Faculty of Sciences and Tecnology

Department of Informatics Engineering

Monitoring systems for diagnosis and prognosis

Application of Condition Based Maintenance in aircraft systems

Daniel Romário Gomes Azevedo

Dissertation in the context of the Master in Informatics Engineering, Specialization in Intelligent System advised by Prof. Alberto Cardoso and co-advised by Prof. Bernardete Ribeiro and presented to the Faculty of Sciences and Technology / Department of Informatics Engineering.

January 2019

UNIVERSIDADE Ɖ COIMBRA

This page is intentionally left blank.

Abstract

Aircraft maintenance is an important subject matter in the aircraft field and, as more useful information is gathered and processed, improvements in this area are beneficial and valuable to the aircraft industry. In particular, Condition Based Maintenance (CBM) can be useful to aircraft maintenance, as it can help predict when a failure will occur based on the condition of the components.

Using CBM, a Prognostics and Health Management (PHM) system can be built with the objective of overseeing the degrading behavior of the aircraft equipment, and previewing replacement at the optimal time, thereby maximizing useful lifetime of such components.

Different Machine Learning techniques integrated in a PHM system, will be developed in this work, with the objective of accurately predicting the aircraft components' Remaining Useful Lifetime (RUL). These techniques will encompass intelligent data analysis, model training and testing. The prognosis calculated from these models will be evaluated through the performance metrics established.

The dataset to be used in order to train, test and validate the techniques implemented will contain real data from specific aircraft subsystems which will be granted by an airline company, possibly KLM Royal Dutch Airlines. Nevertheless, due to the unavailability of the data, in this initial phase, the dataset used is the PHM08 Challenge Dataset [1] in complement with the Turbofan Engine Degradation Simulation Data Set [2] where the main goal will be to predict the RUL of turbofan engines in aircraft.

The scope of the implementation for PHM methods capable of predicting the components' RULs fits within the scope of the H2020 Real-time Condition- based Maintenance for Adaptive Aircraft Maintenance Planning (ReMAP) project [3].

Keywords

Aircraft Maintenance, Artificial Intelligence, Condition Based Maintenance, Machine Learning, Prognostics and Health Man-

agement, Remaining Useful Life

This page is intentionally left blank.

Contents

This page is intentionally left blank.

Acronyms

AI Artificial Intelligence. 8

ARC Abnormal Runway Contact. 2

C-MAPSS Commercial Modular Aero-Propulsion System Simulation. 11

CBM Condition Based Maintenance. iii, 3, 10

CFIT Controlled Flight Into Terrain. 2

CFRP Carbon Fiber Reinforced Polymer. 11

EoF End of Life. 13

FBM Failure Based Maintenance. 5

FF Feedforward. 8

HI Health Indicator. xi, 23–27, 30, 40

HMM Hidden Markov Model. xi, 9

LOC-I Loss of Control in Flight. 2

MLP Multilayer Perceptron. xi, 15, 28–30, 32

MSE Mean Squared Error. 31, 32

NN Neural Network. xi, 8, 31

PCA Principal Component Analysis. 16

PHM Prognostics and Health Management. 3, 6, 10, 13, 14, 35, 40

RBF Radial Basis Network. 8

RCM Reliability Centered Maintenance. 5

RE Runway Excursion. 2

ReMAP Real-time Condition- based Maintenance for Adaptive Aircraft Maintenance Planning. iii, 3

RMSE Root Mean Squared Error. 31, 32

RNN Recurrent Neural Network. 8

RUL Remaining Useful Lifetime. iii, xi, 3, 5–7, 9, 10, 12–16, 18–27, 30–33, 35, 36, 40

SCF System/Component Failure or Malfunction. 2

SVM Support Vector Machine. 9

TPM Total Productive Maintenance. 5

This page is intentionally left blank.

List of Figures

This page is intentionally left blank.

List of Tables

This page is intentionally left blank.

Chapter 1

Introduction

1.1 Context of the Study

Over time, the complexity embedded in the equipment and systems increases in a way that these become more susceptible to failure. These failures can have different impacts, depending on the system affected and on the cruciality of the failure. These types of failures can be more problematic in critical systems, and in particular, in aircraft systems.

In an aircraft, a failure raises problems in respect to reliability, availability and safety. Therefore, significant effort is required in aircraft maintenance in order to certify that the necessary conditions for aircraft operations are obtained. Aircraft maintenance is expected to prevent failures to happen, guaranteeing and improving its safety and its reliability. It is a fact that degradation of the aircraft components happens over time and the goal of aircraft maintenance is to prevent the components reaching a state where they are likely to fail but optimizing the components' lifetime. Improvements in maintenance can avoid several problems like flight delays or cancellations, gate returns, personnel injuries, in-flight shut downs and maintenance rework. This will lead to huge economical savings and will also have a positive impact on the reputation and competitiveness of the airline. One way to accomplish this is to provide more useful information regarding the state of the aircraft systems in order to perform more accurate diagnostics and predictions about the aircraft condition.

In order to understand the importance of maintenance in aircraft, some statistical facts are presented [4]:

1. In case a large aircraft flight, like Boeing 747, is canceled, it can cost the airline US $140,000.
2. If the aircraft takes off but, due to some technical failure, the aircraft need to land to be repaired it can cost up to US $150,000 per hour to the airline.
3. A delay at the gate can have a cost of US $17,000 per hour.
4. In overall, each year, US $75-100 million are wasted in errors per airline.

Even though there are investments and the strict regulations in this area, aircraft maintenance can be still improved.

According to a study carried out by Boeing with respect to aircraft maintenance [5], the maintenance issues are responsible for 20% - 30% of engine in-flight shutdowns and can

cost up to US $500,000 per shutdown.
From 2013 to 2017, IATA concluded that aircraft malfunction had a contribution of 29% to the aircraft accidents [6].

1.2 Problem Statement - Motivation

According to Airbus [7], regarding hull losses (when an aircraft is destroyed or damaged beyond economical repair), the System/Component Failure or Malfunction (SCF) is responsible for 10% - 15% of the accidents, being the second major cause according to Figure 1.1.

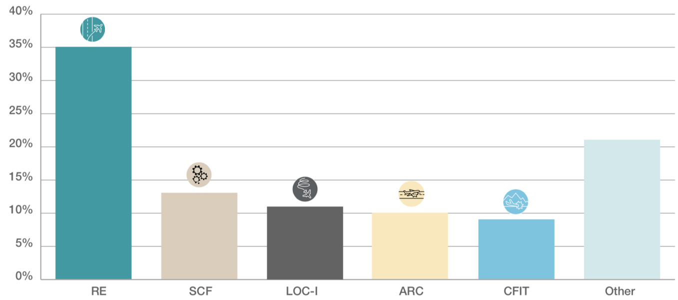

Figure 1.1: Percentage of hull losses by accident category 1998-2017 (from *A Statistical Analysis of Commercial Aviation Accidents 1958-2017* [7])

The accident categories, illustrated in Figure 1.1 are described by Airbus [7] as the following:

- **Runway Excursion (RE):** A lateral or longitudinal deviation that exceeds the surface of the track
- **System/Component Failure or Malfunction (SCF):** Failure or malfunction of an aircraft system or component that leads to an accident. SCF includes the powerplant, software and database systems.
- **Loss of Control in Flight (LOC-I):** Loss of control of the aircraft during flight, not mainly due to SCF.
- **Abnormal Runway Contact (ARC):** Hard landing that leads to an accident. The cause is not primarily due to SCF.
- **Controlled Flight Into Terrain (CFIT):** Collision during flight with some obstacles, like terrain or water, without the indication of loss of control.

Aircraft maintenance is a critical aspect of the aircraft as a small error can lead to dramatic consequences. Therefore, aircraft maintenance should be well planned and executed. With the access of more useful and valuable information regarding the condition of the aircraft components, better diagnostics and planning can be made in aircraft maintenance, turning flights safer, more reliable and more viable regarding the costs associated.

1.3 Aim and Scope

This work attends to develop Prognostics and Health Management (PHM) techniques, following a Condition Based Maintenance (CBM) approach, capable of monitoring the aircraft components' condition and predicting possible failures through the estimation of RULs, in order to better plan and execute aircraft maintenance. Furthermore, the methods implemented will be trained and tested using data retrieved from aircraft sensors, and intelligent data analysis will be performed in order to improve the methods performance, this will include data preprocessing, feature (sensors) selection and data normalization.

This work scope meets the scope of the Real-time Condition- based Maintenance for Adaptive Aircraft Maintenance Planning (ReMAP) project [3]. The ReMAP project is an European project that seeks to develop a open-source solution, the Integrated Fleet Health Management (IFHM), for aircraft maintenance where fixed-interval inspections will be replaced by adaptive condition-based interventions.

Therefore, and in alignment with the ReMAP project proposal regarding this specific Work Package, the main goals established are:

- Perform preprocessing of raw data in order to detect and evaluate eventual imprecise, noisy or irrelevant data.
- Intelligent sensors data analysis in order to diagnose the system condition.
- Develop PHM methodologies for aircraft systems that allow the prognosis of aircraft components' failures.
- Train, test and validate the methodologies implemented with real sensors data retrieved from *specific* aircraft components.

1.4 Structure of Document

In **Chapter 2** some relevant concepts like types of aircraft maintenance, RUL and PHM are detailed. A benchmark of datasets pertinent to the identified problem is also presented. Lastly, performance metrics that will be applied to the implemented techniques, are specified.

In **Chapter 3** the research objectives linked to this work will be described and the approach followed will be defined and explained.

In **Chapter 4** the work developed so far will be presented and discussed through the preliminary results obtained.

In **Chapter 5** the work developed so far, as well as the planning regarding the work for the 2nd Semester will be displayed on a *Gantt Chart*. The risks associated with this work are also described.

In **Chapter 6** a summary of the work performed so far will be elaborated and a brief overview of the work plan for the second Semester will be done.

This page is intentionally left blank.

