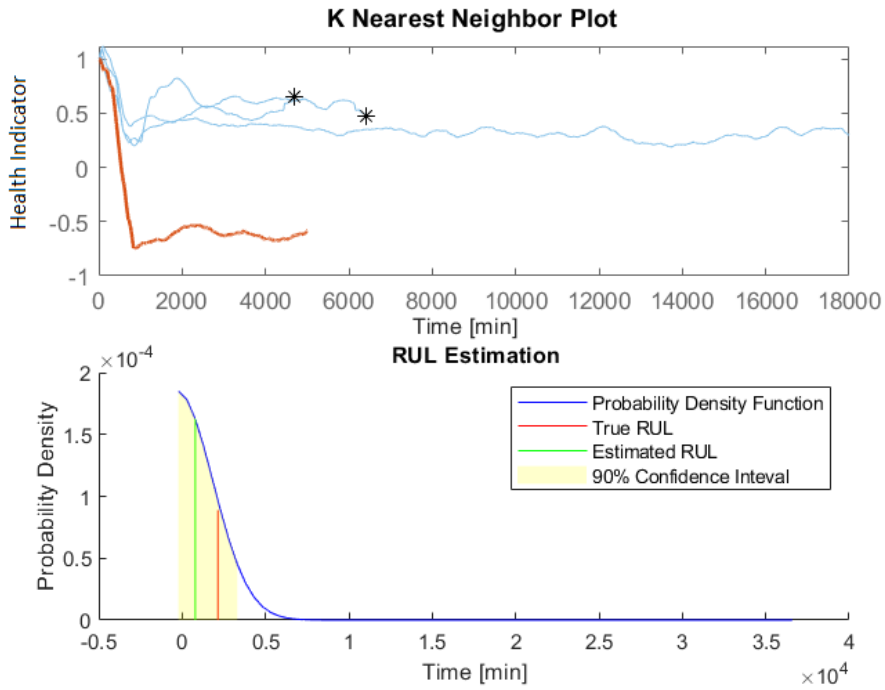


Figure 4.39: RUL estimation results for 70% of time series 4.

similarity models perform estimations on the RUL based on the similarity of the trends of the data (health indicators) and, also, based on the lifetimes of the training time series. Hence, such results are coherent with similarity models theory.

The prediction error was computed as the difference between the true RUL and the estimated RUL. For each global model, the average errors of the three breakpoints are:

- global model 1: average prediction error is 6.4 days;
- global model 2: average prediction error is 3.3 days;
- global model 3: average prediction error is 24.5 days;
- global model 4: average prediction error is 0.9 days;
- global model 5: average prediction error is 29.4 days;
- global model 6: average prediction error is 1.9 days;
- global model 7: average prediction error is 0.5 days.

Considering all global models together, the overall prediction error is 9.6 days. Disregarding the results of global model 3 and global model 5, since these are very deviated from the other models' results, the overall prediction error is 2.6 days.

Despite achieving satisfactory results of RUL estimation of global models for some of the validation time series, the constructed health indicators of such models do not demonstrate the linear degradation trend that was expected. Hence, another approach is proposed.

Global models are generic models since the selection of sensors to construct health indicators was performed based on a degradation model using all the training time series. A contrasting approach is to build models adjusted to each time series, where the sensors' selection for health indicator construction is performed particularly and individually for each time series data. The decision of adapting the models was made aiming to construct health indicators with a more visible degradation trend. Such models are referred to as "Local Models".

Table 4.11: RUL estimation results of global models.

Global Model	Breakpoint	True RUL [min]	Est. RUL [min]	Error
1	30%	12009	21209	overestimate 153.3 h = 6.4 days
	50%	8578	17778	overestimate 153.3 h = 6.4 days
	70%	2148	14346	overestimate 153.3 h = 6.4 days
2	30%	4487	4981	overestimate 8.2 h = 0.3 days
	50%	3205	9691	overestimate 108 h = 4.5 days
	70%	1923	9314	overestimate 123.2 h = 5.1 days
3	30%	29188	-6148	underestimate 589 h = 24.5 days
	50%	20849	-14361	underestimate 586.8 h = 24.5 days
	70%	12509	-22700	underestimate 586.8 h = 24.5 days
4	30%	5012	3651	underestimate 22.7 h = 0.9 days
	50%	3580	2219	underestimate 22.7 h = 0.9 days
	70%	2148	787	underestimate 22.7 h = 0.9 days
5	30%	34751	-7255	underestimate 700.1 h = 29.2 days
	50%	24822	-17656	underestimate 708 h = 29.5 days
	70%	14893	-27440	underestimate 705.6 h = 29.4 days
6	30%	3290	5935	overestimate 44.1 h = 1.8 days
	50%	2350	5239	overestimate 48.2 h = 2 days
	70%	1410	4337	overestimate 48.8 h = 2 days
7	30%	17687	15323	underestimate 39.4 h = 1.6 days
	50%	12634	11484	underestimate 19.2 h = 0.8 days
	70%	7580	10296	overestimate 45.3 h = 1.9 days

Local Models

For Local Models, the selection of sensors to construct the health indicator was performed in an empirical approach. It was tried to select sensors that demonstrated a behavior indicative of a progression until the occurrence of failure, i.e., until the end of the lifetime data.

For each model, one time series was used for the selection of sensors, i.e., one time series was used to train the model, while the other six time series were used to validate the choice of the selected sensors. Additionally, depending on the results of the constructed health indicators of validation time series, RUL estimation was performed to validate the RUL model for Local Models. Thus, 7 local models were trained. The same previous steps explained in section 3.1.2, consisting of data smoothing and data normalization were performed. The difference between Global Models and Local Models is the trended sensors selection approach: for Global Models, the selection was based on the rank of the slope parameters of a linear degradation model (Equation 3.7); for Local Models, the selection was based on empirical observation of the sensors signals.

Following the trended sensors selection, the construction of the health indicators was performed as described in section 3.1.2. Five sensors were selected in each local model and these sensors are listed in Table 4.12.

Table 4.12: Selected sensors in each Local Model.

Local Model	**Sensors**
1	20, 21, 22, 31, 35
2	2, 3, 4, 37, 47
3	1, 3, 4, 11, 49
4	3, 13, 23, 30, 37
5	2, 3, 15, 30, 38
6	3, 4, 27, 35, 37
7	8, 39, 41, 44, 49

In Figure 4.40, the health indicators built with the previously listed sensors for each local model are illustrated. Note that these health indicators refer to the training time series of each local model, that is, these are the health indicators of time series on which the choice of sensors was based.

As noted by observing the curves of the previous figure, when choosing the sensors empirically, the resulting health indicators demonstrate the intended degradation trend. For some of the models, the health indicators of the training time series have in fact a quite evident linear degradation trend – time series 2, 4, and 6. Contrarily, the health indicators of the training time series of the other models do not demonstrate the linear degradation trend so plainly – time series 5 and 7. Furthermore, for the local model 3, the health indicator of the training time series – 3, presents a curve that quickly converges to close to zero and remains practically constant until the end of the time series, which is not advantageous for later estimation of RUL.

However, by selecting the sensors with the empirical observation approach, the resulting health indicators of the validation time series do not demonstrate the expected degradation behavior. Such is possibly due to the fact that a sensor presents curves with distinctive behaviors for different time series. In other words, although a sensor is considered appropriate, based on empirical observation of the training time series, to build the health indicator of the training time series of the local model, that same sensor may not be suitable to build the health indicators of the remaining time series.

In an exemplifying and illustrative way, the results for one of the seven local models are shown. The chosen model is Local Model 2. In Figure 4.41 it can be seen the degradation trend of the sensors selected based on empirical observation of time series 2. The health indicators built with the selected sensors in Local Model 2 are presented in Figure 4.42.

By observing the previous figure, it is noted that, in addition to the health indicator of the time series 2, the health indicator of the time series 6 has a moderately similar trend to what was

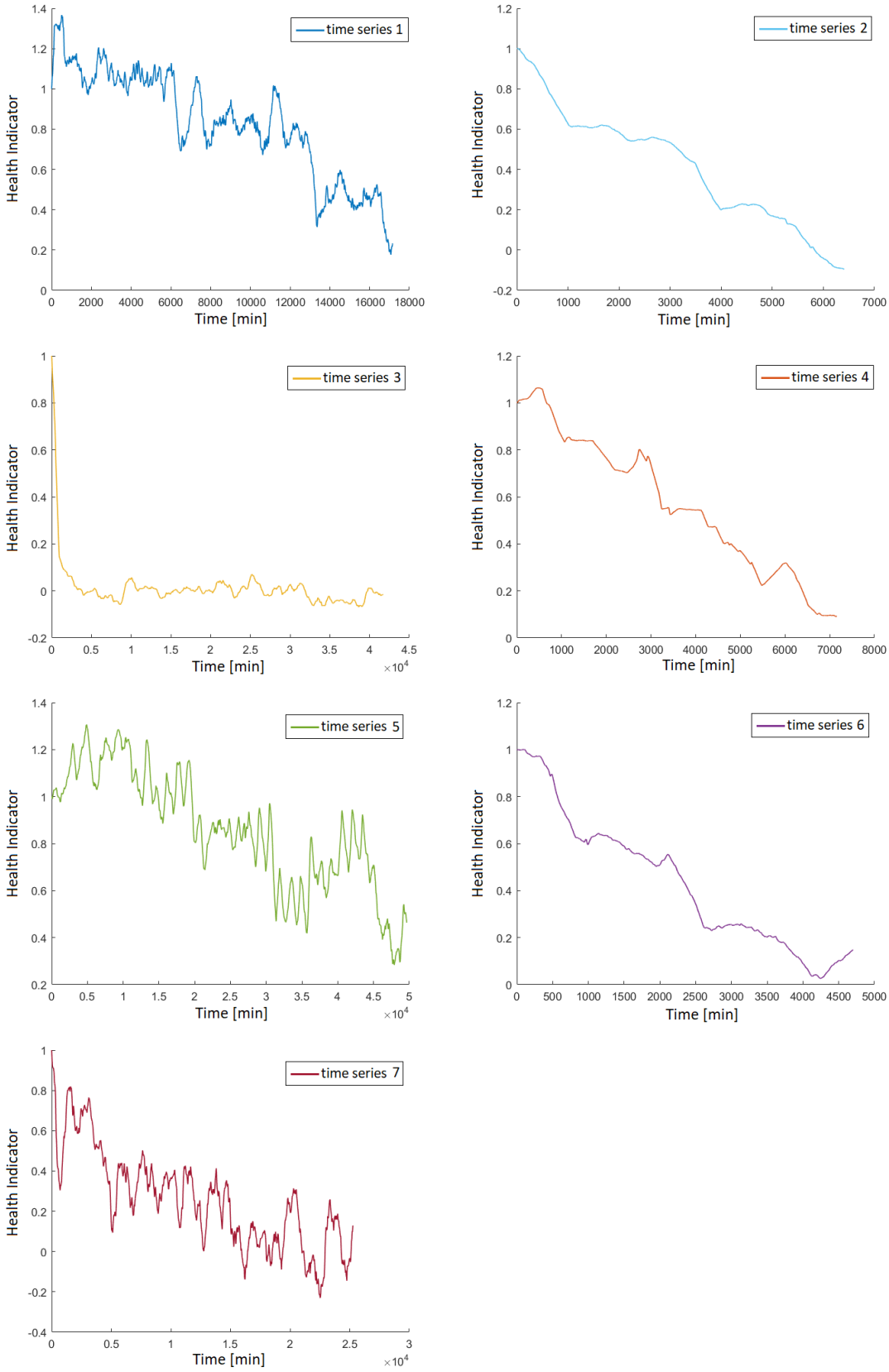


Figure 4.40: Health indicators for training data of each Local Model.

expected. The choice of sensors made for the Local Model 2 produced positive results for time series 2 – training time series, and for time series 6. Therefore, one can say that the empirical choice of sensors for the Local Model 2 works for two of the seven time series.

