Remaining Useful Life Prediction of Aero-engines by Appropriate Utilization of Multi-sensor Signals

Bin Zhang*, Ning Huang, Jiufei Luo, Yi Zhang
School of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China

*Corresponding author: zhangbin@cqupt.edu.cn

Abstract. This paper presents the prediction of remaining useful life (RUL) with appropriate fusion of multi-sensor signals for the aero-engine, which is the heart of an aircraft. With the rapid development of information technology, health condition of one aero-engine is usually monitored with multiple sensors. To properly utilize these multi-sensor condition information for degradation modeling and RUL prediction is one of the key challenges for condition-based maintenance of the whole aircraft. Thus this paper proposes one statistical method based on health indicator (HI) construction and empirical parametric model for aero-engines RUL prediction. The method is validated with run-to-failure data sets of an aircraft gas turbine engine test-bed developed by NASA. Results show that the proposed method can effectively fuse multi-sensor signals to describe the degradation and predict RUL of the aero-engine.

Keywords: Remaining useful life prediction, multi-sensor fusion, aero-engine, health indicator, statistical modeling.

1. Introduction
Health condition of an aero-engine directly affects the operational safety, reliability, and efficiency of the whole aircraft. To monitor operational health of the aircraft engine, multiple sensors of varying types are usually embedded or equipped and thus brings a multiple sensor environment for health analytics of aero-engines. How to utilize the multi-sensor signals for condition-based maintenance is one of the major challenges for aircraft applications [1]. Recently, remaining useful life (RUL) prediction, for its potential in enhancing reliability and maintenance efficiency as well as reducing operational costs, has been attracting attention from both academics and industry.

Practically, RUL prediction aims to prognosticate the left time of a component or system from the current time to the time of functional failure. Then with the estimated RUL, maintenance actions can be prepared beforehand and be done as required. According to the literature, the approaches for the RUL prediction of aero-engines can be classified into model-based and data-driven methods [2]. In model-based methods, a model based on the physics-of-failure knowledge is normally constructed to describe the engine health degradation and do future prediction. Markov model [3] and particle filtering [4] are typical model-based methods for the aero-engine health condition and RUL prediction. However, certain assumptions or simplifications of model-based methods pose limitations on their practical deployment. With the development of big data technique, data-driven methods may be the...
more appealing solution for aero-engines that are with limited knowledge of physics-of-failure but multi-sensor signals. For the data-driven methods, machine learning and statistical approaches are the two popular branches. The machine learning methods based on traditional shallow and emerging deep learning architectures are suitable for prognosis of complex systems as aircraft engines, however the predicted results are hard to explain due to the lack of transparency [5]. On the contrary, the aero-engine monitoring data can be fitted by some parametric or non-parametric statistical models for RUL prediction [6]. Thanks to the good mathematical properties and easy interpretations, RUL prediction of aero-engines using statistical methods have been recently gaining popularity.

In statistically data-driven RUL predictions of one aircraft engine, a health indicator (HI) that describes its condition degradation is constructed by appropriate utilization of the monitoring sensor data. Traditionally, one dominant sensor that reflects the engine health condition is directly chosen as the HI [7]. However, this may be invalid in many cases as single sensor only contains partial or limited information, which is far from enough to comprehensively capture the physical degradation process of the aero-engine. Therefore, combination of multi-sensor signals into a synthesized and virtual HI may be more reasonable. In general, utilization of more sensor signals can be helpful. Unfortunately, it is possible that some sensors are irrelevant to the aero-engine health evolution tracking, due to low sensitivity to the degradation, large interferences or inconsistency among the engine population. To address these issues, sensors that are more degradation-related should be selected. Currently, sensors are mainly selected based on some degradation relevance metrics. Then with the optimally selected sensors, HI can be constructed with one information fusion method, such as PCA [8] and linear data transformation [7]. To secure accurate and efficient RUL predictions, it is highly desired that the constructed HI enables the degradation of one aero-engine be characterized and also the degradation trajectory of each aero-engine from the whole population be consistent. For this purpose, a generic HI construction method to fuse the optimally selected sensors is presented in this work.