Chapter 2

Background and related work/State of the art

2.1 Types of Aircraft Maintenance

There are multiple approaches to aircraft maintenance, the following are highlighted [8]:

- **Failure Based Maintenance (FBM)**: This type of maintenance is performed when a failure occurs, there is no action in order to prevent or detect future failures. This method is risky and has high maintenance costs associated.
- **Condition Based Maintenance (CBM)**: This category is focused on the prevention of the failure. Historical data containing the components' failure behavior is obtained from components condition monitoring and is used to create a decision based model, that can help diagnose and predict future failures. This way, unnecessary replacements are avoided and the aircraft components can run during their full lifetime [9].
- **Reliability Centered Maintenance (RCM)**: This technique focuses on the most important functions of the system, that is, the most cost-effective functions. It cannot prevent all the failures just critical ones, according to the defined criteria. This criteria can be potential human injuries, environmental damage, production loss, etc [10].
- **Total Productive Maintenance (TPM)**: This company-wide approach aims to improve the overall effectiveness of the equipment and to achieve that it defends that all company's departments should be involved, from the maintenance to the project engineering [11].

2.2 Remaining Useful Lifetime

The Remaining Useful Lifetime (RUL) is defined as the time from the current time to the end of the useful life of a certain component. This term can be applied to different contexts, and in this case it will be applied to aircraft. More specifically, it will be used to measure the time from a point where a failure had occurred to the time where that failure evolves to the extent that it provokes the total failure, and thus the malfunctioning of that aircraft component or subsystem. Figure 2.1 presents a general graph of a fault evolution. As can be seen, the main goal is to predict the RUL, which is, the present time (t_p) to the end time (EoL) defined by a specific *failure threshold*. This prediction can be done using different approaches summarized in section 2.3.

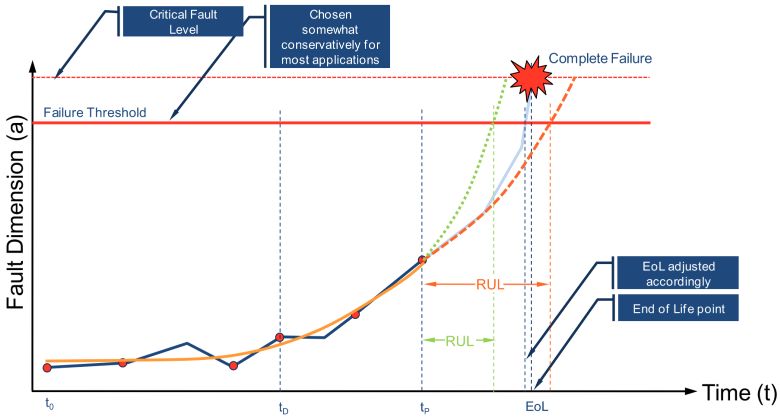

Figure 2.1: Fault behavior (from *Prognostics, The Science of Prediction* [12])

2.3 Prognostics and Health Management

In order to make predictions about the components RUL, PHM methods should be considered. Prognostics and Health Management (PHM) systems are primarily focused on the estimation of component's or subsystem's RUL through the evaluation of the system's current health state [13].

A PHM system is expected to detect possible faults in a subsystem, isolate them and predict their impact on the system through the monitoring of the fault growth. This way a PHM system can prevent a subsystem from total failure, increasing his reliability [13].

Ideally, a PHM system can have a great impact in aircraft maintenance. Using a PHM system it is possible to monitor and estimate the components RUL accurately, this would bring numerous advantages [12]: 1) Decrease damage caused by the failure in that particular subsystem as well as in the healthy surrounding subsystems; 2) Help create a more efficient maintenance plan thereby reducing the logistics costs; 3) Performing aircraft service (equipment replacement) only when needed; 4) Increase reliability and confidence in the system, helping increase airline reputation.

PHM methods are broadly divided into three categories [14], [15] : Data-driven methods, Model-based methods, and a Hybrid approach of the two methods. All methods goal is to accurately predict the system state RUL, using different approaches, for a given time frame in the future, by estimating the system's health state.

2.3.1 Model-based methods

The Model-based methods use physics models to make the predictions. Through these models there is an incorporation of physical understanding requiring human knowledge in order to estimate RUL. The advantages of this type of methods are the fact that estimations don't need historical data and once the model is defined, it can be used in different cases by making small adaptations. A relevant disadvantage is the difficulty of implementing the models due to the need of expert knowledge [15].

2.3.2 Data-driven methods

The Data-driven methods make the predictions of RUL using models learned from the data, normally represented in time series. These models tend to represent the degradation behavior of the system based on historical data. The advantages of this approach are simplicity and the speed in the implementation. However, the need of large amounts of data and the generalization with respect to the components' degradation behavior are some of the drawbacks. The Data-driven methods can be mainly grouped in the following categories:

Regression Based

Regression based models are considered one of the most simple predictive methods. Historical data, in the form of a time series, is used to fit a regression model that can accurately explain the system behavior. The model can then be extrapolated for some points in the future in order to make predictions. The formulation of a regression model is the following [16]:

$$x = \beta_0 + \sum_{p=1}^{P} \beta_p \phi_p(t) + \varepsilon \tag{2.1}$$

where β represents the model parameters, $\phi_p(t)$ is a representative function of t (it can be either linear or non linear) and ε is the noise term. Using regression models it is possible to simulate the degradation behavior of a component or a system in order to predict future failures or anomalies.

Filter Based

Filter based approaches assume that a desired state of a system can't be measured directly but through other measurable variables. One typical example is the Kalman Filter. The Kalman Filter predicts the hidden states for the next time step given the history of estimated states and observing noisy outputs [17].

As explained by Welch G. and Bishop G. [18], a discrete Kalman Filter can be represented by a cycle (Figure 2.2) between two different processes. One is the *Time Update* where the unknown state is estimated in a future time, based on a model defined by mathematical equations. The *Measurement Update* will then provide feedback to the estimation made, adjusting the estimation based on an actual measurement.

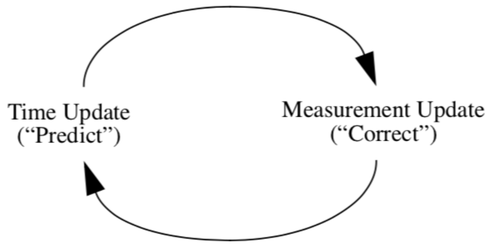

Figure 2.2: Discrete Kalman Filter cycle (from *An Introduction to the Kalman Filter* [18])

Artificial Intelligence (AI) Based

This category includes different type of methods all of them related to Artificial Intelligence. **Neural Network (NN)** based methods are widely used in prognostics methods due to fact that it can receive any type of input and the user doesn't need a significant physical knowledge about the system behaviour in order to model it [19]. These methods establish relationships between the input and a desired output, adjusting the parameter for optimal performance. The most known Neural Network [20] are Feedforward (FF) [21], Radial Basis Network (RBF) [20] and Recurrent Neural Network (RNN) [22].

In Figure 2.3 there is an example of a Feedforward Neural Network. For each Neural Network there are several parameters that should be changed accordingly to the expected output, namely: number of inputs, number of neurons, number of layers and the activation function of each layer.

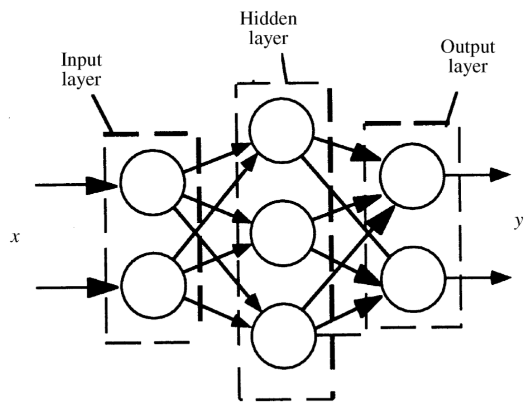

Figure 2.3: example of a Feedforward Neural Network (from *Diagnostic rule extraction from trained feedforward neural networks* [21])

Related to AI, there are also the **Decision Trees**. These are classifiers represented in a tree based diagram that structures all the possible decisions and their consequences [23]. At the end of the process, there will be a decision tree, where the final decisions to be taken are presented in the leaves (end nodes).

Another used technique is **Support Vector Machine (SVM)**. This technique is a supervised learning method for classification, that uses kernel functions to map the feature space in a high dimensional space and finds the surface hyperplane that maximizes the separation between classes (margin), in that high dimensional space [24].

One last group of AI based technique is **Fuzzy Logic**. It is a decision-making method similar to human reasoning in the way that a decision can be made between 0 (No) and 1 (Yes), i.e., there can be different states of truth (partial truth) [8], for example, the temperature in a room can be 0.6*Hot. Through the definition of a fuzzy set and rules, an input space can be computed to an output space. The fact that concepts behind fuzzy reasoning are theoretically simple and that input data can be imprecise and noisy, are the main reasons for the use of Fuzzy Logic.

Markov Based

Markov Based methodologies are based on memoryless Markov processes. There are some relevant variations like **Markov Chain Model** and **Hidden Markov Model (HMM)**.

Markov Chain Model is a stochastic model that describes the probability distribution of the state transitions of a system. It follows the Markov Property, which states that the future system state only depends on the current state, and not on events that occurred in the past. It is mathematically expressed as the following:

$$P(X_{n+1} = x | X_0, X_1, X_2, \dots, X_n) = P(X_{n+1} = x | X_n) \tag{2.2}$$

where x is the system state, and $X_{0...n}$ is a set of random variables, that represent previous states. Thus, in order to compute the probability of a sequence of states s_1, s_2, .., s_n the following formula can be used:

$$P(s_1, s_2, ..., s_n) = \prod_{i=1}^{n} P(s_i | s_{i-1}) \tag{2.3}$$

The **HMM** is also a Markovian based model which assumes: 1) The Markov Property; 2) The system state cannot be observed directly, i.e., it is hidden from the observer. A correlated and observable state is used to infer something about the system's current state, the overall diagram can be seen in Figure 2.4, where $S_{0...T}$ are the hidden states from time 0 to T and $Y_{0...T}$ are the variables observed from time 0 to T [25].

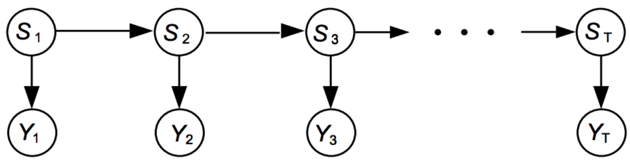

Figure 2.4: HMM diagram (from *An Introduction to Hidden Markov Models and Bayesian Networks* [25])

2.3.3 Hybrid Prognostic approach

As the name suggests, this approach combines both Model-based and Data-Driven methodologies in order to predict the RULs values.

