MDPI

Article

Remaining Useful Life Estimation of Cooling Units via Time-Frequency Health Indicators with Machine Learning

Raúl Llasag Rosero * **, Catarina Silva** **and Bernardete Ribeiro**

CISUC—Centre for Informatics and Systems of the University of Coimbra, Department of Informatics Engineering, Polo II—Pinhal de Marrocos, University of Coimbra, 3030-290 Coimbra, Portugal; catarina@dei.uc.pt (C.S.); bribeiro@dei.uc.pt (B.R.)

* Correspondence: rosero@dei.uc.pt

Abstract: Predictive Maintenance (PM) strategies have gained interest in the aviation industry to reduce maintenance costs and Aircraft On Ground (AOG) time. Taking advantage of condition monitoring data from aircraft systems, Prognostics and Health Maintenance (PHM) practitioners have been predicting the life span of aircraft components by applying Remaining Useful Life (RUL) concepts. Additionally, in prognostics, the construction of Health Indicators (HIs) plays a significant role when failure advent patterns are strenuous to be discovered directly from data. HIs are typically supported by data-driven models dealing with non-stationary signals, e.g., aircraft sensor time-series, in which data transformations from time and frequency domains are required. In this paper, we build time-frequency HIs based on the construction of the Hilbert spectrum and propose the integration of a physics-based model with a data-driven model to predict the RUL of aircraft cooling units. Using data from a major airline, and considering two health degradation stages, the advent of failures on aircraft systems can be estimated with data-driven Machine Learning models (ML). Specifically, our results reveal that the analyzed cooling units experience a normal degradation stage before an abnormal degradation that emerges within the last flight hours of useful life.

Keywords: predictive maintenance; prognostics and health maintenance; remaining useful life; health indicators; machine learning; artificial intelligence

1. Introduction

The interest in developing Prognostics and Health Maintenance (PHM) applications is growing in some industries aiming to reduce maintenance costs and down time [1]. In the aviation industry, the interest has also emerged due to the need for accomplishing strict safety and operational reliability policies [2].

Taking advantage of the profusion of sensors present on modern aircraft, the collection and the analysis of Condition Monitoring (CM) data has been useful to develop data-driven models [1,3]. In consequence, maintenance processes are assisted by such models, which usually leverage the advantages of Artificial Intelligence (AI) and Machine Learning (ML). Particularly, Predictive Maintenance (PM) strategies have been used to determine the advent of a failure by applying Remaining Useful Life (RUL) concepts [4–7], i.e., methods that predict of the remaining time an equipment is estimated to be able to function without failing [8].

In the development of data-driven RUL estimation models, data gathering is mainly focused on (1) getting run-to-failure data [9], and (2) collecting many failure instances [3]. While the first aims to collect data related to the transition from healthy to faulty states, the second aims to identify health degradation patterns.

Data transformations from time or frequency domains have often been used to discover trends that sensor signals usually conceal [6,10]. Such aircraft sensor signals are frequently non-stationary [8,11], i.e., time period and frequency are not constant [12]. Hence, data

transformation on both domains has been proposed [7,13,14]. Unfortunately, there are cases in which the failure advent is generally unable to be observed directly from data. In such cases, the construction of physics-based Health Indicators (HIs) can play a significant role [15,16].

The construction of HIs has been considered an essential stage in PHM, mainly when outputs of a physics-based model are used as inputs for data-driven models. In literature, the integration of two different model types is referred to as the construction of hybrid approaches [15]. When data-driven approaches use ML under a supervised learning approach, data labeling has been challenging because a degradation curve has to be proposed [15,17]. This process is usually known as the health stage division process and its main goal is to identify multiple health stages that could be identified along the equipment lifetime.

Among the main contributions of this work are: (1) the inclusion of physics-based models with data-driven models within a PHM program; (2) a Health Indicator construction approach based on time and frequency domains; (3) the identification of two degradation stages experienced by aircraft cooling units; and (4) the RUL estimation for those units.

This work mainly draws attention to the second contribution because it presents a time-frequency data transformation oriented for non-stationary signals. Concretely, our time-frequency HI construction approach based on the Hilbert spectrum [18,19] is used to accumulate the lowest frequency instances. Our experiments, conducted using data from the Cooling System (CS) of a wide body aircraft fleet [1], suggest that the degradation of Cooling Units (CUs) may be related to the presence of periods with the lowest frequency instances. Our RUL estimations, inferred from a Neural Network (NN) based model have been more precise after prioritizing the health prognostics of the last flight hours.

This article is organized as follows. Section 2 presents the literature review regarding RUL estimation approaches and time-frequency data transformation. Section 3 describes the developing phases and the modules of the proposed PHM program that integrates physics-based and data-driven models. Section 4 details the data gathered from a wide body aircraft fleet and the failure instances associated with the Cooling System. The development of the RUL predictor model is described into three sections: Health Indicator (HI) construction (Section 5), Health Stage division (Section 6) and RUL estimation (Section 7). The RUL estimations obtained by using the proposed HI construction versus using raw data are compared in Section 8. Finally, the conclusions and future lines of research are presented in Section 9.

2. Literature Review and Background

2.1. RUL Prediction Approaches

Several studies have reviewed common RUL prediction approaches and classified them into various categories taking different emphases [12,20–24]. Considering the basic techniques and methodologies, the following four categories have been proposed in [15]: physics-based, statistical-based, data-driven, and hybrid approaches.

Physics-based approaches describe the advent of a failure through mathematical models [24]. Statistical-based estimate the RUL by establishing models based on empirical knowledge and present the RUL prediction as a conditional Probabilistic Density Function (PDF) [22]. Data-driven approaches typically attempt to learn degradation patterns using Artificial Intelligence (AI)/Machine Learning (ML) techniques from run-to-failure data [12,21,22,25]. Finally, hybrid approaches aim to integrate the capabilities of two or more previous approaches [15].

Given the availability of Condition Monitoring (CM) data, usually gathered from digital twin systems (e.g., data from the Commercial Modular Aero-Propulsion System Simulation (CMAPSS) [8,11]), industrial and academic researchers have extended the literature related to developing data-driven models.

Among the main reasons for adopting data-driven approaches is the fact that health degradation can be revealed from data, avoiding PHM practitioners the necessity to delve into the physics of the signals analyzed [10], which is usually complex. That advantage

has allowed researchers to be focused into performing adequate data transformations to learn the health degradation path through AI/ML techniques [6]. Unfortunately, there are cases in which data transformations from time or frequency domains are not sufficient to develop data-driven models.

2.2. Time-Frequency Data Transformations

Data transformations from time or frequency domains, such are kurtosis, mean, peaks, and Fourier transform, among others, have been widely used in the presence of stationary signals [6,7]. Nevertheless, data transformations on a single domain have demonstrated limitations in the analysis of non-stationary signals and non-linear systems [24].