After the fusion of multi-sensor signals of the aero-engine into HI, statistical models are applied to fit the HI for degradation modeling and RUL prediction. Random coefficient models and stochastic process models are the two frequently used groups of statistical approaches to model the available data for prognostic inference. Lasheras et al [9] proposed a hybrid multivariate adaptive regression splines method for the aircraft engine RUL prediction. In [10], an exponential model was proposed for prognosis and validated on degradation data set of an aircraft gas turbine engine. The parametric models are effective when the HI can be well modeled with a specific model, however there are also numerous cases where the aero-engine HI may not be properly fitted by parametric ones. Thus, more flexible degradation modeling and RUL prediction methods are needed for practical aero-engine applications. To partially make up this gap, one functional data analysis method was proposed for health prognostics of aircraft engines in [11]. For its effectiveness and easy of application, the empirical parametric approach is studied to fit the constructed HI for RUL prediction of aero-engines in this paper.

2. The proposed method

In this section, the procedures for fusion of multi-sensor signals into a generic HI and the studied parametric models for RUL prediction of aero-engines are detailed.

2.1. A generic HI construction method

To secure the effectiveness and efficiency of RUL prediction of aero-engines, some pre-processings should be carried out before integration of the available multi-sensor signals into a HI. As not all the original monitoring sensors are relevant to the true health degradation, informative sensors need to be selected. In view that effectiveness of the four metrics of Corr, Mon, Rob and Pre has been validated for relevant sensor selection of aircraft engines prognostics in [11], they are also studied here for informative sensor selection. Nevertheless, these metrics were proposed for the scenario of one operational condition with one failure mode. In this regard, one two-step strategy is presented for multiple operation conditions and multiple failure modes. In the first step, the multi-sensor signals of
each aero-engine under multi-operation conditions are divided into each specific operational condition with clustering or normalization. In the second step, the failure of individual aero-engine is attributed to one mode by clustering the sensor measurements upon failure instant or tracking the varying trend of the multi-sensor signals. Through the above pre-processings, the original multi-sensor signals can be selected using the four metrics. Specifically, denote $s_{mk}$ as the thus selected $k$th sensor for the $m$th failure mode ($m = 1, 2, ..., M; k = 1, 2, ..., K_m$).

In addition, certain kind of reference baseline for construction of a HI should be established. Most often, the initial stage when one aero-engine is deemed to operate at its normal state is chosen as the baseline. Then later aero-engine operational conditions are compared to this baseline, and a HI can be constructed using some kind of metrics, such as Euclidean distance or probability density matching. Actually, the thus constructed HIs measure the deviations from the aero-engine normal operating state. Nevertheless, due to material impurities or manufacturing errors etc., there are large variations in the initial normal stage, which impacts its effectiveness as a baseline. Also, aero-engines that are with similar normal operating status can fail of distinct modes. Thus it is more reasonable to take the sensor measurements upon failure as the baseline.

Based on the above analysis, a generic one-dimensional HI is constructed by fusion of the selected muti-sensor signals with consideration of multi-failure modes as follows.

$$y_i^m = \left( \frac{\sum_{k=1}^{K_m} (s_{mk}^o - s_{mk}^m)^2}{\sum_{n=1}^{N} \sum_{k=1}^{K_m} (s_{mk}^o - s_{nk}^o)^2/(N-1)} \right)^{1/2}$$

Where $y_i^m$ and $s_{mk}^o$ are respectively the HI and the $k$th sensor measurement at the time $t$ for the aero-engine under the $m$th failure mode, $s_{mk}^o$ is the measurement upon failure instant of the $k$th sensor for the $m$th failure mode, and $s_{nk}^o$ is the mean of the measurements upon failure instants of $N$ historically failed aero-engines.