In real-world prognostic scenarios, sometimes the trends of the components' behavior are complex and difficult to predict with a single RUL estimation method. A hybrid approach tries to solve this problem by using multiple techniques of RUL estimations altogether [15].

The aim of this technique is to overcome the limitations of a single prediction method, improving the accuracy of the RUL predicted. The fusion of techniques can be applied in the different stages of the PHM algorithm implementation, like in the data extraction, data analysis and model training.

Types of learning

Regarding the types of learning in the Artificial Intelligence field, there are [19]:

- **Supervised learning**: In this type of learning the data is labeled, which means that the true state of the data instances is known. This way models that map the input into output can be learned from the data. Examples of techniques following this type of learning are *Regression* and *Classification*.
- **Unsupervised learning**: In the unsupervised learning the data is unlabeled as the classes are not known. This type of learning tries to find some pattern or distribution from the data in order to learn more about it. Typical examples are *Clustering* and *Association Rules*.
- **Semi-Supervised learning**: This is a combination of Supervised Learning and Unsupervised Learning due to the fact that some data is labeled and other is not.
- **Reinforcement Learning**: The *Reinforced Learning* is based on actions. The agent is continuously learning from the interactions with the environment and the goal is to take actions that maximize a reward function. One well-known algorithm that uses Reinforcement Learning is *Q-Learning*.

2.4 Dataset

In order to train, test and compare different methodologies developed for PHM, a dataset that comprise useful data about aircraft, ideally sensors data from airplanes, is needed.

2.4.1 Dataset description

In the context of the ReMAP project, an airline company, possibly KLM Royal Dutch Airlines [26], will provide access to authentic data extracted from real aircraft flights, to be used in this work.

The dataset is an important aspect for this work as it is the basis of the Condition Based Maintenance (CBM) system to be implemented, which will allow the monitoring of the condition of particular aircraft subsystems.

In order to perform an intelligent analysis of the data retrieved from the aircraft sensors, with the objective of diagnosing the different aircraft components' state and predicting possible failures, the data presented in the dataset should correspond to raw sensors data and must reflect a specific component/subsystem condition.

Furthermore, the dataset must contain data specific to a discrete and sequential time frame and must be collected from the same aircraft. Ideally, the interested aircraft parts shouldn't be replaced during that specific time frame, so that the degradation behavior of certain components can be analyzed and modeled over time.

In the ReMAP, 12 different aircraft subsystems were identified, namely: Cabin Air Conditioning and Temperature Control System; Cabin Pressurization and Control Systems; Engine Anti-Ice; Power Electronics Cooling systems; Nitrogen Generation System; Common Motor Start Controller; Variable Frequency Starter/Generator; Buss Power Control Unit – Generator Control Unit; Auxiliary Power Unit; Fans; Integrated Cooling System, and Wheels & Brakes.

The dataset should encompass sensors data extracted from specific subsystems of the aircraft in order to analyze, interpret and diagnose the condition of that particular subsystem, and predict possible failures. The aircraft monitoring techniques proposed in this work will be focused on particular subsystems (two or three subsystems) and not the whole aircraft system due to its complexity and comprehensiveness. Possible subsystems to be considered are: Cabin Air Conditioning, Temperature Control System and Common Motor Start Controller.

2.4.2 Dataset Benchmark

For this part of the work, other dataset fitted within this work scope was used in order to gain some sensibility to the data and to experiment and compare the techniques developed so far.

After some research, three interesting datasets were found: 1) The **CFRP Composites Data Set** [27]; 2) The **Turbofan Engine Degradation Simulation Data Set** [2]; 3) The **PHM08 Challenge Data Set** [1] . All the three were taken from the NASA's Prognostics Data Repository [28].

CFRP Composites Data Set

This dataset contains data obtained from test of fatigue aging on Carbon Fiber Reinforced Polymer (CFRP) composites carried out by the Stanford Structures and Composites Laboratory (SACL) in collaboration with the Prognostic Center of Excellence (PCoE) of NASA.

The fatigue tests were made on specific CFRP coupons and with three different layups. The fatigue damage was monitored using two sets of six-PZT-sensors and the resulting data contained the data from the PZT sensors, along with X-Rays of the samples and StrainData. An overall idea of the damage fatigue tests can be seen in Figure 2.5.

where in a) a coupon with the PZT sensors can be seen; in b) the cracks evolution are evidenced; and in c) the failure growth is exhibited in cycles.

Turbofan Engine Degradation Simulation Data Set

In this dataset, the data was generated by an engine degradation simulation [30], carried out using Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) [31], with the objective of simulating a turbofan engine. This dataset was provided by the Prognostics CoE at NASA Ames.

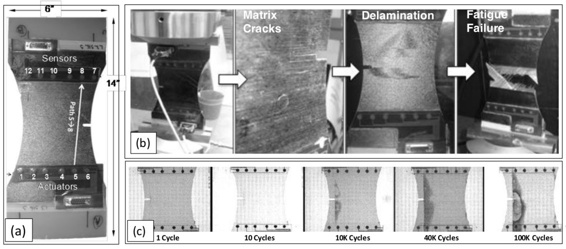

Figure 2.5: CFRP coupon fatigue test (from *Accelerated Aging Experiments for Prognostics of Damage Growth in Composite Materials* [29])

Four datasets are presented and were simulated over different combinations of Operational settings and Fault Modes.

The dataset is made of data from 21 sensors regarding to pressures, temperatures and velocities of different components of the Turbofan engine. These sensors values capture the evolution of a fault and the goal is to predict the number of cycles remaining before the total system failure, i.e., the RUL.

In order to evaluate the predicted RULs , each dataset consists of a *train* dataset and a *test* dataset. The true RULs for the test dataset are provided.

PHM08 Challenge Data Set

This dataset has the same intent and structure of the dataset presented before and was also provided by the Prognostics CoE at NASA Ames.

It was created to a challenge competition held at the 1st international conference on Prognostics and Health Management (PHM08).

The only difference is that the true RULs values for the test dataset are not revealed, instead the competition provided a link were the participants could submit their RULs and a score value was returned. The best approaches were decided based on their score value.

2.5 Performance Metrics

In order to access the accuracy in the predicted RULs, some performance metrics were established. They can be broadly divided into two categories: **Accuracy Based** and **Prognostics Based**.

2.5.1 Accuracy Based

These are traditional performance metrics, that evaluate independently the final estimated RULs. Two of these metrics are described:

- **Score**: The PHM Conference 2008 organization created a formula that calculate the prediction score comparing the estimated $RULs$ and the true $RULs$. This metric is useful to compare with other approaches developed using the same dataset, considering it as a benchmark. The scores of the approaches developed to the PHM 2008 Data Challenge were already published [32]. The Score formula is the following:

$$\begin{cases} \sum_{i=1}^{n} e^{-(\frac{d(i)}{13})} - 1, & d < 0 \\ \\ \sum_{i=1}^{n} e^{(\frac{d(i)}{10})} - 1, & d \geq 0 \end{cases} \tag{2.4}$$

where $d(i) = Estimated\ RUL - True\ RUL$, of unit i.

- **Mean Square Error (MSE)**: This metric calculates the average squared of the difference between the estimated values and true ones. This measure is widely used to calculate errors associated with predictions, although it is not specific for PHM. The formula is the following:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2 \tag{2.5}$$

where n is the number of samples, y_i the estimation and $\widehat{y_i}$ the ground truth.

- **Root Mean Square Error (RMSE)**: This metric calculates the root of the average squared of the difference between the estimated values and true ones. It measures the average amplitude of the error and can be applied in the PHM context. The formula is the following:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{2.6}$$

where n is the number of samples, y_i the estimation and $\widehat{y_i}$ the ground truth.

2.5.2 Prognostics Based

Prognostics metrics are more suitable for evaluating prediction techniques, as they are based on all the prediction series of a data instance instead of single prediction. Next, two of these metrics are explained:

- **Prognostic Horizon**: This metric calculates the difference between the time index when the prediction satisfies a α-bound criteria and the time index corresponding to the End of Life (EoF). The formula is the following [33]:

$$PH = t_{EoL} - t_{i_\alpha} \tag{2.7}$$

where, t_{EoF} is the time index of EoL and t_{i_α} is the first time index where the α-bound criteria is satisfied.

- $\alpha - \lambda$ **Accuracy**: This metric checks if a prediction accuracy is within a specific accuracy interval (specified by α), for a certain point in the time (given by λ). The accuracy interval is defined as a percentage of the true RUL for that given time. The condition to be evaluated is the following [33]:

$$(1-\alpha) * r(t) \leq r(t_\lambda) \leq (1+\alpha) * r(t) \tag{2.8}$$

where α is the accuracy modifier, λ is the time window modifier and r(t) is the true RUL for time t.

Chapter 3

Research Objectives and Approach

3.1 Problem Statement

As discussed before, this work attempts to help improving aircraft maintenance by providing more useful information regarding the condition of the aircraft subsystems or components in a way that better decisions regarding the execution and planning of aircraft maintenance can be made. More specifically, it can help to decide the optimal time to replace a given part by optimizing its useful lifetime. This work will focus on two or three specific aircraft subsystems within the ReMAP scope and will be defined by the location of the sensors providing the data.

In this work, the methodologies to be implemented will be focused on assessing the condition of the aircraft parts of particular subsystems. Using that information, the different methodologies will calculate the RUL, which is a prediction of the useful lifetime remaining of that particular component. The estimation of RUL will be calculated using different techniques according to the methodology followed.

3.2 Research Objectives

The research performed so far was made towards understanding aircraft maintenance and state of art PHM methodologies. The goal of this research was to create a solid knowledge base that will allow the development of PHM systems to prevent aircraft failures through an accurate RUL estimation. Moreover, the defined research objectives were:

- Study the state of the art on aircraft maintenance.
- Benchmark of datasets that fit this work scope.
- Study and development of methodologies to be applied in the PHM context.

3.3 Main Goals

Regarding the objectives of this work, these will be focused on three aspects: raw data preprocessing, sensors data analysis and PHM methodologies implementation, test and validation.

The preprocessing of the dataset data has the objective of removing noise embedded in the data and perform eventual data normalization. Sensor analysis, will target the selection of relevant and useful data for training (feature selection), and apply the necessary transformations to the data such that it is ready to be used in the models training.

The methodologies to be implemented will be focused on the RUL prediction and will follow three different approaches: Distance based prediction, Neural Network based prediction and Extrapolation based prediction. These approaches will be evaluated and compared using the metrics:: PHM 2008 Data Challenge score [30], Mean Square Error (MSE) and Root Mean Square Error (RMSE).