For Local Model 5, as previously mentioned, the health indicator of the training time series – time series 5, presented a moderately evident degradation trend. This trend is justified by the curves of the sensors selected by empirical observation, which despite being the best of time series 5, do not have very evident degradation trends (Figure 4.43). Regarding the remaining time series for Local Model 5, only the health indicator of time series 1 has a roughly similar behavior (Figure 4.44). Concerning the other local models, the empirical choice of sensors produced positive results only for the corresponding training time series, i.e., works for only one of the seven time series.

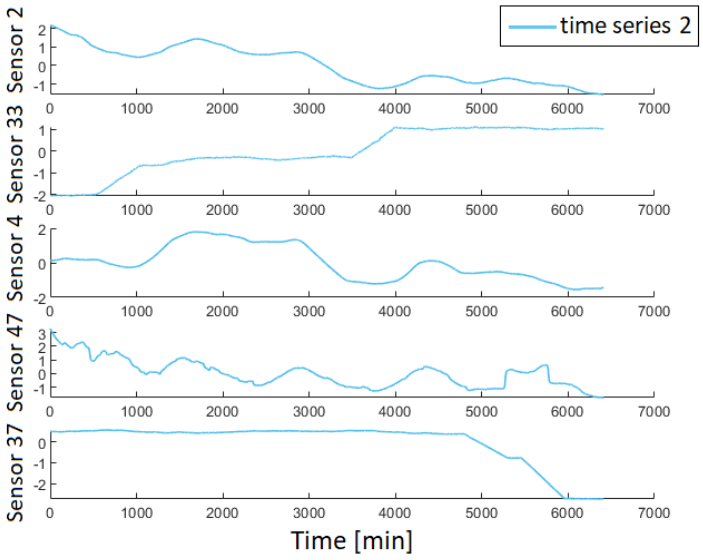


Figure 4.41: Selected sensors for Local Model 2.

In conformity with the results obtained with the construction of health indicators for local models, the results of RUL estimation were validated only for Local Model 2 with the validation time series 6 and for Local Model 5 with the validation time series 1. The referred results are illustrated in Figures 4.45 e 4.46, respectively. Table 4.13 presents in detail the RUL estimation results for the two local models in each of the three breakpoints.

Concerning Local Model 2, using time series 6 as validation, which, as already mentioned, is the only one that presents a relevant health indicator to perform RUL estimation, the error associated with all breakpoints is only 28.5 hours, equivalent to 1.2 days. These are very positive results for RUL estimation since it is possible to predict the remaining time for the pump in time series 6 to fail with an error of just over 1 day.

Concerning Local Model 5, using time series 1 as validation, the error associated with all breakpoints is 441.5 hours, equivalent to 22.6 days. These results are not very useful for predicting the time left until the pump in time series 1 fails, since the forecast has an associated error of more than 22 days. Such is possibly due to the fact that the health indicators of the training time series (time series 5) and time series 1 do not show very similar behaviors and values: the health indicator of time series 5 ends with values near 0.4 whereas the health indicator of time series 1 ends with values close to 0.75 (Figure 4.44). Additionally, the results of RUL estimation may have been negatively affected due to the fact that time series 5 and 1 have very different time duration (time series 1 - 17156 minutes, time series 5 - 49645 minutes).

The possibility of using other groups of sensors was considered, in addition to those that were

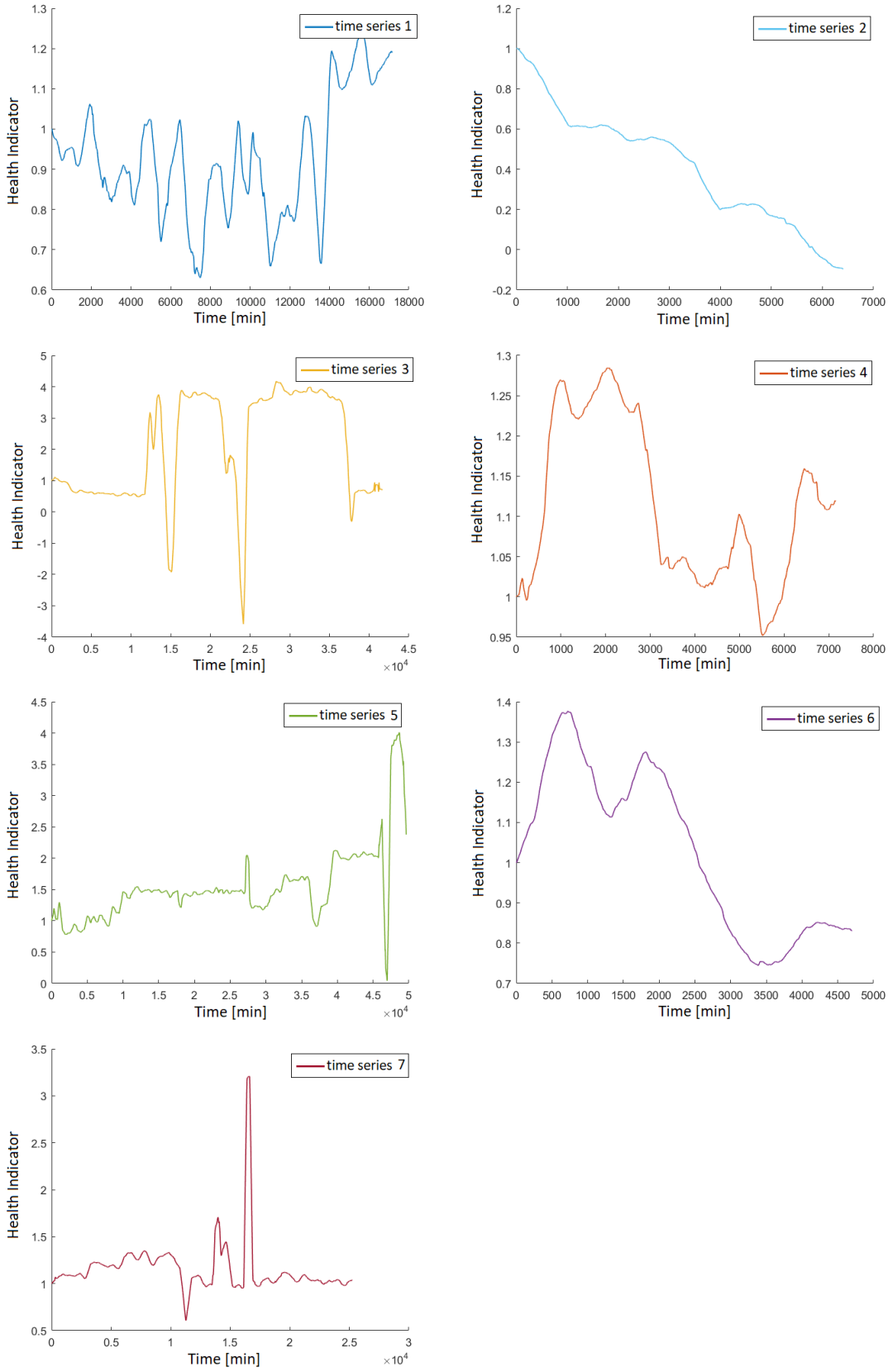


Figure 4.42: Health indicators for the seven time series of Local Model 2.

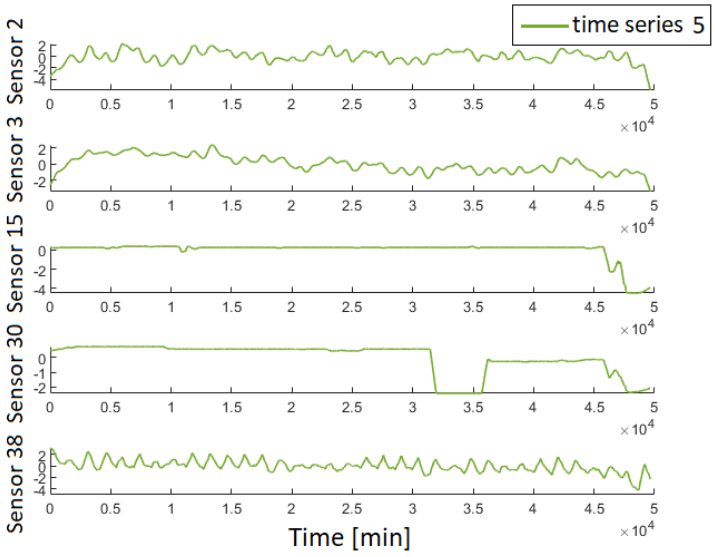
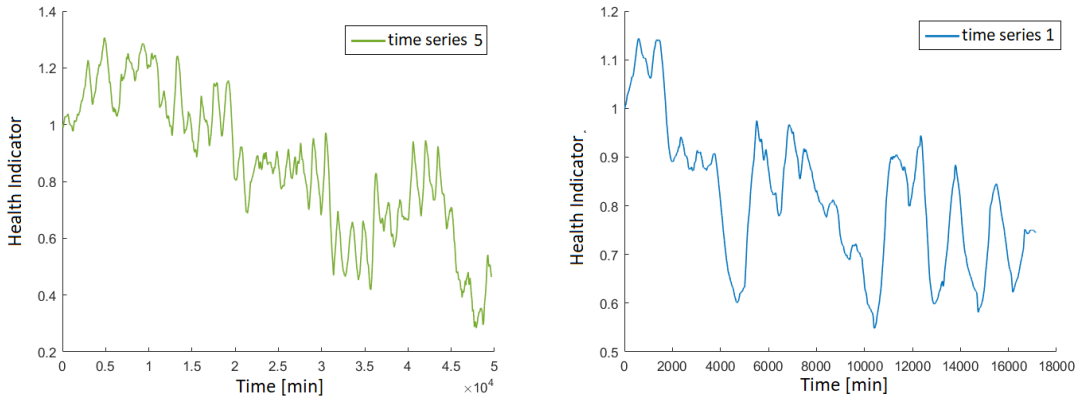


Figure 4.43: Selected sensors for Local Model 5.



(a) (b)

Figure 4.44: Health indicators of Local Model 5 for the time series: (a) 5 and (b) 1.

Table 4.13: RUL estimation results of two local models.

