

A Case for the Use of Data-driven Methods in Gas Turbine Prognostics

Marcia Baptista¹, Cairo L. Nascimento Jr.², Helmut Prendinger³, and Elsa Henriques⁴

^{1,4} *Instituto Superior Tecnico, Universidade de Lisboa, Lisbon, 1049-001, Portugal*
marcia.baptista@ist.utl.pt, elsa.h@ist.utl.pt

² *Instituto Tecnologico de Aeronautica (ITA), 12228-900, Sao Jose dos Campos, SP, Brazil*
cairo@ita.br

³ *National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan*
helmut@nii.ac.jp

ABSTRACT

The goal of data-driven methods is to remove dependence on classical models of structured expert judgment and draw insights to causal relationships directly from the data. This paper investigates the potential of using data-driven methods, namely uni-variate multiple linear regression, k-nearest neighbors, feed-forward neural networks, random forests and linear support vector regression to predict the end of life (EOL) and remaining useful life (RUL) of engineering systems. The algorithms are demonstrated on a real-world large-scale dataset consisting of a multidimensional time series of health monitoring indicators collected from a set of commercial aircraft gas turbine engines. A stratified version of 10-fold cross-validation is used to compare the prognostics performance of the five prognostics models. An experience-based Weibull model is chosen as the baseline method. Models are evaluated according to established metrics in the field including median absolute error, median absolute deviation and relative accuracy. The prediction results indicate that support vector regression and random forests are the most accurate models. Neural networks and k-nearest neighbors also show improved forecast skill compared to the baseline model while beating the more traditional technique of linear regression. In regards to error spread, results are not as expressive even though all the selected data-driven methods provide good results, outperforming the baseline.

1. INTRODUCTION

The importance and validity of incorporating health monitoring (HM) signals (e.g. vibration, noise, and temperature mea-

surements) to improve the prognostics of engineering systems, such as gas turbine engines, are well documented. Examples of such studies include, to name a few, the work of DePold and Gass (1998), who were among the first to present an expert system for gas turbine end of life (EOL) and remaining useful life (RUL) estimation based on HM time series, as well as the work of Brotherton, Jahns, Jacobs, and Wroblewski (2000) who applied neural nets and rule extractors to automatically discover prognostics indicators from gas turbine HM signals, and more recently, Zaidan, Mills, Harrison, and Fleming (2016) who used in-service gas turbine engine data to demonstrate the value of using fleet-specific HM data to better estimate degradation trajectories.

Various modeling approaches have been used to extract relevant information from HM time series in the field of gas turbine prognostics. For instance, Li and Nilkitsaranont (2009) proposed a combined linear and quadratic model with considerable success. These kind of models have the advantage of being simple to understand as well as being cheap from the computational standpoint. However, to incorporate the non-linear dynamics of HM indicators, it is often needed to resort to more advanced machine learning methods such as artificial neural networks or support vector regression. Here, it is of note the work of Xue et al. (2008) who use a nearest neighbor approach to aircraft gas turbine prognostics and the work of Zaidan et al. (2015; 2015; 2016) who focus on Bayesian approaches to jet engine EOL and RUL estimation.

Despite the existing literature on data-driven models for gas turbine prognostics, a comprehensive study comparing the several approaches seems to be lacking. The aim of this study is hence to investigate the potential of using different data-driven, i.e. purely empirical techniques, to forecast the overhauls of a modern jet engine using HM indicators. The

Marcia Baptista et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

selected models include a univariate multiple linear regression (LR) model, a feed-forward neural network (NN), random forests (RF), and linear support vector regression (SVR). The primary purpose of this study is to examine the ability of the chosen data-driven methods to capture the relationships between the HM series and the maintenance events of the jet engines. Concretely, we aim to reinforce the notion that data-driven methods based on HM series can be considered a robust modeling alternative to conventional prognostics models, such as experience-based methods.

The remainder of this paper is organized as follows. We start in Section 2 with related work and a description of the case study in Section 3. In Section 4 we describe the methodology and in Section 5 we compare experimentally the selected data-driven methods to an experience-based Weibull baseline model. Section 6 concludes the paper.

2. THEORY AND RELATED WORK

Modern day maintenance programs in aeronautics are starting to be based on the *on-condition* concept, where maintenance activities occur when the equipment condition demands it (Jardine, Lin, & Banjevic, 2006). The idea is that if equipment can be evaluated while still in service, the overall cost of maintenance goes down. On-condition maintenance reduces the need for prescribed “hard-time” intervals but requires routine monitoring of performance parameters of the equipment such as the temperature, pressure, vibration, fuel flow, oil consumption, and rotor speed. Changes in any of these parameters beyond specified limits can warrant a maintenance intervention.

A side effect of on-condition maintenance is greater reliance on statistical analysis and machine learning methods to predict the frequency and timing of maintenance events and their corresponding costs. These methods aim to replace the experience-based models of traditional preventive maintenance. Here, by *experience-based* methods we mean models that use life usage data gathered during a significant period of time to adjust the parameters of reliability models such as exponential and Weibull (Tobon-Mejia, Medjaher, Zerhouni, & Tripot, 2012). Historically, the most used experience-based model is the Weibull three-parameter (α , β , γ) probability density function (PDF) (Weibull, 1951), which can be formally defined as

$$f_T(t) = \begin{cases} \frac{\beta}{\alpha} \left(\frac{t-\gamma}{\alpha}\right)^{\beta-1} e^{-\left(\frac{t-\gamma}{\alpha}\right)^\beta} & t \geq 0, \\ 0 & t < 0, \end{cases} \quad (1)$$

where α is the scale parameter (or characteristic life), β is the shape parameter (or slope), and γ is the location parameter (or failure free time).

Figure 1 illustrates a typical experience-based model with

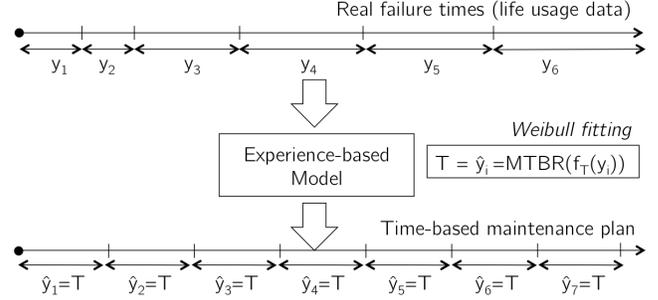


Figure 1. Experience-based model. Hard-time intervals (T) expressed in the second timeline are estimated from the life usage data ($\{y_i\}_{i=1}^n$) in the first timeline.

the Weibull distribution. As shown, the set of removal times is used to derive the mean time between repairs. In reliability models based on the Weibull distribution the mean time between repairs (MTBR) is computed as

$$MTBR = \gamma + \alpha \Gamma\left(\frac{1}{\beta} + 1\right) \quad (2)$$

where $\Gamma()$ is the gamma function defined by

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx.$$

The time prescribed between two preventive maintenance actions, \hat{y}_i , is computed from the mean time between repairs (MTBR) according to the criticality and risk of the equipment such that

$$\hat{y}_i = \delta MTBR \quad (3)$$

where $\delta \in [0, 1]$ measures the degree of criticality and risk of the equipment

Several authors apply experience-based methods to model jet engine reliability. For instance, Stranjak, Dutta, Ebden, Rogers, and Vytelingum (2008) propose a model where engine reliability is determined by combining individual component distributions, approximated by the Weibull function. Here, the whole-engine reliability is a function of the individual reliability of the most critical modules of the engine. A finite mixture model is used to capture the combination of individual performances.