3.4 Approach

Some simple machine learning techniques for RUL prediction were reproduced for preliminary experiments. These experiments were performed in order to gain sensibility and knowledge about the data.

The techniques chosen for replication purposes were selected taking into account the RUL prediction methods developed for the PHM08 Challenge Competition. Moreover the technical report *Benchmark of methods applied to PHM08 Challenge Dataset* [34] was written with the goal of analysing the RUL prediction methods applied to that specific dataset.

An overview of the work developed for the competition is presented in *Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets* [32], from which the following categorization can be extracted: **Neural Network based** methods, **Extrapolation based** methods and **Similarity based** methods.

Neural Network based

In the Neural Network based methods, as the name suggests, the main techniques around this approach are based on Neural Networks. The overall idea of this approach is to create mappings between a set of Inputs (from the dataset) and the final RULs. To achieve this, two steps are carried out, namely: the data preprocessing step in which the training data is converted into a labeled feature space and the second step, is to map the feature vectors obtained, into RULs. Along with that, feature selection, data normalization and noise filtering are recommended in order to obtain better results and some examples of these techniques are: Differential Evolution, Normalization based on Operational settings, Kalman Filter, etc. The main advantage of this approach is that it can outperform any other model in the right conditions as it is a powerful tool. Some disadvantages may be the lack of systematic approach and sensibility to select the right parameters and lack of intuition to improve performance continuously.
Some examples include: Recurrent Neural Network (RNN) [35], Artificial Neural Network (ANN) [36] and Multilayer Perceptron (MLP) [37].

Extrapolation based

Another approach is Extrapolation based method, following this concept, a mapping between *Health Index (HI)* and RUL is made. In order to achieve this, there are two main steps: the first is to do a mapping between the data (sensor measurements) and *HI* for each training unit, the result of this mapping will be a set of degradation model (similar to the

best classified method). The second step consists of a mapping between the *HI*, calculated using the degradation models, and the final RUL. This can be done by fitting polynomial regressions or using averaging methods. The accuracy of the results will depend on the mapping functions; the first mapping is more difficult and challenging because the health index is deeply inserted in the dataset contaminated with noise. Regarding the second mapping, normally regression models are used, more specifically, non linear methods, since it can simulate better the degradation trends.
The noise filtering or adaptation can be a critical issue because of its complex nature (it is multimodal and dependent on the Operational settings). As this is an Extrapolation based approach, noisy data can have a negative impact in the RUL estimated. This is why noise filtering is an important step in the data preprocessing. Some examples include: Logistic Regression [38] and Linear and Bayesian updating [39]

Similarity based

The last approach is Similarity based methods. Using this procedure, a library with instances (sensor measurements labeled) is created, which will be used as a comparison base. Then for each test instance, the given instances will be compared with the instances in the library, and the most similar will generate a set of RUL estimations that will then be used to obtain a final RUL. The similarity criteria can be sensor space [40] or *Health Index* [41]. In summary, three tasks can be defined: instances library creation, predictions of RUL through local models and aggregation of local predictions of RUL. As an example, the PHM08 Challenge Competition winner [41] followed this approach, using Linear and Exponential Regressions to create the models and a weighted method for the RULs aggregation.
The advantages of this approach are that new instances can be easily incorporated, have a good generalization capability and can be parallelized to reduce computational efforts. The disadvantages are the influence of the noise and fault modes in the final result. Some methods to overcome these problems could be feature selection, Principal Component Analysis (PCA) and data partitioning.

A method for RUL prediction was developed for each of the three approaches described above, the implementations of which are detailed in Section 4.2. The implementations might need an adaptation depending on data of the dataset used.

In the second Semester, the current implementations will be improved with the objective of predicting the RUL more accurately. New approaches may be explored, focusing on other Machine Learning techniques like, for example, *Support Vector Machine (SVM).*

This page is intentionally left blank.

Chapter 4

Current work and preliminary results

The work developed so far can be split into two main phases: 1) the search for a dataset suited for this work and within the ReMAP context and 2) the implementation through reproduction and comparison of methodologies for aircraft condition monitoring through the components RUL prediction using different approaches.

4.1 Dataset selection

The dataset benchmark described in Section 2.4.2 was made with the intention of finding a dataset with reliable aircraft sensors data, where the techniques developed in this work could be applied in order to predict future failures. As such, with respect to the 3 datasets found, the last two will be used, that is, the **Turbofan Engine Degradation Simulation Data Set** and the **PHM08 Challenge Data Set**.

As these two datasets have the same structure, both will be used.

The selection of these datasets was based on their particular characteristics regarding the developing prognostics algorithms, namely [32]: 1) The data was obtained from a high fidelity simulation that models very closely a real system; 2) The high presence of noise in the data approximating it to real data; 3) The masking of fault effects due to the Operational settings, very common in these type of systems; 4) Multiple units (trajectories) in the dataset which allow the identification and extraction of trends and other useful information in order to predict RUL. Also, due to similarities of scope, it is expected that these datasets can be applied within the ReMAP project.

Each of the two datasets chosen have specific characteristics that can be useful to this work: the **Turbofan Engine Degradation Simulation Data Set** provides the true test RULs which can be applied to the established accuracy based metrics, in order to compare different approaches. The **PHM08 Challenge Data Set** is useful because the score values obtained by the approaches developed during competition were made available online [32], and so they can be used to compare with the score values obtained in this work.

Regarding the first dataset, **CFRP Composites Data Set**, some adverse characteristics were found, namely: lack of data related to the StrainData in the dataset; poor documentation, without great level of detail; difficulty in understanding the data available due to its complexity, dataset was not intended to be used in the context of aircraft testing, and little work (papers, technical reports, ...) done with this dataset.

4.1.1 Dataset Description

As explained before, in this work, the **Turbofan Engine Degradation Simulation Data Set** (referred from now on as **Turbofan Dataset**) and the **PHM 2008 Data Challenge Data Set** (referred from now on as **PHM08 Challenge Dataset**) will be used.

These datasets contain data generated by C-MAPSS, a model created by Prognostics CoE at NASA Ames to simulate a turbofan engine [30]. In Figure 4.1 it is possible to visualize a simplified diagram of C-MAPSS with all components considered.

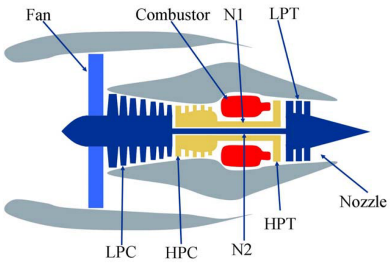

Figure 4.1: Simplified model of C-MAPSS (from *Damage Propagation Modeling for Aircraft Engine Prognostics* [30])

Although the data presented in the datasets is modeled and not extracted from real aircraft systems, there is a big approximation to real sensors data. Some relevant and important aspects of real aircraft flights, like initial components wearing and presence of noise in the system were taken in consideration when creating the model.

These datasets were created with the goal of, based on data sensors, predicting aircraft failures through the estimation of RUL. Each dataset contains several aircraft trajectories and a RUL value should be predicted for each one, as a failure was inserted in each trajectory.

4.1.2 Dataset structure

Table 4.1, represents the structure of the two datasets chosen. As can be observed, each dataset was simulated over different number of Operational settings combinations and Fault Modes, which influences the complexity in the RUL prediction. The number of trajectories in the training dataset and in the testing dataset is also different for each dataset.

Table 4.1: Datasets Structure and Details

Datasets		Train Trajectories	Test Trajectories	Operational Settings	Fault Modes
Turbofan	**FD001**	100	100	1	1
	FD002	260	259	6	1
	FD003	100	100	1	2
	FD004	248	248	6	2
PHM08 Challenge		218	218	6	2

Each dataset has a *train* dataset and a *test* dataset. The PHM08 Challenge Dataset provides also a *final_ test* dataset for validation, which was used to rank the participants.

Each dataset contains data from several flight trajectories represented by time series. In the start of each time series the engine operate in normal conditions, i.e., without any failure, then at some point of the time series a fault is generated and grows from there. In the *train* datasets the fault grows until a maximum threshold is reached where the turbofan engine cannot operate anymore (total failure), this marks the end of the time series. In the *test* datasets the fault grows and the time series finish before the threshold is reached. The objective here is to predict, based on the sensors data, the number of cycles remaining until the total failure of the engine, i.e., the RUL.

The data is provided in text files containing 26 columns and each row represents one operational cycle.

The information in each column is the following:

1) Unit number
2) Time, in cycles or hours
3) Operational setting 1
4) Operational setting 2
5) Operational setting 3
6) Sensor measurement 1
7) Sensor measurement 2
...
26) Sensor measurement 21

For each row there is information about: *Unit number*, that is, the identifier of the trajectory; *Time*, represented in operational cycles (or hours); 3 different *Operational settings* and 21 *Sensors measurements.*

The three Operational settings, also referred as Operational regimes, refers to: *Altitude*; *Mach Number* and *TRA* and they have impact in the engine performance. The *Altitude*, as the name suggests, refers to the plane altitude in each moment of the flight, the *Mach Number* is the ratio of the aircraft speed and the speed of sound in the surrounding environment [42] and the *TRA*, which stands for *Throttle Resolver Angle*, is defined as the angular deflection of the pilot's power lever and it is used to control the thrust applied to the aircraft engine [43].

As illustrated in the dataset structure, each operational cycle is influenced by one of the six different combinations of three settings generated by the C-MAPSS model. The combinations applied in the PHM 2008 Challenge Dataset are presented in Table 4.2.