Local Model	**Breakpoint**	**True RUL [min]**	**Est. RUL [min]**	**Error**
2 validated with time series 6	30%	3290	5000	overestimating 28.5 h = 1.2 days
	50%	2350	4060	
	70%	1410	3120	
5 validated with time series 1	30%	12009	44498	overestimating 441.5 h = 22.6 days
	50%	8578	41067	
	70%	5146	37635	

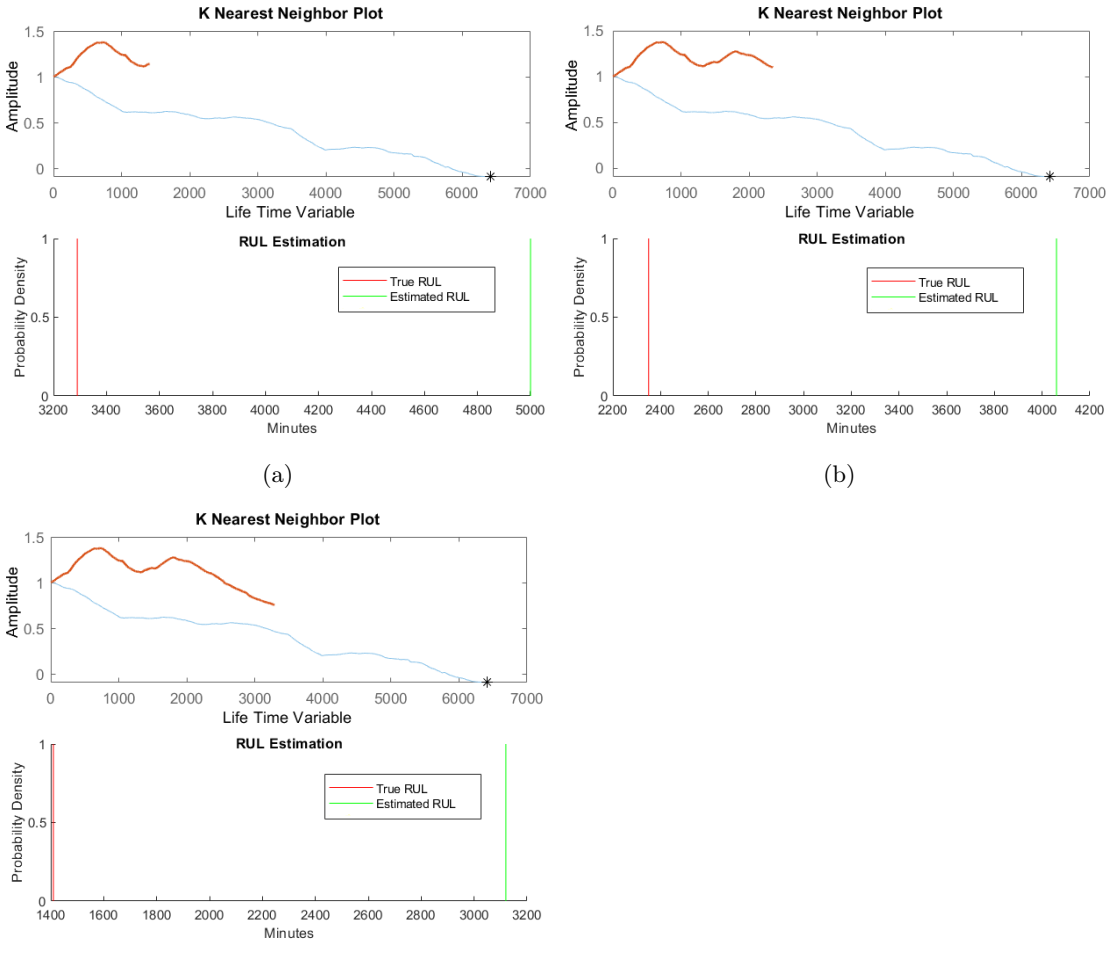
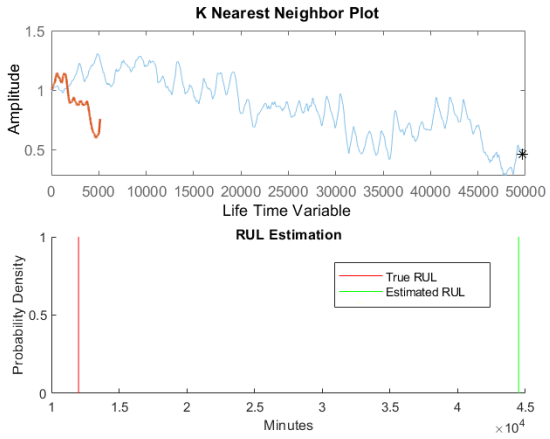
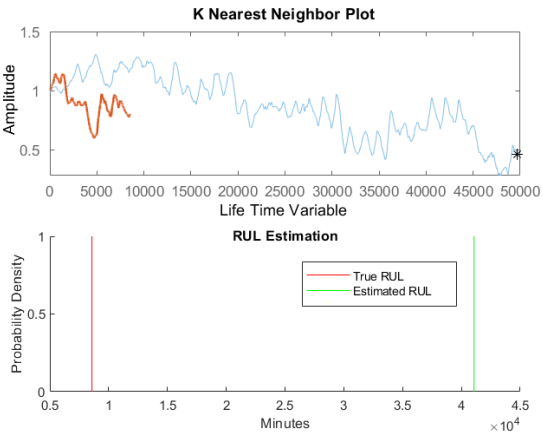


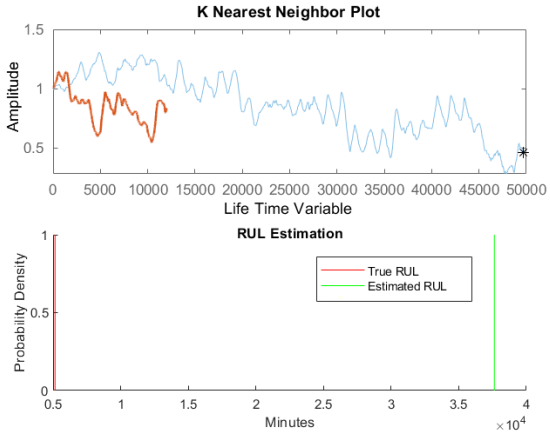
Figure 4.45: RUL estimation results validating Local Model 2 with time series 6 at (a) 30% breakpoint, (b) 50% breakpoint and (c) 70% breakpoint.



(a)



(b)



(c)

Figure 4.46: RUL estimation results validating Local Model 5 with time series 1 at (a) 30% breakpoint, (b) 50% breakpoint and (c) 70% breakpoint.

initially selected. From Table 4.12 one can verify that sensor 3 is common to five of the seven time series and sensors 4 and 37 are common to three of the seven time series. Additionally, sensors 2, 30, 35, and 49 are common to two of the seven time series. The potentiality of using the common sensors 3, 4, and 37 in all local models was tested, however, the results obtained for the health indicators were worse compared to the results obtained with the five sensors initially chosen by empirical observation.

The possibility of including the remaining common sensors with the previous ones was also evaluated. However, as with the previous common sensor group, the resulting health indicators did not exhibit the intended behavior. Other combinations of sensors were tested, but none produced significant improvements in the results. Therefore, the best group of sensors chosen for the local models are the five sensors chosen initially by empirical observation of the sensor curves.

Although the approach of the local models fulfills the initial purpose of building health indicators with more clear degradation trends compared to the health indicators of the global models, such approach is not more advantageous for RUL estimation purposes, since the health indicators for validation time series did not present trends similar to the corresponding training time series.

4.3 Economic Analysis Applied to a Set of Results of Kaggle Case Study

To validate the reasoning behind the economic assessment explained in section 3.1.3, the same was applied to the set of results of Global Model 4 of Kaggle case study.

From the values relating to Global Model 4 present in Table 4.11, the chart presented in Figure 4.47 was elaborated. In this figure, the red line illustrates the evolution of the true RUL and the blue line corresponds to the evolution of the estimates calculated by the model over the recorded observations, which correspond to the time in minutes. It can be seen that, despite having a certain deviation from the real value, the RUL estimates of the model maintain practically the same trend throughout the evaluation.

The chart in Figure 4.48 shows the costs for a normal operation of the pump until failure - at 119 hours of pump operation -, illustrated by the red line, and also, the costs for a pump operation with maintenance actions applied at the time of 50% of estimated RUL - at 48.3 hours of pump operation -, illustrated by the green line.

The pump acquisition costs were considered to be $Ca = 8000€$, and, according to the proportions mentioned in the methodological explanation of the economic analysis in section 3.1.3, it was considered that the costs of repairing a serious failure are $Cf = 3000€$ and the costs of applying maintenance actions are $Cm = 400€$. For the calculation of operating costs, only costs related to energy consumption were considered: $cost = energy\ consumed$ (kWh) $\times\ tariff$ (€/kWh). It was considered that the pump would be operating without interruption, with an energy consumption of 100kWh and that the cost of the tariff is 0.1371€/kWh [54]. From the mentioned values, the curves of Figure 4.48 were drawn. With this chart, it is possible to have a more visual perception of the advantage in terms of costs and operating time in applying an RUL estimation model in order to plan maintenance actions. In this case, with this set of results, it would be possible to reach the moment when the pump would fail (at 119 hours) with a saving of more than 2500€, with the potential failure already prevented by the maintenance actions applied at the moment of 50% of the estimated lifetime and with the lifetime of the pump extended beyond the supposed.

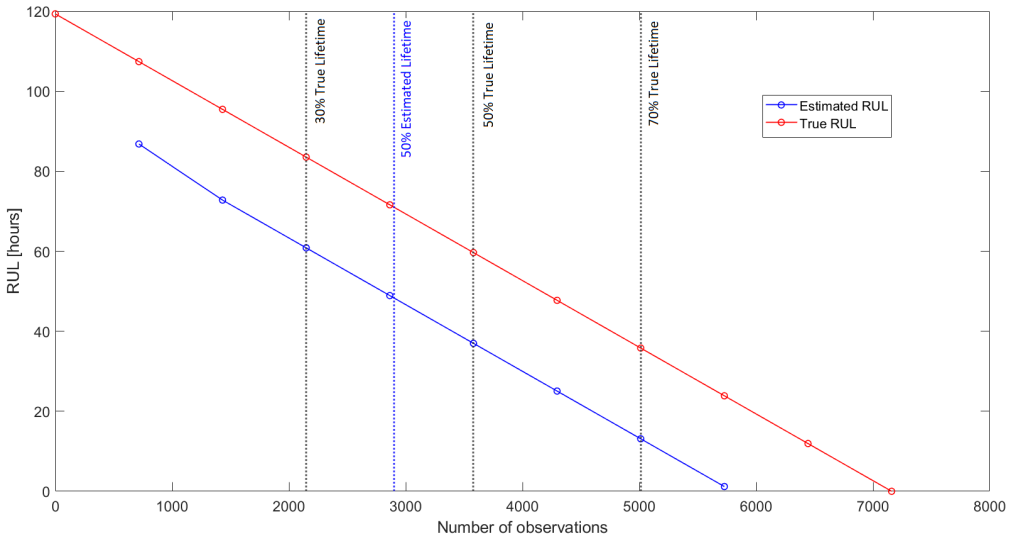


Figure 4.47: True and Estimated RUL of Global Model 4.

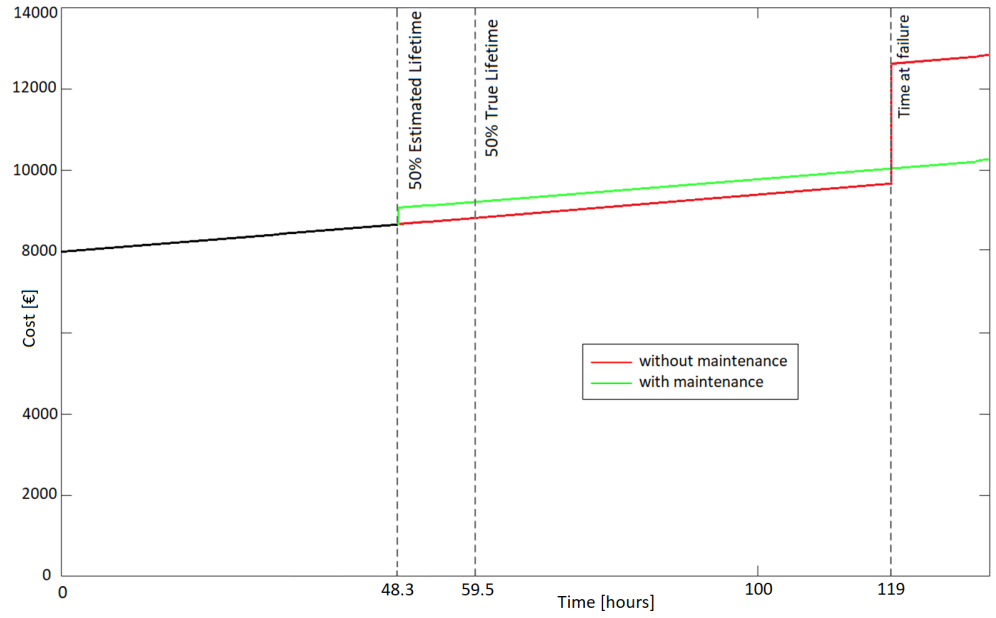


Figure 4.48: Cost comparison between the two scenarios.