2.2. The empirical parametric models
With the accumulation of damages (such as wear and fatigue), the health condition of an aero-engine gradually degrades. Therefore, properly tracking the degradation and accurately predicting RUL is of great significance for condition-based maintenance of aero-engines. That is the constructed HI should be modeled and be extrapolated for future prediction.

For a statistical method, the HI trajectory is usually fitted with one kind of empirical model for degradation modeling and prediction. The exponential (Exp) and power law (Pow) are two empirical parametric models that have been frequently researched in statistical modeling. They are utilized to fit the aero-engine HI with the functional form as Eq. (2) in this work. Additionally, the aero-engine HI trajectory is also fitted by the general $p$th polynomials (Poly).

$$y_i = x(t_i) + \epsilon$$

Exp: \( x(t) = a \times \exp(bt) + c \)

Pow: \( x(t) = a \times t^b + c \)

where $y_i$ and $x(t_i)$ is the observed and latent true HI at the time $t_i$ of one aero-engine, $\epsilon$ is the noise term that is assumed to be normally distributed as $N(0, \sigma^2)$, and $a$, $b$, and $c$ are fitting coefficients.

To fairly compare performance of the above empirical parametric models for the HI modeling, the three commonly used evaluation criteria as in Eq. (3), i.e. sum of square error (SSE), root mean square error (RMSE) and the coefficient of determination (R-square), are employed in the following analysis.

$$\text{SSE} = \sum_{i=1}^{T} (y_i - \hat{y}_i)^2,$$
$$\text{RMSE} = \left( \frac{\sum_{i=1}^{T} (y_i - \hat{y}_i)^2}{T} \right)^{1/2},$$
$$\text{R-square} = 1 - \frac{\text{SSE}}{\text{SST}}, \quad \text{SST} = \sum_{i=1}^{T} (y_i - \text{mean}(y_i))^2\]
Where \( y_i \) and \( \hat{y}_i \) are respectively the true and estimation at the time \( t_i \) of one aero-engine, and \( I \) is the total number of monitoring time instants.

3. Case studies
To validate the proposed method, run-to-failure data sets of turbofan engine was simulated using the dynamic tool of C-MAPSS that was developed by NASA [12]. In the output data sets, each individual run-to-failure instance of the same aero-engine unit was represented by multi-sensor signals. To be specific, 21 sensor measurements were recorded for every cycle of each aero-engine run-to-failure instance. The multi-sensor signals were contaminated with measurement noises and each engine started with different initial health conditions and manufacturing variations which were unknown. In this paper, FD003 data sets containing 100 training and 100 testing data of aero-engines is studied to illustrate and demonstrate the effectiveness and performance of the presented method. These engines were simulated under one operational condition with two failure modes (i.e. HPC and fan). More details about the simulation and data set can be found in [12]. Also, note that deterioration of the aero-engine with multiple usage conditions can also be analyzed by the method with the pre-processing as discussed in Subsection 2.1.

3.1. Results and analysis of the constructed HI
As there are two failure modes in the aero-engine data set, clustering of the failures is conducted. The results indicate that there are respectively 56 and 44 training engines for failure modes of HPC and fan. Then the run-to-failure multi-sensor signals of the training aero-engines from each failure mode are evaluated by the four metrics of Corr, Mon, Rob and Pre. Since the sensor selection for the HPC failure mode has already been studied in [11], only the results for the fan failure mode are presented here in Table 1. For the metrics of Corr, Mon and Rob that are defined on sensors from one engine, statistics of the their means are given based on results of the 44 engines.