Table 4.2: Operational Settings

Altitude (K ft.)	Mach Number	TRA (°)
0	0	100
10	0.25	20
20	0.70	0
25	0.62	80
35	0.84	60
42	0.84	40

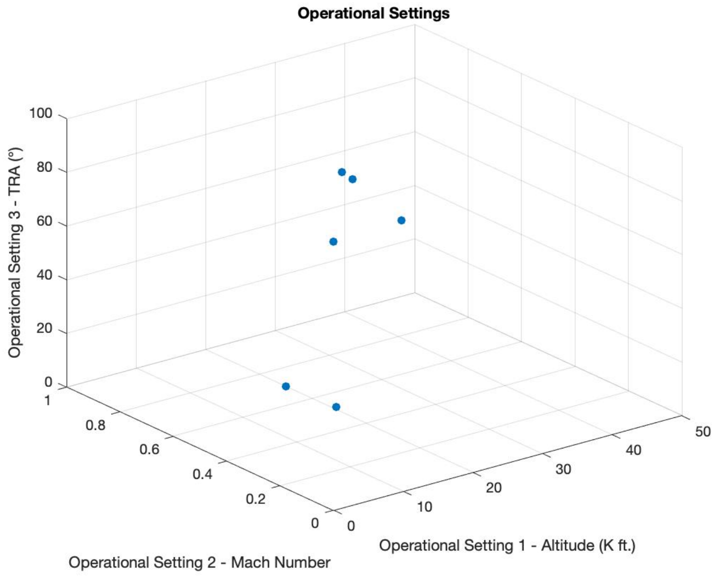

Figure 4.2: Operational settings relation

In Figure 4.2 the Operational settings values are represented. As can be observed, there is no relation between each of the Operational settings points as these are dispersed in the 3D space without any direct correlation amongst them. As such, and because they have a significant influence in the engine performance, the degradation behavior of the failures should be different and independent according to the Operational setting in use. Therefore, this will need to be take into account, when developing the methods for the RUL prediction.

The details regarding the 21 sensors are described in Table 4.3.

4.2 Techniques for RUL prediction implemented

With the objective of creating a base of comparison for the algorithms to be implemented and to gain some sensibility with the dataset values, three data driven prognosis methods implemented for the PHM08 Challenge Competition were reproduced, namely:

1. The winner method (*A Similarity-Based Prognostics Approach for Engineered Systems*) [41].
2. An adaptation of the winner method using a Extrapolation based RUL prediction.
3. An adaptation of the winner method using a Neural Network based RUL prediction.

All the approaches were implemented using MATLAB R2018a [44].

Table 4.3: Sensors description

Description	Units
Total temperature at fan inlet	ºR
Total temperature at LPC outlet	ºR
Total temperature at HPC outlet	ºR
Total temperature at LPT outlet	ºR
Pressure at fan inlet	psia
Total pressure in bypass-duct	psia
Total pressure at HPC outlet	psia
Physical fan speed	rpm
Physical core speed	rpm
Engine pressure ratio (P50/P2)	–
Static pressure at HPC outlet	psia
Ratio of fuel flow to Ps30	pps/psi
Corrected fan speed	rpm
Corrected core speed	rpm
Bypass Ratio	–
Burner fuel-air ratio	–
Bleed Enthalpy	–
Demanded fan speed	rpm
HPT coolant bleed	lbm/s
LPT coolant bleed	lbm/s

4.3 Similarity based Prognostics Approach

This approach was selected because it was the implementation that better predicted the RULs of the test dataset. As the name suggest, a similarity based approach was followed [41].

This method has two main stages: training and testing. Figure 4.3 shows the generic steps executed in each stage.

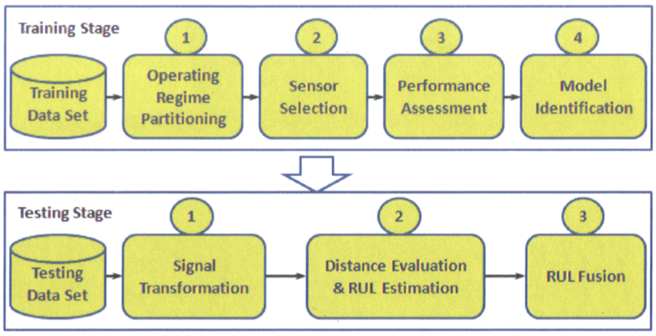

Figure 4.3: Steps taken in each stage (from *A Similarity-Based Prognostics Approach for Engineered Systems*)

4.3.1 Training

Operation Regime Partitioning

In this step some data preprocessing was performed, before using the data.

As referred to before, the three Operational settings in each cycle can be different. Analyzing the overall sensor data, it is not possible to identify a clear trend without dividing sensor data by the different combinations of the Operational settings. Therefore, it was required to identify to which combination of Operational settings each cycle belongs to. This was achieved using the k-means method, as the number of combinations is known by the dataset description.

The cycle values C in the training dataset were also modified. The data from each training unit was rearranged so that the last unit cycle corresponded to 0, to accomplish this the following formula was applied to all training data: $C^{adj} = C$ - *Unit Life*.

Sensor Selection

Next, feature selection was performed with the objective of choosing the most relevant features/sensors in estimating the RUL. This selection was made by observing the behavior of each of the 21 sensors over time. If a clear and continuous trend was found for a specific sensor, that one, should be selected. Otherwise, if some inconsistent trends were displayed it might indicate that that sensor is not useful in predicting the RULs and so it is discarded. At the end, a set of important features should be used during the training of the model. The set of sensors used in this technique (and in the other 2 approaches) were: 2, 3, 4, 7, 11, 12 and 15.

Performance Assessment

In the Performance Assessment step the features extracted from the data were fused in order to produce the Health Indicator (HI) values, which reflect the system health condition. These were calculated through linear regression models, more specifically, a regression model per each operational regime was generated. The models were fitted using filtered training data specific for each regime.

The linear regression formulation used was the following:

$$y = a + \sum_{i=1}^{N} \beta_i x_i + \varepsilon \tag{4.1}$$

where $x = (x_1, x_2, ..., x_N)$ is a feature vector of dimension N, $(a, \beta_1, \beta_2, ..., \beta_N)$ are the model parameters and ε refers to the noise term. The y is a health indicator and is binary: it has the value of 0 if, in the training unit, the cycle is near the end of life; and it has the value of 1 if, in the training unit, the cycle corresponds to an early life of the units. The remaining cycles are not in the extremes of the units' life time and so they were not considered for the fitting of the models. The y values can be calculated, using the following formulation:

$$y = \begin{cases} 0, & C_i^{adj} > C_{max} \\ 1, & C_i^{adj} < C_{min} \end{cases} \tag{4.2}$$

where i is the number of the cycle in the training unit. For this implementation C_{max} has the value of -5 and C_{min} the value of -300.

Having now one regression model per each regime, a HI time series can be calculated for each unit according to the operational regime of each cycle. The HI time series, obtained during the implementation, for training unit 1 is illustrated in Figure 4.4.

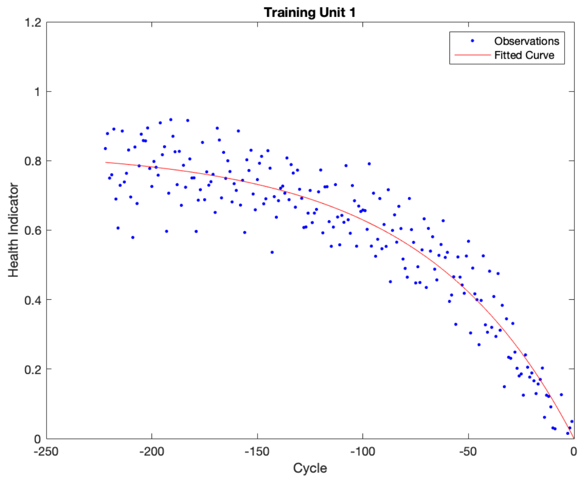

Figure 4.4: HI time series obtained for training unit 1

Model Identification

In this step an exponential regression model, M_i, was created for each training unit i. These models simulate the different degradation behavior of each training unit and were fitted using the respective unit training data.

The exponential regression formulation used was the following:

$$y = a(e^{bC^{adj}+c} - e^c) + \varepsilon \tag{4.3}$$

where C^{adj} is the cycle index, (a, b, c) are the model parameters of the training unit i, y is the HI and ε is the noise term. Figure 4.5 shows the curves corresponding to the exponential models obtained for the first 10 training units generated.

The library of exponential models created in this step was then used to calculate the RULs of the test units.

4.3.2 Testing

In the testing stage, the set of models representing the degradation behaviors of the training units, were used to infer the RULs of each test unit.

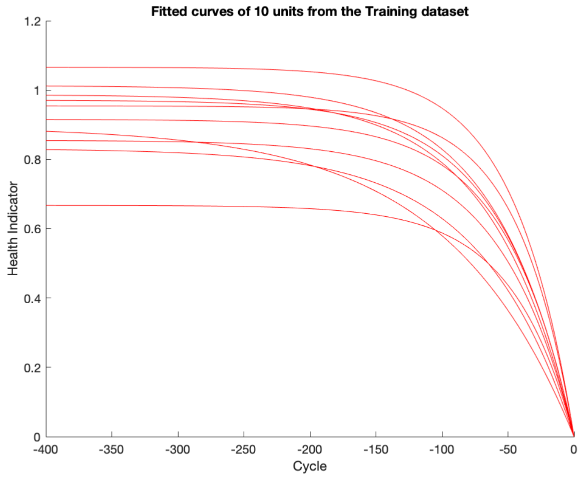

Figure 4.5: HI curves produced for the first 10 training units

Signal Transformation

Before calculating the RULs, the test data was preprocessed, similarly to the preprocess applied to the training data.

For each unit in the testing data set, the selected sensors data were classified by operating regimes, transformed by the linear models for performance assessment obtained during training, and merged to obtain an HI sequence Y [41].

Distance Evaluation & RUL Estimation

In this step, a moving average method was applied to the HI series.

Next, the distance between a test unit and each of the training models, with respect to r consecutive cycles, was calculated applying an Euclidean Distance, as represented in following formula:

$$d(\tau, Y, M_i) = \sum_{j=1}^{r} (Y_j - f_i(-\tau - r + j))^2 / \sigma_i^2 \tag{4.4}$$

where τ represents the number of cycles that sequence Y of the test unit is left shifted from the last cycle of model M_i, Y is the HI sequence of the test unit, f_i is the value of the exponential model of training unit i, r is the number of consecutive cycles analyzed from the test unit and the σ_i^2 is the variance associated with the fitting of model M_i.

In order to find the cycle from which the test unit fits the most in the training unit, the value of τ should be tested in the following interval: $0 \leq \tau \leq T_i - r$.

For each model M_i in the library, the RUL estimation for a test unit i is:

$$RUL_i = arg\ min_\tau\ D_i \tag{4.5}$$

Where the distance score D_i regarding that training unit i is:

$$D_i = min_\tau\ d_i(\tau, Y, M_i) \tag{4.6}$$

As a result, a pool of RULs was obtained for each test unit.