When the degree of a time-frequency moment can be interpreted and moments can be useful to predict health states, obtaining the following moments is suggested: conditional spectral moment, conditional temporal moment and joint time-frequency moment [3,26]. Otherwise, time-frequency data transformations, such as, Short-time Fourier Transform (STFT) [27], Wigner-Ville Distribution (WVD) [14], Empirical Mode Decomposition (EMD) [13] and Hilbert spectrum [18] can be promising alternatives.

In time-frequency analysis, Fast Fourier (FF) transform has been widely used [28]. In prognostics, STFT has been used to identify the advent of failures from the power spectrum. However, FF has physical meaning for linear systems, i.e., systems in which the change of the outputs is somehow proportional to the change of the inputs [10,13,29].

As an alternative, WVD has been considered as the most prominent time-frequency representation since it satisfies an exceptionally large number of mathematical properties [14]. Moreover, its applicability can be extended to multidimensional vectors. However, in the presence of large amounts of data, WVD is less powerful given that the computational cost can be high due to redundancies in the Fourier Transform (FT) computation [30]. As an alternative, the time-frequency distribution, Hilbert spectrum, can be promising, since it is a distribution of signal amplitude that helps in distinguishing a mixture of moving signals [18]. However, when the advent of a failure cannot be directly observed from time-frequency data transformation, even using AI/ML techniques, the need for developing hybrid approaches gains importance [12,15,20].

3. Approach

The hybrid PHM program, proposed and implemented in this work, constructs HIs from a physics-based model and uses them as input signals of a data-driven model. Since the implementation of hybrid approaches is recommended when a single approach is not able to satisfactorily predict the RUL, the need of describing a general prognostics program gains interest. In Figure 1, we illustrate the development stages and the modules for constructing physic-based and/or data-driven RUL predictors.

The backbone of our PHM program is composed of three prognostics stages presented in [4]. Those main stages (dotted line boxes) are: (1) Data acquisition; (2) Data pre-processing; and, (3) Prognostic model phase. This prognostics process flow has been conceived to estimate the RUL by applying only data-driven approaches. For that reason, we adopted the PHM modules processes (gray boxes) presented in [15] to construct hybrid approaches, namely: *data acquisition*; *HI construction*; *Health Stage division*; and, *RUL estimation*.

Our PHM program suggests adopting the data pre-processing modules (yellow boxes) when possible. In case the data-driven models do not achieve prognostics expectations, such is the case of the present work, the HI construction module can be adopted when health degradation is unable to be illustrated from data. In the listed modules, presented in the next sections, we performed the following technical processes, A to D, detailed in the following sections:

A. **Data acquisition** : refers mainly to constructing the experimental scenario. The main goal of this stage is to acquire run-to-failure data (Section 4.1). This stage integrates data transformations (Section 4.2), such as, data noise reduction, and data cleaning,

before applying feature engineering techniques [6]. In this technical process, handling imbalanced data (Section 4.3) and defining training and testing samples (Section 4.4) supports the prognostics model stage.

B.   **HI construction**: refers to data transformations conducted by signal processing (Sections 5.1–5.3) or AI/ML techniques [15]. Feature engineering modules, such as, feature extraction, and feature selection, aim to extract and select the indicators that better describe the machinery degradation [31,32].

C.   **Health Stage division**: considers different health stages, according to the varying degradation trends of HIs [15]. Considering two health stages, the problem of finding the point that separates them is well known as *elbow point* detection [17].

D.   **RUL estimation**: refers to the development of an AI/ML technique capable of predicting the RUL within health stages that present obvious degradation trends.
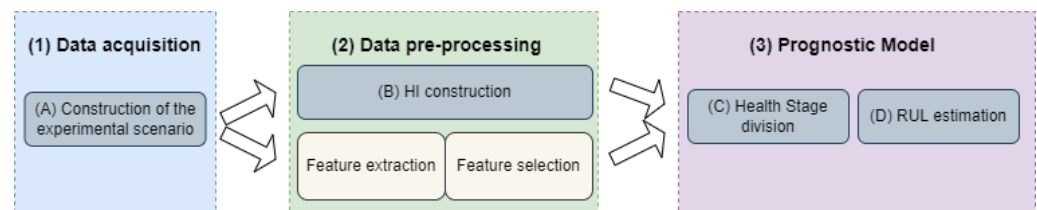

Figure 1. Prognostics and Health Maintenance program.

4. Experimental Scenario

This section refers to the tasks performed on the *Data acquisition* stage. In this case, run-to-failure data was acquired from the Cooling System (CS) of a wide body aircraft fleet. The data was obtained from the project Real-Time Condition-based Maintenance for Adaptative Maintenance Planning-ReMAP of the European Union's Horizon 2020 [33].

We consider using ReMAPs data to develop health prognostics of real aircraft systems and assist technicians in maintenance decision-making related to the CS. This cooling system senses the temperature of the wide body aircraft to maintain its operability. Specifically, the temperature is regulated by transporting liquid cooling through pumps, cooling units (CUs) and motor controllers.

As the Remaining Useful Life (RUL) concept was adopted to predict the amount of time left before an aircraft system fails [8], we considered measuring the time in terms of remaining flight hours. Flight hours as an unit of time has been used in aerospace studies [2,34,35]. Particularly, we are interested in predicting the RUL of CUs because the personal staff have documented the failure instances of these units.

For the ReMAP project, CSs of 17 different aircraft were monitored along 14 different flight phases using a sample rate of 1 Hz. During 29 months, data from 18,295 flights were collected, gathering 26.4 GB of data. Among sensor data, the personal staff has also documented maintenance interventions of three types of failures, namely: Flight Deck Events (FDE), Aircraft Technical Log (ATL) and Predictive Maintenance (PM) provided by technicians.

Only PM failures were used to monitor the health status because FDE and ATL failures were not confirmed by maintenance technicians. Despite 40 PM failures on CUs having occurred, only the analysis of 21 failures was considered because 2 failures were labeled as "dubious", 10 were labeled as "likely", while the remaining 7 did not present sensor data or presented constant measurements on sensors.

The CS of a single aircraft is monitored through 36 sensors located at four equivalent CUs. Therefore, each CU is monitored by analyzing 9 sensor signals. Among these signals, informative fields such as flight ID, departure data time, flight phase, row number, altitude and airspeed were also considered to monitor the CU health status along the different flights phases.

4.1. Construction of Run-to-Failure Data

The retrieved data does not provide run-to-failure data by default; CS data has to be ordered first. The flight ID, departure data time, flight phase and row number fields are used to order the data. To get run-to-failure data, we assume that technicians intervened CUs just when their useful life ended. That is, the failure threshold was defined as the last operating measurement before a maintenance intervention.

Under this assumption, we can get a run-to failure trajectory per aircraft CU. To get the different trajectories, we get the flights (flight ID), a previous maintenance intervention of each aircraft tail, and then sort the sensor measurements by row number. For easy understanding, each run-to-failure trajectory has been linked to the associated failure (failure ID). In Table 1, the aircraft tail, the number of measurements and the number of flight hours for each trajectory are detailed. Note that the 21 trajectories presented are sorted by flight hours.