Chapter 5

Conclusions

The main focus of the work of this dissertation was related to the development of a predictive maintenance methodology applied to hydraulic pumps, contemplating the following tasks: (i) detection and classification of the pump condition and any associated failures; (ii) prediction of the remaining service life of a pump in operation.

Although the work developed did not cover the elaboration of a complete predictive maintenance methodology, ready to use in an industrial setting, and the fact that it was not implemented in an experimental study case, it was still possible to explore and explain in detail the operation of relevant techniques that can be integrated into an eventual complete maintenance plan for a hydraulic pump.

The aspects investigated in the course of the research part of the work focused on techniques and approaches used in the development of preventive and predictive maintenance methodologies for pumps and other hydraulic systems: methods for identifying potential failures and the components most susceptible to failure, different types of analysis to assess the state of the systems, construction of condition indicators following different criteria, prediction of the systems' behavior using machine learning algorithms, types of models used in the component health prognosis, among others.

Regarding the methodology used in the Fault Detection and Classification task, it was found that:

- Regarding the studied features, those referring to speed signals indicate no relevance for fault classification. In general, it is concluded that the most relevant features are those of pressure and flow signals, and only later, those of torque;

- In both case studies, the algorithms that demonstrated the best performances were the decision tree algorithms and the ensemble algorithms. Taking into account what has been evaluated, these two types of algorithms indicate that they are the most relevant for classifying failure modes from pump data;

- In the first study case, with an average of only 30 observations recorded in the data set for each failure mode, the best classification was obtained of 82.9%, while in the second case of study, with 500 observations recorded in the data set for each failure mode, the best classification was 94.6%. The amount of data used in the development of techniques that employ the use of machine learning algorithms influences the quality of the classification of the models. Thus, it appears that, as stated in the literature, the more data available, the better the performance of the classification models;

- In the second case study, the Fine Decision Tree algorithm maintained the same performance levels using the three different groups of features - the first and second groups that differed in the amount of best features used, according to the ANOVA metrics, and, the third group, where only the features of the flow and head signals were included. This

algorithm indicates to be a good candidate for a quick, simple, viable, and economically interesting solution for the classification of failures with few measured variables.

Regarding the methodology adopted for the task of RUL Estimation, it was found that:

- The proposed approach works well for the first case study, however, it fails in some aspects when applied to the second case study. It appears that the quality of the results of this methodology depends not only on the techniques followed for the construction of health indicators and RUL estimation models but also on the suitability of the data analyzed for the purpose in question;
- The second case study of RUL demonstrated the critical need to have access to information regarding the equipment and the type of associated failure, to successfully implement the proposed methodology. In fact, this case study had the particularity of not knowing much about the data set, what variables the sensors were associated with, and the nature of the failures that led to pump failures during the data measurement period;
- Regarding the second case study, the decisions made in the initial data analysis - filling in the missing values, dividing the data set into 7 time series, and considering that the failures could be related in some way - were important since they enabled RUL estimation results to be achieved;
- Still regarding the second case study of RUL, concerning global models, those who obtained the worst results of RUL estimation were those whose time series had periods much more deviated from the average periods of the other time series. This was due to how it was decided to compute the RUL estimation - using similarity models, which estimate values based on similarities in the data trend of the other time series and also based on the periods of those same series.

In general, it appears that the proposed techniques and their methodologies prove to be viable options to be integrated into a predictive maintenance plan applied to hydraulic pumps. Through the simple economic analysis presented in this work, it was found that the application of approaches related to predictive maintenance is economically viable and advantageous. When it is possible to estimate, even with a certain associated error, the time remaining until the pump fails, it is possible to plan and schedule maintenance actions to mitigate and postpone the occurrence of the potential failure in time. Even though it was not included in the economic analysis with associated values, the fault classification technique also proved to be advantageous in a predictive maintenance methodology since, through sensor data, it is possible to identify and classify whether the pump is in healthy operation or defective and, also, the associated fault.

The work developed had some limitations, such as: (1) the impossibility of having collected experimental data and the following difficulty in finding pertinent data sets in online repositories; (2) having used data sets where it was not possible to guarantee their quality and suitability for the analyzes carried out; and (3) the fact that the techniques described in this work have been applied to only two case studies each.

To improve the approaches proposed as integral parts of a future predictive maintenance methodology, the following are proposed as future work:

- Application of methodologies with data sets that guarantee their quality and quantity, with relevant information about their systems and accompanied by a complete history of associated failures, thus ensuring the suitability of the data for the objective under study;
- Implementation of the models in an experimental context, adding to the steps proposed in this work the data collection step, to better understand the viability of the proposed methodology in the face of a real situation;
- Application of the fault classification methodology including other types of features, such as machinery rotating features, which seem to have potential considering the purpose of this study;

- Exploration of other types of RUL estimation models in addition to the residual comparison based similarity model studied;
- Elaboration of an economic analysis that seeks to evaluate the feasibility of implementing a predictive maintenance plan in an industrial context, taking into account the cost of applying the proposed approaches and considering the factors necessary for its success: the acquisition of sensors and equipment, the costs related to the computation of models and the costs associated with human resources;
- Improve and enhance the future predictive maintenance methodology, including other maintenance tasks, in addition to those proposed here.

Intentionally blank page.

Bibliography

[1] B. Coelho, *Energy efficiency of water supply systems using optimisation techniques and micro-hydroturbines*. PhD thesis, Universidade de Aveiro, 2016.

[2] R. Santos Coutinho and A. Kepler Soares, "Simulação de bombas com velocidade de rotação variável no EPANET," *Eng Sanit Ambient*, vol. 22, no. 4, pp. 797–808, 2017.

[3] C. R. de Mello and T. Y. Jr., "Escolha De Bombas Centrífugas," *Revista Ciência e Agrotecnologia*, no. 29, pp. 1–27, 1999.

[4] R. Mackay, "Understanding positive displacement pumps," 2019. http://www.pumpscout.com/articles-expert-advice/understanding-positive-displacement-pumps-aid89.html, Last accessed on 2020-06-02.

[5] C. Bipat, "Main Types of Pumps: Centrifugal and Positive Displacement," 2018. https://www.ny-engineers.com/blog/main-types-of-pumps, Last accessed on 2020-06-03.

[6] B. Nesbitt, *Handbook of Pumps and Pumping*. No. December, Elsevier Science & Technology Books, 2006.

[7] M. Sahdev, "Centrifugal Pumps: Basic Concepts of Operation, Maintenance, and Troubleshooting (Part-I)," *The Chemical Engineers' Resources*, pp. 1–31, 2012.

[8] R. S. Beebe, *Predictive Maintenance of Pumps Using Condition monitoring*. Elsevier, 1 ed., 2004.

[9] D. Kernan, "Pumps 101 : Operation, Maintenance and Monitoring Basics," *ITT White Paper*, pp. 1–10, 2009.

[10] K. Mckee, G. Forbes, I. Mazhar, R. Entwistle, and I. Howard, "A review of major centrifugal pump failure modes with application to the water supply and sewerage industries," in *ICOMS Asset Management Conference*, (Gold Coast, QLD, Australia), pp. 1–12, 2011.

[11] T. F. Wnek, "Pressure Pulsations Generated by Centrifugal Pumps," tech. rep., Warren, Massachusetts, 1987.

[12] EFNMS, "European Federation of National Maintenance Societies." http://www.efnms.eu/, Last accessed on 2020-06-13.

[13] P. Leite, E. Vivas, L. Valente, F. Ferreira, J. Costa, and M. Teixeira, "Avaliação de desempenho de grupos eletrobomba através da realização de testes de eficiência," *12.º Congresso da Água*, no. 1, p. 15, 2014.

[14] M. Jensen, "Is it worth reducing unplanned downtime with machine learning?," 2019. https://neurospace.io/blog/2019/02/is-it-worth-reducing-unplanned-downtime-with-machine-learning/, Last accessed on 2020-06-14.

[15] M. Walsh, "Predictive & Preventive Maintenance," 2011. https://www.pumpsandsystems.com/predictive-preventive-maintenance, Last accessed on 2020-06-14.

[16] A. Budris, "How to Get More From Pump Preventive Maintenance," 2008. https://www.waterworld.com/technologies/pumps/article/16189801/how-to-get-more-from-pump-preventive-maintenance, Last accessed on 2020-06-13.

[17] S. Peters, "Creating Predictive and Preventive Maintenance Plans for Pumps," 2019. https://blog.craneengineering.net/difference-between-predictive-and-preventative-pump-maintenance, Last accessed on 2020-06-14.

[18] M. B. J. Navega and O. E. F. D. Júnior, "Manutenção Preditiva em bombas," *Bolsista de valor*, vol. 2, pp. 173–177, 2012.

[19] P. V. Senthiil, V. S. Mirudhuneka, and A. Shirrushti, "Predictive Maintenance Model Development Using Life Prediction Methodology," *An International Journal (ESTIJ)*, vol. 4, no. 3, pp. 2250–3498, 2014.

[20] F. Ahmed, "Condition Based Maintenance– CBM," 2017. https://www.fiixsoftware.com/condition-based-maintenance/, Last accessed on 2020-06-14.

[21] R. J. U. d. M. Duarte Ferreira, *Condition-Based Maintenance Framework Development for Several Applications*. PhD thesis, Universidade do Minho, 2014.

[22] J. F. Leonard, "What is FMEA," 2016. http://www.jimleonardpi.com/blog/what-is-fmea/, Last accessed on 2020-07-20.

[23] E. Lisowski and J. Fabiś, "Prediction of potential failures in hydraulic gear pumps," *Archives of Foundry Engineering*, vol. 10, no. 3, pp. 73–76, 2010.

[24] H. Shen, Z. Li, L. Qi, and L. Qiao, "A method for gear fatigue life prediction considering the internal flow field of the gear pump," *Mechanical Systems and Signal Processing*, vol. 99, pp. 921–929, 2018.

[25] D. L. Rivera, M. R. Scholz, M. Fritscher, M. Krauss, and K. Schilling, "Towards a Predictive Maintenance System of a Hydraulic Pump∗," *IFAC PapersOnline*, vol. 51, no. 11, pp. 447–452, 2018.

[26] J. L. Parrondo, S. Velarde, and C. Santolaria, "Development of a predictive maintenance system for a centrifugal pump," *Journal of Quality in Maintenance Engineering*, vol. 4, no. 3, pp. 198–211, 1998.

[27] D. Bansal, D. J. Evans, and B. Jones, "A real-time predictive maintenance system for machine systems," *International Journal of Machine Tools & Manufacture*, vol. 44, pp. 759–766, 2004.

[28] S. Farokhzad, H. Ahmadi, A. Jaefari, M. R. A. A. Abad, and M. R. Kohan, "897. Artificial neural network based classification of faults in centrifugal water pump," *Journal of Vibroengineering*, vol. 14, no. 4, pp. 1734–1744, 2012.

[29] R. Zouari, S. Sieg-Zieba, and M. Sidahmed, "Fault Detection System for Centrifugal Pumps Using Neural Networks and Neuro-Fuzzy," *Surveillance 5 Cetim Senlis*, no. October, 2004.

[30] W. Han, L. Nan, M. Su, R. Li, and X. Zhang, "Research on the Prediction Method of Centrifugal Pump Performance Based on a Double Hidden Layer BP Neural Network," *Energies*, vol. 12, no. 14, pp. 1–14, 2019.

[31] S. Zhang, M. Hodkiewicz, L. Ma, and J. Mathew, "Machinery condition prognosis using multivariate analysis," *Engineering Asset Management*, 2006.

[32] F. Calabrese, A. Regattieri, L. Botti, C. Mora, and F. G. Galizia, "Unsupervised fault detection and prediction of remaining useful life for online prognostic health management of mechanical systems," *Applied Sciences*, vol. 10, no. 12, 2020.

[33] Y. Peng, M. Dong, and M. J. Zuo, "Current status of machine prognostics in condition-based maintenance: A review," *International Journal of Advanced Manufacturing Technology*, vol. 50, no. 1-4, pp. 297–313, 2010.