RUL Fusion

The last step is to merge the pool of RULs in one final RUL for each test unit. It is advised by the authors to, first, do some filtering in the pool of RULs. They presented some techniques for filtering:

- **Candidate Selection**: Some candidates of the RUL pool can be discarded, namely the ones whose distance score is 25% higher than the smallest distance score presented in the pool.
- **Outlier Removal**: The RULs particularly long (for example, larger than 190 cycles) or particularly short (for example, shorter than 125 cycles in total lifetime) should be removed.

  In case more outliers need to be removed, the following constraint can be applied to the RULs:

  $$Q_{0.5} - 3(Q_{0.5} - Q_{0.25}) < RUL_i < Q_{0.5} + 2(Q_{0.75} - Q_{0.5}) \tag{4.7}$$

  , in which $Q_{0.25}$, $Q_{0.5}$ and $Q_{0.75}$ are the first, second and third quartiles, respectively.

Finally for the RUL determination, the following weighted average was performed:

$$RUL = (13/23).min_i(RUL_i) + (10/23).max_i(RUL_i) \tag{4.8}$$

4.4 Extrapolation based Prognostics Approach

In order to experiment other category of prognostics methodologies, an Extrapolation based method was implemented. The implementation is, again, divided in two main stages: Training and Testing.

Training

With the objective of comparing this approach with the one implemented previously, the training steps will be identical. The training will include **Operation Regime Partitioning**, **Sensor Selection** and **Performance Assessment**. In the end, a regression model per regime should be created and they should be the same as the ones generated in the previous implementation.

Testing

In the testing phase, three steps need to be achieved: **Signal Transformation**, **HIs calculation** and **RUL extrapolation**.

Signal Transformation

Before calculating the RULs by extrapolation, the test data was preprocessed, similarly to the preprocess applied to the training data. This preprocessing includes **Operation Regime Partitioning** and **Sensor Selection**.

HIs calculation

In this step the HI values were calculated regarding the test dataset. Based on the regression models created previously and on the operating regime of each cycle in the test dataset, a HI sequence was obtained for each test unit.

In order to remove noise, a moving average filter with window length of 7 was applied to the HI sequence of each unit. The length of the window determined was based on the attenuation in the HI values. With smaller windows, the smoothing was not significant and thus a window length of 7 was applied.

RUL Extrapolation

For each test unit, a polynomial fitting curve of the HI values was generated and the RUL regarding that test unit was extrapolated. The choice of the polynomial order that better fits the HI values was done by analysing the graph with the HI values and respective polynomial curves (from the 1st order to the 4th) and concluding which order was more appropriated to that specific values.

In Figure 4.6 the RUL extrapolation performed for test unit 13, is demonstrated. As can be observed, the RUL corresponds to the difference between the cycle corresponding to the EoL of that unit and the last cycle presented on that unit history.

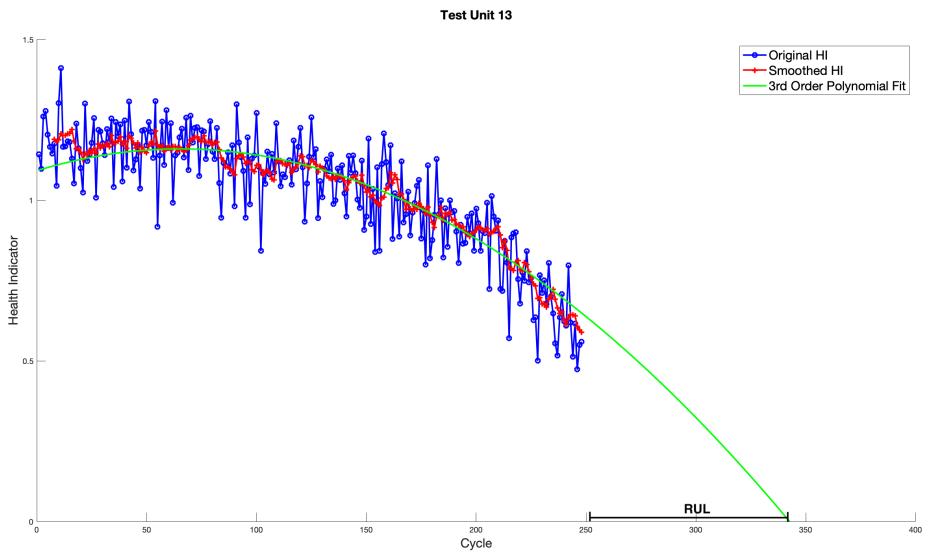

Figure 4.6: RUL Extrapolation generated for the test unit 13

4.5 Neural Network Based Approach

Following the same intention of comparing different approaches for the same dataset, a Neural Network based method was implemented.

The Neural Network consisted in a Multilayer Perceptron (MLP) and its design was inspired in an approach submitted for the competition [37].

4.5.1 Data preprocessing

Before training the MLP some preprocessing steps were performed.

Data normalization

The data from the training and test datasets was normalized in order to uniformize the range of values. The formula applied was the following:

$$N(x^d) = \frac{x^d - \mu^d}{\sigma^d} \tag{4.9}$$

where x^d corresponds to the original values for sensor d, and μ^d and σ^d represents the mean and standard deviation of the values for sensor d, respectively.

Data smoothing

In order to smooth the data from the datasets, a moving average was applied, with a window length of five. In Figure 4.7 it is possible to see the difference between the original and the smoothing data.

Figure 4.7: Smoothing performed on Sensor 6 of training unit 1

Historical Run Indicator

The author suggested the addition of six features to the dataset, representing the total number of cycles spent in each operational regime since the beginning of the run. This could enhance the knowledge of the Neural Network. Figure 4.8 illustrates the evolution of the run indicators in the training unit 1.

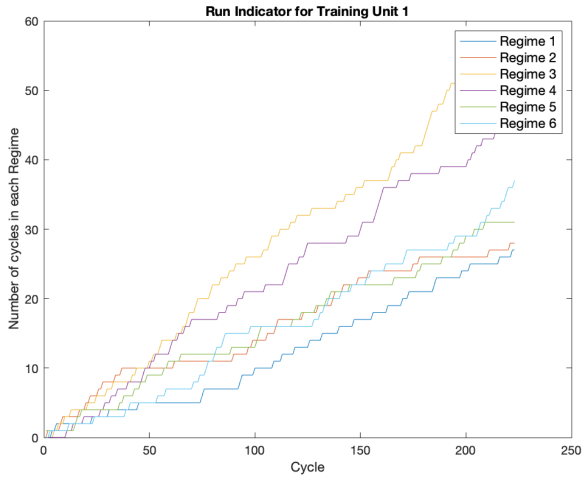

Figure 4.8: Evolution of the Run Indicators for the 6 Operational Regimes

Calculation of the Health Index of the training dataset

Using the regression models created previously and the operating regime of each training cycle, a HI value was calculated for each cycle. In order to smooth the values due the presence of noise, a moving average with window length of 5 was used. In Figure 4.9 it is possible to observe the HI sequence obtained for the training unit 1 and the respective smoothing.

4.5.2 Training

In the training stage, using the regression models obtained in the previous implementation, for each regime, a MLP was trained using the training dataset.

The MLP was generated with 16 **inputs** and **two layers**: a hidden layer and an output layer. The hidden layer had 40 neurons and a *tan-sigmoid* as the activation function. The output layer used a *linear transfer* function as the activation function.

The training data, that is, the MLP *input*, had the following columns:

- 3 Operating Regimes
- 7 Sensors selected
- 6 Run Indicators

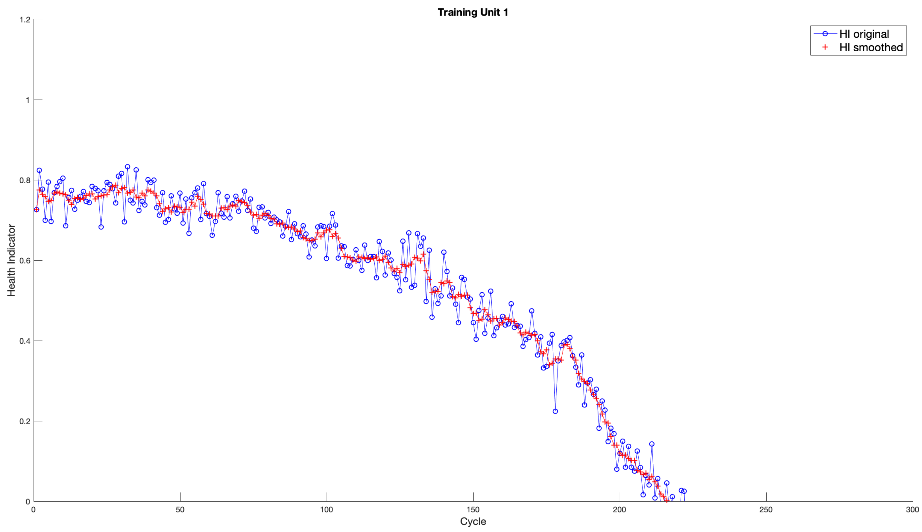

Figure 4.9: HI sequence obtained for training unit 1

The MLP *target*, was the HI sequence calculated previously.

The overall MLP structure is illustrated in the Figure 4.10.

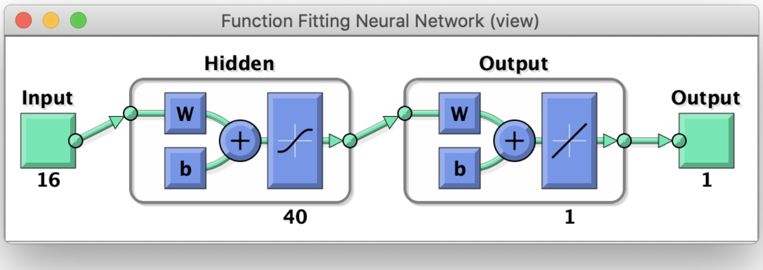

Figure 4.10: MLP Architecture

4.5.3 Testing

On the completion of the MLP training, the test dataset was simulated in the trained net. The output corresponded to the HI values of each cycle in the test data.

For each test unit, a moving average filter with window length of 7 was applied in order to smooth the HI values, reducing the noise associated. The reason of the choice of the window length was the same as before.

In order to calculate the RUL, the next step was extrapolating, using the most suited polynomial, the HI sequence for each test unit, until the point it reaches the 0. This point corresponds to End of Life of that unit. The order of the polynomial was chosen by graphic analysis.

