Elsevier Editorial System(tm) for Journal of Computational and Applied Mathematics
Manuscript Draft

Manuscript Number: CAM-D-18-00096R1

Title: A hybrid ARIMA-SVM model for the study of the remaining useful life of aicraft engines

Article Type: VSI:Conference IMM_2017

Section/Category: VSI:Conference IMM_2017

Keywords: Remaining Useful Life (RUL); aircraft engines; Vector Autoregression Moving-Average (VARMA); Support Vector Machines (SVM); Genetic Algorithms (GA)

Corresponding Author: Dr. Fernando Sánchez Lasheras, Ph.D.

Corresponding Author's Institution: University of Oviedo

First Author: Celestino Ordoñez

Order of Authors: Celestino Ordoñez; Fernando Sánchez Lasheras, Ph.D.; Javier Roca Pardiñas; Francisco Javier de Cos

Abstract: In this research, an algorithm is presented for predicting the remaining useful life (RUL) of aircraft engines from a set of predictor variables measured by several sensors located in the engine. RUL prediction is essential for the safety of those aboard, but also to reduce engine maintenance and repair costs. The algorithm combines time series analysis methods to fore-cast the values of the predictor variables with machine learning techniques to predict RUL from those variables. First, an auto-regressive integrated moving average (ARIMA) model is used to estimate the values of the predictor variables in advance. Then, we use the result of the previous step as the input of a support vector regression model (SVM), where RUL is the response variable. The validity of the method was checked on an extensive public database, and the results compared with those obtained using a vector auto-regressive moving aver-age (VARMA) model. Our algorithm showed a high prediction capability, far greater than that provided by the VARMA model.

A multivariate time series based approach for the study of the remaining useful life of aircraft engines

HIHGLIGHTS

- A method to forecast the remaining useful life of aircraft engines is proposed.
- The predictor variables were obtained from sensors located in the engine.
- The proposed method combines ARIMA and SVM models.
- Results of our method unsurpassed those obtained using a VARMA model.

A hybrid ARIMA-SVM model for the study of the remaining useful life of aicraft engines

Celestino Ordóñez[a], Fernando Sánchez Lasheras[b*], Javier Roca-Pardiñas[c], Francisco Javier de Cos Juez[a]

[a]*Department of Mining Exploitation and Prospecting, University of Oviedo, c/Independencia 13, 33004 Oviedo, Spain*

[b]*Department of Mathematics, University of Oviedo, c/Federico García Lorca 18, 33007 Oviedo, Spain*

[c]*Department of Statistics and Operations Research, University of Vigo, 32608 Vigo, Spain*

*Corresponding author. Tel.: + 34 984 833 135

E-mail address: sanchezfernando@uniovi.es (F. Sánchez Lasheras)

keywords: Remaining Useful Life (RUL); aircraft engines; Vector Autoregression Moving-Average (VARMA); Support Vector Machines (SVM); Genetic Algorithms (GA).

Abstract

In this research, an algorithm is presented for predicting the remaining useful life (RUL) of aircraft engines from a set of predictor variables measured by several sensors located in the engine. RUL prediction is essential for the safety of those aboard, but also to reduce engine maintenance and repair costs. The algorithm combines time series analysis methods to forecast the values of the predictor variables with machine learning techniques to predict RUL from those variables. First, an auto-regressive integrated moving average (ARIMA) model is used to estimate the values of the predictor variables in advance. Then, we use the result of the previous step as the input of a support vector regression model (SVM), where RUL is the response variable. The validity of the method was checked on an extensive public database,

and the results compared with those obtained using a vector auto-regressive moving average (VARMA) model. Our algorithm showed a high prediction capability, far greater than that provided by the VARMA model.

1. Introduction

Nowadays aircraft engines are mainly either lightweight piston engines or gas turbines. Compared to steam turbines, gas turbines have hardly any cooling needs. In addition, their low thermal inertia allows them to reach their full load in very short time. This simplicity, in comparison with steam turbines and reciprocating engines, gives them two additional advantages when compared to other thermal machines: simple maintenance and high reliability. In fact, the reduction of lubrication and refrigeration requirements, the continuity of the combustion process and the absence of alternative motions means that the probability of failure decreases. However, gas turbines also have a significant number of drawbacks, including high rotational speed and low performance (30-35%) when compared to diesel alternative engines (almost 50%) or to steam turbines (values of 40% are common).