| Sensor  | Corr   | Mon  | Rob  | Pre  | Sensor  | Corr   | Mon  | Rob  | Pre  |
|---------|--------|------|------|------|---------|--------|------|------|------|
| T2      | NaN    | 1.000| 1.000| NaN  | phi     | 0.8509 | 0.2956| 0.9995| 0.9500|
| T24     | 0.7699 | 0.0517| 0.9997| 0.8086| NRf     | 0.8458 | 0.2233| 0.9996| 0.9310|
| T30     | 0.8066 | 0.0548| 0.9982| 0.8219| NRc     | 0.8524 | 0.1742| 0.9997| 0.9547|
| T50     | 0.7932 | 0.0755| 0.9979| 0.8729| BPR     | 0.8518 | 0.0680| 0.9982| 0.8833|
| P2      | NaN    | 1.000| 1.000| NaN  | farB    | NaN    | 1.000| 1.000| NaN  |
| P15     | 0.5039 | 0.6569| 0.9999| NaN  | htBleed | 0.8144 | 0.2527| NaN  | 0.8459|
| P30     | 0.8519 | 0.2605| 0.9994| 0.9502| Nf_dmd  | NaN    | 1.000| 1.000| NaN  |
| Nf      | 0.8443 | 0.2251| 0.9998| 0.9352| PCNfR_dmd| NaN    | 1.000| 1.000| NaN  |
| Nc      | 0.8508 | 0.1660| 0.9997| 0.9418| W31     | 0.8226 | 0.0767| 0.9981| 0.8413|
| epr     | 0.7254 | 1.000| 1.000| 0.8300| W32     | 0.8235 | 0.0610| 0.9982| 0.8190|
| Ps30    | 0.8076 | 0.1190| 0.9984| 0.9072|         |        |      |      |      |

From the results, it can be observed that there are default values (i.e. NaN) or 1 for eight sensors (i.e. T2, P2, P15, epr, farB, htBleed, Nf_dmd and PCNfR_dmd). A further analysis shows that the above sensor measurements are constant values or having a step-wise trend with the engine operating cycle. These sensors are not helpful for the RUL prediction, so they are excluded. For the other 13 sensors, there seems no significant difference. The Nf and NRf are the same kind of sensors, and Nf is shown in Fig. 1. It fluctuates largely at the early operating stage and is with a very small varying range. Thus the two sensors should be excluded for the aero-engine HI construction. Compared to the failure mode of HPC researched in [11], the two sensors Nc and NRc demonstrate a consistently increasing tendency. This may be because that the Nc and NRc can well capture the engine degeneration under the fan failure mode while these two sensors are poor for sensing the HPC failure mode. Considering that it is difficult to identify the true failure mode of an in-service aero-engine at its early operational
stage, the Nc and NRc are not selected to have a balance between the two failure modes. That is the remaining nine sensors are finally selected for HI construction of the fan failure mode.

With the selected nine-dimensional sensors and clustering of the sensor measurements upon failure for each failure mode, the aero-engine HI is constructed following Eq. (1). Results of the 100 training engines under the two failure modes are shown in Fig. 2, which depicts a general decreasing tendency. Fig. 2 (a) shows that under the HPC failure mode, the trend of the constructed HI for the 56 engines is relatively consistent and the degrading rate increases gradually during the whole life. While for the fan failure mode, the health of the 44 engines degrades slowly at the beginning and exacerbates at the later stage. It can also be observed that there is large dispersion for the engine lives even under the same failure mode.

3.2. Results and analysis of the RUL prediction

As stated in Subsection 2.2, the constructed HI is fitted using the empirical parametric models of Exp, Pow and Poly. Specifically, 3rd polynomials (Poly3) from the Poly type is considered. The HI modeling results for the two failure modes of FD003 training engines are summarized in Table 2. Statistically, the means for the three criteria of SSE, RMSE and R-square are given. To have a view of the curve fitting, the modeling for one engine HI is plotted in Fig. 3.

