An unsupervised approach for health index building and for similarity-based remaining useful life estimation

Sébastien Schwartz[a,*], Juan José Montero Jiménez[a,c], Rob Vingerhoeds[a], Michel Salaün[a,b]

[a] ISAE-SUPAERO, 10 Avenue Edouard Belin, Toulouse 31400, France
[b] Institut Clément Ader, 3 Rue Caroline Aigle, Toulouse 31400, France
[c] Tecnológico de Costa Rica, Calle 15, Avenida 14, Cartago 30101, Costa Rica

A B S T R A C T

Predictive maintenance techniques attempt to trigger a maintenance intervention at the right moment by estimating the life expectation. Predictive maintenance is increasingly implemented by automated approaches able to perform diagnostics and prognostics. The main part of recent research in these approaches is focused in machine learning structures whose reasoning is implicit and cannot be easily explained. This poses a problem for their implementation in highly constrained area such as aeronautics. To overcome this constraint, explicit reasoning approaches such as the Similarity-Based Model (SBM) can be implemented. The SBM has been widely used for fault diagnostics and the remaining useful life (RUL) estimation, but the development of SBM includes tasks that often rely on high skilled experts. For instance, data reduction techniques required for SBM are often performed by experts judgment whose outcomes are not always consistent. The produced features from these techniques are used to build the Health Index that can be used to create the degradation trends that serve as a reference for the SBM. To overcome these difficulties, an automatic and unsupervised approach based on the Kernel Principal Component Analysis is proposed to enhance the Health Index creation. It preserves as much of the sensor information as possible improving the similarity-based RUL estimation. Additionally, when estimating the RUL of a system, the most similar degradation trends stored in the SBM library are used to compute individual RULs, the final RUL is obtained by a fusion rule technique that combines all these individual RULs into a consolidated value. For the fusion rule techniques, a self-adaptive method that does not rely on human expertize is proposed. This fusion rule can benefit of the accumulated knowledge over the SBM operation. This unsupervised approach to develop a SBM is validated with promising results against an equivalent and supervised algorithm that came out best in the 2008 prognostic health management challenge.

Keywords:
Kernel principal component analysis
Predictive maintenance
Prognostic
Similarity-based
Remaining useful life

1. Introduction

Predictive maintenance refers to a specific maintenance strategy, aiming at identifying incipient faults, forecasting future failures and at triggering the maintenance actions accurately when needed (Montero Jimenez et al., 2020). Maintenance strategies are often classified into three categories: (1) corrective maintenance where actions are performed to restore a system after a breakdown or a deteriorated functional behavior, (2) preventive maintenance where maintenance actions are performed at a fixed operational interval, and (3) predictive maintenance where specific maintenance actions are based on measurements on the concerned system. Remaining Useful Life (RUL) estimation plays an important role in predictive maintenance; it provides insight on the system deterioration due to faults appearance or wear, and tries to show when the system would no longer perform its intended function. As such, accurate RUL estimation improves safety, reliability and availability of the system. It attempts to avoid sudden breakdowns minimizing unnecessary maintenance time / cost (Gu and Chen, 2016). Therefore, it is important to focus on techniques that can improve RUL predictions, especially on high-risk engineering systems (Li et al., 2018; Yu et al., 2019; Zhao et al., 2017).

According to Ramasso and Saxena (2014), prognostics for maintenance can be divided into three broad categories as: functional mappings between set of inputs and RUL, such as Li et al. (2019), Jiang et al. (2019), functional mapping between Health Index (HI) and RUL, such as Hassani et al. (2019), Climente-Alarcon et al. (2017), and similarity-based matching, such as Bleakie and Djurdjanovic (2013), Liang et al. (2019), Zhang et al. (2019). The main focus of this work is on the use in aeronautics, where it is important that the

* Corresponding author.
E-mail address: sebastien.Schwartz@isae-supaero.fr (S. Schwartz).

https://doi.org/10.1016/j.compind.2022.103716

approaches can be perfectly understood and explained fully. The last category, similarity-based matching, is therefore interesting as it contains approaches that are more explainable than machine learning approaches. The current study focuses on similarity matching for RUL computation. A Similarity-Based Model (SBM) uses operational records to represent historic degradation trends for a given system (e.g. an aircraft, a vehicle, etc.), that are later used to assess the degradation of systems of the same type (Wang et al., 2008). These degradation trends are built upon a representative set of run-to-failure system instances (e.g. engines, bearings, battery cells, etc.). A physical degradation trend should be generated for each system instance. The RUL for these degradation trends is known and they are stored in a library. Prognostic assessments are performed by comparing the degradation evolution of a system against those stored in the SBM library. The degradation trends can be represented by direct measures such as sensor records, or indirect measures such as health indexes computed from sensor records. The use of HI is often privileged over raw sensor data as it can consolidate several data sources in a single value that facilitates the understanding and visualization of a system degradation. Another important aspect in the SBM for RUL estimation is the uncertainty management in the prognostics. A SBM can retrieve multiple reference degradation trends from its library that are similar to the current degradation of a system under test, but with different RUL estimations. A fusion rule is often used to obtain a consolidated value of RUL, considering the contribution of the different reference degradation trends that have been retrieved by the SBM.