The safe and reliable operation of aircraft engines is one of the main principles of the aeronautic industry [1,2]. The maintenance of aircraft engines in an operational condition and the early detection of possible failures is a fundamental requirement. In the case of aircraft engines and for safety reasons, forecasting of remaining useful life (RUL) has become an important issue in the last decade [3-5]. Condition-Based Maintenence (CBM) is a maintenance approach in which the maintenance of a piece of equipment is made based on its current status and not simply on the time passed. In order to put CBM maintenance systems in place, reliable models of Remaining Useful Life (RUL) are mandatory. The CBM approach to maintenance started in the aerospace industry in the mid – 1990s and now is a well-known methodology [6].

The future health of an engineering system is predicted by a discipline called Prognostics and Health Management (PHM). The approach to the prediction of a system's health can be made in three ways [7]:

- Using machine learning and data recognition techniques that can predict the RUL of the system without any prior knowledge of the problem. This is called the data-driven approach [8].
- Understanding the physical process and the interrelationships of the variables in the RUL of the system. This approach is called model-based [9].
- Finally, combining both the data-driven and the model-based approaches in order to predict the RUL of the system. This is called a hybrid approach [10].

In the present research, the prediction of the RUL is performed by a data-driven approach.

2. Material and methods

2.1. Materials

Fig. 1 shows a diagram of the type of engine used to validate our method, with its main elements. In this kind of device, the air is introduced into the low-pressure compressor through the fan. In the following step, it travels to the high-pressure compressor. This is when the air is heated in the combustor; also in the combustor, the air is mixed with fuel and ignited. The fuel combustion raises the HPC (high-pressure turbine) discharge air velocity to drive the HPT (high pressure turbine) and the LPT (low-pressure turbine). The engine employed for the present study is based on a low-frequency, transient, performance model of a high-speed ratio, dual-

spool, low by-pass, variable cycle, turbofan engine with a digital controller. The controller has a 50 Hz frequency.

Data employed in this research refers to an aircraft engine of 90,000 lb thrust class. The information retrieved from the aircraft corresponds to different operating conditions that range from sea level to 40,000 ft, with temperatures ranging from -51 ºC to 39 ºC. In fact, the data comes from the MAPSS software. This program has revisions for both civil and military applications. The present research used the military version, which can perform realistic simulations according to the Standard Full Authority Digital Engine Controllers (FADEC) [11]. A more in-depth explanation can be found in the C-MAPSS User's Guide [12].

Our database is composed of information relating to a total of 100 different turbines with a total number of observations that varies from 128 to 362. Each record of the engine state has a total of 24 variables. Three of these are operational settings, while the other 21 represent values for engine performance measurement. We would like to remark that these measurements are contaminated by noise.

The descriptive statistics of the response and the predictor variables are summarized in Table 1. The information of the data base corresponds to six different flight conditions with altitudes from 0 to 42,000 feet, speeds from 0 to 0.84 Mach and a throttle resolver angle from 20 to 100 degrees. According to the results of Table 1, the values of variables Demanded fan speed, Demanded corrected fan speed and Total temperature at fan inlet are constant, while the variation of other variables like Operational Setting 3, Presure at fan inlet, Engine pressure ratio and Burner fuel-air ratio are very small, with a standard deviation almost equal to zero. Constant and 'almost constant' variables were discarded for the study. Afterwards, the correlation coefficients of the remaining variables were calculated. Those variables whose correlation coefficient was over 0.8 were analyzed and removed from the study in order to avoid problems associated to collinearity. In our case, the variables removed were Total temperature at LPT out-

let and Total pressure at HPC outlet, because their correlation coefficients with Physical fan speed and Static pressure at HPC outlet were 0.830 and -0.823, respectively. Static pressure at HPC outlet (psia) was removed because its correlation coefficient value with Ratio of fuel flow to Ps30 was -0.847. Finally, Static pressure at HPC outlet was removed, as its correlation coefficient with Corrected fan speed was 0.826.

2.2. Methods

Our approach makes use of several statistical techniques to estimate RUL from a set of covariates measured by the sensors located in the engine. A brief summary of these techniques is provided in order to facilitate the comprehension of this work.

2.2.1 Vector Autoregressive Moving-Average (VARMA)

The Vector Autoregressive Moving-Average (VARMA) methodology is a multivariate time series method that modelizes several dependent time series together, taking into account the correlations within each one of the time series and also across them [13,14]. In this work, VARMA was used for both forecasting the predictor variables and contrasting the results of the proposed method.

Let $X_t = \left(x_{1t},...,x_{Kt}\right)^t$ be a k-dimensional stationary time series, the VARMA(p,q) model has the following expression:

$$A_0 x_t = A_1 x_{t-1} + ... + A_p x_{t-p} + M_0 u_t + M_1 u_{t-1} + ... + M_q u_{t-q} \quad (1)$$

where *t* represents the time, $A_0, A_{1,...,} A_p$ are $\left(K \times K\right)$ autoregressive parameter matrices and $M_0, M_1, ..., M_q$ are $\left(K \times K\right)$ moving average parameter matrices. Likewise, $u_t, u_{t-1}, ..., u_{t-q}$

are terms of Gaussian error with zero mean and time invariant covariance matrix. The zero order matrices A_0 and M_0 are assumed to be nonsingular and they are often equal to the identity matrix.