Referring to the results in Table 2, it can be concluded that the three empirical parametric models fit the trend of the noisy HI accurately. When it comes to the two kinds of failure modes, the two
Table 2. HI fitting results of the 100 training engines.

| Failure mode | Method | SSE     | RMSE   | R-square | Failure mode | Method | SSE     | RMSE   | R-square |
|--------------|--------|---------|--------|----------|--------------|--------|---------|--------|----------|
| HPC          | Exp    | 88.83   | 0.6650 | 0.7964   |              | Exp    | 102.72  | 0.5814 | 0.8864   |
|              | Pow    | 88.87   | 0.6669 | 0.7955   | fan          | Pow    | 103.43  | 0.5834 | 0.8855   |
|              | Poly3  | 87.86   | 0.6648 | 0.7976   | Poly3        | 103.61  | 0.5857  | 0.8852  |

Figure 3. HI fitting results for one training engine from the failure mode of: (a) HPC; (b) fan.

The criteria of RMSE and R-square show that better degradation modeling results are obtained for the fan failure mode, which may be explained by the fact that more number of monitoring HIs are generally available to fit the three models for aero-engines have much longer operating cycles (lifetimes) under this failure mode. This is also in accordance with the bigger SSE values.

To further study the effectiveness of the proposed method for RUL prediction, original monitoring sensor data sets of the testing aero-engines from FD003 are processed following the same procedure as for the training units. With pre-processing the multi-sensor signals of the 100 testing engines, it is found that 39 engines fail of the HPC mode and 42 engines fail of the fan mode, while the remaining 19 ones can not be classified to either of the two failure modes. Then HI that describes the health degeneration of each testing engine is constructed by fusing the selected nine-sensor signals.

In statistical prognostics, the HI is extrapolated to the future for RUL prediction, thus a failure threshold requires to be predefined. Although there are some ISO standards that may be helpful for

Figure 4. Illustration for RUL prediction of one testing engine.
Table 3. Comparisons of RUL prediction results for the 100 testing engines.

| Failure mode | Method | RMSE  | Range of prediction errors |
|--------------|--------|-------|----------------------------|
| Known HPC    | Exp    | 36.83 | [-96, 110]                 |
|              | Pow    | 41.78 | [-111, 113]                |
|              | Poly3  | 29.74 | [-111, 43]                 |
| fan          | Exp    | 40.20 | [-91, 113]                 |
|              | Pow    | 47.93 | [-68, 113]                 |
|              | Poly3  | 49.31 | [-178, 150]                |
| Unknown Treated as HPC | Exp | 60.20 | [-126, 65] |
|              | Pow    | 57.90 | [-91, 113]                 |
|              | Poly3  | 53.97 | [-97, 120]                 |
| Treated as fan | Exp | 128.60 | [-100, 287] |
|              | Pow    | 125.62 | [-45, 200]                |
|              | Poly3  | 95.58 | [-83, 168]                 |

determination of the failure threshold, they are documented for general applications and may not be directly applicable for specific situations, especially for complex aircraft engine systems whose health deterioration is tracked with the constructed virtual HI as in this work. Thus the threshold for the two failure modes of HPC and fan are respectively determined through a statistical analysis of the constructed HI from the training engines. In Fig. 4, the RUL prediction for one testing engine from the HPC failure mode is illustrated.

With the constructed HI and the failure threshold, RULs of the 100 testing engines are predicted with the three empirical parametric models. The results are listed in Table 3, where RMSE and Rang of prediction errors (i.e. predicted RUL – true RUL) are calculated to demonstrate the performance of the methods. For the 19 engines with unknown failure mode, they are artificially treated as one of the two failure modes to predict RULs.

The statistical results in Table 3 indicate that RUL can be predicted more accurately when the failure mode information is known. Further, the prediction RMSE of the HPC failure mode is smaller than that of the fan failure mode. This may partially because the different varying patterns of the HI for the two failure modes, as shown in Fig.2 and discussed in Subsection 3.1. Among the three empirical models, the Poly3 model generally gives the best RMSE performance for the testing engines. However, all the three models are not with satisfactory performance in view of the long range of prediction errors. To make a sense of that, the RUL prediction result for each testing engine is shown in Fig. 5 and Fig. 6.