Ebden, Stranjak, and Roberts (2010) also propose a finite mixture model to describe jet engine failure modes. Here, the mixed Weibull distribution is estimated from a large data set of around 300 jet engines. Estimation is subject to censoring at various times. Parametric uncertainty is derived analytically from the inverse Fisher information matrix and is mapped visually onto the functions of use in reliability theory such as the hazard function and survival function.

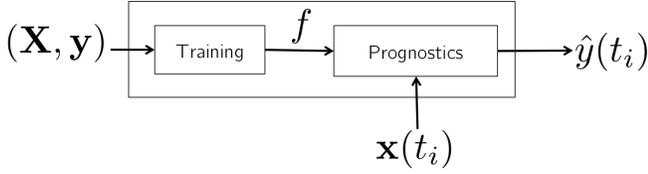


Figure 2. Direct data-driven prognostics architecture. Data-driven prognostics proceed in two steps: a training stage where data are used to create a learning model f and a prediction stage where model f is applied to new observations to generate EOL or RUL predictions.

An alternative to experience-based methods are on-condition *data-driven* models (Schwabacher, 2005). In these approaches, statistical and machine learning techniques are used on large sets of performance and degradation data to forecast the equipment future state. The most direct data-driven methods divide the on-condition prognostic problem into two sequential stages: i) a training stage, in which, a model f is fit to past observations (\mathbf{X}, \mathbf{y}) such that

$$\mathbf{X}_{m,n} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{pmatrix}, \mathbf{y}_{m,1} = \begin{pmatrix} y_{1,1} \\ y_{2,1} \\ \vdots \\ y_{m,1} \end{pmatrix} \quad (4)$$

where $\mathbf{X}_{m,n}$ denotes the matrix of dependent (covariates) variables and $\mathbf{y}_{m,1}$ the vector of independent (target) variables such as the equipment EOL/RUL

and an ii) estimation stage in which, using the model f , the EOL or the RUL of the system is determined based on new observations as

$$\hat{y}(t) = f(t, \mathbf{x}(t), \boldsymbol{\theta}(t)) \quad (5)$$

where $t \in \mathcal{R}$ is the discrete time variable, $\mathbf{x}(t) \in \mathbb{R}^n$ is the input vector, $\boldsymbol{\theta}(t) \in \mathbb{R}^{n_\theta}$ is the parameter vector, f is the model's output function, and $\hat{y}(t) \in \mathbb{R}^{n_y}$ is the output target (EOL or RUL).

This data-driven architecture is shown in Figure 2. In the first procedural step the system is provided with inputs $\mathbf{X}(t)$ and corresponding measured outputs $y(t)$. With these data and the parameter vector $\boldsymbol{\theta}(t)$ the system is able to estimate function f . After this, the prognostics module determines the EOL/RUL estimate, $\hat{y}(t_i)$, based on a set of new observations $\mathbf{x}(t_i)$. It is important to note that the modeled function f is only trained on a subset of the observations with model evaluation being performed on a test set of out-of-model observations. The 10-fold cross-validation evaluation scheme is usually used at this stage.

A range of data-driven methods has been applied to prog-

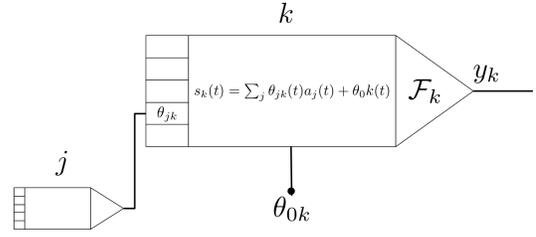


Figure 3. The basic components of an artificial neural network. The propagation rule used here is the 'standard' weighted summation.

nostics, from multivariate statistical methods to neural networks and Markovian processes (Baptista et al., 2016a, 2017; Schwabacher, 2005). With respect to statistical methods, linear regression methods have been investigated for gas turbine prognostics in a number of works (Li & Nilkitsaranont, 2009; Weckman, Marvel, & Shell, 2006). In their simplest form, regression for RUL estimation works by fitting the available data on component degradation (\mathbf{X}, \mathbf{y}) and then by extrapolating the equipment evolution up to failure as

$$\hat{y}(t) = f(t, \mathbf{x}(t), \boldsymbol{\theta}(t)) = \mathbf{x}(t)\boldsymbol{\theta} + \theta_0 \quad (6)$$

where the parameter vector $\boldsymbol{\theta}$ and θ_0 are obtained from fitting the training data \mathbf{X} and output vector \mathbf{y} to a linear function f using an optimization algorithm such as gradient descent.

The main advantage of methods such as linear or quadratic regression is that their corresponding function f can be plotted in the coordinate plane. This allows to more easily study the correlations between the input $\mathbf{x}(t)$ and the output $\hat{y}(t)$. These statistical models have however the disadvantage of being unable in many situations to capture the non-linear dynamics of HM indicators (Riad, Elminir, & Elattar, 2010).

With respect to artificial intelligence techniques, artificial neural networks is one of the most popular approach to damage estimation (Di Maio & Zio, 2013). A NN consists of a pool of simple processing units which communicate by sending signals to each other over a large number of weighted connections (Kröse, Krose, van der Smagt, & Smagt, 1993). Each unit performs a relatively simple function: receive input from its neighbors or external sources and use these to compute and propagate a signal to other units. Figure 3 schematizes a typical neuron architecture. In its simplest form, the total input to a neuron unit k is the weighted sum of the separate outputs $a_j(t)$ from each of the connected units plus a bias term θ_{0k} :

$$s_k(t) = \sum_j \theta_{jk}(t) a_j(t) + \theta_{0k}(t) \quad (7)$$

where θ_{jk} is the weight of neuron j on unit k

The output of a neuron unit is computed from an activation function \mathcal{F} (e.g sigmoid function), which determines the level of activation based on the input $s_k(t)$ and the current level of activation $y_k(t)$ of the unit

$$y_k(t+1) = \mathcal{F}_k(y_k(t), s_k(t)) \quad (8)$$

Within neural systems three types of units exist: input units which receive data from outside the network, output units which send data out of the network, and hidden units whose input and output signals are transferred within the network. In the case of prognostics, each input a_j of the neural network corresponds to a signal from the input vector $\mathbf{x}(t)$ and the output unit corresponds to $\hat{y}(t)$. The parameters $\theta(t)$ of the neural network f are learned from a training algorithm that attempts to minimize a cost function dependent on the training data (\mathbf{X}, \mathbf{y}) .