This model can be written in lag operator notation as follows:

$$A(L)x_t = M(L)u_t \tag{2}$$

where $A(L) = A_0 - A_1L - \cdots A_pL^p$ and $M(L) = M_0 + M_1L + \cdots M_qL^q$

Stationarity and invertibility are assumed, which requires that the roots of $|A(L)| = 0$ and $|M(L)| = 0$ are outside the unit circle.

If $M(L) = I$, we obtain a pure vector autoregressive model (VAR) of order *p*. If $A(L) = I$, we obtain a pure vector moving average model (VMA) of order *q*.

Autoregressive models, such as VARMA, are very flexible in handling a wide range of patterns in the time series by changing the parameters.

2.2.2 Autoregressive integrated moving average (ARIMA)

The ARIMA model is a generalization of the autoregressive moving average (ARMA) model [15]. The autoregressive part of ARIMA indicates that the evolving variable of interest is regressed on their previous values while the moving average part indicates that the regression error is a linear combination of errors terms that occurred at different times in the past.

The initial set up of an ARIMA model is based on the observation of the time series graph and in the analysis of autocorrelations for different time delays. In order to perform this procedure in a systematic way, a well-known methodology called Box-Jenkins is applied [16]

A general ARIMA model can be expressed as $ARIMA\ (p, d, q)$ where p represents the number of autoregressive terms, d is the number of non-seasonal differences needed for stationary and q is the number of lagged forecast errors in the prediction equation.

The $ARIMA\ (p, d, q)$ model is as follows:

$$\varphi(L)(1-L)^d Y_t = \Theta(L)\varepsilon_t$$

Where:

Y_t is the actual value.

ε_t is the random error at time period t.

L is the lag operator that is defined by $LX_t = X_{t-1}$.

$\varphi(L)$ is the autoregressive operator, represented as a polynomial in the backshift operator, $\varphi(L) = 1 - \varphi_1\ B - \cdots - \ \varphi_p\ B^p$

2.2.3 Support Vector Machines (SVM)

Support Vector Machines (SVM) were developed as a methodology to be applied in binary classification problems. In addition, it was extended for the solution of other problems such as regression. Support vector machines (SVM) for regression were developed by Vapnik and co-workers at AT&T Bell Laboratories [17-19]. Given their good performance, SVM has become a standard for both classification and regression in a wide range of machine-learning software. We used SVM to estimate RUL from the forecasted values of the predictor variables.

Let $\{\mathbf{x}_i, y_i\}_{i=1}^n$ be a training dataset, where $\mathbf{x}_i \subset X \in \mathbb{R}^d$ represents the predictor covariates and $y_i \in \mathbb{R}$ the response variable. In the ε-SV linear regression [18] the aim is to find a function $f(x) = \langle \mathbf{w}, \mathbf{x} \rangle + b,\ \mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R}$ that has at most a deviation ε from *y* for all training

data. Analytically speaking, the solution of this problem is formulated as the following minimization problem with restrictions

$$\begin{cases} \min\limits_{\mathbf{w},b} \dfrac{1}{2}\|\mathbf{w}\|^2 + C\sum\limits_{i=1}^{n}\left(\xi_i + \xi_i^*\right) \\ y_i - \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) \le \varepsilon + \xi_i,\ i \in \{1,\ldots,n\} \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i \ge \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{3}$$

where $\|\cdot\|$ is the Euclidean norm, ξ_i and ξ_i^* the so-called slack variables and $C > 0$ determines the trade-off between the flatness of *f* and the value of such deviations. The flatness of *f* depends on $\|\mathbf{w}\|$ (the smaller the elements of **w** are, the flatter *f* is).

The quality of the estimation is measured by the ε- insensitive loss function L_ε proposed by Vapnik:

$$L_\varepsilon = \begin{cases} 0 & \text{if } |\xi| < \varepsilon \\ |\xi| - \varepsilon & \text{otherwise} \end{cases} \tag{4}$$

The slack variables account for the deviations of the solution beyond the ε-sensitive zone.

If *C* is too large, then the objective is to minimize the average loss (empirical risk), without regard to model complexity.