The HI development and the selection of suitable reference degradation trends to estimate a RUL are complex tasks that rely on experts' knowledge and judgment. This study attempts to simplify these processes by the implementation of data engineering techniques such as the Kernel Principal Component Analysis (KPCA) for the HI development and a self-adaptative fusion rule for RUL estimation. The KPCA can help to automate and simplify the feature extraction and data processing tasks that are exclusively performed by experts. Performing manual data reduction requires human knowledge and expertize in order to produce a proper feature subset. Manual tuning on feature techniques is done if the case study needs it. Automatic approaches, that are not specific to a specific case study, are promising, as they make the approach more consistent, as opposed to humans that may or may not always perform work in the same way. In addition, automatic approaches allow to estimate the loss of the data and try to come up with something as close as possible to reality. The proposed fusion rule also attempts to automate the selection of the reference degradation trends retrieved by the SBM to compute the final RUL of a system unit under test. RUL can be presented as a discrete value or as an interval. Both options are studied in the proposed SBM. The goal is to facilitate the implementation of SBM for RUL computation even for those professionals who are not specialists in HI development, health/degradation modeling, and fusion rules for RUL estimation using SBM.

The paper is organized as follows: Section 2 presents the state of the art in SBM for RUL computation and justifies the current work objectives. Section 3 presents the methodology leading to the RUL estimation using the KPCA technique and a SBM. Section 4 presents the case study selected to assess the proposed approach, the 2008 Prognostic Health Management (PHM08) challenge Wang et al. (2008). Section 5 presents the results and discussion of the proposed approach implementation on the case study. Section 6 concludes the paper and proposes future work perspectives.

2. Related work

Prognostics in maintenance can be carried out by different types of models (Montero Jimenez et al., 2020; Berri et al., 2021). These models can be classified in three main families: data-driven models, knowledge-based models, and physics-based models. Models from these families can be combined and these multi-model combinations are often called hybrid models. Within the data-driven models the Similarity-Based Models (SBM) are found. SBM have been widely used for prognostics purposes as they are relatively simple to implement and provide results that are explainable.

One of the main challenges in prognostics models is related to the Health Index (HI) generation that allows the study of degradation trends of a system of interest. For a SBM, these degradation trends are stored in the SBM library and are used as reference to assess the degradation of other systems of the same type. There have been some attempts to automate the creation of HI from raw data for SBM. For example, in Wang et al. (2008), sensor reduction is performed, followed by a linear regression on the remaining sensor set to compute the HI. Then, an exponential nonlinear regression model is used to generate the degradation trends. In Huang et al. (2019), HIs are produced by multi-linear regressions and trajectory similarity-based prediction is used for real-time estimation of remaining useful life. In order to improve RUL estimation, several variables can be considered to build the HI, increasing the complexity in the creation of the HI. For example, Cai et al. (2020) uses a similarity matching procedure to query similar run-to-failure profiles from historical data library. The selection is performed by the technique of the kernel two sample test (Gretton et al., 2013). The probability distribution of the RUL is obtained by a Weibull analysis. Other approaches use artificial neural networks with similarity-based approaches such as in Yu et al. (2020), or Bektas et al. (2019).

The level of complexity increases with the data size for training the SBM. Large databases can demand high computational resources to compute the HI and obtain the degradation trends. To overcome these obstacles, data reduction is crucial. However, reducing the data and keeping the relevant information poses a challenge to predictive maintenance systems developers (Kumar and Galar, 2018). Two main techniques are often implemented for the data reduction: feature selection and feature extraction. Feature selection techniques reduce a data set by discarding inappropriate or redundant variables (features), but may result in a loss of information from discarded features (Guyon and Elisseeff, 2003). Feature extraction techniques attempt to extract new features from an initial set of data measurements. Feature extraction tends to preserve more information than feature selection after the data reduction. It involves a transformation that is often not reversible due to information loss during the process, an example of feature extraction in the principal component analysis Pearson (Pearson, 1901). An example of this can be found in Chen et al. (2017), where sensors showing evident trends are manually selected and a Principal Component Analysis (PCA) is later performed to create a new basis generated by the two principal components. In this new basis, a failure center is determined. It corresponds to a point reached by system instances when a failure occurs. The distance between a projected instance and the failure center is used to represent the system health state. An extension of the PCA is the Kernel Principal Component Analysis (KPCA) which can be used to generate a degradation indicator by combining KPCA outputs (i.e. principal components) with Arrhenius and Eyring models (Feng et al., 2016). Model parameters are estimated with the maximum likelihood estimation. As part of the current study contributions, a KPCA is implemented to build a HI, reducing the amount of data that allows a reduction of computation power to be processed, but also avoiding the loss of important information during the data reduction.

Another challenge of SBM for prognostics in maintenance is related to the RUL computation. When a new system unit is tested with the SBM, several reference degradation trends can be retrieved from the SBM library with a close similarity in terms of the degradation evolution but with different RUL estimations. To overcome this uncertainty problem, fusion rules are often implemented in SBM to estimate a consolidated value from a contribution of the individual RUL_i obtained from comparing a testing unit against the reference degradation trends in the SBM library. Khelif et al. (2014) and Bektas et al. (2019) propose a fixed number of individual RUL_i to be fused which is fixed by a human expert while Wang et al. (2008) considers all available RUL_i and proposes some expert rules to discard non-representative ones. However, the selection of the individual RUL_i that will be used to estimate the final RUL is not a trivial task and has a high dependence on human expert judgments that will select the number of RUL_i to be considered or will discard manually non-representative reference degradation trends from the available options. As these fusion rules highly rely on human experts, there is no possibility to incorporate new knowledge acquired during the SBM operation. The current study proposes a self-adaptive method that autonomously proposes a set of individual RUL_i to compute the final RUL. The goal is not only to avoid the dependence on the human experts to discard non-representative degradation trends, but also to incorporate the accumulated knowledge during the SBM operation that can be used to refine the fusion rule.