Figure 5. RUL prediction results for testing engines of the known failure modes: (a) HPC; (b) fan.
Figure 6. RUL prediction results for testing engines with unknown failure modes: (a) treated as HPC; (b) treated as fan.

Note that the testing engine No. is sorted with the increasing RULs in both Fig. 5 and Fig. 6. From these two figures, similar conclusions as those of Table 3 can be firstly drawn. Furthermore, it can be seen that more accurate RUL predictions can be achieved with the approaching of failure (i.e. small RULs), especially with the case of known failure mode as in Fig. 5. To artificially assign engines with certain failure modes may obtain both relatively accurate and poor predictions (Fig. 6). Thus the failure mode information is helpful to secure the accuracy.

4. Conclusions
The aero-engine is the heart of one aircraft and its RUL is one of the major issues for condition-based maintenance of the aircraft or the whole fleet. For RUL prediction with multi-sensor signals, one statistical method that appropriately fuses the multi-sensor information into one generic HI and then fits the HI with empirical parametric models is proposed in this study. Case studies are carried out with the simulation data set of one aircraft gas turbine engine. The results demonstrate that the presented method can effectively utilize the selected sensors to model and predict the RUL of the aircraft engine. In this work, the scenario of single operational condition with two failure modes is studied, validation of the presented method for multiple operating conditions under multiple failure modes will be pursued in the future.

Acknowledgments
This work was financially supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJ1704109) and the Open Found of The State Key Laboratory of Mechanical Transmissions (Grant No. SKLMT-KFKT-201809).

References
[1] H. Hanachi, C. Mechefske, J. Liu, A. Banerjee, Y. Chen. Performance-based gas turbine health monitoring, diagnostics, and prognostics: A survey. IEEE Transactions on Reliability, 2018, 67, 1340-1363.
[2] Z. X. Li, K. Goebel, D. Z. Wu. Degradation modeling and remaining useful life prediction of aircraft engines using ensemble learning. Journal of Engineering for Gas Turbines and Power, 2019, 141, 041008.
[3] S. Al-Dahidi, F. Di Maio, P. Baraldi, E. Zio. Remaining useful life estimation in heterogeneous fleets working under variable operating conditions. Reliability Engineering and System Safety, 2016, 156, 109-124.
[4] J. B. Yu. Aircraft engine health prognostics based on logistic regression with penalization
regularization and state-space-based degradation framework. Aerospace Science and Technology, 2017, 68, 345-361.

[5] C. G. Huang, H. Z. Huang, Y. F. Li. A Bi-Directional LSTM prognostics method under multiple operational conditions. IEEE Transactions on Industrial Electronics, 2019, 66, 8792-8802.

[6] X. S. Si, W. B. Wang, C. H. Hu, D. H. Zhou. Remaining useful life estimation – a review on the statistical data driven approaches. European Journal of Operational Research, 2011, 213, 1-14.

[7] P. F. Wang, B. D. Youn, C. Hu. A generic probabilistic framework for structural health prognostics and uncertainty management. Mechanical Systems and Signal Processing, 2012, 28, 622-637.

[8] K. Le Son, M. Fouladirad, A. Barros. Remaining useful lifetime estimation and noisy gamma deterioration process. Reliability Engineering and System Safety, 2016, 149, 76-87.

[9] S. F. 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, 7062-7083.

[10] K. B. Liu, S. Huang. Integration of data fusion methodology and degradation modeling process to improve prognostics. IEEE Transactions on Automation Science and Engineering, 2016, 13, 344-354.

[11] B. Zhang, K. Zheng, Q. Q. Huang, S. Feng, S. Q. Zhou, Y. Zhang. Aircraft engine prognostics based on informative sensor selection and adaptive degradation modeling with functional principal component analysis. Sensors, 2020, 20, 920.

[12] A. Saxena, K. Goebel, D. Simon, N. Eklund. Damage propagation for aircraft engine run-to-failure simulation. In Proceedings of the International Conference on Prognostics and Health Management, Denver, CO, USA, 6-9 October 2008.