Other machine learning methods are starting to be applied to EOL and RUL estimation. For instance, Huang, Wang, Li, Zhang, and Liu (2015) argue that support vector machines (SVM) represent promising approaches to prognostics. In support vector regression (SVR) (Drucker, Burges, Kaufman, A, & Vapnik, 1997) a set of training data (\mathbf{X}, \mathbf{y}) is assumed while the goal is to define f as a function that has at most a ε deviation from targets \mathbf{y} in the training data. For a linear SVR, this function takes a form similar to that of linear regression:

$$\hat{y}(t) = f(t, \mathbf{x}(t), \boldsymbol{\theta}(t)) = \mathbf{x}(t)\boldsymbol{\theta} + \theta_0 \quad (9)$$

A requirement of this function is its *flatness*, which means a small $\boldsymbol{\theta}$. This small w can be obtained by minimizing the norm, $\|\boldsymbol{\theta}\|^2 = \langle \boldsymbol{\theta}, \boldsymbol{\theta} \rangle$, as a convex optimization problem:

$$\text{minimize } \frac{1}{2} \|\boldsymbol{\theta}\|^2 C \left(\sum_i \xi_i + \xi_i^* \right) \quad (10)$$

where ξ_i and ξ_i^* are relaxation factors and C is a penalty factor

The applications of SVR to RUL prediction are still mostly restricted to the prognosis of bearings and batteries (Huang et al., 2015). An effort to apply this technique to prognostics in aeronautics is the work of Baptista et al. (2016b). In this study, SVR is shown to be an efficient method for engine bleed valve prognostics.

Other techniques which could benefit from further exploration are tree-based methods. In these methods the predictor space is stratified or segmented into a number of simple regions whose splitting rules can be summarized in a tree. The more popular tree-based methods grow multiple trees which are then combined to yield a single consensus prediction. Formally, such a tree-based approach assumes the existence of f

such that

$$\hat{y}(t) = f(t, \mathbf{x}(t), \boldsymbol{\theta}(t)) = \frac{1}{B} \sum_{b=1}^B f_b(\mathbf{x}(t), \boldsymbol{\theta}(t)) \quad (11)$$

where B is the number of decision or regression trees f_b trained on a sample b of (\mathbf{X}, \mathbf{y}) .

Another approach which has yielded promising recent results is instance-based regression (Xue et al., 2008; Khelif, Malinowski, Chebel-Morello, & Zerhouni, 2014). In instance-based learning, new problem instances $(\mathbf{x}(t))$ are compared with instances seen in training (\mathbf{X}, \mathbf{y}) using similarity-distance metrics.

Besides the above referred techniques, a large number of machine learning methods can be found in the literature (Schwabacher & Goebel, 2007). As an exhaustive comparison of all these methods is not feasible we selected the techniques that showed the most promise for prognostics (Schwabacher & Goebel, 2007) or that are considered among the most representative algorithms for machine learning (Wu et al., 2008).

3. CASE STUDY

In this section, the case study is introduced by describing its background and data.

3.1. Background

Engine prognostics or the forecasting of engine degradation which supports the on-condition maintenance is and will continue to be a challenging task. This follows mostly due to the uncertainty associated with gas turbine design, environmental and operating conditions. This study aims to explore the incorporation of statistical information to develop a data-driven approach to EOL and RUL estimation of modern aircraft gas turbine engines. To show the effectiveness of our approach we present a field application. In the application, the varying degradation conditions and maintenance actions that happen to the engines are considered.

3.2. Data

Common to data-driven approaches is the modeling of the desired response variable using large volumes of historical data (Goebel, Saha, & Saxena, 2008). In this case study, our data set describes the evolution of performance of a set of commercial jet engines between approximately ten years in different intervals of time for each engine. Concretely, the data consists of a cross-sectional time series in the sense that for each engine, and as exemplified in Figure 4, we have a multi-variate series that represents the temporal progression of the engine HM signals. In the figure, each horizontal line represents a possible health monitoring variable such as

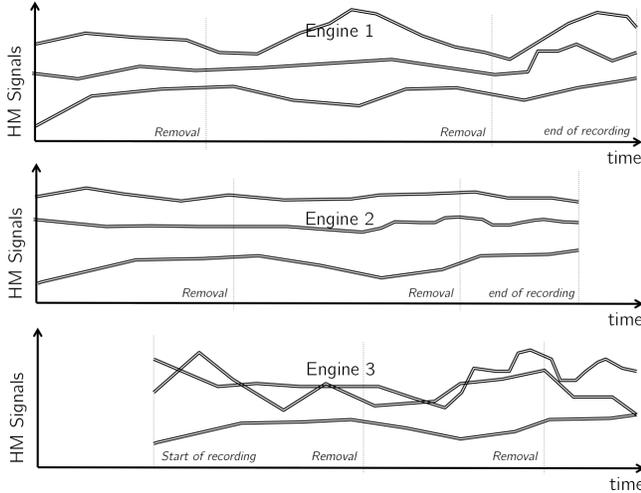


Figure 4. Performance data. For each engine, the evolution of performance is described using a set of engine health monitoring indicators such as temperature, vibration, rotor speed. There is also information about the engine maintenance events.

temperature, vibration, pressure. The vertical dashed lines represent possible failure/maintenance events. These signals are measured at three different flight phases: one measurement is taken at take-off, another at climb and three other at cruise. Overall, we analyze data of 1.8×10^6 flights (i.e cycles) (average of $3.35 \times 10^3 \pm 10^3$ cycles per engine). This adds up to around 3GB of raw data.

In addition to performance signals, there is also information about the engine overhauls. An engine overhaul can be defined as a comprehensive inspection that involves removing and disassembling the engine, testing all its sub-systems, cleaning and replacing parts as needed and then reassembling the engine (Seemann, Langhans, Schilling, & Gollnick, 2010). For a better sense of the data, please also consider the histogram of total removal time shown in Figure 5. In the plot, the data is skewed to the right, indicating that the mean (354 ± 285 days) is higher than the median (306 ± 285 days). This implies that short overhauls are more frequent than medium-to-long overhauls. The graph also illustrates how the empirical data sample is well fit to the theoretical Weibull distribution (black fitted line). This finding indicates that, the Weibull distribution is most likely the best fit for the experience-based model.