This study attempts to improve RUL estimation using a SBM. The contributions of the current study can be seen as:

1. An unsupervised building of a Health Index with limited information loss by the implementation of a KPCA in the data preprocessing to train the SBM.
2. A self-adaptive fusing rule to estimate the final RUL without the dependence of human experts manipulation that is also capable to incorporate accumulated knowledge acquired from the tested units over the SBM operation.

3. Proposed methodology

3.1. Overview

A Similarity-Based Model (SBM) development includes two main parts: the creation of the library of known references which in the case of predictive maintenance are the known degradation references, and the similarity measures that will compare a target data record against the known references in the SBM library. In the current study, this comparison will allow to determine the RUL of a system. Fig. 1 summarizes the proposed SBM model for RUL estimation. It has been divided in three main steps: data processing, performance assessment and RUL estimation.

On the left side of Fig. 1 the training of the SBM is explained. The raw training data is assessed with the Kernel Principal Component Analysis (KPCA) to obtain the principal components to build the health index with a reduced data but keeping as much degradation information as possible. The health index for each operational cycle of each run-to-failure system in the training data can be used to build the degradation reference trends.

On the right side of Fig. 1 the RUL estimation process is presented. The data from the target system to be assessed is processed (the same way as for the training of the SBM) in order to obtain the actual health index. Then, by accumulating a health index from each operational cycle of the target system a degradation trend can be obtained. This degradation trend is compared against those in the SBM Library to obtain individual RULs from each comparison. At this point, several comparisons with a very high similarity can be identified, but often these multiple comparisons provide very different RUL estimations. To overcome this uncertainty problem a fusion rule is proposed in order to obtain the actual RUL of the evaluated system. This chapter addresses the theoretical background of the proposed SBM for RUL estimation and extends the explanations of the main steps presented on the right side of Fig. 1.

3.2. Similarity-based model library

A vital component in a Similarity-Based Model (SBM) for prognostics in predictive maintenance is the library that stores known degradation trends, that are later compared against the current degradation of a system to compute the RUL. These degradation trends can be built from raw sensor data in form of time series. A data processing step is often performed on the sensor data to select features that can describe the system behavior and to eliminate unusable and redundant data (Guyon and Elisseeff, 2003). Variance offers a good mathematical tool to automatically discard features with a low evolution. Performing variance sensors with a custom threshold produces a subset of selected features.

In the proposed methodology, the Kernel Principal Component Analysis (KPCA) (Schölkopf et al., 1997) is used. The approach is able to represent the high order statistics of data thanks to its nonlinear nature (Datta et al., 2018), compared to the principal component analysis, which is a linear transformation. The KPCA is an unsupervised technique that reduces the data set and leads to a lower feature set, with a normalization on the interval [-1;1]. A min-max normalization method is adopted favouring the KPCA to find a better subset representation with all variables in the same scale. No knowledge about system specificities nor system lifetime is used. KPCA might be confused with an auto-encoder approach. Despite the similarities, it is important to recall that auto-encoders rely on a machine learning structure that requires training and testing phases that are not required in a KPCA approach (Fournier and Aloise, 2019). In addition, because of its structure, an auto-encoder approach may lead to reasoning that can not be explained. Most of the time, feature

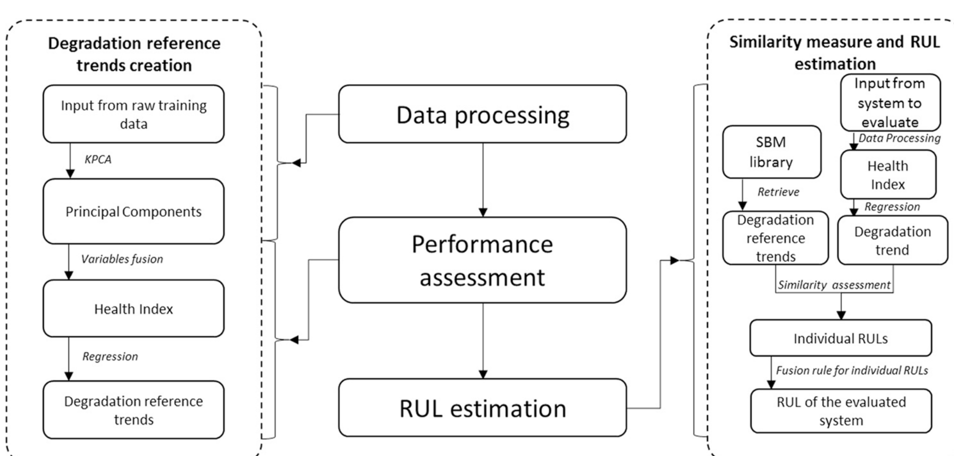


Fig. 1. Proposed similarity-based model for RUL estimation.

selection and feature extraction techniques are used to reduce the amount of information so to perform prognostic estimation efficiently. The proposed methodology aims at automating these feature selection and extraction processes using KPCA in such a way the approach becomes consistent, supported by more formal approaches that allow to estimate the loss of the data and keeping as much information as possible in the reduced subset of data.