The optimization problem in (3) is computationally simpler to solve in its Lagrange dual formulation. The solution is a linear combination of a subset of sample points called support vectors.

$$f(x) = \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)\langle \mathbf{x}_i, \mathbf{x} \rangle + b;\ \alpha_i, \alpha_i^* \ge 0 \tag{5}$$

being $\mathbf{w} = \sum\limits_{i=1}^{n}\beta_i \mathbf{x}_i = \sum\limits_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)\mathbf{x}_i$, and α_i, α_i^* the Lagrange multipliers.

The support vectors correspond to the observations for which α_i or $\alpha_i^* \neq 0$.

The Lagrange dual formulation allows extending the solution to nonlinear functions by replacing the dot product $\langle \mathbf{x}_i, \mathbf{x} \rangle$ with a positive definition function $k(\mathbf{x}_i, \mathbf{x})$ (kernel) as follows:

$$k(\mathbf{x}_i, \mathbf{x}) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle \tag{6}$$

where $\phi : X \subset \mathbb{R}^d \rightarrow \mathbb{R}^r$ is a transformation that maps **x** into a high dimensional space (feature space). The explicit coordinates in the feature space and even the mapping function ϕ become unnecessary when we define a kernel. The advantage of this procedure, known as the kernel trick, is that the complexity of the optimization problem remains dependent only on the dimensionality of the input space, and not on the feature.

The solution of the optimization problem is analogous to

$$f(x) = \sum_{i=1}^{n} \beta_i k(\mathbf{x}_i, \mathbf{x}) + b \tag{7}$$

Using this method, nonlinear SVM finds the optimal function in the transformed predictor space.

There are many types of kernels in existing literature, polynomial and tangent hyperbolic kernels being two of the most cited. Selecting a particular kernel type and kernel function parameters is usually based on application-domain knowledge.

In the present research, four different kinds of kernels were employed: linear, polynomial, radial basis and sigmoidal.

2.2.4 Genetic Algorithms (GA)

Genetic algorithms are well-known evolutionary algorithms developed by Holland [20]. This methodology mimics the natural evolution of a population, considering the proposed solutions

of a problem as genetic chains that are combined in order to improve their performance. The fitness of the proposed solutions is evaluated, and the individuals with the highest performance are preserved and combined in order to create the next generation set of proposed solutions. In this study GA has been used to find optimum values for the SVM parameters.

The genetic algorithms methodology is based on three genetic operations that are present in most biological systems:

- **Reproduction:** the set of solutions of a generation is created using the individuals of the previous generation as the base. That is the reason why the individuals of one generation have most of the characteristics of the individuals of the previous generation.
- **Crossover:** the crossover operator of the GA mimics the biological mechanism of reproduction, using two individuals of a solution set in order to create an individual of the next generation.
- **Mutation:** the mutation operator is employed in order to maintain the diversity of solutions. The behavior of this operator is very similar to biological mutation, in which new offspring are born with random changes that have nothing to do with their parents' chromosomes.

There is another operator characteristic of the GA methodology that accelerates the improvement of the fitness function, although it is not present in biological systems. It is called *elitism*, and allows some of the better genomes from one population set to survive (they are cloned in the next generation).

Finally, it must be remarked that one of the most valuable characteristics of GA is its ability to avoid falling into local optimus. A more in-depth explanation of GA can be found elsewhere [21, 22].

The GA algorithm can be mathematically formulated as follows. Let D be a finite domain and $f : D \to R$ the function to be optimized. It can be said that f has a local minimum in $\hat{x} \in D$ when in a certain domain of x called $N(x), \forall y \in N(x) \to f(x) \leq f(y)$.

Let P_0 be an initial population of μ elements, $P_0 = \{I_0^1, ..., I_0^\mu\}$

The population obtained after a certain number of steps γ is represented by P_γ .

The function $fsel$ randomly selects , and with replacement, a set of individuals $y \in P_\gamma$.

$$fsel : (\alpha, x) \to y$$

where α is a vector of dimension μ made up of random values.

A global reproduction function $freprod$ is defined as one which creates a population of off-spring $z \in P_\gamma$ using some selected individuals by means of the crossover operator:

$$freprod(\beta, y) \to z$$

with β a vector of dimension $\mu/2$ of random integer numbers that belongs to the set $\{1, ..., l-1\}$.

The individual mutation function $find\_mut$ is applied to any individual I and creates a new mutated individual called MI . The function individual mutation is defined as follows:

$$find\_mut(I) = MI \;\; \forall j \in \{1, ..., l\}; \; P(sm_j = s_j) = 1 - p_m$$

where p_m represents the mutation probability of each one of the elements that constitute the individual I.

Finally, the function $fext$ creates a new population $n \in P_{\gamma+1}$ starting from two populations $x, z \in P_\gamma$.

$$fext:(x,z) \rightarrow n N_i \begin{cases} X_i \, if \, i \leq \mu \\ Z_{i-\mu} \, if \, i > \mu \end{cases}$$

where N_i with $i = 1, \ldots, 2\mu$ is the ith individual of population n.