4. METHODOLOGY

In order to understand the reliability patterns of jet engines, we perform several experiments using field data, where each experiment analyzes a different data-driven method. This section describes the methodology followed to validate the main hypothesis of this study:

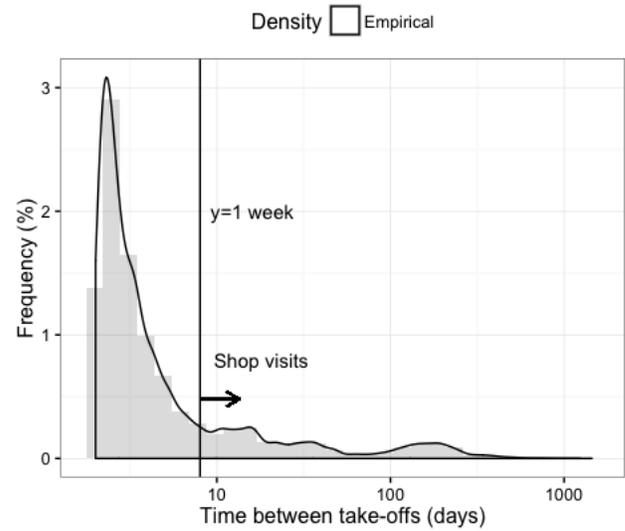


Figure 5. Data exploration. This study analyses performance level and full overhaul worksopes and ignores maintenance events lasting less than 1 week. The selected algorithms are used on a right-sided skewed data set of removals shown in the plot. As shown, the Weibull distribution is a good fit for the data.

H: Condition-based predictive data-driven models based on maintenance data and HM signals are able to outperform preventive experience-based models.

To investigate this hypothesis we compare two reliability approaches: (a) experience-based modeling based on the Weibull distribution (I) and (b) data-driven methods based on overhaul data and HM signals (II) We aim to show that despite the good fit of the Weibull model to our maintenance data, data-driven models can derive better estimates of the equipment EOL and RUL.

4.1. Experience-based modeling

The experience-based approach consists in using 10-fold cross-validation to measure the fit of the Weibull distribution to the data set of overhaul times. We chose cross-validation because this sampling technique has been well studied as a basis for measuring predictive accuracy (Kohavi et al., 1995). In the implemented cross-validation, the whole data set of overhaul times $\{y_i\}_{i=1}^n$ is randomly partitioned into k equal sized subsamples (i.e testing folds). For each fold k , a Weibull distribution (Equation 1) is fitted to the remaining $k - 1$ subsamples. From this fit, the hard-time interval T_k of fold k is computed with δ set to one (Equations 2 and 3). The performance of the k model is then computed using different

Table 1. Description of covariates.

Type	Covariate	Unit	Alias
Environmental	Altitude	Feet	ALT
	Outside Air Temperature	C	OAT
	Vibration Intercase	IPS	BB
Mechanical	Nominal Shaft Speeds	%	N1, N2, N3
	Marginal Shaft Speeds	%	N1_MW, N2_MW, N3_MW
	Shaft Vibrations	IPS	VB1, VB2, VB3
Performance	Delta pressure in several stations	%	DP160, DP20, DP25, DP3
	Delta temperatures in several stations	C	DT160, DT25, DT3
	Turbine Cooling Air Front and Rear	C	TCAF, TCAR
	Delta Fuel Flow	%	DFP
Systemic	Engine Pressure Ratio	Unit	EPR
	Oil Pressure	PSI	OIP
	Oil Temperature	C	OIT
	Margin in High-Pressure (HP) turbine temperature	C	TGT_MW
Others	Time since removal	Take-offs	TIME
	Number of past removals	Unit	N_REMOVALS

quantitative measures of fit such as the mean error

$$Ac_{f_{T_k}} = \frac{1}{v_k} \sum_{i=1}^{v_k} \varepsilon_{y_i, \hat{y}_i} = \frac{1}{v_k} \sum_{i=1}^{v_k} (y_i - \delta T_k) \quad (12)$$

where v_k is the number of testing observations in fold k and $\varepsilon_{y_i, \hat{y}_i}$ its the model bias given by the difference between the real value y_i and the predicted value \hat{y}_i of observation i . T_k is computed from $\gamma + \alpha \Gamma(\frac{1}{\beta} + 1)$ where α, γ and β are parameters of the Weibull distribution $f_{T_k}(t)$ fit to fold k

The results from the k folds can then be averaged to produce a single performance estimation

$$Ac_{f_T} = \frac{1}{k} \sum_{i=1}^k Ac_{f_{T_k}} \quad (13)$$

Algorithm 1 presents the pseudo-code of the approach.

4.2. Data-driven modeling

The data-driven approach (II) is based on maintenance data and HM signals of several aircraft engines. Five distinct models are constructed from state-of-the-art techniques: univariate multiple linear regression (LR), K-nearest neighbors (K-NN), feed-forward neural networks (NN), random forests (RF) and linear support vector regression (SVR). Here, the target of prediction is the estimation of the remaining time to an overhaul (\hat{y}_i) at the i^{th} flight (i.e. cycle).

Formally, we work on a set of past data (\mathbf{X}, \mathbf{y}) that consists of m observations. Each observation characterizes a flight for

a given engine – consisting of a set of n covariates as

$$\mathbf{X}_{m,n} = \begin{matrix} & \text{Cov}_1 & \text{Cov}_2 & \cdots & \text{Cov}_n \\ \text{Cycle}_1 & \left(\begin{matrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{matrix} \right) \\ \text{Cycle}_2 & & & & \\ \vdots & & & & \\ \text{Cycle}_m & & & & \end{matrix} \quad (14)$$

$$\mathbf{y}_{m,1} = \begin{matrix} & \text{RUL} \\ \text{Cycle}_1 & \left(\begin{matrix} y_{1,1} \\ y_{2,1} \\ \vdots \\ y_{m,1} \end{matrix} \right) \\ \text{Cycle}_2 & \\ \vdots & \\ \text{Cycle}_m & \end{matrix} \quad (15)$$

where $\mathbf{X}_{m,n}$ denotes the matrix of dependent (covariates) variables and $\mathbf{y}_{m,1}$ the vector of RUL values for each cycle in matrix \mathbf{X}

Table 1 presents the list of considered covariates. The selection of these covariates was steered by discussion with domain experts within limits of the available data. The predictors are grouped into five domains according to the type of effect exerted by the parameter on the engine: maintenance-related variables, environmental variables, mechanical effects related to moving engine parts (e.g rotation, vibration), performance effects related to the thermodynamics and fluid mechanics of the engine (e.g temperature, pressure), and systemic effects that affect the whole engine.