Let us now describe briefly how KCPA works. KPCA is an extension of Principal Component Analysis (PCA) that generalizes it to nonlinear dimensionality reduction by using kernel methods. Considering n as the number of samples and m as the number of features, the input space of data $x_{(k,\ l)} \in R^{(m\times n)}$ is mapped to the feature space F: $\Phi(x_k)$ with $k \in [1, m]$ by the non-linear function Φ as:

$$\Phi\colon \mathbb{R}^n \to F \subset \mathbb{R}^m \tag{1}$$

It is assumed that feature data are centered, i.e. $\sum_{k=1}^{m}\Phi(x_k) = 0$. The covariance matrix $\overline{C}$ is:

$$\overline{C} = \frac{1}{m}\sum_{j=1}^{m}\Phi(x_j)\Phi(x_j)^T \tag{2}$$

The Principal Component (PC) of the feature space F are determined through the eigenvalues λ≥0 and the unitary eigenvectors V from the relation:

$$\lambda.\ V = \overline{C}\cdot V \tag{3}$$

This formula is equivalent to:

$$\lambda\langle\Phi(x_k), V\rangle = \langle\Phi(x_k), \overline{C}V\rangle \text{ for all } k \in [1, m] \tag{4}$$

where $\langle x, y\rangle$ is the dot product operation of x and y. Eigenvectors V can be written as:

$$V = \sum_{i=1}^{m}\alpha_i\Phi(x_i) \tag{5}$$

where α_i is the corresponding eigenvector coefficient. Substituting Fig. 2 and Fig. 5 into Fig. 4 and defining the kernel $m \times m$ matrix K as:

$$K_{ij} = \langle\Phi(x_i), \Phi(x_j)\rangle,\ \ i \in [1, m],\ j \in [1, m] \tag{6}$$

the following expression is exhibited:

$$m\lambda K\alpha = K^2\alpha \tag{7}$$

where $\alpha = [\alpha_1, ..., \alpha_m]^T$. The resolution of this relation is performed through the sub-system:

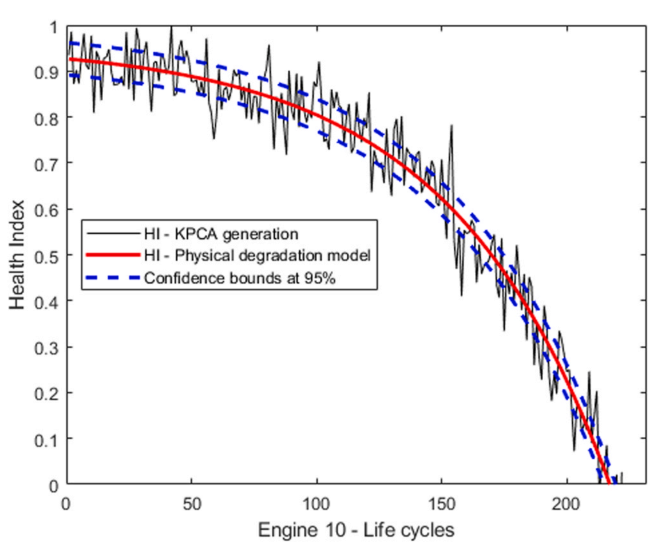


Fig. 2. Confidence interval for the Health Index/degradation function of a jet-engine example.

$$m\lambda\alpha = K\alpha \tag{8}$$

In general, the $\phi(x_i)$ are not centered. In this case, matrix K is replaced by the kernel-centered matrix $\overline{K}$(Schölkopf et al., 1997), which is then computed by:

$$\overline{K} = K - I_m\cdot K - K.\ I_m + I_m\cdot K\cdot I_m \tag{9}$$

where I_m is the diagonal $m \times m$ squared matrix with $1/m$ as value on the diagonal. There exist several kernel non-linear functions to compute the matrix K. A Gaussian Radial Basis Function (RBF) kernel is used here. For a given couple (x_i, x_j), the RBF kernel, named G, is defined as:

$$\begin{aligned} G(x_i, x_j) &= exp\left(\frac{\parallel x_i - x_j \parallel^2}{2\sigma^2}\right) \\ K_{ij} &= (G(x_i, x_j))_{ij},\ \ i \in [1, m],\ j \in [1, m] \end{aligned} \tag{10}$$

where σ is the dispersion coefficient of the RBF. Thus, the k^{th} KPCA feature is expressed as:

$$PC_k(x) = \langle V_k, \Phi(x)\rangle = \sum_{i=1}^{m}\alpha_{i,k}\langle\Phi(x_i), \Phi(x)\rangle \tag{11}$$

where V_k is the k^{th} eigenvector, $\Phi(x)$ is a test point and PC_k is the k^{th} principal component (PC). The number of PCs has an influence on the computed health index (HI). First, PCs are ordered in a descending order on the corresponding eigenvalue and each PC contribution is evaluated using:

$$Contr_k = \frac{\lambda_k}{\sum_{i=1}^{m}\lambda_i} \times 100\% \tag{12}$$

where $Contr_k$ is the contribution of the k^{th} PC. Then, the number of Principal Components (nPC) is evaluated using the Cumulative Contribution Percent (CCP) with a custom threshold (CCP_{th}) as:

$$CCP = \sum_{i=1}^{nPC} Contr_k \geq CCP_{th} \tag{13}$$

The Health Index (HI) is then computed by linear combination weighted with the percentage of contribution as:

$$HI(x) = \sum_{i=1}^{nPC} Contr_k\cdot PC_k(x) \tag{14}$$

The generated HI is rescaled in the interval [0;1] and represents the system health state. A high value equal or close to 1 represents a healthy system whereas a low value close to 0 corresponds to a faulty system. In addition to the physical degradation trends, confidence bounds are estimated with bootstrap techniques (Carpenter and Bithell, 2000). Confidence bounds are the lower and upper intervals that indicate the likelihood to predict correctly the remaining useful life within a given certainty.