Table 1. CU Trajectories by Failure ID.

Failure ID	Aircraft Tail	Measurements	Flight Hours
110	iilvtkok	880,285	245
107	ibauqnx	814,507	226
30	enwslczm	758,600	211
25	lbhkyjhi	744,231	207
111	dlkzncgy	724,419	201
23	dlkzncgy	620,916	174
18	wnjxbqsk	589,841	164
22	iilvtkok	537,029	149
21	iefywfmy	520,490	145
2	ibauqnxj	482,575	134
15	trmblwny	439,745	122
113	lbhkyjhi	313,901	87
28	tjyjdtaf	272,910	76
19	iefywfmy	235,371	65
17	wnjxbqsk	170,628	47
31	ekzlmbdx	167,440	47
27	tjyjdtaf	104,288	29
26	tjyjdtaf	103,828	29
116	iefywfmy	103,393	29
7	ekzlmbdx	71,143	20
1	ibauqnxj	34,994	10

4.2. Noise Reduction and Data Cleaning

Cooling Units are monitored by 9 equivalent sensors. However, 3 of them presented a lot of infinite and `NaN` measurements, hampering their analysis. The fields, which where discarded from the first CU, correspond to `xtooix*`, `wqybtv*` and `ytauwt*`. Equivalent sensors where removed from the remaining CUs.

After data discarding, there were identified sensors that usually presented outliers in all the CUs. The outliers were removed by applying a Linear Interpolation with a threshold factor (*mean* = 1.5), configuration that was applied to sensors `yucqij*` and `wqgsfv*` sensors, and their equivalent ones in the remaining CUs. We adopted this threshold factor to clean those sensor signals because the non-noisy values were located below the *mean* = 1.5. So, after generalizing this global threshold, the same procedure was conducted to equivalent sensors of the remaining units.

4.3. Handling Imbalanced Data

Data from the Cooling System is unbalanced respecting to their 14 flight phases, presenting most of the instances in the 8th phase. For that reason, those initial phases have

aggregated into the following ones, namely: start (1, 2), climb (3–7), cruise (8), descent (9–13) and finish (14).

The alteration between flight phases over time enables the analysis of the degradation of CUs. Nevertheless, the period in which the RUL has meaning needs to be defined through a degradation function f. To do that, we considered the *EndOfLife* time or or maximum useful life time when technicians intervened in a determined CU. Then, according to Equation (1), we counted back the flight hours (t) that a determined aircraft experienced to simulate a constant degradation. At this stage, the degradation function is defined as follows:

$$f(t) = \lfloor EndOfLife - t \rceil \tag{1}$$

where floor and ceil operators restrict to RUL $\in \mathbb{N}$.

4.4. Training and Testing Data

From the 28 trajectories of the CS data, some of them presented the maintenance intervention within the first 30 h while others surpassed the 100 h. Since we are interested in analyzing the degradation in periods with more than 48 flight hours, we selected the first 14 trajectories (above the double horizontal line) described in Table 1.

For the next stages, we considered the following trajectories:

- Training= $[2, 15, 18, 21, 22, 25, 28, 107, 110, 111, 113]$
- Testing= $[19, 23, 30]$

Criteria used to define training and testing trajectories corresponds to the F-test because this parametric method tests two population variances. In our case, we test the variance of flight hours of training and testing samples by analysing the F value obtained as follows:

$$F = \frac{S_{Train}^2}{S_{Test}^2} = \frac{3033.65}{5761} = 0.526.$$

when the similarity of variances is commonly assumed in literature when $0.5 < F < 2$. Using this selection criterion, we also ensure that training and testing trajectories present a similar length of health stages. This similarity gains interest in the *Health Stage division* process, mainly because health stages have to be separated by a determined point (e.g., the *elbow point*, when two health stages are identified [17]).

5. Health Indicators Construction

Commonly, in the *Data pre-processing* stage, PHM practitioners may identify the following problems: (1) data is composed of non-stationary signals [13,19] or (2) signals represent nonlinear processes [19,28]. In the presence of linear or stationary signals, the degradation of equipment may be easily extracted from features of time or frequency domains [3,28,36]. However, when the analysis of a single domain is not enough, both domains must be analyzed together [3]. In the case where health degradation is not grasped by data-transformations, the construction of Health Indicators (HIs) from physics-based approaches has been suggested [12,15].

5.1. Short-Time Fourier Transform (STFT)

Historically, STFT, expressed in Equation (2), has been one of the most used time-frequency representations of time signals [27]. Although the Fourier transform has physical sense under the presence of linear systems and stationary signals [10,13]. Since the basic idea of STFT is a moving window $w(\tau - t)$ Fourier Transform centered at t, where τ represents the time index, the time domain moving window over the signal generates a 2D time-frequency distribution (t, f) called spectrogram.

$$STFT(t, f) = \int_{-\infty}^{\infty} x(\tau)w(\tau - t)e^{-j2\pi f\tau}d\tau \tag{2}$$

Selecting an adequate w implies an extra effort, mainly because choosing from Hamming, Hanning, Kaiser-Bessel and Gaussian moving windows depend on the application problem [37].

In our experimental scenario, we applied a STFT to each sensor signal with a Hann window of size 128 and 50% of overlapping. After that, we found that some periods with the fewer magnitude frequencies appeared could be related with CUs degradation. In Figure 2, in which the SFTF of sensor `wqgsfv*` corresponding to the `failure ID` = 23 is illustrated, we observed that some of these (blue) periods appeared just before a maintenance intervention (vertical red line).
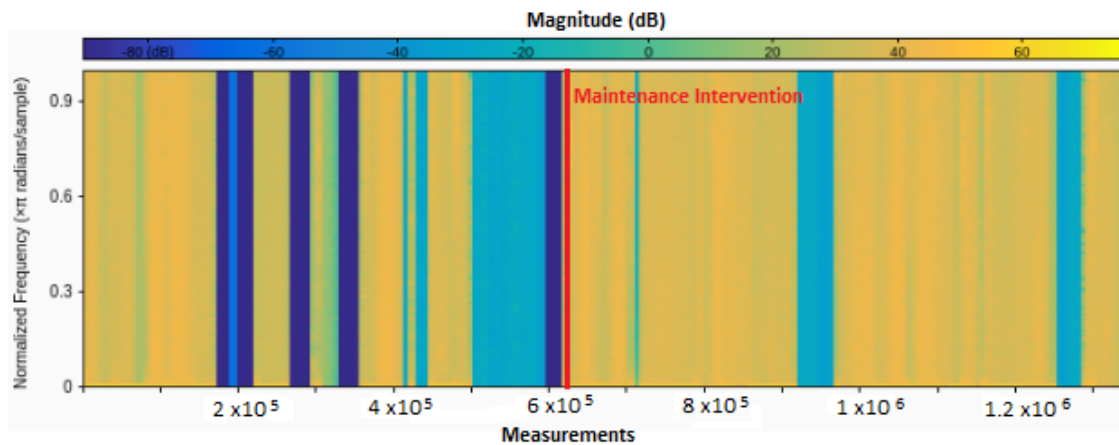

Figure 2. Degradation periods with low magnitude values.