Before its use, tuple (\mathbf{X}, \mathbf{y}) was submitted to a pre-processing procedure. First, and before any data transformation, there was a cleaning stage: the techniques of Tukey's boxplot (Tukey, 1977) and Medcouple-based outlier detection method (Brys, Hubert, & Struyf, 2004) were used to detect overhauls extremely long or short which were consid-

Algorithm 1 Cross-validation of model *I***Precondition:** $n_k = 10$ represents the number of folds and y_i represents overhaul time at index i

```

1: function CROSS-VALIDATION( $n_k, \{y_i\}_{i=1}^n$ )
2:    $Ac_{f_T} \leftarrow 0$  ▷ Let performance be zero
3:   for  $k \leftarrow 1$  to  $n_k$  do
4:      $\text{trainSet}_k, \text{testSet}_k \leftarrow \text{PARTITION}(\{y_i\}_{i=1}^n, n_k, k)$  ▷ Randomly split  $\{y_i\}_{i=1}^n$  into  $n_k$  equal-sized bins
5:      $\alpha, \beta, \gamma \leftarrow \text{fitWeibull}(\text{trainSet}_k)$ 
6:      $T \leftarrow \gamma + \alpha\Gamma(\frac{1}{\beta} + 1)$ 
7:      $Ac_{f_T} \leftarrow Ac_{f_T} + \frac{1}{v_k} \sum_{i=1}^{v_k} (y_i - \delta T), y_i \in \text{testSet}_k$  ▷ Compute error rate using bias formula
8:   end for
9:   return  $\frac{1}{n_k} Ac_{f_T}$ 
10: end function

```

Algorithm 2 Cross-validation of model *II***Precondition:** $n_k = 10$ represents the number of folds and y_i represents cycle i

```

1: function CROSS-VALIDATION( $n_k, \{y_i\}_{i=1}^n$ )
2:    $Ac_f \leftarrow 0$  ▷ Let performance be zero
3:   for  $k \leftarrow 1$  to  $n_k$  do
4:      $\text{trainSet}_k, \text{testSet}_k \leftarrow \text{PARTITION}(\{y_i\}_{i=1}^n, n_k, k)$  ▷ Randomly split  $\{y_i\}_{i=1}^n$  into  $n_k$  equal-sized bins so that each removal is in a different bin
5:      $f(t, \mathbf{x}(t), \boldsymbol{\theta}(t)) \leftarrow \text{fit}(\text{trainSet}_k)$ 
6:      $Ac_{f_T} \leftarrow Ac_{f_T} + \frac{1}{v_k} \sum_{i=1}^{v_k} (y_i - \hat{y}_i), y_i \in \text{testSet}_k$  ▷ Compute error rate using bias formula and
        $\hat{y}_i = f(t_i, \mathbf{x}(t_i), \boldsymbol{\theta}(t_i))$ 
7:   end for
8:   return  $\frac{1}{n_k} Ac_f$ 
9: end function

```

ered as missing or incorrect data. The medcouple (MC) is a robust skewness measure that allows calculating the nominal data range:

$$[Q1ce^{a\text{MC}}\text{IQD}, Q3 + ce^{b\text{MC}}\text{IQD}], \text{MC} \geq 0 \quad (16)$$

$$[Q1ce^{a\text{MC}}\text{IQD}, Q3 + ce^{-a\text{MC}}\text{IQD}], \text{MC} < 0 \quad (17)$$

where Q1 and Q3 represent the lower and upper quartiles, respectively, of the data distribution, and $\text{IQD} = Q3 - Q1$ is the interquartile distance, a measure of the spread of the data similar to the standard deviation. The rule threshold parameter is taken as $c = 1.5$, while the other two parameters are taken as $a = -4$ and $b = 3$ (typical values).

Please note that the Medcouple method is appropriate for our data as we deal with skewed univariate data (Figure 5) while other outlier detection methods such as the interquartile range only apply to normally distributed data.

After the cleaning procedure, data were standardized to zero mean and unit variance. Principal component analysis (PCA) (Dunteman, 1989) was then used for all methods except the neural networks model. We selected the number of principal component features of the PCA according to the cumulative contribution of the first components. The selected number

of principal components varied per machine learning algorithm. Here, the threshold for the cumulative contribution was that of 90%. After the pre-processing, data (\mathbf{X}, \mathbf{y}) were used to evaluate the performance of the different regression schemes to engine overhaul estimation. The regression techniques were only trained on a subset of the observations to evaluate how well the techniques could generalize to unseen observations. Evaluation on out-of-model observations was performed on the test set. Here, and as for the Weibull model, we compared the algorithms using a 10-fold cross-validation setup.

The hyper-parameters of the algorithms were selected by hand, based on preliminary random experiments. Examples of hyper-parameters tuned include the number of neighbors k in the k -Nearest Neighbor algorithm or the distance metric/similarity function, the number of trees in the random forests or the number of layers and neurons in the neural networks.

In order to evaluate the predicted values, we compute accuracy using established metrics in the field such as mean error or bias (ME) or median absolute deviation (MdAE) (Saxena et al., 2008). The mean error expresses error as

$$\text{ME} = \frac{1}{v_y} \sum_{i=1}^{v_y} (y_i - \hat{y}_i) \quad (18)$$

where v_k is the number of testing observations in fold k and ME its the model bias for fold k given by the difference between the real value y_i and the predicted value \hat{y}_i of observation i .

The median absolute error expresses absolute error as

$$\text{MdAE} = \text{median}(\{|y_i - \hat{y}_i|\}_{i=1}^{v_k}) \quad (19)$$

while relative accuracy is a customized metric from the prognostics field (Saxena et al., 2008) computed as

$$\text{RA}(\%) = \frac{1}{v_k} \sum_{t=1}^{v_y} 100 \times \left(1 - \left| \frac{\hat{y}_i - \text{median}\{y_i\}_{i=1}^{v_y}}{y_i} \right| \right) \quad (20)$$

For error spread we compute the median absolute deviation (MAD) as:

$$\text{MAD} = \text{median}(\{|y_i - \text{median}(\{y_i\}_{i=1}^{v_k})|\}_{i=1}^{v_k}) \quad (21)$$

To sum up, the used data-driven architecture is in Figure 2. The architecture is provided with inputs \mathbf{X} and corresponding measured outputs \mathbf{y} . With this data and the parameter vector θ the system is able to estimate function $f(t, \mathbf{x}(t), \theta(t))$. After this, the prognostics module determines the RUL estimates represented as $\hat{y}(t_i)$. The scheme of 10-fold cross-validation is used to estimate and validate the model.