The reason why GA are employed in the present research is because of their ability for finding quasi-optimum values without been trapped in local optimums [22]. One of the drawbacks of this methodology is that, like the rest of the evolutive methodologies, it is not able to guarantee that the optimum value will be found.

2.3. The algorithm

In the present research, a new algorithm is proposed in order to calculate the RUL of aircraft engines. This algorithm, in a first stage, performs a forecast of the input variables by means of an ARIMA model that is explained in section 2.2.2. In a second stage, the RUL is estimated with the help of an SVM model (please see the details of this methodology in section 2.2.3) using as input values those calculated with the help of the ARIMA. Please note that the algorithm proposed is new and in its first stage makes use of the ARIMA process, whose complexity is lower than that of VARMA, a multivariate time series method detailed in section 2.2.1. Figure 2 presents the flowchart of the algorithm. As has been stated before the Figure shows how the variables of the different sensor are forecasted by means of an ARIMA model and, afterwards, these values are employed as input information for the SVM model.

As has already been explained, this research employs four different kinds of kernels which were tested for the training of the SVM model: linear, polinomial, radial basis and sigmoidal. The choice of the best kernel function for each problem, as well as the parameters of the hyperparameters, are important steps in the training procedure of the SVM model. One of the most common methodologies for parameter optimization is the grid search and the most elaborated and systematic technique of the gradient descent method. However, as the number of variables to optimize becomes larger, this technique becomes inviable because of its computational cost. Instead, in this research the selection is performed by means of genetic algorithms as in general, strategies based on evolutionary algorithms are more efficient as intelligent tuning strategies than the grid search. The relative importance of the prediction variables in the result was also analyzed.

3. Results and discussion

In a first attempt to modelize the aircraft RUL, VARMA models were trained for each aircraft engine. In practice, due to the results obtained, the modeling was restricted to the VAR process as the best model found was VAR(42) $(p = 0.028)$. As the VARMA results (R^2 = 0.5436, RMSE = 47.6320 and MAE = 37.6152) did not improve upon previous research for the prediction of RUL [2, 22] in $t+1$ the results obtained for the predictions in $t+2, t+3, t+4$ and $t+5$ are not detailed.

Using the hybrid algorithm presented in Section 2.3, the RUL for different aircraft engines was predicted in $t+1, t+2, t+3, t+4$ and $t+5$. Figure 3 shows the RUL values versus those estimated by means of the hybrid algorithm for four times lags. As can be appreciated, there is a

good fit that, as expected, decreases with time. Please note that the range of dispersion of differences between real and predicted values changes significantly from one aircraft engine to another, a pattern found in previous works [3]. For the sake of brevity, the results for $t+5$, which shown the same pattern, are not shown.

The SVM parameters were tuned using GA for SVM with v-regression type. Accordingly, value v was set to 1, the chosen kernel was polynomial, with degree 3, $\gamma = 2$ and independent parameter $\alpha_0 = 5$. Regarding the general parameters of the SVM, the cost was set at $C = 1.12$ and $\varepsilon = 0.1$.

Table 2 shows the values of the statistics used to estimate the accuracy of the algorithm. In this research the performance is measured by means of the determination coefficent (R^2), root mean square error (RMSE) and mean absolute error (MAE). The high correlation for $t+1$ that is visually observed in Fig. 3, is confirmed by a determinantion coefficient value of 0.9315 that improves on our previous research [3,23]. In addition, the results obtained for two and more units of time ahead allow us to be optimistic about the usefulness of our algorithm in predicting RUL.

Table 2. R^2, RMSE and MAE obtained for the RUL forecast with the proposed hybrid algorithm for times in advance from 1 to 5.

Finally, Table 3 shows the variables importance ranking according to their contribution to the R^2 of the proposed hybrid algorithm of the RUL for $t+1$. In order to perform this calculus, variables were removed from the database one by one and the whole process of model training repeated. The R^2 of the algorithm for $t+1$ without this variable was compared with that of the algorithm calculated using all the variables. As may be observed, the most important variable is the LPT coolant bleed, followed by the HPT coolant bleed, Ratio of fuel flow to Ps30 and

Total pressure at HPC outlet. The influence of only five variables is above 0.1, while 11 are under 0.05, of which 8 are under 0.01.

Although the VARMA modelling technique allows several dependent time series to be modelled together and account for both cross and within-correlations of the series, its performance in the present research was worse than the performance obtained by previous studies by the authors and by the hybrid algorithm proposed. From the authors' point of view, this could be attributed to the fact that VARMA are a multivariate generalization of autoregressive moving average (ARMA) [26] based on the stationary nature of the data, while univariate ARIMA models combine differencing of non-stationary time series with the ARMA model.