[34] S. K. Gajawada, "ANOVA for Feature Selection in Machine Learning," 2019. https://towardsdatascience.com/anova-for-feature-selection-in-machine-learning-d9305e228476, Last accessed on 2020-11-20.

[35] Mathworks, "MATLAB Documentation." https://www.mathworks.com/help/, Last accessed on 2021-01-05.

[36] R. J. MacKenzie, "One-Way vs Two-Way ANOVA: Differences, Assumptions and Hypotheses," 2018. https://www.technologynetworks.com/informatics/articles/one-way-vs-two-way-anova-definition-differences-assumptions-and-hypotheses-306553, Last accessed on 2020-11-20.

[37] P. Gupta, "Decision Trees in Machine Learning," 2017. https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052, Last accessed on 2020-11-15.

[38] X. Yang, "Linear Discriminant Analysis, Explained," 2020. https://towardsdatascience.com/linear-discriminant-analysis-explained-f88be6c1e00b, Last accessed on 2020-11-15.

[39] R. Gandhi, "Support Vector Machine — Introduction to Machine Learning Algorithms," 2018. https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47, Last accessed on 2020-11-15.

[40] O. Harrison, "Machine Learning Basics with the K-Nearest Neighbors Algorithm," 2018. https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761, Last accessed on 2020-11-15.

[41] T. Acharya, "Advanced Ensemble Classifiers," 2019. https://towardsdatascience.com/advanced-ensemble-classifiers-8d7372e74e40, Last accessed on 2020-11-15.

[42] S. Narkhede, "Understanding Confusion Matrix," 2018. https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62, Last accessed on 2020-12-18.

[43] S. Chakraborty, N. Gebraeel, M. Lawley, and H. Wan, "Residual-life estimation for components with non-symmetric priors," *IIE Transactions (Institute of Industrial Engineers)*, vol. 41, no. 4, pp. 372–387, 2009.

[44] T. Wang, J. Yu, D. Siegel, and J. Lee, "A similarity based prognostics approach for remaining useful life estimation of engineered systems," *International Conference on Prognostics and Health Management*, pp. 4–9, 2008.

[45] E. Zio and F. Di Maio, "A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system," *Reliability Engineering and System Safety*, vol. 95, no. 1, pp. 49–57, 2010.

[46] Hydraulic Institute, Europump, and Office of Industrial Technologies - US Department of Energy, "Pump Life Cycle Costs: A Guide to LCC Analysis for Pumping Systems - Executive Summary," tech. rep., 2001.

[47] N. Aleixo, "LCC – Custo do Ciclo de Vida Útil de uma Bomba," 2013. https://www.ksb.com/ksb-pt/Informacoes_tecnicas-noticias_ch/Arquivo/2013-info-tecnicas-e-noticias/lcc-custo-do-ciclo-de-vida-util-de-uma-bomba-/177614/, Last accessed on 2020-12-19.

[48] J. Wang, L. Zhang, L. Duan, and R. X. Gao, "A new paradigm of cloud-based predictive maintenance for intelligent manufacturing," *Journal of Intelligent Manufacturing*, vol. 28, no. 5, pp. 1125–1137, 2017.

[49] Mathworks, *Predictive Maintenance Toolbox™ User's Guide*. 2020.

[50] R. Isermann, *Fault-Diagnosis Applications Model-Based Condition Monitoring*. 2011.

[51] National Aeronautics and Space Administration, "Prognostics Center of Excellence - Data Repository." https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/, Last accessed on 2021-01-10.

[52] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation," *2008 International Conference on Prognostics and Health Management*, pp. 1–9, 2008.

[53] N. Phantawee, "pump_sensor_data | Kaggle," 2018. https://www.kaggle.com/nphantawee/pump-sensor-data, Last accessed on 2021-01-11.

[54] "Preços da electricidade para utilizadores domésticos e industriais," 2018. https://www.pordata.pt/Europa/Preços+da+electricidade+para+utilizadores+domésticos+e+industriais+(Euro+ECU)-1477, Last accessed on 2020-12-20.

Appendices

Appendix A

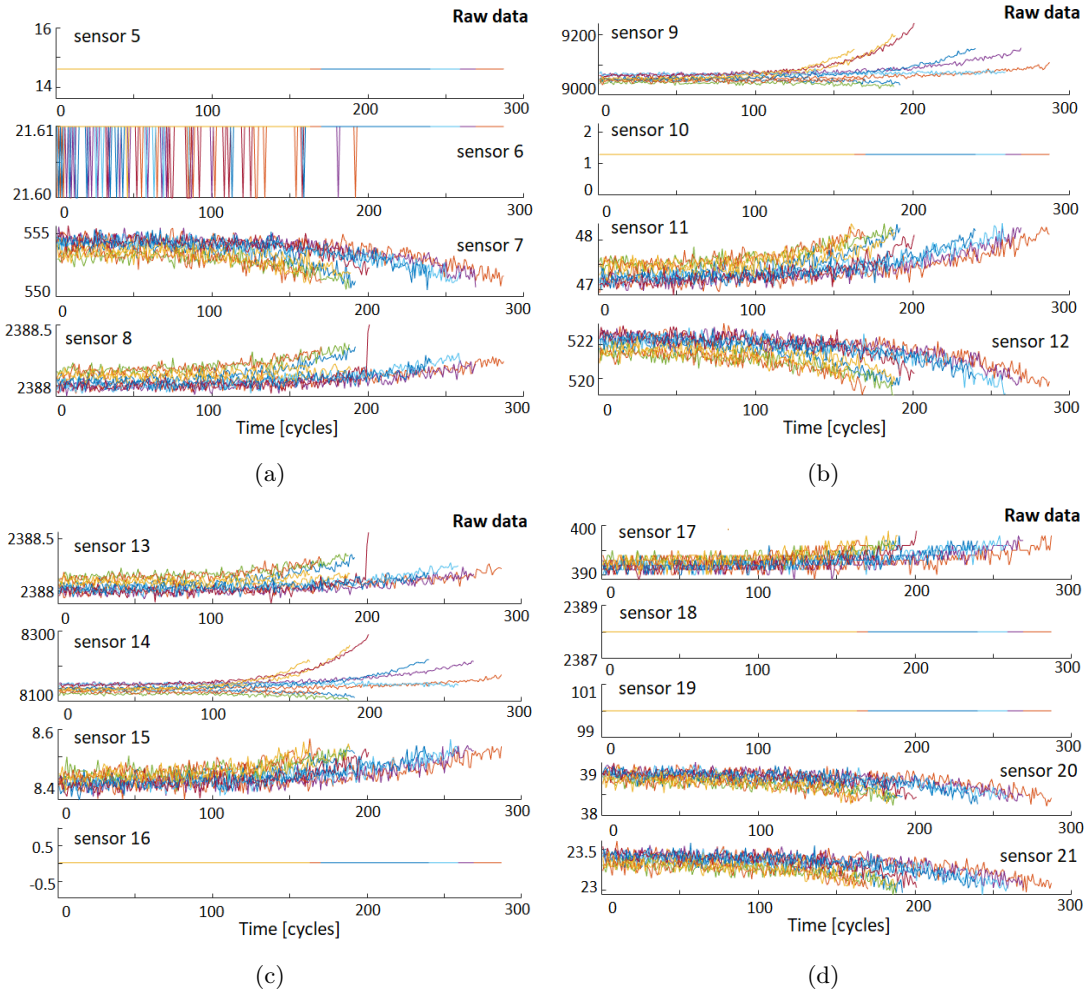


Figure A.1: Sample of 10 time series of raw training data set of sensors: (a) 5-8, (b) 9-12, (c) 13-16, (d) 17-21.

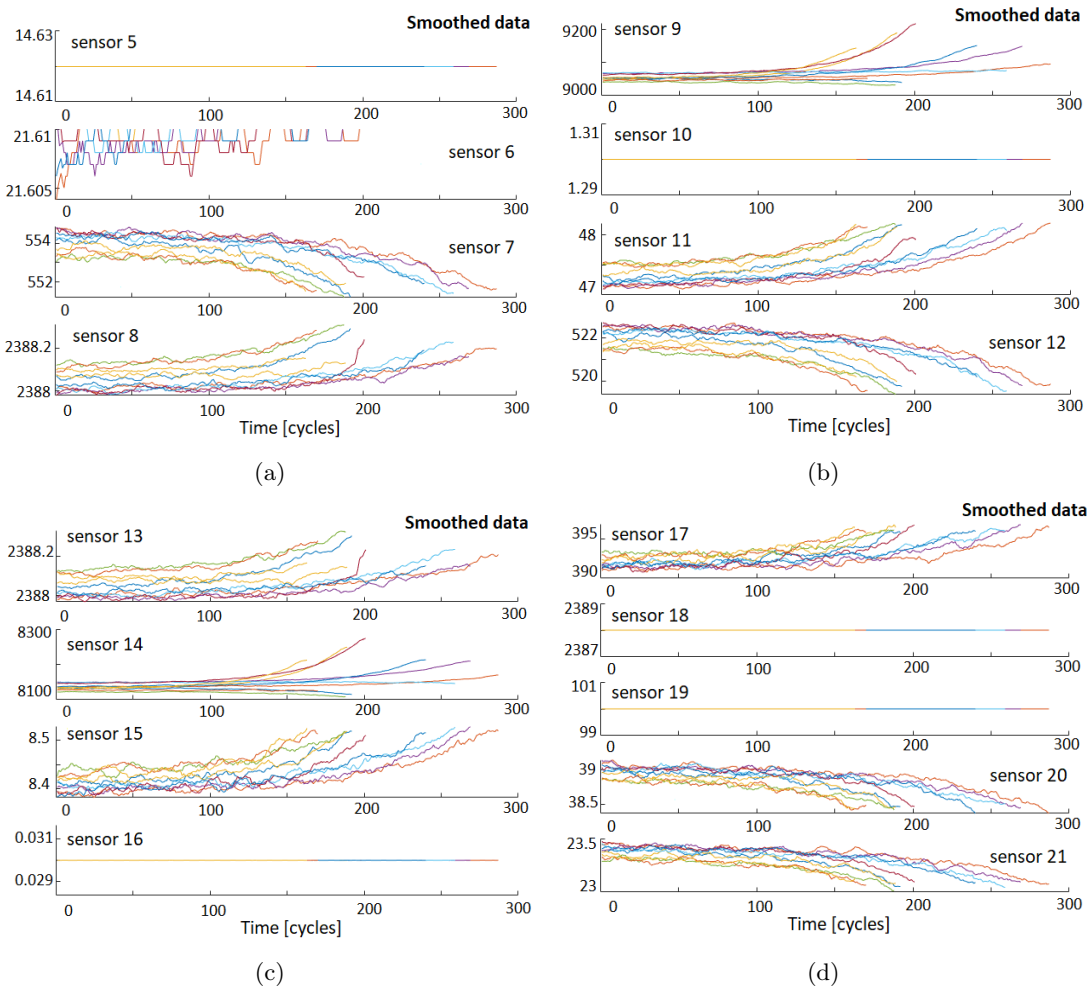


Figure A.2: Sample of 10 time series of smoothed training data set of sensors: (a) 5-8, (b) 9-12, (c) 13-16, (d) 17-21.