5. RESULTS

After describing the data pre-processing and modeling approach we report our findings in this section. Concretely, we report the results of using the comparative research method to test the main hypothesis H. The goal here is to find evidence that the use of advanced data-driven methods can benefit the field of jet engine prognostics.

5.1. Numerical Results

In Table 2 we present the numerical results of this case study, namely of the baseline method (Weibull analysis) and the data-driven models of linear regression (LR), K-nearest neighbors (K-NN), random forests (RF), support vector machines (SVM) and neural networks (NN). Please note that the experience-based model and the data-driven model output different predictions but they essentially target the same output. For the experience-based model, the returned Weibull random variable is the predicted length of an overhaul. For the data-driven model it is the remaining time to an overhaul or RUL. Accordingly, both models attempt to predict the same RUL at different times: the Weibull model immediately after a removal, and the data-driven methods every time a health monitoring signal is generated. The two models (I and II) and their errors are therefore comparable.

The first finding from Table 2 is that the results seem to indicate that all algorithms can outperform in accuracy – as

measured by the metric of MdAE, experience-based methods (Weibull analysis). Concretely and in our case, the median absolute error (MdAE) of the experience-based method was clearly surpassed by the data-driven algorithms of NNET, LR, SVM, K-NN and RF. Here, the best performance was attained by the RF and the SVM with an average median error of 141 calendar days.

The results regarding the metric of relative accuracy (RA) were not as expressive as for the median absolute error (MdAE). The higher accuracies were attained by the RF and the SVM models with 72% of accuracy. It is worthwhile to discuss why the models relative accuracy did not differ much. Please note that RA is a metric closely related to median absolute percentage error (MAPE) (Saxena et al., 2008), in fact, it is its inverse ($\text{RA} = 1 - \text{MAPE}$). Since this latter is scale sensitive it is important to note that RA will give more importance to incorrect predictions of small actual values than large values, especially when working with low-volume data. Notice that because ‘actual’ is in the denominator of the equation, when the actual value is not zero, but quite small, the MAPE and consequently RA will often take on extreme values. Since in our case the RA values are reasonable, its values reinforce the notion that all the algorithms can come up with a favorable RUL estimation (also when close to the removal).

In regards to the precision of the methods, measured by the median absolute deviation (MAD), results were better for the data-driven methods. As the Weibull model outputs the same single output for every observation – the mean time between repairs (MTBR) of the training data set, predictions of the removal overall time deviate considerably from each other, leading to a high MAD. The remaining data-driven models were hence able to beat this score having lower errors, except for the K-nearest neighbor algorithm.

To sum up, the data-driven models were better in absolute accuracy and also, but not as expressively, in relative accuracy. Results were also promising in regards to error spread. These results suggest that there is enough evidence to support hypothesis H.

5.2. Illustrative Example

In this section we provide an illustrative example of how the different tested algorithms compute the RUL of a jet engine. As an example, consider Figure 6 which shows the application of the algorithms to 4 different removals. In the Figure, the time index i is shown on the x-axis and the predicted residual life is on the y-axis. The diagonal RUL* depicts the true RUL. Also depicted in the plot are the predictions of the several methods, LR, K-NN, RF, SVM and NN. The closer these predictions to the true RUL* curve, the better the model accuracy. The less dispersed the predictions around the RUL* curve the higher the model precision.

Table 2. Performance results.

	Experience-Based (Weibull)	Linear Re- gression (LR)	K-Nearest Neighbors (K-NN)	Neural Networks (NN)	Random Forests (RF)	Support Vector Machines (SVM)
Mean Error (ME)	113.16 ± 49.92	-153.18 ± 56.48	-113.29 ± 48.84	-175.23 ± 74.63	-164.57 ± 53.12	-31.97 ± 52.26
Median Absolute Error (MdAE) (days)	206.26 ± 49.92	171.20 ± 11.39	166.91 ± 14.59	165.97 ± 8.19	141.57 ± 11.64	141.26 ± 11.27
Relative Accuracy (RA)(%)	63.88 ± 0.26	68.21 ± 0.06	68.08 ± 0.04	68.94 ± 0.14	72.07 ± 0.03	71.89 ± 0.05
Median Absolute Deviation (MAD) (days)	306.09 ± 49.92	178.09 ± 25.98	216.46 ± 23.20	175.15 ± 36.49	208.49 ± 27.25	149.42 ± 24.93

Analyzing Figure 6 it can be seen that all algorithms can come up with reasonably good estimates of RUL although these estimates can vary considerably from one algorithm to another. For instance, the LR, the simplest approach, has a tendency to output similar predictions to similar points – here, by similar points we mean observations close in time. This tendency results in the LR predictions usually following a more or less well-defined horizontal trajectory.

In contrast with the LR, which works by finding the best fitting straight line, the K-NN method uses a similarity search approach to find the best neighbors. Here, predictions tend to have a higher degree of noise as the output of the algorithm is the average of the k-nearest neighbors. Nevertheless, the overall absolute accuracy of the model is considerably higher than that of the LR.

The predictions of the RF model have a distinct pattern from the remaining models. The method works by having more dispersed predictions along the RUL* line than the remaining methods. This follows from the way regression decision trees operate. Since decision trees work by a series of local decisions and a random forest is an ensemble of randomly selected decision trees – to avoid overfitting – the output of the RF model can deviate considerably. Nevertheless, the model accuracy is good and tends to improve near the removal, as the error spread diminishes.

The results of the NN algorithm in Figure 6 are also promising. The model is especially good at long-term predictions even though it often fails to provide accurate estimation near the engine removal.

Fig. 6 also shows the predictions of SVR, which also follow the targets well. As illustrated, this is the best model among the five, as its predictions deviate the least from the perfect correlation line of RUL*.

6. CONCLUSION

Five different methods for jet engine prognostics models were tested on a proprietary data set. The used methods were

linear regression (LR), K-nearest neighbors (K-NN), neural networks (NN), random forests (RF) and support vector regression (SVR). All methods showed a reasonable good performance. The predictions of the SVR method resulted in the best values for the median absolute error and median absolute deviation metrics. The good agreement between the modeled remaining life and the observations for the tested techniques lead to the conclusion that data-driven models can potentially be used as an alternative to traditional experience-based methods such as Weibull analysis. It can be said that data-driven techniques are worth further exploration in the field of engineering prognostics.

We hope that with this work we can promote the widespread adoption of data-driven techniques not only for gas turbine prognostics but also for other industrial fields. Capital intensive assets such as engines requiring predictive maintenance are common in the industry and could well benefit from the use of advanced machine learning. This work hence aims to contribute to a better understanding of traditional machine learning methods and its utility to prognostics in general.