4. Conclusions

The main contribution of the present research is the proposal of a new algorithm that performs very satisfactorily in predicting RUL for more than one period ahead. Predicting RUL in advance is important in order to detect a reduction in the RUL that would affect the operation of the aircraft engine.

In the proposed method, each input variable measured by a specific sensor was forecasted using an ARIMA model, and the results used as covariates of an SVM model where the dependent variable is the RUL. The predictive capacity of our proposal is much greater than that obtained solving the problem once using a multivariate VARMA model, and also than that reported in previous research using different approaches. An analysis of the relative importance of the predictor variables reveals that the RUL is most influenced by only five variables.

From the authors' point of view, the results obtained are promising, and would be applicable to the optimization of maintenance planning, not only of aircraft engines but also of un-

manned aerial vehicles (UAVs), the optimal control of which is an important research topic nowadays [22]. Indeed, our method would be even more especially relevant for UAVs due to the lack of physical presence of human pilots inside the vehicles who might be able to detect a possible breakdown in advance.

In our future research, we will try to improve the prediction of RUL and also to predict more time units ahead with the help of deep learning methodologies and also of the Multivariate Autoregressive Forests (mv-ARF) [23], which employs tree-based ensemble learners with auto-regressive components.

Acknowledgments

This work supported by project FC-15-GRUPIN14-033 of the Fundación para el Fomento en Asturias de la Investig11ación Científica Aplicada y la Tecnología (FICYT) (Spain), with FEDER support.

References

[1] Y. M. Hussain, S. Burrow, Leigh Henson, P. Keogh, Benefits Analysis of Prognostics & Health Monitoring to Aircraft Maintenance using System Dynamics. Third European Conference of the Prognostics and Health Management Society 2016 Publication Year: 2016 Publication Volume: 7 Publication Control Number: 023 Page Count: 13.
http://www.phmsociety.org/node/1883/ (accessed the 1st December 2017)

[2] E. Zio. Reliability engineering: old problems and new challenges, Reliab. Eng. Syst. Saf. 94 (2009) 125–41.

[3] P.J.G. Nieto, E. Garcia-Gonzalo, F.S. Lasheras, FJ de Cos Juez, Hybrid PSO–SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability, Reliab. Eng. Syst. Saf. 138 (2015) 219-231.

[4] Z. Wei, T. Tao, D. ZhuoShu, E. Zio, A dynamic particle filter-support vector regression method for reliability prediction, Reliab. Eng. Syst. Saf. 119 (2013) 109–16.

[5] C. Okoh, R. Roy, J. Mehnen, J., L. Redding, Overview of Remaining Useful Life Prediction Techniques in Through-life Engineering Services. Procedia CIRP 16 2014 158-163.

[6] I. Bazovsky, Reliability theory and practice. NewYork, NY, Dover Publications, 2004.

[7] P.P. O'Connor, A. Kleyner, Practical reliability engineering, Chichester(UK), Wiley, 2012.

[8] S. Xiao-Sheng, W.B. Wang, H. Chang-Hua, Z. Dong-Hu, Remaining useful life estimation – A review on the statistical data driven approaches, Eur. J. of Oper. Res. 213(1) (2011) 1-14.

[9] Y. Lei, L. N. Li, S. Gontarz, J. Lin, S. Radkowski, J. Dybala, A Model-Based Method for Remaining Useful Life Prediction of Machinery, IEEE Trans. Rel. 65(3) (2016) 1314 – 1326.

[10] Y. Song, D. Liu, Ch. Yang, Yu Peng. Data-driven hybrid remaining useful life estimation approach for spacecraft lithium-ion battery, Microelectron, Reliab. 75 (2017) 142-153.

[11] F. Dean, J. de Castro, J., J. Litt, User's Guide for the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS), NASA/ARL; Technical Manual TM2007-215026, NASA Center for AeroSpace Information: Cleveland, OH, USA, 2007.

[12] T. Wang, J. Yu, D. Siegel, J. Lee, A similarity-based prognostics approach for remaining useful life estimation of engineered systems, In Proceedings of the IEEE International Conference on Prognostics and Health Management (PMH 2008), Denver, CO, USA, 6–9 October 2008; pp. 1–6.

[13] R.S. Tsay. Multivariate Time Series Analysis with R and Financial Applications, John Wiley, 2013

[14] G.E.P. Box, G. M. Jenkins, Time Series Analysis. Holden-Day, 1970.