5.2. Hilbert Spectrum

After noticing that degradation of CUs may be related to the presence of periods with fewer magnitude frequencies, we were interested in using a time-frequency representation capable to catch those periods.

The Hilbert spectrum, defined by Equation (3), is a time-frequency distribution of signal amplitude that helps in distinguishing a mixture of moving signals [18]. The given spectrum is computed by: (1) mathematically decomposing a signal into Intrinsic Mode Functions (IMF); and (2) applying the Hilbert transform to compute the instantaneous frequency of each component by obtaining $w = d\theta(t)/dt$ (the derivative of the phase θ w.r.t the time t) [13]. After performing the Hilbert transform on n IMF signals, the Hilbert spectrum is obtained. In practice, the Hilbert-Huang transform (HHT) has been widely used to obtain $X(t)$ [13]:

$$X(t) = Re\left[\sum_{k=1}^{n} a_k(t)e^{j \int w_k(t)dt}\right] \tag{3}$$

where $Re[]$ represents the real part of terms within brackets and both a_k and w_k are constants.

This transform uses the Empirical Model Decomposition (EMD) to get the IMF signals, for then, apply the Hilbert transform to each signal. Despite the implementation of the EMD depends on how IMFs are defined and the stop criterion used to find them [19], the implemented through a sifting process is composed of the following steps:

1. Identify all the local extrema of the input signal $x(t)$.
2. Separately connect all the maxima and minima with natural cubic spline lines to form the upper, $u(t)$, and lower, $l(t)$, envelopes.
3. Compute the mean of the envelopes as $m(t) = [u(t) + l(t)]/2$.
4. Find a tentative IMF $h(t)$ as the difference between the $x(t)$ and $m(t)$, $h(t) = x(t) - m(t)$.
5. Check if $h(t)$ satisfy the IMF definition and stoppage criterion.
6. If $h(t)$ does not satisfy the definition, repeat steps [1–5] many times til it satisfies the definition.

7. If $h(t)$ does satisfy the definition, assign the tentative IMF as an IMF component, $c(t)$.
8. Repeat steps [1–7] on the residue, $r(t) = x(t) - c(t)$, as the data.
9. Stop when the residue contains no more than one extremum.

5.3. Physics Model-Based Approach

Our proposed approach, based on Hilbert spectrum modeling, aims to emphasize the periods in which sensor signals experience the lowest frequencies. Our approach constructs a single HI form each sensor signal $x(t)$ of a determined CU. In [38], this procedure aims to capture the physics of the failure.

In the construction of the Hilbert spectrum, the accumulation of periods with similar frequencies is proposed by using IMFs. We are interested on the first IMF because it identifies and decomposes the finest component of the shortest period at each instant [27].

To perform the EMD, the following definition of Intrinsic Mode Function was adopted:

IMF: *any function having the same numbers (or at most differing by one) of zero-crossings and extrema, and also having symmetric envelopes defined by local maxima and minima, respectively* [19].

To get the first IMF $c_1(t)$, the following mathematical operations are performed:

$$X(t) - m_{1,1}(t) = h_{1,1}(t);$$

$$m_{1,1}(t) - m_{1,2}(t) = h_{1,2}(t);$$

$$m_{1,k-1}(t) - m_{1,k}(t) = h_{1,k}(t);$$

$$h_{1,k}(t) = c_1(t),$$

where indices refer an iteration k of each EMD step. In our case, the EMD applied to get the first IMF c_1. In this implementation, k iterations are performed while the difference $SD \nless \epsilon$ or while $h_{1,k}(t)$ does not satisfy the IMF definition. In this implementation, SD, defined by equation 4, is defined by the Cauchy type criterion [13].

$$SD = \sum_{t=0}^{T} \left[\frac{|h_{1,k-1}(t) - h_{1,k}(t)|^2}{h_{1,k-1}^2(t)} \right] \tag{4}$$

Once $c_1(t)$ is obtained from a single $X(t)$ time series, the $HI(t)$ signal that corresponds to the evolution of health conditions is defined in Equation (5):

$$HI(t) = \int |c_1(t)| dt \tag{5}$$

Given that the application of the integral operator implies obtaining large values, we performed a Z-score normalization process to the $HI(t)$ signal. The same normalization procedure was applied to the `altitude`, the `airspeed` and the `flight phase` to use them as inputs for the following technical processes.

6. Health Stage Division

Once Health Indicators were obtained, we considered using two health stages proposed in [11,17,34]. That consideration, illustrated in Figure 3, implies that CUs experience an imperceptible degradation before experiencing a abnormal degradation [8,11]. In that sense, the following piecewise function, defined by Equation (6), was formulated to identify the point (*elbow point*) that separates both degradation stages by conducting a binary classification. In Equation (6), a measurement labeled with 0 means that it corresponds to the imperceptible degradation, otherwise, it corresponds to the abnormal degradation stage.

$$f(t) = \begin{cases} 0 & if \quad t \leq elbowPoint \\ 1 & \end{cases} \tag{6}$$

Although statistical procedures such as t-test have been applied to identify the *elbow point* (red point), we adopted ML algorithms used in [17], since they have provided an accurate distinction between health degradation stages. Note that the lengths of imperceptible and abnormal degradation stages can have a large number of combinations, as in Figure 3. For that reason, a sliding window (blue vertical line) technique has been adopted to identify the best partition.
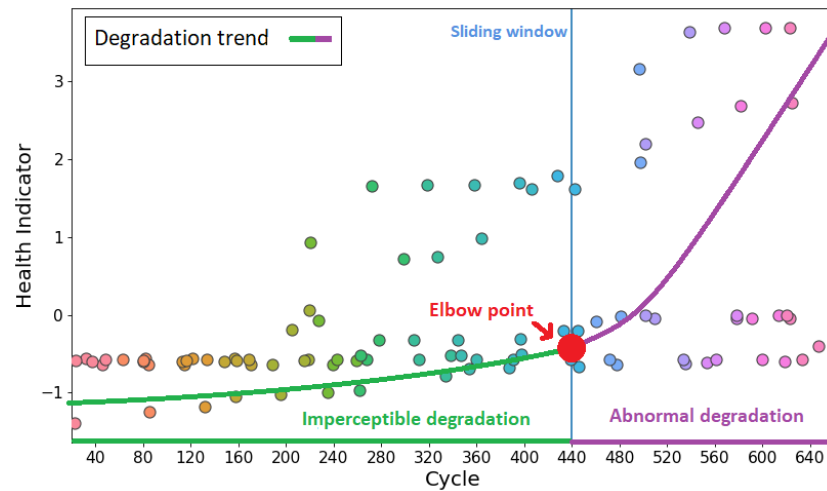
Figure 3. Elbow point detection based on sliding window.