The training of the similarity-based model is performed using all available run-to-failure instances of a system. The generated physical degradation trends expand the similarity database in order to be used by the similarity method itself for the RUL estimation.

3.3. Similarity assessment for RUL estimation

A degradation trend M_i from each training unit (i.e. a system instance used to train the similarity-based model) is created and stored in the model library. These trends can be described as functions y in terms of a time t:

$$M_i\colon y = f(t),\ \ t \in [0, T_i] \tag{15}$$

where T_i is the final time associated to the trend M_i. The physical degradation trend shape depends on the application. Different physical models could be used to create the function that describe the system health state.

The similarity degree d_i between a testing unit (i.e. a system instance under investigation) and training units is determined with a distance function $d(Y_j, M_i)$ where Y_j is the output of the testing unit j of T_j instances as $Y_j = [y(0), ..., y(T_j)]^T$. In this paper, Y_j represents the health index of a testing unit j. A small distance means high similarity whereas the opposite represents no relation. This function can be defined in various ways such as the Manhattan distance. The Euclidean distance is the simplest one (Wang et al., 2008):

$$d_i = d(Y_j, M_i) = \|Y_j - M_i\|^2 \tag{16}$$

The two vectors Y_j and M_i are respectively of length T_j and T_i. The distance is then computed on the minimum of the length as $\min(T_j, T_i)$. A first attempt to estimate the RUL could be performed by assessing the most similar degradation reference trend to the current degradation of the system under study. It is important to note that the age at which data is recorded, as well as the initial degradation of time series, has no influence in proposed approach. In contrast to other approaches such as a Cox model, in which the initial risk of failure is required to estimate a RUL, the proposed similarity-based degradation assessment and RUL estimation can be performed on systems with unknown initial degradation. Incidentally, the case study in the next section is composed of degradation records with an unknown initial point that helped confirm the proposed approach advantages. Different reference trends from the SBM library might show the highest similarity with a different RUL estimation. This phenomena is often observed when the system is in the incipient phase of degradation and the process until failure could last a few or several cycles. To overcome this uncertainty problem, SBM for predictive maintenance often use fusion rules to estimate the RUL based on several retrieved references from the SBM library. For the proposed model, a distance measure d_i is estimated for each degradation reference trend M_i, and using a fusion rule the individual RUL_i are consolidated in a final RUL value:

$$RUL = \sum_i \bar{w}_i.\ RUL_i,\ \sum_i \bar{w}_i = 1,\ \bar{w}_i = \frac{w_i}{\sum_k w_k} \tag{17}$$

where w_i and w_k are respectively a weight that is estimated by the inverse of the degree of similarity d_i and d_k, using Fig. 16. The weights $\bar{w}_i$ represent the percentage of similarity between a testing unit j with a unit i in the SBM library. The weight sum of all considered units i are then equal to 1 (i.e. 100%). It means that a unit with a high degree of similarity will have a weight value near 1. The number of unit i used to compute the final RUL is addressed by a fusion rule (see Fig. 3.4). This number is the same for the sum of index k. At the minimum, only 1 engine in the SBM library is considered and at the maximum, all of them. Fig. 17 provides a guideline for estimating the RUL as single value in terms of operational cycles.

In real life applications, the RUL estimation is often provided in intervals. This practice is intended to deal with uncertainties and facilitate the decision-making process based on the RUL estimation. Confidence bounds are a tool that is often implemented to obtain a confidence interval from a RUL estimation. It is a custom threshold that estimates the lower and upper bounds of health index/degradation functions. The built-in Matlab function 'predint' allows the user to compute the confidence bounds of a fitted function with a custom confidence threshold. Fig. 2 represents the confidence bounds of a degradation function example of a jet-engine that shows a exponential degradation behavior. This example is part of the case-study in section 4 and shows a confidence interval of 95% for the degradation of the jet-engine. In order to provide an alternative option to present the RUL estimation of the proposed SBM, each degradation reference trend in the SBM library is assigned with a confidence interval of 95% leading to lower and upper RUL estimation (respectively RUL_{lower_i} and RUL_{upper_i}). The fusion rule is applied on each bound estimation from each trend in M_i of a training unit i in order to estimate the RUL as follows:

$$RUL_{lower} = \sum_i \bar{w}_i.\ RUL_{lower_i},\ \sum_i \bar{w}_i = 1,\ \bar{w}_i = \frac{w_i}{\sum_k w_k} \tag{18}$$

$$RUL_{upper} = \sum_i \bar{w}_i.\ RUL_{upper_i},\ \sum_i \bar{w}_i = 1,\ \bar{w}_i = \frac{w_i}{\sum_k w_k} \tag{19}$$

where w_i are the same weight than from Fig. 17, RUL_{lower_i} and RUL_{upper_i} are respectively the lower and the upper final RUL estimations.