The tested techniques have great potential for improvement especially in the fields of parameter selection and generalization to other prognostics situations. As future research, we intend to study formal methodologies for validation of data-driven approaches, and investigate fusion of prognostic estimates such as ensemble methods. It is also our intention to perform a deeper analysis of the tested data-driven models with metrics related to prediction horizon, sensitivity to damage state estimation, confidence distribution, evolution of distribution around actual time of failure, and stability/robustness of the prediction.

ACKNOWLEDGMENT

This work was partially supported partly by the MIT Portugal Program and FCT - Fundacao para a Ciencia e a Tecnologia (SFRH/BD/52347/2013).

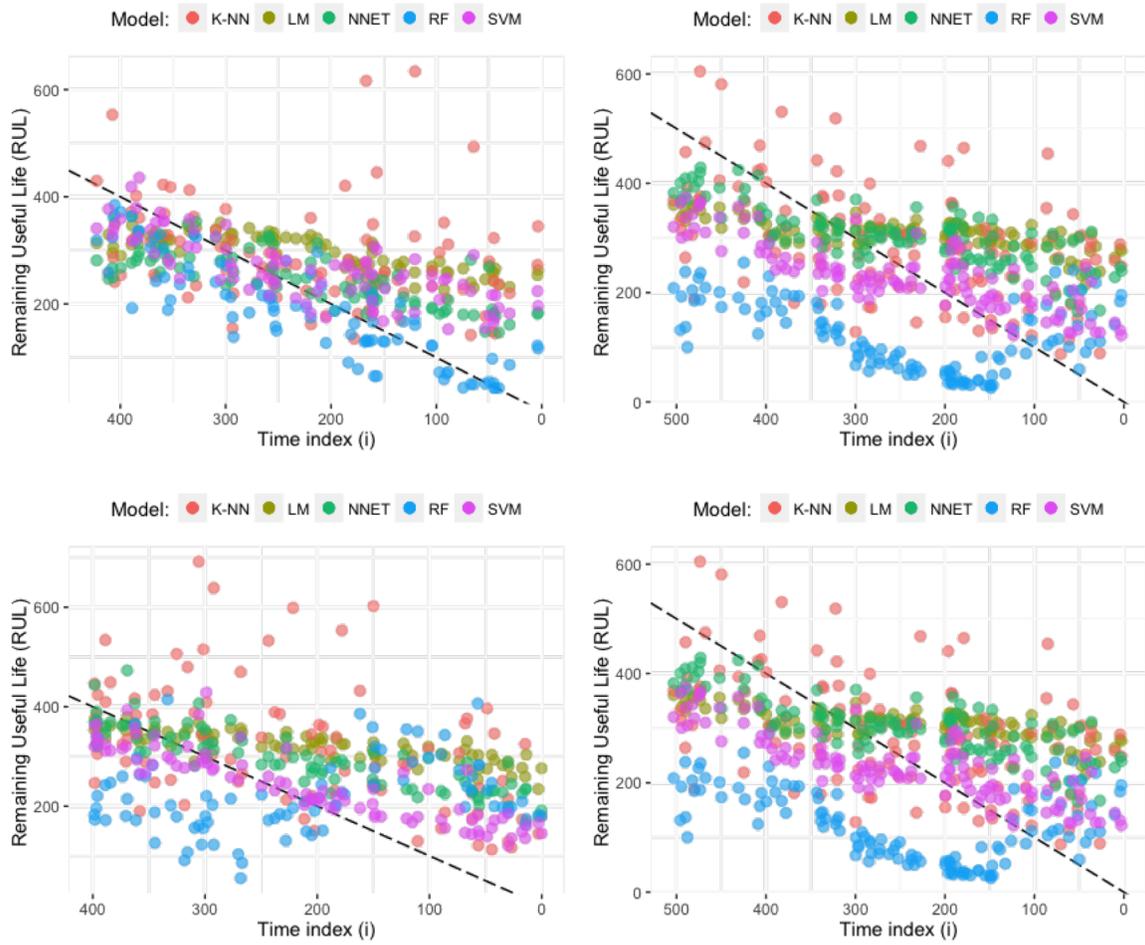


Figure 6. Remaining useful life prediction trajectory of the 5 data-driven algorithms for 4 removals.

NOMENCLATURE

Nomenclature used in paper follows.

EOL	End of Life
HM	Health Monitoring
K-NN	K-Nearest Neighbors
LR	Linear Regression
MAD	Median Absolute Deviation
MAPE	Mean Absolute Percentage Error
MdAE	Median Absolute Error
MTBR	Mean Time Between Repairs
NN	Neural Networks
RA	Relative Accuracy
RF	Random Forests
RUL	Remaining Useful Life
SVR	Support Vector Regression