[15] F. Sánchez Lasheras, F. J. de Cos Juez, A. Suárez Sánchez, A. Krzemień, P. Riesgo Fernández. Forecasting the COMEX copper spot price by means of neural networks and ARIMA models. Resour Policy (2015) 45, 37-43.

[16] A. Krzemień, P. Riesgo Fernández, A. Suárez Sánchez, F. Sánchez Lasheras. Forecasting European thermal coal spot prices. Journal of Sustainable Mining (2015) 14, 203-210.

[17] B. E. Boser, I. M. Guyon, V. N. Vapnik, A training algorithm for optimal margin classiers. In D. Haussler editor, 5th Annual ACM Workshop on COLT, pages 145-152, Pittsburgh, PA 1992. ACM Press.

[18] I. Guyon, B. Boser, V. Vapnik, Automatic capacity tuning of very large VC-dimension classiers. In Stephen Jose Hanson, Jack D. Cowan, and C. Lee Giles editors. Advances in Neural Information Processing Systems, volume 5, pages 147-155. Morgan Kaufmann, San Mateo, CA, 1993.

[19] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.

[20] J. Holland, Adaption in Natural and Artificial Systems, first ed., Michigan Press, Ann Arbor, 1975.

[21] J.R.A. Fernández, C.D. Muñiz, P.J.G. Nieto, F.J. de Cos Juez, F.S. Lasheras, Forecasting the cyanotoxins presence in fresh waters: A new model based on genetic algorithms combined with the MARS technique, Ecol. Eng. (2013) 53, 68-78

[22] C.O. Galán, F.S. Lasheras, F.J. de Cos Juez, A.B. Sánchez. Missing data imputation of questionnaires by means of genetic algorithms with different fitness functions. J. Comput. Appl. Math. 311 (2017) 704-717.

[23] F.S .Lasheras, P.J.G. Nieto, F.J. de Cos Juez, R.M. Bayón, V.M.G. Suárez, A hybrid PCA-CART-MARS-based prognostic approach of the remaining useful life for aircraft engines, Sensors (2015) 15 (3), 7062-7083.

[24] P. Bader, S. Blanes, E. Ponsoda. Structure preserving integrators for solving (non-)linear quadratic optimal control problems with applications to describe the flight of a quadrotor. J. Comput. Appl. Math. (2014) 262, 223-233.

[25] K. S. Tuncel, M. G. Baydogan. Autoregressive forests for multivariate time series modeling. Pattern Recognition (2018) 73, 202-215.

[26] P. J. García Nieto, F. Sánchez Lasheras, E García-Gonzalo, F. J. de Cos Juez. PM 10 concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP, VARMA and ARIMA: A case study. Sci. oTotal Environ (2018) 621, 753-761.

Figure

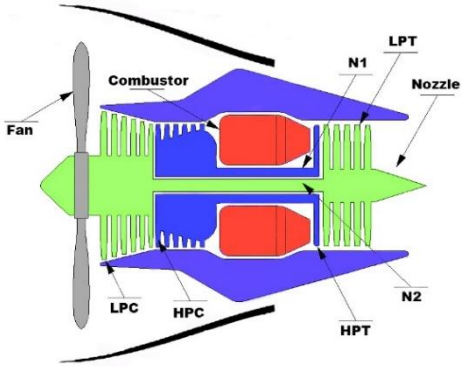

Figure 1. Simplified diagram of the simulated engine (LPC: low-pressure compressor; HPC: high-pressure compressor; LPT: low-pressure turbine; HPT: high-pressure turbine; N1 turbine axis; and N2: turbine shaft).

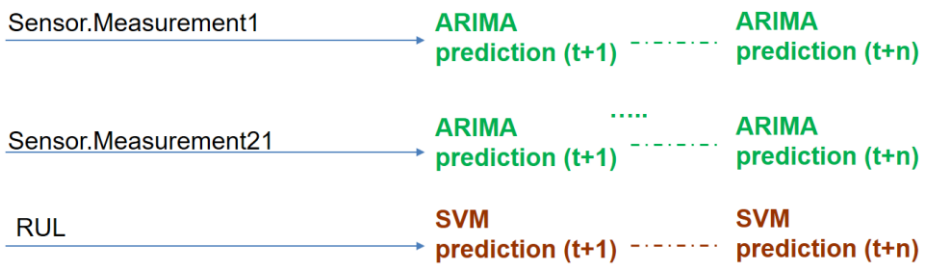

Figure 2. Diagram of the proposed hybrid algorithm.