3.4. Fusion rule for RUL estimation

The final RUL is estimated using a fusion rule that takes into consideration the contribution of individual RUL_i that are obtained from the comparison of the current degradation of a system under study and the retrieved degradation reference trends of the SBM library.

The proposed fusion rule is based on the principle that data acquired from the tested units with the SBM can be used to improve the fusion rule. The number of the most similar Retrieved Reference Degradation Trends (RRDT) to be considered to compute the individual RUL_i changes depending on the accuracy of the final RUL estimation on previous tested units.

Given a SBM library with n reference trends and a testing unit j at a specific lifespan point with known true RUL:

1. The final RUL_j of the testing unit j is computed using as number of RRDT all possible options from 1 until n.
2. Each estimated final RUL_j is assessed against the true RUL using a performance indicator (see Fig. 3.5).
3. Best number of RRDT is obtained from best performance of RUL_j against the true RUL, and it is selected for the next testing unit.
4. The testing unit with known true RUL is stored in the SBM library to improve the reference trends records.
5. Any time a testing unit reaches its end of life and its true RUL information becomes available it can be used to update the number of RRDT following the previous steps.

It is important to notice that the number of RRDT to be considered for the final RUL estimation can benefit of the acquired knowledge from tested units only when they have been run-to-failure and the true RUL becomes available. The number of RRDT will be updated once all this knowledge is assessed and incorporated into the SBM library. If the tested units are replaced before failure, the knowledge is incomplete and cannot be used to update the number of RRDT. For the validation of the proposed fusion rule, two scenarios are being considered:

1. With the assumption that the tested units do not reach the end-of-life, then the number of RRDT remains static from the first proposal.
2. With the assumption that the tested units reach the end-of-life and the true RUL is available. The accumulated knowledge is used to reevaluate and refine the number of RRDT.

The main advantage of this fusion rule is a reduction of human interaction in the RUL estimation. It is an automatic approach that self-adapts to any application and can benefit from accumulated knowledge acquired over the SBM operation.

3.5. Prognostics evaluation

In order to assess the impact of the KPCA for the degradation reference trends generation, and the self-adaptive fusion rule, the Remaining Useful Life (RUL) should be assessed. A first step to perform the prognostics evaluation is by comparing the estimated RUL against the true RUL. For a given testing unit j, this comparison can be estimated as follows:

$$\Delta_j = \text{Estimated } RUL_j - \text{True } RUL_j \tag{20}$$

where Δ_j is the difference between the estimated and the true RUL of a testing unit j (i.e. the error in number of cycles, see Fig. 3). This comparison can also be performed if the RUL estimation is performed using a confidence interval instead of a single RUL value. The upper and lower bounds RUL create a confidence interval. If the true RUL belongs to this interval, the estimation is considered to be true. Hence, Fig. 20 is modified as follows:

$$\Delta_j = \begin{Bmatrix} 0 \; if \; RUL_{lower} \le RUL_{true_j} \le RUL_{upper} \\ RUL_{upper} - RUL_{lower} \; if \; RUL_{upper} \le RUL_{true_j} \\ RUL_{lower} - RUL_{upper} \; if \; RUL_{true_j} \le RUL_{lower} \end{Bmatrix} \tag{21}$$

This comparison between the estimated RUL and the true rule can be used in different indicators calculation so that the performance of the SBM can be assessed. One example of these performance indicators can be found in Saxena et al. (2008a), in which a custom score that provides an absolute linear penalty for RUL estimations (see Fig. 3). It is computed as follows:

$$score = \sum_k S_j, \; S_j = \begin{Bmatrix} e^{-\frac{\Delta_j}{b_1}} - 1, \; \Delta_j \le 0 \\ e^{\frac{\Delta_j}{b_2}} - 1, \; \Delta_j > 0 \end{Bmatrix} \tag{22}$$

where S_i is the computed score for the tested unit i, b_1 and b_2 are both determined constants according to the required system criteria. In Saxena et al. (2008a) these constants are set as $b_1 = 13$ and $b_2 = 10$, for the case study used in this paper. In the current paper, the same values were used. This performance indicator will be referred to as *PHM08 score* as it was original used to assess the results in the Prognostics and Health Management (PHM) prognostics challenge in 2008. The PHM08 score has been adopted as it is the performance indicator that has been mostly used in the case study used in the current research (see Fig. 4).

The Root Mean Square Error (RMSE) is also a performance indicator that is often used to assess prediction models for RUL estimation. It is less restrictive than the PHM08 score for the estimations that are close to the true RUL but as the RMSE uses an exponential penalty, the error is higher than the one obtained from the PHM08 score for estimations that are far from the true RUL (see Fig. 3). The RMSE is calculated as follows:

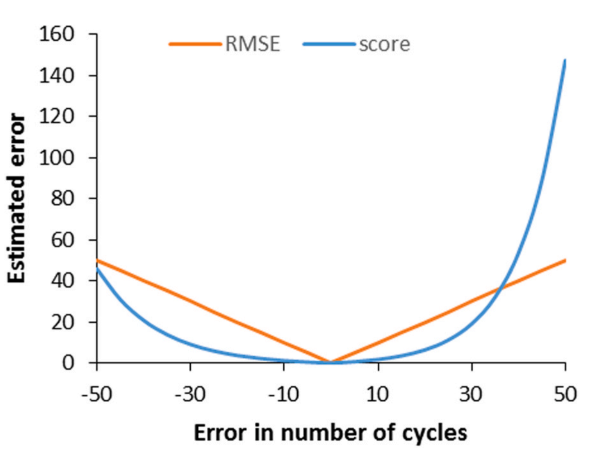


Fig. 3. Prognostic evaluation metrics.