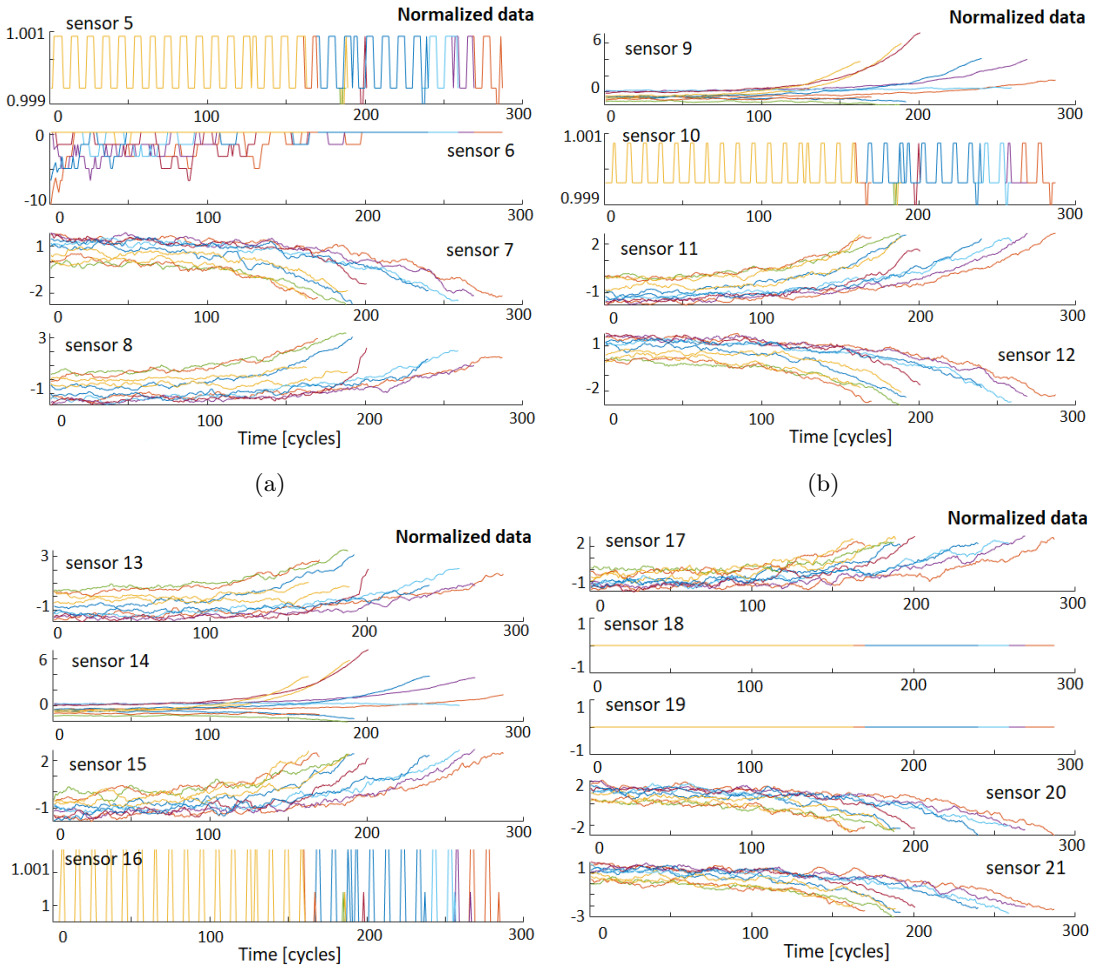


Figure A.3: Sample of 10 time series of normalized training data set of sensors: (a) 5-8, (b) 9-12, (c) 13-16, (d) 17-21.

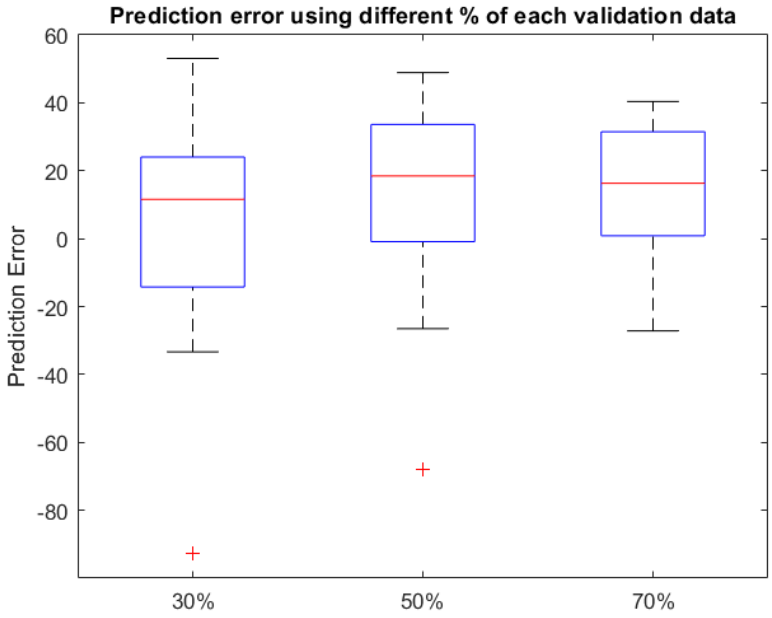


Figure A.4: Box plots with median, 25th and 75th percentiles and outliers of prediction error for each breakpoint of validation data set.

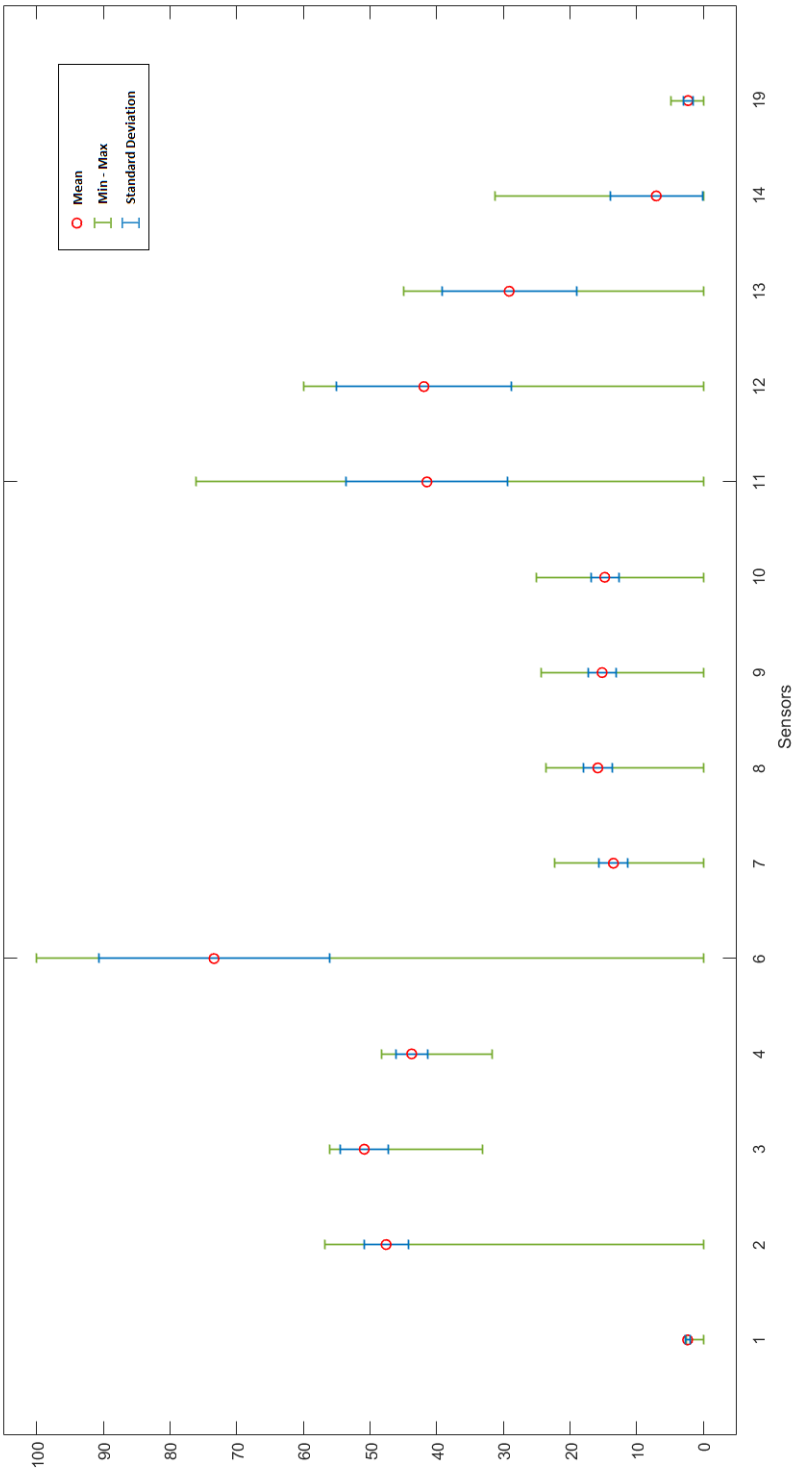


Figure A.5: Statistical values for each sensor (A).

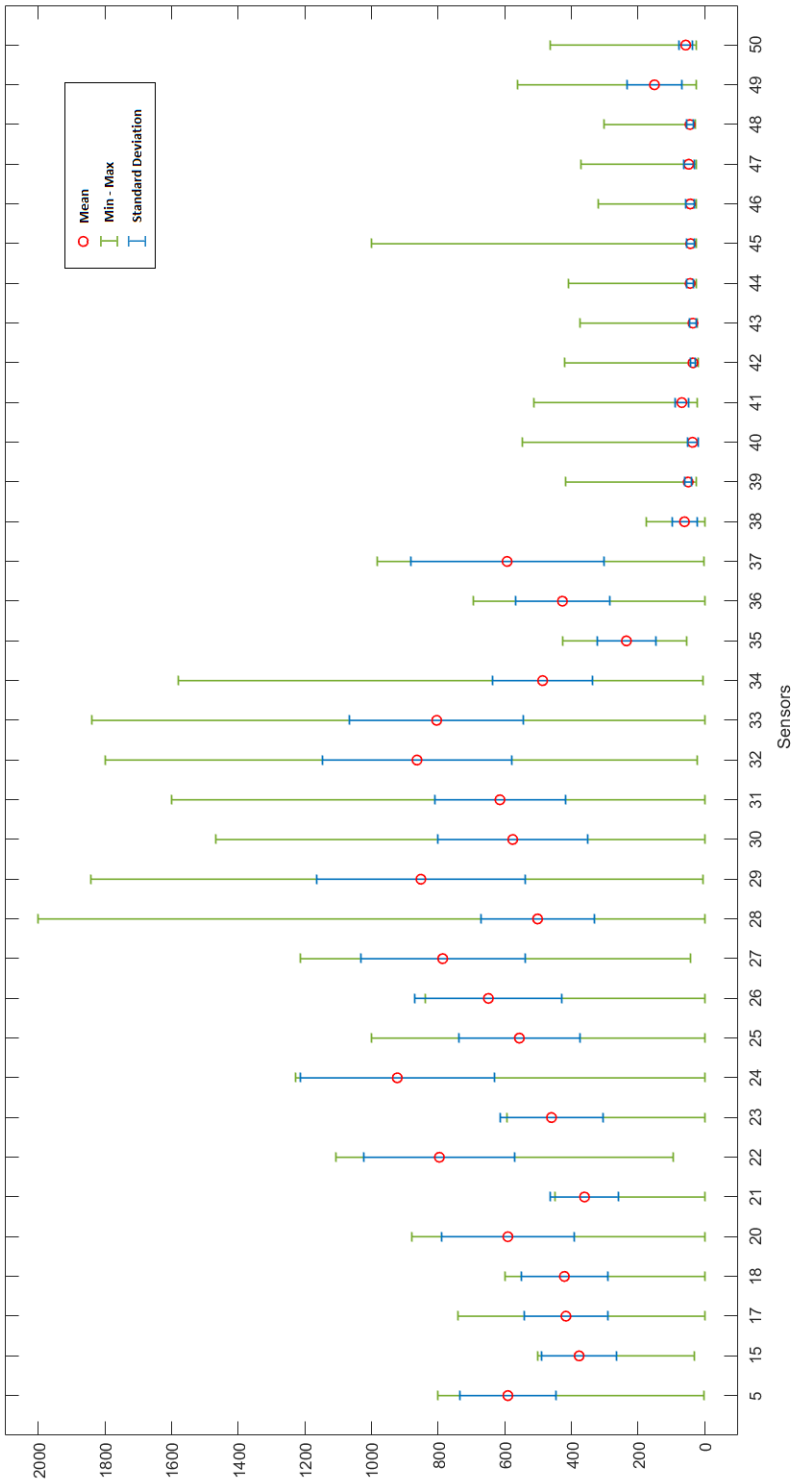


Figure A.6: Statistical values for each sensor (B).