Because in [17], the classification health stages has obtained promising results using CMAPSS data and the following algorithms: Multi-Later Perceptron (MLP), Random Forest Classifier (RFC) and Support Vector Classifier (SVC), we used them to distinguish the two health stages (0: imperceptible and 1: abnormal). The underlined configuration settings for these algorithms were selected by evaluating the hyperparameters detailed in Table 2.

Table 2. Elbow point hyperparameter tuning.

Algorithm	Hyperparameter	Search Range
Multi-Layer Perceptron	Neuron node	Dense
	Hidden Layers	2, 3
	Nodes 1st layer	20..30
	Nodes 2nd layer	20... 30... 50
	Optimizer	Adam, SGD
	Activation	Relu, Tanh
	Output Activation	Relu, Sigmoid
	Loss	Binary Cross Entropy
	Step-size	0.001, 0.01
	Early stopping	Yes, No
	Epochs	100, 200
Random Forest Classifier	Decision Trees	10, 25, 50, 75
	Min Samples to Split	2, 5, 10
	Min Samples at Leaf	1, 2, 4
	Max Features	Auto, Sqrt
	Nodes per layer	25, 50, 100
	Max Tree Depth	10, 25, 50, 75
	Boostrap Training	True, False
	Quality of a split	Gini, Entropy
Support Vector Classifier	Kernel	Linear, Rbf, Poly
	coef0	0.0, 0.1, 0.2
	gamma	Scale, Auto
	cs	0.01, 0.02, 0.05
	degree	0, 1, 2, 3, 4

Once hyperparameter tuning was conducted by fixing the elbow point into the last 48 flight hours, the parameters in bold were selected when models presented the higher classification performance. In Figure 4, we observe that those hyperparameters were selected when the classifiers obtained the higher values for the Area Under the Curve (AUC).
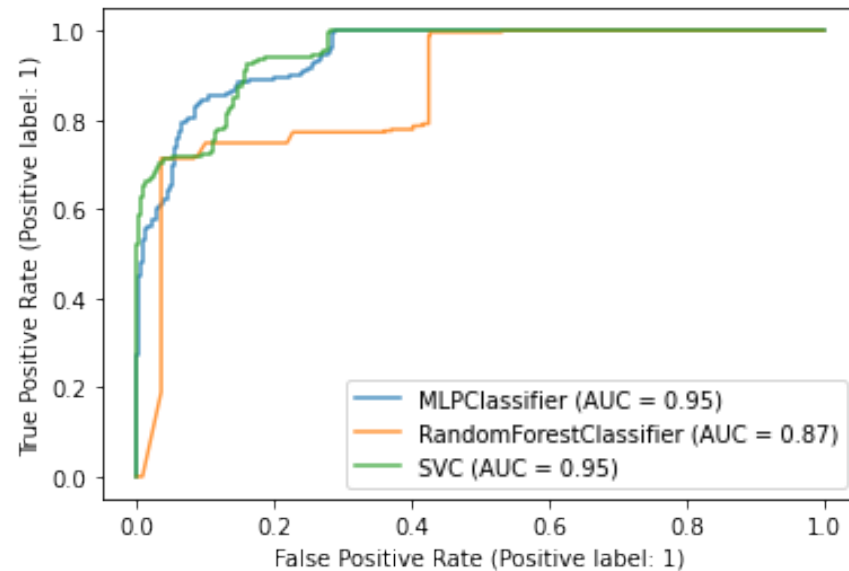
Figure 4. Hyperparameter tuning based on AUC when elbow point= 48.

To identify the point that better separates the health stages, we used a window of 4 flight hours. In Table 3, we observed that the elbow point took values from 16 to 72 with a step size of 4. After comparing the classification performance of models trained with different elbow point, we identified the best classifications when that point is configured between the last 28 and 44 flight hours.

Since we are comparing the performance of groups of three classifiers (RFC, SVC and MLP), we averaged their accuracy, F1-score and AUC to used it as comparison criterion. We used the average Accuracy as a selection criterion to highlight the period in which classification takes the best performances. In Figure 5, in which the average of classification metrics is illustrated for different elbow point values, we see that the best classifications were achieved when that point takes a value from 44 to the last 28 flight hours.

Figure 5. Elbow point detection performance.

Table 3. Elbow point detection by evaluating different point values.

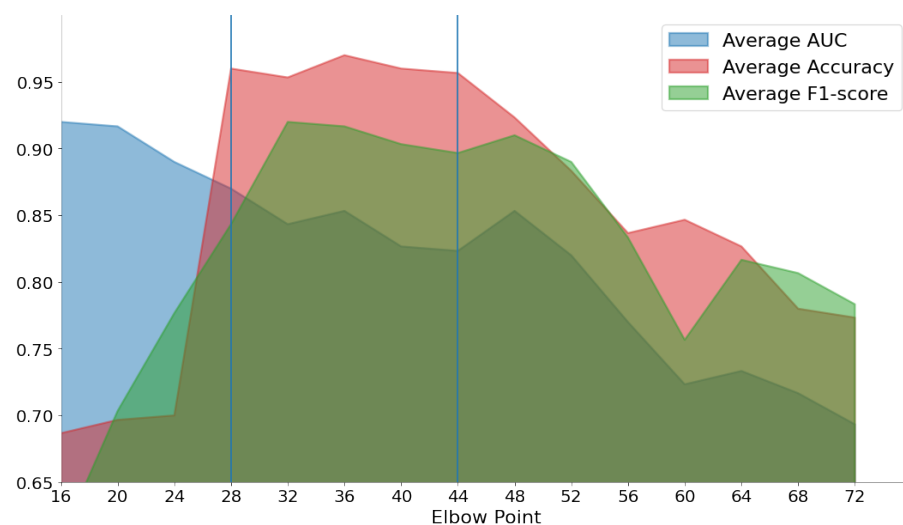| Point Point	RFC			SVC			MLP			Average		
	Accuracy	F1	AUC	Accuracy	F1	AUC	Accuracy	F1	AUC	Accurracy	F1-Score	AUC
72	0.59	0.54	0.72	0.72	0.7	0.77	0.77	0.76	0.83	0.77	0.78	0.69
68	0.64	0.57	0.73	0.75	0.72	0.81	0.76	0.74	0.85	0.78	0.8	0.72
64	0.65	0.57	0.74	0.77	0.74	0.86	0.78	0.76	0.88	0.83	0.82	0.73
60	0.56	0.5	0.77	0.80	0.76	0.88	0.81	0.78	0.89	0.85	0.76	0.72
56	0.66	0.54	0.75	0.82	0.79	0.86	0.83	0.8	0.90	0.84	0.83	0.77
52	0.76	0.67	0.82	0.85	0.82	0.92	0.85	0.82	0.91	0.88	0.89	0.82
48	0.82	0.76	0.87	0.87	0.83	0.95	0.87	0.83	0.95	0.92	0.91	0.85
44	0.86	0.8	0.92	0.82	0.73	0.99	0.79	0.68	0.96	<u>0.96</u>	<u>0.89</u>	<u>0.82</u>
40	0.84	0.75	0.91	0.82	0.7	1.00	0.82	0.7	0.97	<u>0.96</u>	<u>0.90</u>	<u>0.83</u>
36	0.91	0.86	0.95	0.82	0.67	1.00	0.83	0.7	0.96	<u>0.97</u>	<u>0.92</u>	<u>0.85</u>
32	0.84	0.69	0.90	0.84	0.68	0.99	0.85	0.7	0.97	<u>0.95</u>	<u>0.92</u>	<u>0.84</u>
28	0.87	0.72	0.92	0.87	0.7	0.99	0.87	0.87	0.97	<u>0.96</u>	<u>0.84</u>	<u>0.87</u>
24	0.95	0.91	0.97	0.89	0.73	1.00	0.83	0.45	0.13	0.70	0.78	0.89
20	0.98	0.96	0.96	0.95	0.91	1.00	0.86	0.46	0.13	0.69	0.7	0.92
16	0.98	0.96	0.99	0.89	0.47	0.94	0.89	0.47	0.13	0.68	0.6	0.92