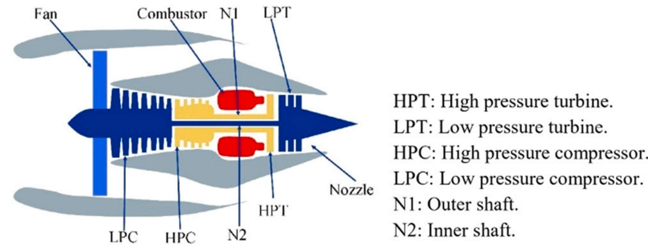


Fig. 4. Simplified diagram of the 90K engine Saxena et al. (2008a).

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N} \Delta_j^2} \tag{23}$$

where N is the number of evaluated samples.

The mean error indicator is also used to have an overview of the RUL estimation on the database. This traditional error provides a global overview of the testing phase. As the mean error relies on the True RUL for its computation is not simple to plot the penalty behavior as for the RMSE and the PHM08 score.

$$Mean \quad error = \frac{1}{N}\sum_{j=1}^{N} \frac{|\Delta_j|}{\text{True } RUL_j} \tag{24}$$

A prognostic approach could provide good results with a specific performance indicator, and bad results with another one. Therefore, having good results using multiple performance indicators simultaneously is a challenge. It improves the approach validation. For this research effects, the three performance indicators are then used in the validation of the proposed approach as they provide a wider overview of the its performance. Using only one of these indicators may not be representative to evaluate the results.

4. Case study

4.1. Overview

To illustrate the proposed approach, a case study using the Commercial Modular Aero-Propulsion System Simulation (CMAPSS) is proposed (Saxena et al., 2008a; Saxena et al., 2008b). This case study is a NASA benchmark and is composed of simulated run-to-failure data of commercial jet engines and has been widely used in academic research to test predictive maintenance models. CMAPSS is a tool for the simulation of a realistic large commercial turbofan engine (Fig. 4) for the 90,000 lb thrust class. Thanks to editable input parameters, it is possible to specify operational profile, closed-loop controllers, environmental conditions such as altitude, etc. Furthermore, various degradation profiles can be managed in different components of the engine.

The CMAPSS case study available in the NASA repository NASA (NASA, 2007) is composed of five data sets (see Table 1), one of which for the 2008 Prognostic Health Management (PHM08) challenge. In these data sets, the simulated engines have one or six operational conditions driven by engine control settings (Altitude, Mach number and Throttle Resolver Angle) and one or two fault modes. Each engine has different initial wear, degradation rate, the exact moment of first use is not known and in addition, all data have process and measurement noises. In the current study, there are three datasets with one fault and two with two faults (see Table 1).

Each dataset consists of multi-variable time series, divided into a training and a testing file. Each data set contains three input

Table 1
C-MAPSS datasets characteristics.

Id	Name	Operational conditions	Fault Modes	Failed System Part	Number of Engines
#1	FD001	1	1	HPC	100
#2	FD002	6	1	HPC	260
#3	FD003	1	2	HPC, Fan	100
#4	FD004	6	2	HPC, Fan	549
#5	FD005 (PHM08)	6	1	HPC	218

Table 2
List of 21 output variables from CMAPSS tool.

Sensor id	Symbol	Description	Units
1	T2	Total temperature at fan inlet	°R
2	T24	Total temperature at LPC outlet	°R
3	T30	Total temperature at HPC outlet	°R
4	T50	Total temperature at LPT outlet	°R
5	P2	Pressure at fan inlet	psia
6	P15	Total pressure in bypass-duct	psia
7	P30	Total pressure at HPC outlet	psia
8	Nf	Physical fan speed	rpm
9	Nc	Physical core speed	rpm
10	epr	Engine pressure ratio (P50/P2)	–
11	Ps30	Static pressure at HPC outlet	psia
12	Phi	Ratio of fuel flow to Ps30	pps/psi
13	NRf	Corrected fan speed	rpm
14	NRc	Corrected core speed	rpm
15	BPR	Bypass Ratio	–
16	farB	Burner fuel-air ratio	–
17	htBleed	Bleed Enthalpy	–
18	Nf_dmd	Demanded fan speed	rpm
19	PCNfR_dmd	Demanded corrected fan speed	rpm
20	W31	HPT coolant bleed	lbm/s
21	W32	LPT coolant bleed	lbm/s

variables representing the control/environment settings and 21 outputs sensors (see Table 2).

4.2. Case study adaptation

For this research study, only the three data sets with one failure mode out of the five available ones have been selected: FD001, FD002 and FD005. Each CMAPSS data set is composed of two main files, the first one is intended for training of prognostic models and the second one for testing. The training file includes several engines that have been run-to-failure, meaning that the true RUL for each engine cycle is known. For the testing file the engines have not been run-to-failure and their true RUL is not provided. This testing file cannot be directly used to test prognostic models. To overcome this limitation and obtain faster results the training file has been split in two parts: 80% will remain as training data and 20% will be used for testing with a known RUL value.