REFERENCES

- Baptista, M., de Medeiros, I. P., Malere, J. P., Nascimento, C., Prendinger, H., & Henriques, E. M. (2017). Comparative case study of life usage and data-driven prognostics techniques using aircraft fault messages. *Computers in Industry*, 86, 1–14.
- Baptista, M., de Medeiros, I. P., Malere, J. P., Prendinger, H., Nascimento Jr, C. L., & Henriques, E. (2016a). A comparison of data-driven techniques for engine bleed valve prognostics using aircraft-derived fault messages. In *Annual European Conference of the Prognostics and Health Management Society* (Vol. 7, p. 13).
- Baptista, M., de Medeiros, I. P., Malere, J. P., Prendinger, H., Nascimento Jr, C. L., & Henriques, E. (2016b). Improved time-based maintenance in aeronautics with regressive support vector machines. In *Annual Conference of the Prognostics and Health Management Society 2016* (Vol. 7, p. 10).
- Brotherton, T., Jahns, G., Jacobs, J., & Wroblewski, D. (2000). Prognosis of faults in gas turbine engines. In *Aerospace IEEE Conference* (Vol. 6, pp. 163–171).
- Brys, G., Hubert, M., & Struyf, A. (2004). A Robustification of the Jarque–Bera Test of Normality. In *COMPSTAT 2004 Symposium, Section: Robustness*.
- DePold, H. R., & Gass, F. D. (1998). The application of expert systems and neural networks to gas turbine prognostics and diagnostics. In *ASME 1998 International Gas Turbine and Aeroengine Congress and Exhibition*.
- Di Maio, F., & Zio, E. (2013). Failure prognostics by a data-driven similarity-based approach. *International Journal of Reliability, Quality and Safety Engineering*, 20(01), 1350001.
- Drucker, H., Burges, C. J., Kaufman, L., A, S., & Vapnik, V. N. (1997). Support vector machines. *Advances in Neural Information Processing Systems*, 9, 155–161.
- Dunteman, G. H. (1989). *Principal Components Analysis* (No. 69). Sage.
- Ebden, M., Stranjak, A., & Roberts, S. (2010). Visualizing uncertainty in reliability functions with application to aero engine overhaul. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(1), 163–173.
- Goebel, K., Saha, B., & Saxena, A. (2008). A comparison of three data-driven techniques for prognostics. In *62nd Meeting of the Society for Machinery Failure Prevention Technology (MFTP)* (pp. 119–131).
- Huang, H.-Z., Wang, H.-K., Li, Y.-F., Zhang, L., & Liu, Z. (2015). Support vector machine based estimation of remaining useful life: Current research status and future trends. *Journal of Mechanical Science and Technology*, 29(1), 151–163.
- Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7), 1483–1510.
- Khelif, R., Malinowski, S., Chebel-Morello, B., & Zerhouni, N. (2014). Rul prediction based on a new similarity-instance based approach. In *2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE)* (pp. 2463–2468).
- Kohavi, R., et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI* (Vol. 14, pp. 1137–1145).
- Kröse, B., Krose, B., van der Smagt, P., & Smagt, P. (1993). *An introduction to neural networks*. Citeseer.
- Li, Y., & Nilkitsaranont, P. (2009). Gas turbine performance prognostic for condition-based maintenance. *Applied Energy*, 86(10), 2152–2161.
- Riad, A., Elminir, H., & Elattar, H. (2010). Evaluation of neural networks in the subject of prognostics as compared to linear regression model. *International Journal of Engineering & Technology*, 10(1), 52–58.
- Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., & Schwabacher, M. (2008). Metrics for Evaluating Performance of Prognostic Techniques. In *Prognostics and Health Management Conference* (pp. 1–17).
- Schwabacher, M. (2005). A survey of data-driven prognostics. In *Proceedings of the AIAA Infotech@ Aerospace Conference* (pp. 1–5).
- Schwabacher, M., & Goebel, K. (2007). A survey of artificial intelligence for prognostics. In *AAAI fall symposium* (pp. 107–114).
- Seemann, R., Langhans, S., Schilling, T., & Gollnick, V. (2010). Modeling the life cycle cost of jet engine maintenance. *Hamburg: Technische Universität Hamburg-Harburg*.
- Stranjak, A., Dutta, P. S., Ebden, M., Rogers, A., & Vytelingum, P. (2008). A multi-agent simulation system for prediction and scheduling of aero engine over-

haul. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems: Industrial track* (pp. 81–88).

- Tobon-Mejia, D. A., Medjaher, K., Zerhouni, N., & Tripot, G. (2012). A data-driven failure prognostics method based on mixture of gaussians hidden markov models. *IEEE Transactions on Reliability*, 61(2), 491–503.
- Tukey, J. W. (1977). *Exploratory Data Analysis*.
- Weckman, G. R., Marvel, J. H., & Shell, R. L. (2006). Decision support approach to fleet maintenance requirements in the aviation industry. *Journal of Aircraft*, 43(5), 1352–1360.
- Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 103, 293–297.
- Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... others (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1–37.
- Xue, F., Bonissone, P., Varma, A., Yan, W., Eklund, N., & Goebel, K. (2008). An instance-based method for remaining useful life estimation for aircraft engines. *Journal of Failure Analysis and Prevention*, 8(2), 199–206.
- Zaidan, M. A., Harrison, R. F., Mills, A. R., & Fleming, P. J. (2015). Bayesian hierarchical models for aerospace gas turbine engine prognostics. *Expert Systems with Applications*, 42(1), 539–553.
- Zaidan, M. A., Mills, A. R., Harrison, R. F., & Fleming, P. J. (2016). Gas turbine engine prognostics using bayesian hierarchical models: A variational approach. *Mechanical Systems and Signal Processing*, 70, 120–140.
- Zaidan, M. A., Relan, R., Mills, A. R., & Harrison, R. F. (2015). Prognostics of gas turbine engine: An integrated approach. *Expert Systems with Applications*, 42(22), 8472–8483.

BIOGRAPHIES



Marcia Lourenco Baptista (BS and MSc. in Informatics and Computer Engineering. Instituto Superior Tecnico, Lisbon, Portugal, September 2008) is a PhD candidate student at the Engineering Design and Advanced Manufacturing (EDAM) - Leaders for Technical Industries (LTI) program under the umbrella of the MIT Portugal

Program. Her research focuses on the development of data-driven prognostics techniques for the aeronautical sector. Her research interests include machine learning applied to health monitoring systems and structural health management.



Cairo Lucio Nascimento Junior received the B.S. degree in electrical engineering from the Federal University of Uberlandia (UFU, Uberlandia, Brazil) in 1984, the M.S. degree in electronics engineering from the Instituto Tecnológico de Aeronautica (ITA, Sao Jose dos Campos, Brazil) in 1988 and the Ph.D. degree in electrical engineering from the UMIST, Control Systems Centre (Manchester, UK) in 1994. Since 1986 he has been a lecturer with the Division of Electronic Engineering, ITA, and has supervised (or co-supervised) 9 Ph.D. and 27 M.Sc./Prof. M. students. He is the co-author of a book on intelligent systems and was the chairman for the 1999 Fourth Brazilian Conference on Neural Networks. His current research interests include autonomous systems, artificial intelligence, mobile robotics, control engineering, the development of internet-based remote-access laboratories for engineering education and data mining for knowledge extraction and application on fraud detection and PHM (Prognostics and Health Management).



Helmut Prendinger received his Master and Doctoral degrees in Logic and Artificial Intelligence from the University of Salzburg in 1994 and 1998, respectively. Since 2012, he is a full professor at the National Institute of Informatics (NII), Tokyo, after joining NII in 2004 as Associate Professor. Previously, he held positions as research associate (2000–2004) and JSPS postdoctoral fellow (1998–2000) at the University of Tokyo, Dept. of Information and Communication Engineering, Faculty of Engineering. In 1996–1997, he was a junior specialist at the University of California, Irvine. His research interests in artificial intelligence include machine learning, intelligent user interface, cyber-physical systems, and the melding of real and virtual worlds, in which areas he has published more than 220 peer-reviewed journal and conference papers. His vision is to apply his research to establishing the IT infrastructure for Unmanned Aerial Vehicles, or “drone”. He is a member of IEEE and ACM.

Elsa Maria Pires Henriques has a doctorate degree in Mechanical Engineering and is associated professor at Instituto Superior Tecnico in the University of Lisbon.



She is responsible for the “Engineering Design and Advance Manufacturing (LTI/EDAM)” post-graduation. During the last fifteen years she has participated and/or coordinated several national and European R&D projects in collaboration with different industrial sectors, from tooling to automotive and aeronautics, mainly related to manufacturing, life cycle based decisions and management of complex design processes. She has a large number of scientific and technical publications in national and international conferences and journals. She was a national delegate in the 7th Framework Programme of the EU.