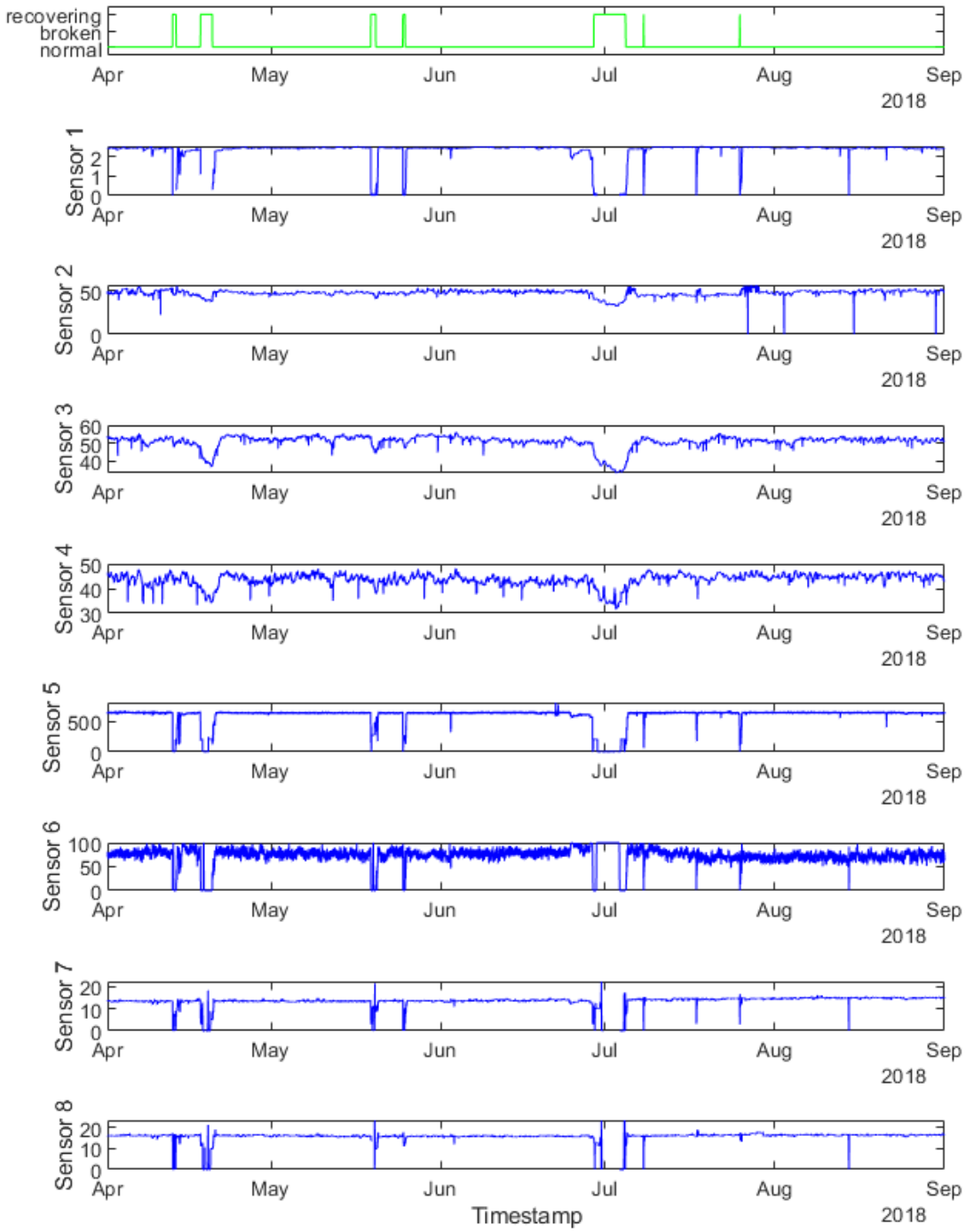


Figure A.7: Data for sensors 1, 2, 3, 4, 5, 6, 7, 8.

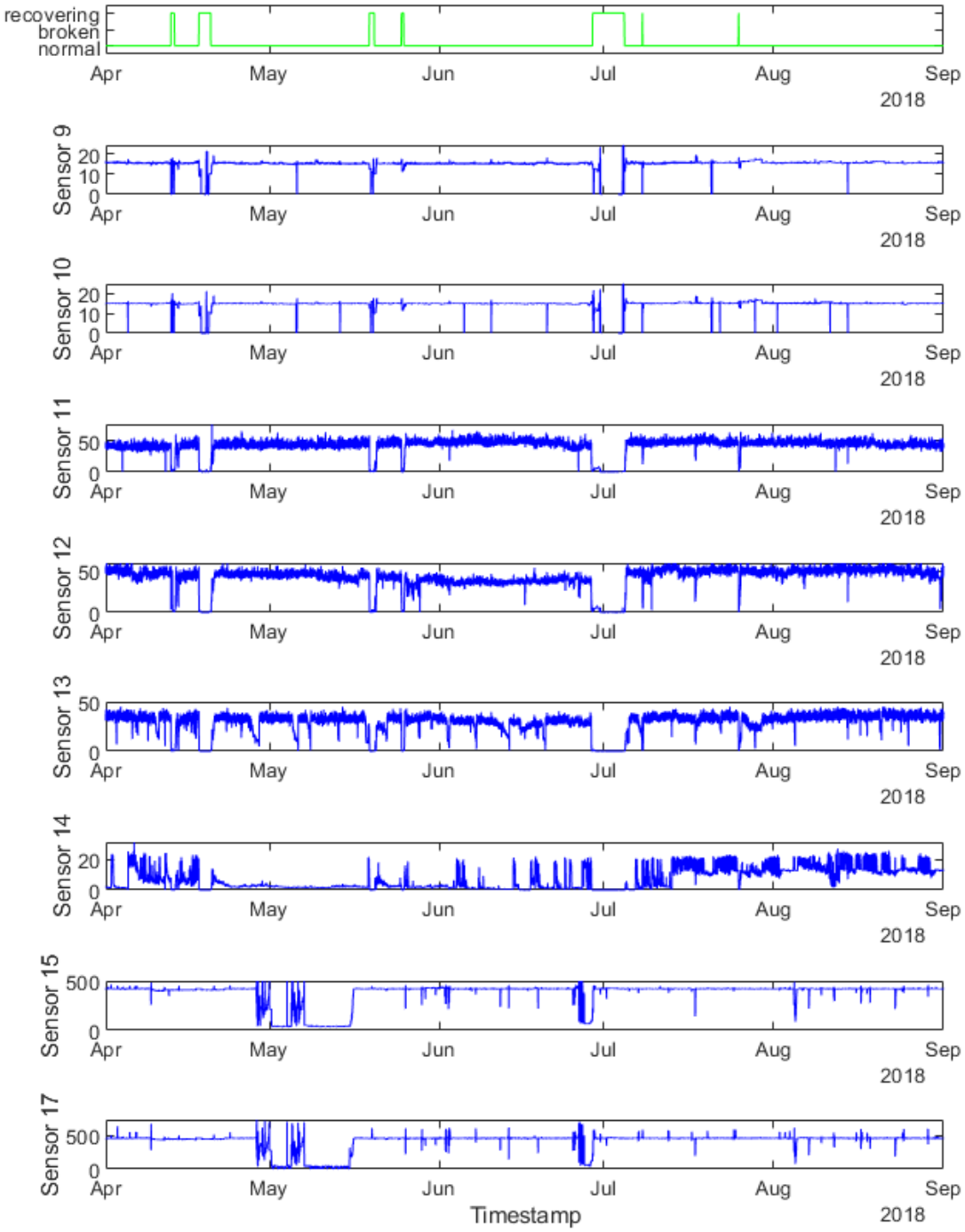


Figure A.8: Data for sensors 9, 10, 11, 12, 13, 14, 15, 17.

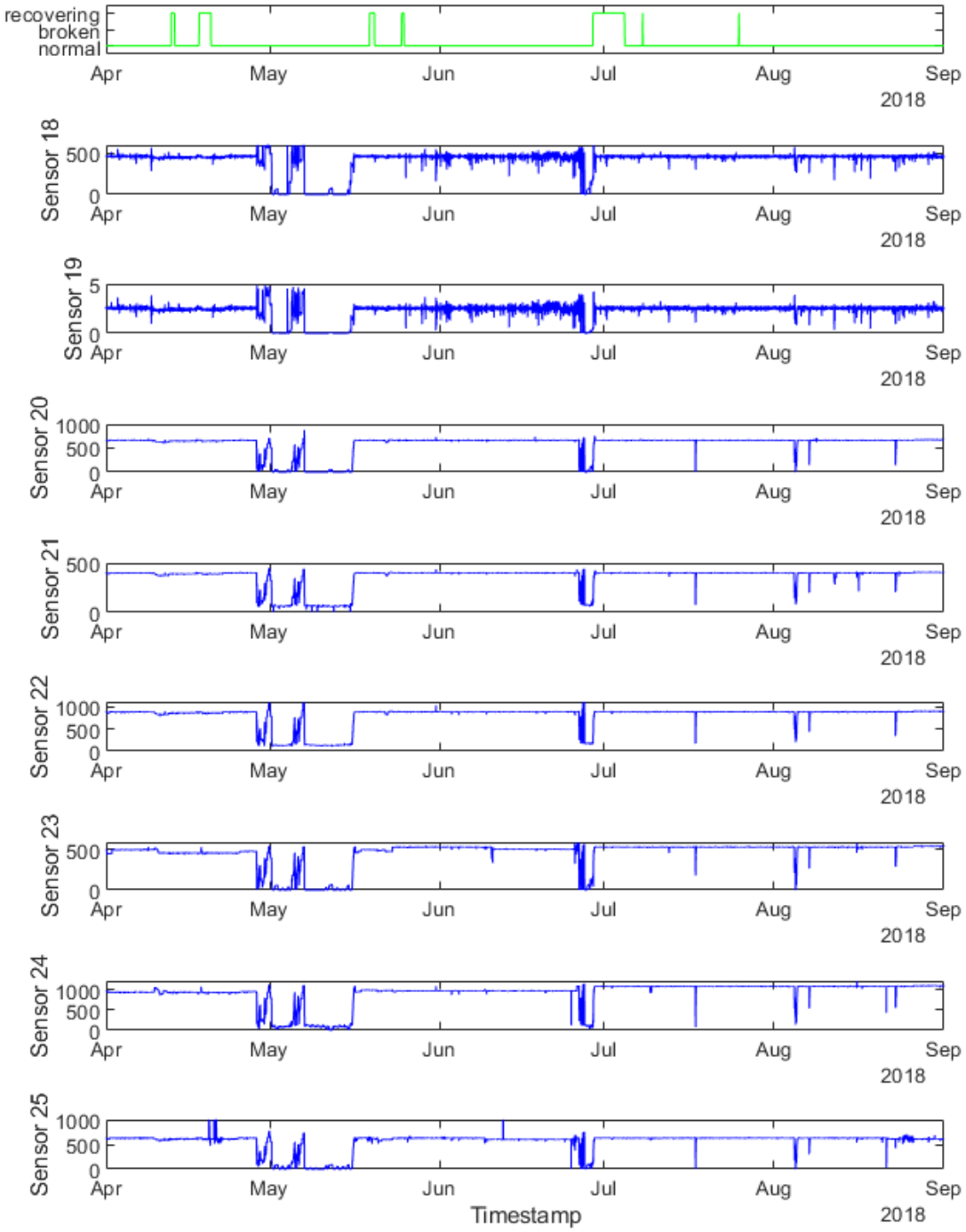


Figure A.9: Data for sensors 18, 19, 20, 21, 22, 23, 24, 25.