The RUL for that unit corresponds to the difference between the cycle corresponding to the engine EoL and the last cycle presented on that unit test history. Figure 4.11 shows the extrapolation achieved for the test unit 13.

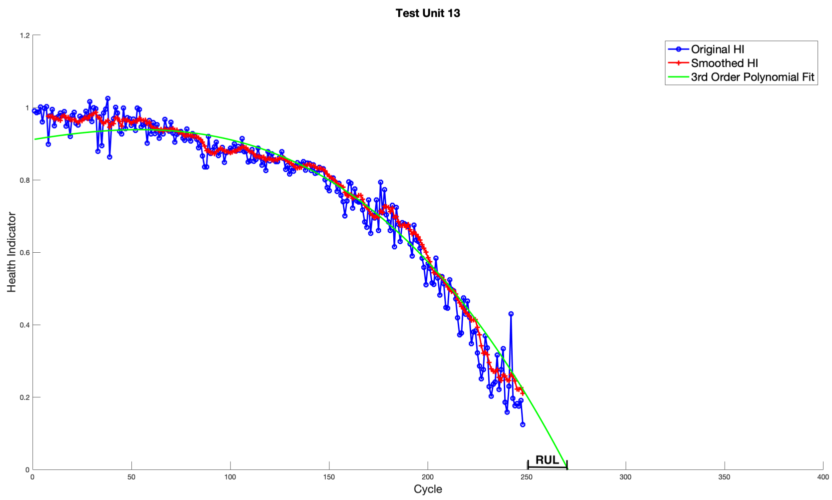

Figure 4.11: RUL Extrapolation performed using a NN approach for test unit 13

4.6 Comparison of techniques developed

The three developed approaches were evaluated with specific metrics in order to compare amongst them to determine which is the most promising approach.

These were tested in the Turbofan Dataset and in the PHM08 Challenge Dataset.

These approaches are expected to be used in other datasets that follow the same structure, this means, datasets which also contain aircraft sensors data, and share the same goal, i.e., the aircraft components' failure prediction through the estimation of components/subsystems RUL. It is also expected that these approaches can be used with data concerning other aircraft subsystems considered in the ReMAP project.

Furthermore, an adaptation and an adjustment of these approaches, will be required in order to be applied in a future dataset, that will contain real sensors data from aircraft provided by an airline company, as established in the ReMAP project. This is the dataset expected to be used during the rest of this work.

4.6.1 Results

In order to compare the different approaches, the used metrics were: Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Score Error.

The Table 4.4 shows the comparison, in terms of MSE, of the different developed approaches. As the true RULs for the PHM08 Challenge Dataset were not provided, it is not possible to compute the MSE for that dataset.

Table 4.4: MSE Error

	Turbofan Dataset				PHM08 Challenge Dataset
	FD001	**FD002**	**FD003**	**FD004**	
Similarity Approach	**395.02**	513.14	480.80	708.36	-
NN Approach	**656.96**	1104	825.27	1222.66	-
Extrapolation Approach	**632.89**	1503.50	3574.84	1922.83	-

The Table 4.5 compares the different approaches, in terms of RMSE. Again, as the true RULs for the PHM08 Challenge Dataset were not provided, the RMSE value was not computed for that dataset.

Table 4.5: RMSE Error

	Turbofan Dataset				PHM08 Challenge Dataset
	FD001	**FD002**	**FD003**	**FD004**	
Similarity Approach	**19.87**	22.65	21.92	26.62	-
NN Approach	**25.63**	32.65	28.72	34.95	-
Extrapolation Approach	**25.15**	38.77	59.78	43.85	-

The Table 4.6 compares the different approaches, in terms of Score Error. This metric was created by the PHM08 Challenge Competition. The results obtained for the PHM08 Challenge Dataset were obtained through the submission of the results in the link provided by the Competition for evaluation of the results.

Table 4.6: Score Error

	Turbofan Dataset				PHM08 Challenge Dataset
	FD001	**FD002**	**FD003**	**FD004**	
Similarity Approach	**654.73**	3979.52	2164.34	7242.26	1036.56
NN Approach	**1693.37**	644970.70	599436.03	82043.89	9628.34
Extrapolation Approach	**1738.89**	1.166e+6	1.439e+14	612177.72	32527.17

The analysis of the results of MSE, RMSE and Score Error presented in tables 4.4, 4.5 and 4.6, allows one to conclude that the Similarity based approach is the method that more accurately estimates the RUL for the testing dataset. The second best approach is Neural Network based, and the approach with least accurate results is the Extrapolation based approach.

In terms of dataset, the FD001 was the one with better results. This was expected as the data from this dataset was generated with just one failure embedded and only one operational regime in use. These facts simplified the diagnose of the failure and the prediction of the respective RUL.

The Similarity based approach provided good results regarding the RUL prediction, which was expected, given that the Similarity based approach followed was the one with the best score in the PHM 2008 Challenge Competition.

The results of the Neural Network based approach were below the expected. Some possible reasons for not obtaining the best results of this approach might be some important preprocessing step missing or not well executed, or a wrong decision regarding the choice of the order of the regression that better described the degradation behaviour of the test instances. Another possible reason was the deficient definition of the MLP layout, namely the number of neurons and layers.

The Extrapolation based also produced poor results. This was expected due to the fact that the RUL calculation was based on a simple regression. In particular, this type of approach can drastically weaken results when the dataset size is small, which was the case of some of the aircraft trajectories presented in the dataset.

It was expected that both the Similarity based and Neural Network based approaches would give better results due to the techniques used for RUL prediction. The Extrapolation based approach was expected to give worse results due to the simple methodology assumed for RUL prediction.

Comparing the Score error obtained for the PHM08 Challenge Dataset with the ones achieved by the authors, the Similarity approach got relatively closer to the expected, as the authors score was 512.12 and the score obtained in this work was 1036.56. The reasons for this difference might be the way the moving average was applied, including the considered window length, or the definition of the r value in the equation 4.4, which influences the overall RUL calculation.

Analysing the results and bearing in mind the second Semester, the improvement of the Similarity and Neural Network approach seems possible and legitimate and it will be performed. In particular, the Neural Network results suggests that they can be significantly improved by performing extra preprocessing steps, or improving the current ones (ex: feature selection) or defining a different structure for the Neural Network by changing the number of neurons, layers or the activation functions. In case it becomes viable, some new Machine Learning techniques will be explored in order to improve the RUL prediction accuracy.

This page is intentionally left blank.

Chapter 5

Work Plan

In this section will be presented a summary of work developed in this first Semester, as well as, a work plan for the second Semester.

5.1 Work performed in the first Semester

With the objective of summarizing and review all the work developed until this time, in Figure 5.1 there are illustrated the tasks executed so far, in the form of a *Gantt chart*.

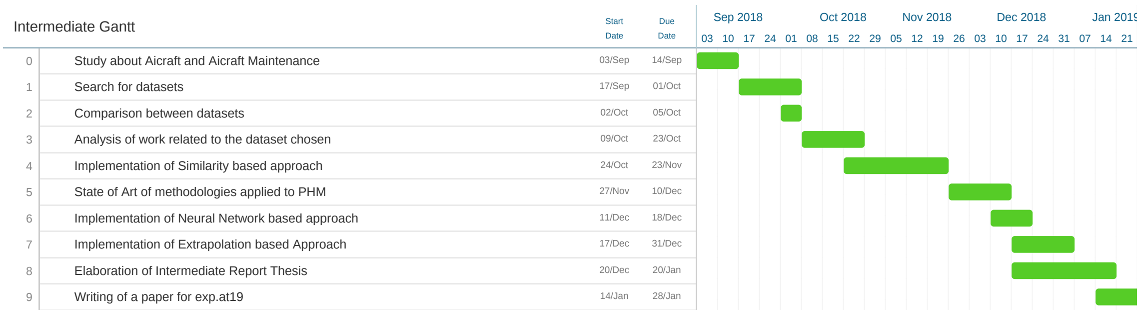

Figure 5.1: *Gantt chart* of work developed during First Semester

During the months of September and October, the work developed was associated with the study and learning of new concepts regarding aircraft maintenance and the search and comparison of possible datasets to use, as well as, an analysis of methodologies applied on them, namely in the PHM08 Challenge Dataset. During the months of November and December the developed work concerned the study of state of art of regarding PHM methodologies and the implementation and improvement of different approaches of RUL prediction using the chosen dataset. In January, the main focus was in the write of the intermediate report and a paper for the *Experiment@ International Conference* [45] were a web interface for the comparison of the implemented methodologies will be presented.

5.2 Work plan for the second Semester

Regarding the Work Plan for the second Semester, Figure 5.2 illustrates the set of tasks to be achieved.

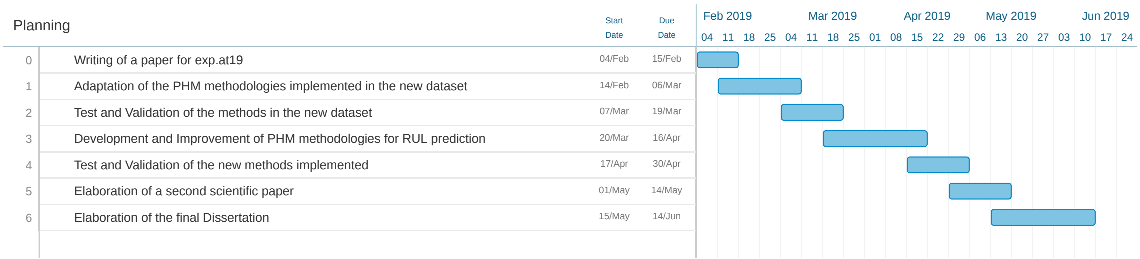

Figure 5.2: *Gantt chart* of work plan for Second Semester

In the second Semester the work plan begins with the conclusion of the scientific paper for the *Experiment@ International Conference*, initiated in the first Semester. Next will be performed an adaptation and adjustment of the methods already implemented in the new dataset, that will be provided in the ReMAP context by the KLM Royal Dutch Airlines [26]. This dataset will be more realistic in terms of data, as it will consist of data retrieved from real aircraft sensors. These methods will then be validated and tested.

The task on which more time will be spent, is the exploration of new Machine Learning approaches regarding RUL estimation and the improvement of the existing ones. These will be also tested and validated in order to assess if there was improvement or not.