7. RUL Estimation

Predictive maintenance aims to estimate the equipment lifetime, usually since it started to be monitored. Detecting an abnormal degradation stage when a aircraft achieves its 44 last operating hours gives technicians time to minimize Aircraft On Ground (AOG) times. The worst-case could be detecting the abnormal degradation when the system is very close to failing. According to our piecewise degradation function, defined by Equation (7), large values for the elbow point are preferred in our case. While the imperceptible degradation stage takes a RUL equal to the elbow point, the abnormal degradation stage takes a RUL similar to Equation (1), where the floor and ceiling operators are limited to RUL $\in \mathbb{N}$. To use this signal as input for the ML algorithm, its range was also reduced using the Z-score normalization process.

$$f(t) = \begin{cases} elbowPoint & if \quad t \leq elbowPoint \\ \lfloor EndOfLife - t \rceil \end{cases} \tag{7}$$

A multi-layer perceptron algorithm demonstrated to be useful to estimate the RUL. From the list of hyperparameters detailed in Table 4, the options in bold were demonstrated to reduce the estimation error.

Table 4. Remaining useful life hyperparameter tuning.

Algorithm	Hyperparameter	Search Range
	Neuron node	<u>Dense</u>
	Hidden Layers	**2**, 3
	Nodes per 1st layer	2, 3, 4, 5, 6, <u>7</u>, 8, 9
	Nodes per 2nd layer	2, 3, <u>4</u>, 5, 6, 7, 8, 9
Multi-Layer Perceptron	Optimizer	Adam, <u>SGD</u>
	Activation	Relu, <u>Tanh</u>
	Output Activation	<u>Relu</u>, Sigmoid
	Loss	<u>RMSE</u>
	Step-size	0.001, <u>0.01</u>

The loss function used to train the model and evaluate its achieved performance corresponds to the Root Mean Squared Error (RMSE) metric:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(RUL_i - R\hat{U}L_i)^2}$$

where, m is the number of measurements, RUL is the ground truth for the measure i and $R\hat{U}L_i$ is the remaining useful life inferred. Given that the sampling rate of 1 Hz in data collection implies the analysis of a large m, we considered evaluating four downsampling rate configurations, namely: $\frac{1}{1800}$ Hz, $\frac{1}{720}$ Hz, $\frac{1}{360}$ Hz and $\frac{1}{240}$ Hz.

Our experiments demonstrate that a down sampling rate of $\frac{1}{360}$ Hz is able to significantly reduce the time on the training process and the estimation error. We conclude that after obtaining 10 models by fixing the elbow point at 44, and then, calculating the mean and the standard deviation for the RMSE. The best performance was achiever by the down sampling rate $\frac{1}{360}$ Hz, which obtained a RMSE of 9.99 ± 1.29, while the other down sampling rates presented significant larger performances.

Implementing an *early stopping* strategy to prevent overfitting and reduce the training time, we were able to train the RUL prediction model in less than 2 min. That strategy considered training the model with a maximum of 580 epochs and evaluating the minimization on the loss function at 10 epochs.

After configuring the early stopping strategy and defining an adequate down sampling rate, we evaluated the influence of the elbow point, the number of nodes of hidden layers and the activation function used in the output layer. In Figure 6, we can observe that the elbow point and the activation function (0: sigmoid and 1: linear) really compromise the estimation error. For each combination of the factors analyzed, 10 models were obtained to evaluate them by comparing confidence intervals.

Figure 6. Hyperparameter Tuning of the RUL estimation

We considered evaluating three values for the elbow point, namely: 72, 48, and 36. Those values correspond to the assumptions that the variance of failures may be predicted 3, 2 and 1.5 operating days before experiencing a critical failure. In Figure 6, we can observe that RUL predictions are accurate when elbow point is set into the last operating flight hours. Predicting the health status into a large failure degradation implies increasing the estimation error. However, there are cases in which early predictions are preferred.

Observing the influence of the activation function of the output layer, we are able to notice that the linear activation function on the output layer aims to present accurate results. At the other hand, observing the influence of number of nodes of the layers, we were able to identify the more accurate predictions when the first and second hidden layers have 7 and 4 nodes, respectively.

8. Results and Discussion

The PHM program described in Section 3 has illustrated the construction path of physics-based and data-driven approaches. In case of data transformations from time or frequency domains not being appropriate to identify health degradation patterns directly from data, our PHM program suggests constructing physics-based Health Indicators and using them as inputs of data-driven approaches.

To evaluate the performance gaining by construction the time-frequency HIs with the proposed approach, we decided to train RUL estimators using raw data and using the HIs. In Table 5, we observe the performance obtained with three different elbow points. Recall that the health stage division allowed the construction of a piece-wise degradation, in which the RUL lineally decrease into the last flight hours. The elbow points presented in Table 5 correspond to the points that better divided the proposed two health stages.

To choose the better elbow point, we calculated the mean and the standard deviation of the RMSE after getting 10 RUL predictors per point. All the models were trained using the same downsampling rate, degradation function, Z-score normalization, MLP hyperparameters and early stopping settings, but differing in input data. Results presented fewer prediction errors by setting a linear degradation within the last 48 flight hours.

Table 5. RUL estimation using raw data vs. HI construction.