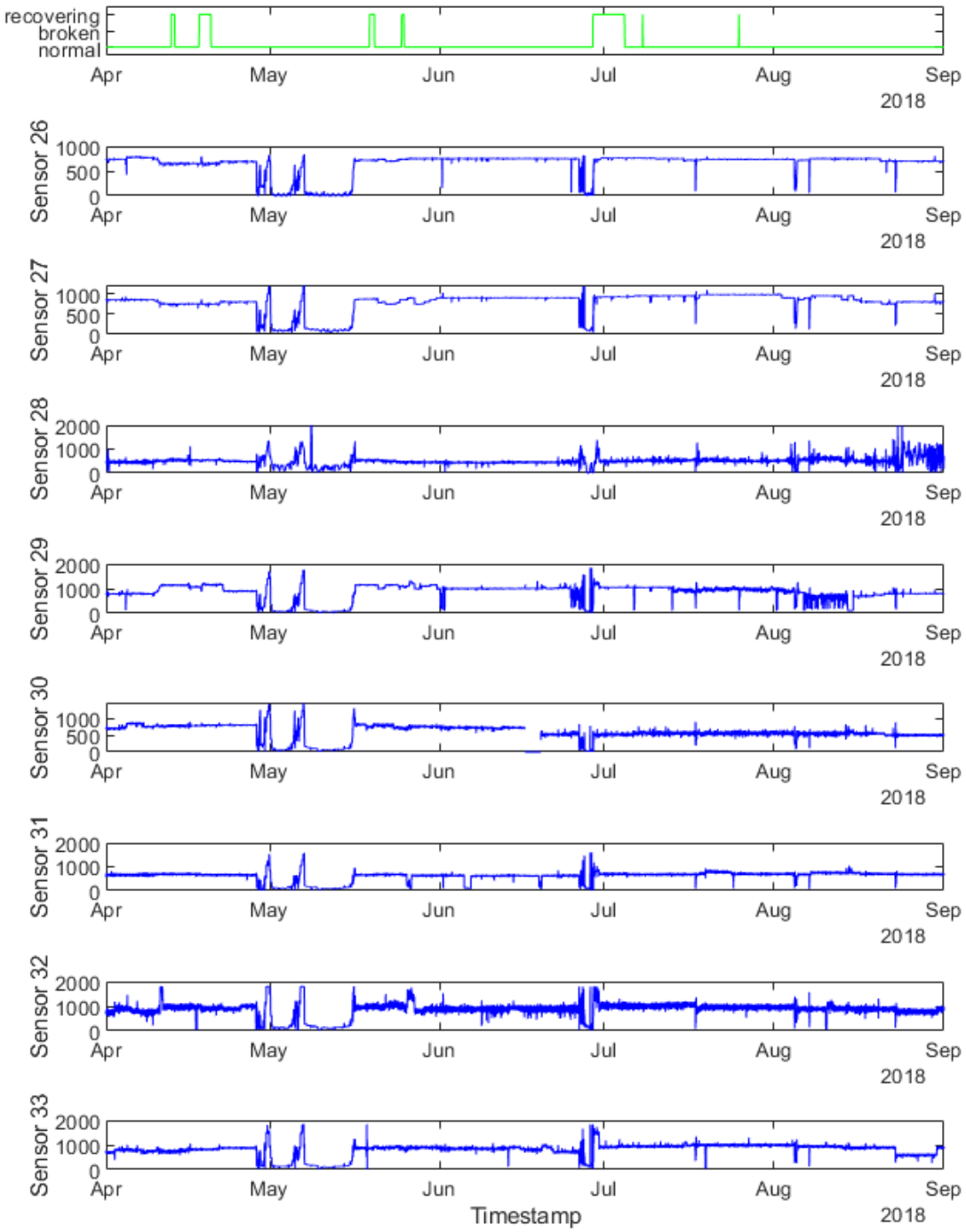


Figure A.10: Data for sensors 26, 27, 28, 29, 30, 31, 32, 33.

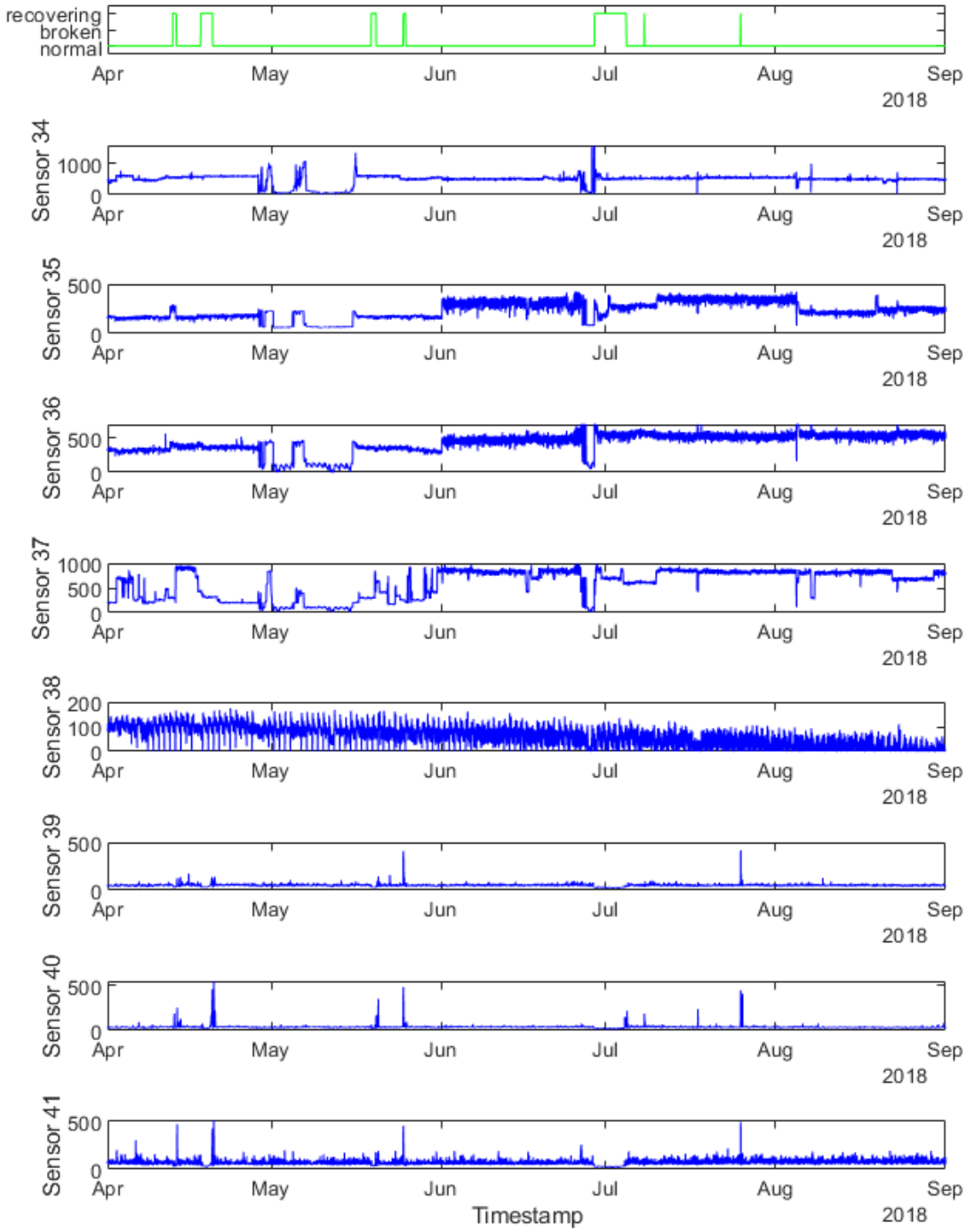


Figure A.11: Data for sensors 34, 35, 36, 37, 38, 39, 40, 41.

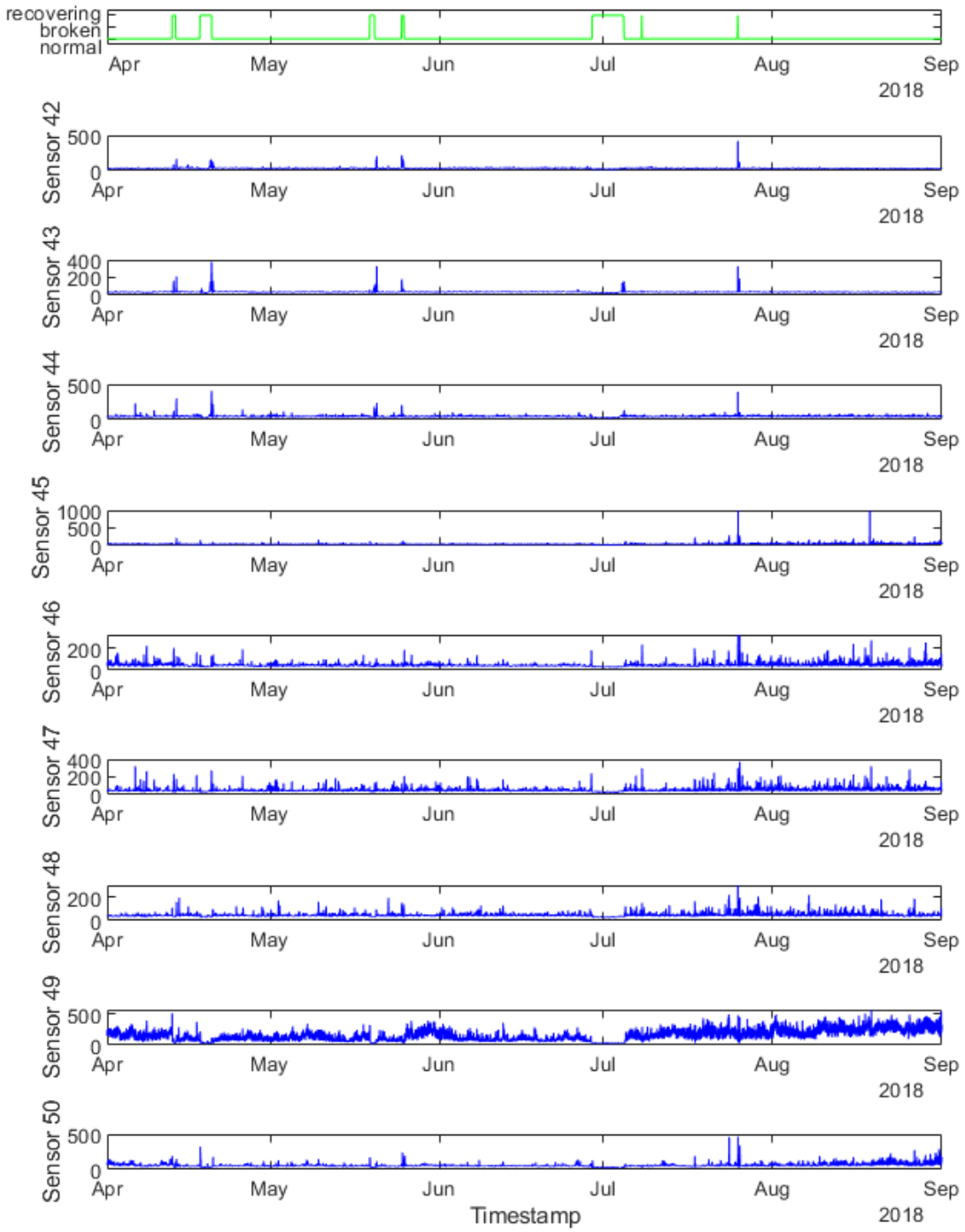


Figure A.12: Data for sensors 42, 43, 44, 45, 46, 47, 48, 49, 50.