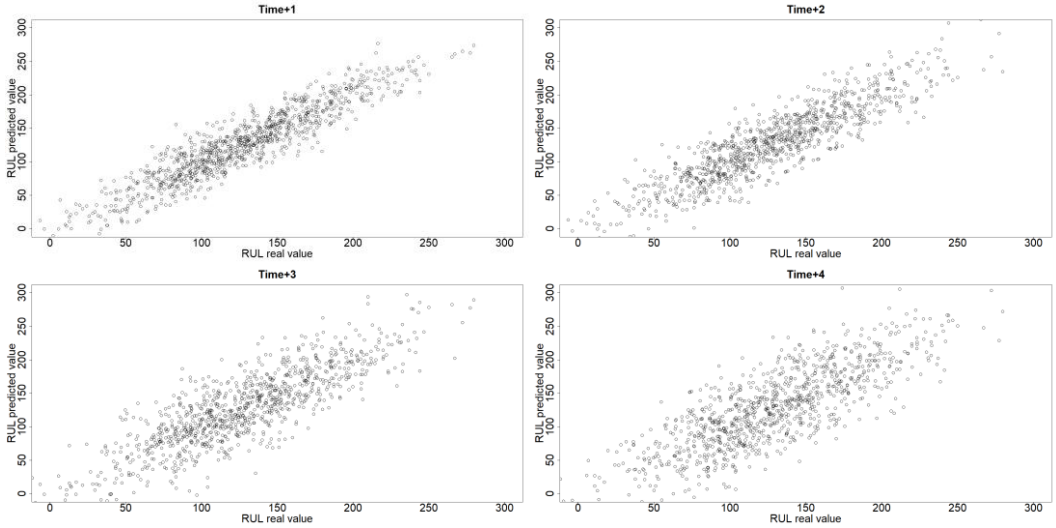

Figure 3. Scatter plots showing real versus predicted RUL a) $t+1$, b) $t+2$ c) $t+3$ and d) $t+4$.

Table

Table 1. Descriptive statistics of all the variables of the database.

Output and input variables	**Mean**	**Sd.**
Remaining useful life	108.808	68.881
Operational setting 1	$-8.870\ 10^{-6}$	0.003
Operational setting 2	$2.350\ 10^{-6}$	0.003
Operational setting 3	100.002	10^{-6}
Total temperature at fan inlet (ºR)	518.670	0
Total temperature at LPC outlet (ºR)	642.681	0.500
Total temperature at HPC outlet (ºR)	1590.523	6.131
Total temperature at LPT outlet (ºR)	1408.934	9.000
Pressure at fan inlet (psia)	14.620	3.390×10^{-6}
Total pressure in bypass-duct (psia)	21.609	0.001
Total pressure at HPC outlet (psia)	553.368	0.885
Physical fan speed (rpm)	2388.097	0.070
Physical core speed (rpm)	9065.243	22.082
Engine pressure ratio (P50/P2)	1.300	4.66×10^{-13}
Static pressure at HPC outlet (psia)	47.5412	0.267
Ratio of fuel flow to Ps30 (pps/psi)	521.414	0.738
Corrected fan speed (rpm)	2388.096	0.719
Corrected core speed (rpm)	8143.753	19.076
Bypass ratio	8.442146	0.038
Burner fuel-air ratio	0.0300	1.56×10^{-14}
Bleed enthalpy	393.212	1.549
Demanded fan speed (rpm)	2388.000	0
Demanded corrected fan speed (rpm)	100.000	0
HPT coolant bleed (lbm/s)	38.8163	0.181
LPT coolant bleed (lbm/s)	23.279	0.108

Table 2. R^2, RMSE and MAE obtained for the RUL forecast with the proposed hybrid algorithm for times in advance from 1 to 5.

time	R^2	RMSE	MAE
$t+1$	0.9315	39.6843	27.6837
$t+2$	0.8979	41.3629	28.7352
$t+3$	0.8469	45.2593	29.0939
$t+4$	0.7456	47.6721	31.7868
$t+5$	0.6662	50.8108	32.6096

Table 3. Variable importance according to their influence on the R^2 of the model.

Input variables	R^2 influence
LPT coolant bleed (lbm/s)	0.1505
HPT coolant bleed (lbm/s)	0.1420
Ratio of fuel flow to Ps30 (pps/psi)	0.1379
Total pressure at HPC outlet (psia)	0.1146
Bypass ratio	0.1075
Total temperature at LPC outlet (ºR)	0.0638
Bleed enthalpy	0.0547
Corrected fan speed (rpm)	0.0469
Corrected core speed (rpm)	0.0365
Physical core speed (rpm)	0.0239
Engine pressure ratio (P50/P2)	0.0076
Total pressure in bypass-duct (psia)	0.0067
Operational setting 1	0.0065
Operational setting 3	0.0054
Burner fuel-air ratio	0.0028
Remaining useful life	0.0021
Operational setting 2	0.0017
Pressure at fan inlet (psia)	0.0006