Elbow Point	Raw Data		HI Construction	
	Mean RMSE	Std. RMSE	Mean RMSE	Std. RMSE
72	20.790	0.352	17.111	2.924
48	14.019	1.248	8.623	0.774
36	9.384	0.858	6.480	1.630

In Figure 7, the RUL estimation of the first CU of the aircraft `enwslczm` is compared by using different input, namely: (a) raw data and (b) HIs constructed from the time-frequency domain. In (a), we can observe that the RUL estimated curve (light blue curve) oscillates around the value defined in the imperceptible health stage (blue curve). However, fails in predicting the RUL in the abnormal health stage. On the other hand, in (b), the HI construction seems to solve that problem because the RUL estimated curve (light blue curve) follows a similar tendency into the last flight hours.
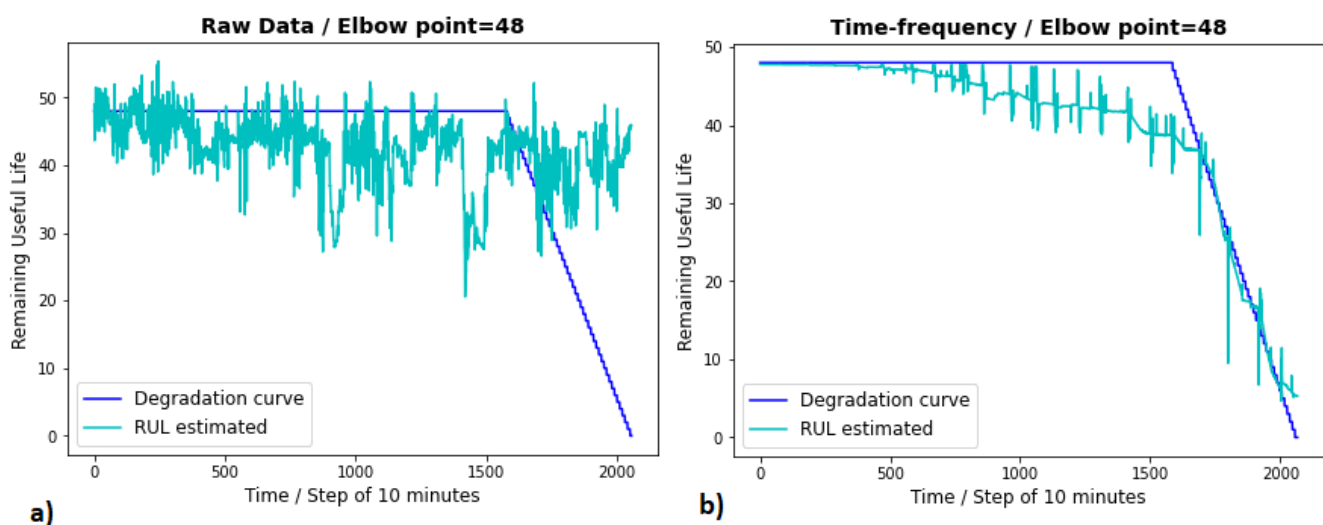

Figure 7. RUL of first CU of `aircraft=enwslczm` with an Elbow Point = 28: (**a**) raw data vs. (**b**) Health Indicators from time-frequency domain.

Comparing the results of applying HI construction versus using raw data trained with an *elbow point* of 48, we notice that (b) improved the predictions of (b). While the raw data approach obtained a RMSE of 14 ± 1.24, the HI construction obtained a RMSE of 8.623 ± 0.77. These performance gains are mainly related to RUL estimations on the abnormal degradation stage, where the estimations of (b) are closer to the proposed degradation curve than (a).

9. Conclusions and Future Work

The Health Indicators constructed from the time-frequency domain were demonstrated to be useful to describe the degradation of the cooling units of the wide body aircraft. The construction of HIs was not only useful to predict the Remaining Useful Life of aerial subsystems, but was also useful to identify two health stages of CUs, namely: an imperceptible degradation stage and a abnormal degradation stage.

After estimating the RUL of aerial systems, we believe that our HI construction approach could be used when machinery degradation may be related to the presence of periods with the lowest power spectrum. In general, we believe that our work can be easily replicated. While our HI construction approach applies the well-known Empirical Mode Decomposition (EMD) method used to obtain Intrinsic Mode Functions (IMFs), our prognostics program should be adopted in the presence of stationary and non-stationary signals. Before reusing our approach in other PHM applications, we recall that is important to evaluate the development of physics-based and data-driven models separately. The development of hybrid approaches is recommended when individual approaches do not satisfy the PMH practitioner's expectations. In the near future, we are interested in providing our estimations to aerospace maintenance technicians in order to evaluate the potential of our approach in practice.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
CBM	Condition-Based Maintenance
CM	Condition Monitoring
CS	Cooling System
CU	Cooling Unit
EMD	Empirical Mode Decomposition
FF	Fast-Fourier
HI	Health Indicator
IMF	Intrinsic Mode Function
ML	Machine Learning
MLP	Multi-Layer Perceptron
PHM	Prognostics and Health Maintenance
PM	Predictive Maintenance

RFC	Random Forest Classifier
RUL	Remaining Useful Life
STFT	Short-Time Fourier Transform
SVC	Support Vector Classifier
WVD	Wigner-Ville Distribution