In the last period of the Semester the writing of another scientific paper aiming the work developed during the Thesis is planned, as well as the writing of the final Dissertation report.

5.3 Risks

In this initial stage, it is important to anticipate and predict possible problems associated with this work. For that, a risk analysis was performed and the results are presented next.

For the **Probability** assessment, the following scale was defined:

- **1 - Very Low**
- **2 - Low**
- **3 - Medium**
- **4 - High**
- **5 - Very High**

For the **Impact** assessment, the following scale was used [46]:

- **1 - Minimal**
- **2 - Minor**
- **3 - Moderate**
- **4 - Significant**
- **5 - Severe**

5.3.1 Risk 1 - Unavailability of data

- **Name:** Unavailability of data
- **Description:** Real data from aircraft may not be provided for this work.
- **Impact:** 4
- **Probability:** 2
- **Mitigation action:** Use the backup datasets already identified.

5.3.2 Risk 2 - High execution time of the used methods

- **Name:** High execution time of the used methods
- **Description:** As Machine Learning models will be created, trained and tested, the execution time of the methods can be high.
- **Impact:** 3
- **Probability:** 4
- **Mitigation action:** Use computers with more computational power, or select just a part of the dataset to be used.

5.3.3 Risk 3 - Lack of quality of the data in the new dataset

- **Name:** Lack of quality of the data in the new dataset
- **Description:** As the new dataset may arrive, it can be noisy and not represent the aircraft failures in the data
- **Impact:** 5
- **Probability:** 2
- **Mitigation action:** Use more complex preprocessing techniques or, in last case, use the backup dataset

The summary regarding **Impact vs Probability** of the risks identified is presented next, in the form of a table:

Table 5.1: Impact & Probability Analysis

Impact / **Probability**	**1 - Minimal**	**2 - Minor**	**3 - Moderate**	**4 - Significant**	**5 - Severe**
5 - Very High					
4 - High			**Risk 2**		
3 - Medium					
2 - Low				**Risk 1**	**Risk 3**
1 - Very Low					

This page is intentionally left blank.

Chapter 6

Conclusion

In this first part of the work, some important steps were achieved. The study in the field of the evaluation of aircraft condition allowed the understanding of some important concepts like RUL and HI and stimulated the study of the different methodologies that can be applied in PHM systems. The benchmark of datasets also enabled the finding of a work related dataset, that was used in order to do some preliminary experiments regarding the RUL predictions based on different data driven approaches. The three different approaches implemented were: a Distance based, an Extrapolation based and a Neural Network based approach. Comparing the results, it can be concluded that the Similarity based approach was the one which more accurately estimated the RUL for the test instances in the dataset.

The work planned for the second Semester involves the improvement of the approaches tested in this first Semester and the possible exploration of other promising machine learning approaches regarding the RUL estimation. The access to a new dataset with real sensor data is expected, and thus an analysis of the dataset will be required and the adaptation of the methodologies developed so far will be performed.

References

[1] A. Saxena and K. Goebel. PHM08 Challenge DataSet. *NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository), NASA Ames Research Center, Moffett Field, CA*, 2008.

[2] A. Saxena and K. Goebel. Turbofan Engine Degradation Simulation Data Set. *NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository), NASA Ames Research Center, Moffett Field, CA*, 2008.

[3] ReMAP - Real-time Condition-based Maintenance for Adaptive Aircraft Maintenance Planning. `https://h2020-remap.eu/`. Accessed: 2018-11-26.

[4] Serdar Dalkilic. Improving aircraft safety and reliability by aircraft maintenance technician training. *Engineering Failure Analysis*, 82(May):687–694, 2017.

[5] William Rankin, Rebecca Hibit, Jerry Allen, and Robert Sargent. Development and evaluation of the maintenance error decision aid (meda) process. *International Journal of Industrial Ergonomics*, 26(2):261 – 276, 2000.

[6] Safety Report. *IATA - Health safety report 2017*. IATA, 2017.

[7] Airbus. A Statistical Analysis of Commercial Aviation Accidents 1958-2017. 2017.

[8] Basim Al-Najjar and Imad Alsyouf. Selecting the most efficient maintenance approach using fuzzy multiple criteria decision making. *International Journal of Production Economics*, 84(1):85 – 100, 2003.

[9] Jong Ho Shin and Hong Bae Jun. On condition based maintenance policy. *Journal of Computational Design and Engineering*, 2(2):119–127, 2015.

[10] Atiq Waliullah Siddiqui and Mohamed Ben-Daya. Reliability centered maintenance. *Handbook of Maintenance Management and Engineering*, 60:397–415, 2009.

[11] C. J. Bamber, J. M. Sharp, and M. T. Hides. Factors affecting successful implementation of total productive maintenance: A UK manufacturing case study perspective. *Journal of Quality in Maintenance Engineering*, 5(3):162–181, 1999.

[12] A. Saxena. Prognostics, the science of prediction. *Annual Conference of the PHM Society (PHM2010)*, 2010.

[13] Jay Lee, Fangji Wu, Wenyu Zhao, Masoud Ghaffari, Linxia Liao, and David Siegel. Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications. *Mechanical Systems and Signal Processing*, 42(1-2):314–334, 2014.

[14] Mark Schwabacher and Kai Goebel. A Survey of Artificial Intelligence for Prognostics. The Intelligence Report. *Association for the Advancement of Artificial Intelligence AAAI Fall Symposium 2007*, pages 107–114, 2007.

[15] Farzaneh Ahmadzadeh and Jan Lundberg. Remaining useful life estimation: review. *International Journal of Systems Assurance Engineering and Management*, 5(4):461–474, 2013.

[16] T. Wang. *Trajectory Similarity Based Prediction for Remaining Useful Life Estimation*. PhD thesis, University of Cincinnati, 2010.

[17] Pin Lim, Chi Keong Goh, Kay Chen Tan, and Partha Dutta. Estimation of Remaining Useful Life Based on Switching Kalman Filter Neural Network Ensemble. *PHM Society*, pages 1–8, 2014.

[18] Greg Welch and Gary Bishop. An Introduction to the Kalman Filter. pages 1–16, 2006.

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

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

[21] Yimin Fan and C. James Li. Diagnostic rule extraction from trained feedforward neural networks. *Mechanical Systems and Signal Processing*, 16(6):1073–1081, 2002.

[22] Arnaz Malhi, Ruqiang Yan, and Robert X. Gao. Prognosis of defect propagation based on recurrent neural networks. *IEEE Transactions on Instrumentation and Measurement*, 60(3):703–711, 2011.

[23] David McSherry. Strategic Induction of Decision Trees BT - Research and Development in Expert Systems XV. *Research and Development in Expert Systems XV*, (Chapter 2):15–26, 1999.

[24] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, Sep 1995.

[25] Z. Ghahramani. An introduction to hidden markov models and bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, pages 9–42, 2001.

[26] KLM Royal Dutch Airlines. `https://www.klm.com`. Accessed: 2019-01-5.

[27] Cecilia C. Larrosa Abhinav Saxena, Kai Goebel and Fu-Kuo Chang. CFRP Composites Data Set. *NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository), NASA Ames Research Center, Moffett Field, CA*, 2008.

[28] NASA Prognostics Data Repository. `https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/`. Accessed: 2018-11-17.

[29] Abhinav Saxena, Kai Goebel, Cecilia Larrosa, V. Janapati, and F.-K. Chang. Accelerated aging experiments for prognostics of damage growth in composite materials. 2011.

[30] Abhinav Saxena, Member Ieee, Kai Goebel, Don Simon, and Neil Eklund. Damage Propagation Modeling for Aircraft Engine Prognostics. *Proceedings of IEEE International Conference on Prognostics and Health Management*, pages 1–9, 2008.

[31] Propulsion System Simulation C-mapss. User ' s Guide for the Commercial Modular Aero- User ' s Guide for the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS). 2(January 2007), 2016.

[32] Emannuel Ramasso and Abhinav Saxena. Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets, 2014.

[33] Abhinav Saxena, Jose R Celaya, Bhaskar Saha, Sankalita Saha, and Kai Goebel. On Applying the Prognostic Performance Metrics. *Annual Conference of the Prognostics and Health Management Society*, pages 1–16, 2009.

[34] D. Azevedo, B. Ribeiro, and Alberto Cardoso. Benchmark of methods applied to PHM08 Challenge Dataset. Technical report, University of Coimbra, 2018.

[35] Felix O. Heimes. Recurrent neural networks for remaining useful life estimation. *2008 International Conference on Prognostics and Health Management*, pages 1–6, 2008.

[36] Manzar Abbas. System Level Health Assessment of Complex Engineered Processes. *Lancet neurology*, 10(3):209, 2011.

[37] A. M. Riad, Hamdy K. Elminir, and Hatem M. Elattar. Evaluation of neural networks in the subject of prognostics as compared to linear regression model. *International Journal of Engineering & Technology*, 10(06):52–58, 2010.

[38] Wen Ben Jone, Carla Purdy, and Hal Carter. Master of Science Committee Chair :. *Computer Engineering*, 2009.

[39] Jianzhong Sun, Hongfu Zuo, Wenbin Wang, and Michael G. Pecht. Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance. *Mechanical Systems and Signal Processing*, 28:585–596, 2012.

[40] Emmanuel Ramasso, Michèle Rombaut, and Noureddine Zerhouni. Joint prediction of observations and states in time-series based on belief functions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 43(1):37–50, February 2013.

[41] T. Wang, Jianbo Yu, D. Siegel, and J. Lee. A similarity-based prognostics approach for remaining useful life estimation of engineered systems. In *2008 International Conference on Prognostics and Health Management*, pages 1–6, Oct 2008.

[42] Mach Number. `https://www.grc.nasa.gov/www/k-12/airplane/mach.html`. Accessed: 2019-12-5.

[43] Throttle Resolver Angle. `https://en.wikipedia.org/wiki/Thrust_lever`. Accessed: 2019-12-5.

[44] The MathWorks, Inc. *MATLAB and Statistics Toolbox Release 2018a*, Natick, Massachusetts, United States, 2018.

[45] Experiment@ International Conference. `http://expat.org.pt/home/`. Accessed: 2019-1-19.

[46] Risk impact assessment and prioritization. https://www.mitre.org/publications/systems-engineering-guide/acquisition-systems-engineering/risk-management/risk-impact-assessment-and-prioritization. Accessed: 2019-01-16.