References

1. Basora, L.; Bry, P.; Olive, X.; Freeman, F. Aircraft Fleet Health Monitoring with Anomaly Detection Techniques. *Aerospace* **2021**, *8*, 103. [CrossRef]
2. Chen, X.; Yu, J.; Tang, D.; Wang, Y. Remaining useful life prognostic estimation for aircraft subsystems or components: A review. In Proceedings of the 10th International Conference on Electronic Measurement and Instruments, Chengdu, China, 16–19 August 2011.
3. Adhikari, P.; Rao, H.G.; Buderath, M. Machine Learning based Data Driven Diagnostics & Prognostics Framework for Aircraft Predictive Maintenance. In Proceedings of the 10th International Symposium on NDT in Aerospace, Dresden, Germany, 24–26 October 2018.
4. Javed, K.; Gouriveau, R.; Zemouri, R.; Zerhouni, N. Features Selection Procedure for Prognostics: An Approach Based on Predictability. *IFAC Proc. Vol.* **2012**, *45*, 25–30. [CrossRef]
5. Patrick, R.; Smith, M.; Zhang, B.; Byington, C.; Vachtsevanos, G.; Del Rosario, R. Diagnostic enhancements for air vehicle HUMS to increase prognostic system effectiveness. In Proceedings of the 2009 IEEE Aerospace Conference, Big Sky, MT, USA, 7–14 March 2009.
6. Atamuradov, V.; Medjaher, K.; Dersin, P.; Lamoureux, B. Prognostics and Health Management for Maintenance Practitioners-Review, Implementation and Tools Evaluation. *Int. J. Progn. Health Manag.* **2017**, *12*, 31–62. [CrossRef]
7. Biggio, L.; Kastanis, I. Prognostics and Health Management of Industrial Assets: Current Progress and Road Ahead. *Front. Artif. Intell.* **2020**, *3*, 88. [CrossRef] [PubMed]
8. Saxena, A.; Celaya, J.; Balaban, E.; Goebel, K.; Saha, B.; Saha, S.; Schwabacher, M. Metrics for evaluating performance of prognostic techniques. In Proceedings of the International Conference on Prognostics and Health Management, Denver, CO, USA, 6–9 October 2008.
9. Jones, D.; Snider, C.; Nassehi, A.; Yon, J.; Hicks, B. Characterising the Digital Twin: A systematic literature review. *CIRP J. Manuf. Sci. Technol.* **2020**, *29*, 36–52. [CrossRef]
10. Titchmarsh, E.C. *Introduction to the Theory of Fourier Integrals*; Oxford University Press: Oxford, UK, 1938; Volume 141, p. 183.
11. Arias, C. M.; Kulkarni, C.; Goebel, K.; Fink, O. Aircraft Engine Run-to-Failure Dataset under Real Flight Conditions for Prognostics and Diagnostics. *Data* **2021**, *6*, 5. [CrossRef]
12. Kan, M.S.; Tan, A.C.C.; Mathew, J. A review on prognostic techniques for non-stationary and non-linear rotating systems. *Eur. J. Oper. Res.* **2015**, *62–63*, 1–20. [CrossRef]
13. Huang, N.E.; Zheng, S.; Steven, R.; Manli, C.W.; Hsing, H.S.; Quanan, Z.; Nai-Chyuan, Y.; Chi, C.T.; Liu H.H. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. *Proc. R. Soc. London Ser. A Math. Phys. Eng. Sci.* **1971**, *03*, 903–995. [CrossRef]
14. Matz, G.; Hlawatsch, F. Wigner distributions (nearly) everywhere: Time-frequency analysis of signals, systems, random processes, signal spaces, and frames. *Signal Process.* **2003**, *83*, 1355–1378. [CrossRef]
15. Lei, Y.; Li, N.; Guo, L.; Li, N.; Yan, T.; Lin, J. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mech. Syst. Signal Process.* **2018**, *104*, 799–834. [CrossRef]
16. Azevedo, D.; Cardoso, A.; Ribeiro, B. Estimation of Health Indicators using Advanced Analytics for Prediction of Aircraft Systems Remaining Useful Lifetime. *Int. J. Progn. Health Manag.* **2020**, *5*, 1–10.
17. Baptista, M.L.; Henriques E.M.P.; Goebel, K. More effective prognostics with elbow point detection and deep learning. *Mech. Syst. Signal Process.* **2021**, *146*, 106987. [CrossRef]
18. David, H. Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen. In *Integralgleichungen und Gleichungen mit unendlich vielen Unbekannten*; Vieweg + Teubner Verlag, Ed.; Publishing House: Wiesbaden, Germany, 1989; pp. 8–171.
19. Wang, G.; Chen, X.; Qiao, F.; Wu, Z.; Huang, N.E. On Intrinsic Mode Function. *Adv. Adapt. Data Anal.* **2011**, *2*, 277–293. [CrossRef]
20. Jardine, A.K.S.; Lin, D.; Banjevi, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [CrossRef]
21. Sikorskaab, J.Z.; Hodkiewiczb, M.; Ma, L. Prognostic modelling options for remaining useful life estimation by industry. *Mech. Syst. Signal Process.* **2011**, *25*, 1803–1836. [CrossRef]
22. Si, X.S.; Wang, W.; Hua, C.H.; Zhou, D.H. Remaining useful life estimation—A review on the statistical data driven approaches. *Eur. J. Oper. Res.* **2011**, *16*, 1–14. [CrossRef]
23. Basora, L.; Olive, J.; Dubot, T. Recent Advances in Anomaly Detection Methods Applied to Aviation. *Aerospace* **2019**, *6*, 117. [CrossRef]
24. Heng, A.; Zhang, S.; Tan, C.C.A.; Mathew, J. Rotating machinery prognostics: State of the art, challenges and opportunities. *Mech. Syst. Signal Process.* **2009**, *23*, 724–739. [CrossRef]

25. Bieber, M.; Verhagen W.J.C. ; Santos, B.F. An Adaptive Framework for Remaining Useful Life Predictions of Aircraft Systems. *Int. J. Progn. Health Manag.* **2021**, *6*, 60–70.
26. Tracer, B.; Loughlin, P.J. What are the time-frequency moments of a signal? In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, Atlanta, GA, USA, 9 May 1996.
27. Seif, H.; Duff, L.; Guy, P.; Laurent, S.; Rachid, G.; Ghozlen, H.B.M. Hilbert-Huang Transform versus Fourier Based Analysis for Diffused Ultrasonic Waves Structural Health Monitoring in Polymer Based Composite Materials. *PU-S05: Structural Health Monitoring.* **2012**, *1*, 2417–2422.
28. Zhang, B.; Zhang, L.; Xu, J. Degradation Feature Selection for Remaining Useful Life Prediction of Rolling Element Bearings. *Qual. Reliab. Eng. Int.* **2015**, *32*, 547–554. [CrossRef]
29. Scott, A.C. *Nonlinear Biology*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 181–276.
30. Cao, J.-L.; Chen, G.-H. Introduction to the Theory of Fourier Integrals. *Shanghai Univ.* **2003**, *7*, 265–269. [CrossRef]
31. Khalil, S.; Khalil, T.; Nasreen, S. A Survey Of Feature Selection And Feature Extraction Techniques In Machine Learning. In Proceedings of the Science and Information Conference, London, UK, 27–29 August 2014.
32. Hu, Z.; Baoa, Y.; Xionga, T.; Chiongb, R. Hybrid filter–wrapper feature selection for short-term load forecasting. *Eng. Appl. Artif. Intell.* **2015**, *40*, 17–27. [CrossRef]
33. H2020 Remap. Available online: https://h2020-remap.eu/ (accessed on 3 March 2022).
34. Llasag Rosero, R.; Silva, C.; Ribeiro, B. Remaining Useful Life Estimation in Aircraft Components with Federated Learning. *Int. J. Progn. Health Manag.* **2020**, *5*, 1–9.
35. Li, X.; Ding, Q.; Sun, J. Q. Remaining Useful Life Estimation in Prognostics Using Deep Convolution Neural Networks. *Reliab. Eng. Syst. Saf.* **2018**, *172*, 88. [CrossRef]
36. Nguyen, T.P.K.; Amor, K.; Medjaher, K.; Picot, A.; Maussion, P.; Tobon, D.; Chauchat, B.; Cheron, R. Analysis and comparison of multiple features for fault detection and prognostic in ball bearings. *Int. J. Progn. Health Manag.* **2018**, *4*, 1.
37. Kehtarnavaz Nasser, L. DSP System Design: Cochlear Implant Simulator. In *Digital Signal Processing System Design*, 2nd ed.; Academic Press: Burlington, VT, USA, 2008; pp. 303–320. Available online: https://www.sciencedirect.com/book/9780123744906/digital-signal-processing-system-design (accessed on 3 March 2022).
38. Hu, C.; Youn, B.D.; Wang, P.; Joung, T.Y. Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. *Reliab. Eng. Syst. Saf.* **2012**, *103*, 120–135. [CrossRef]