The extracted testing subset (i.e. the 20% of the original training file) has been adapted in order to have multiple engine data records with known RUL that can be used for the SBM validation. For each engine in the testing subset, partial sections of the lifetime records have been extracted. These sections cover the 40%, 50%, 60%, 70%, 80%, and 90% of the total life time records available for each engine in the extracted testing subset. For example, considering an engine with a run-to-failure life of 100 cycles: a life time records percentage of 40% will contain the 40 first cycles of the engine and the true RUL for this engine would be 60 cycles. This adaptation of the testing subset will also allow to evaluate how the proposed SBM behaves with different amounts of data records.

5. Results and discussion

5.1. Physical degradation reference trends generation

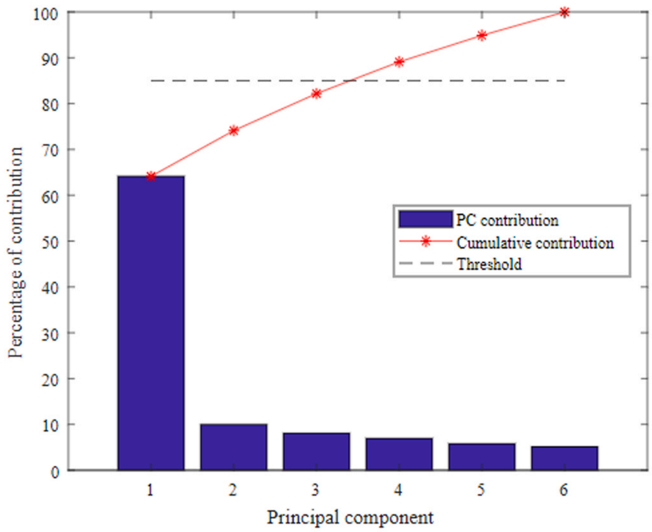
For the implementation of the proposed approach on the CMAPSS case study, the feature selection is performed by using variance analysis with a threshold of 0.001; this aims at removing sensors with constant or binary values that do not provide much information for RUL estimation (Li et al., 2018). All sensors with a variance below or equal to this threshold are discarded. From the list of 21 sensors (see Table 2), the number 1, 5, 6, 10, 16, 18, and 19 are removed. The remaining 14 sensors are used as input for the KPCA feature reduction after a re-scaling in the interval [0,1] for the minimal and maximal values of each variable. The 14 input sensors result in 6 principal components after performing the KPCA. Fig. 5 shows the individual and the cumulative contribution of these six main principal components for datasets FD001 and FD002.

For the training phase of the SBM, the KPCA is used on each engine and on each operational mode to obtain the degradation reference trends. These degradation reference trends are stored in the SBM library. As a custom threshold in this research, it has been decided to keep 85% of the original sensor data to compute the health indexes of the engine. This threshold is motivated by observations in the literature such as Feng et al. (2016) on the same case study and Jiang and Yan (2013) on a nonlinear chemical process monitoring. It means that based on the information provided in Fig. 5 and using Fig. 13 with CCP_{th} = 85%, the four first principal components are selected. Then, the selected principal components are fused using Fig. 14 to obtain the degradation reference trends (i.e. the Health Indexes (HIs)). These reference trends are first of all obtained as series of points. To facilitate their analysis, these series of points go through a curve fitting process. It allows to manage noise and to have smoothed degradation trends. For the CMAPSS case study, the degradation trends are fitted using a generic exponential regression model, inspired by the fitting model used by Saxena et al. (2008a):

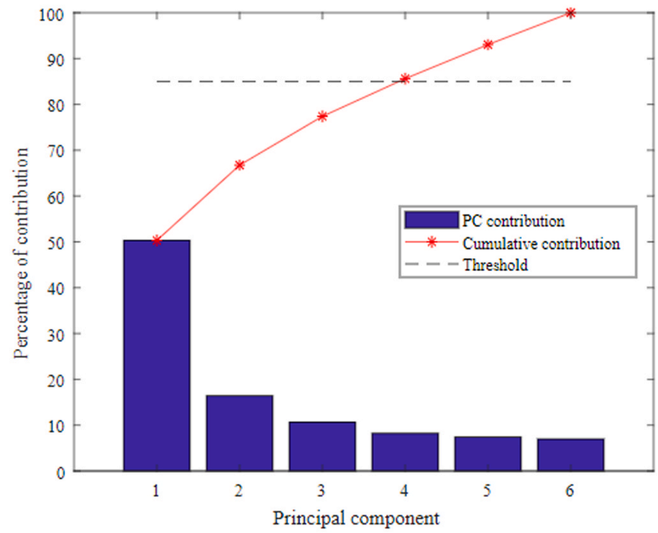
$$f(t) = a - e^{b \cdot t + c} \tag{25}$$

where a, b and c are constants. These constants are obtained by a regression process on the reference trend of a unit. The resulting curve is the HI curve of the unit.

For each testing unit, the data processing must be done following the same way as for the training instances of the SBM. A degradation trend of the testing unit is obtained from the same four principal components and using Fig. 14 and Fig. 25. Once the degradation trend of the testing unit has been obtained, the degree of similarity between the testing engine and each reference degradation trend (from SBM library) is estimated with Fig. 16. The RUL is estimated using equation Fig. 17 for a single value RUL, or using Fig. 18 and Fig. 19 for a confidence interval of RUL. For the validation of the Proposed Approach (PA) of a SBM for RUL estimation, a series of tests are performed and the results are compared against those of a Reference Approach (RA) from Wang et al. (2008). It is important to note that the reference approach was the winner at the PHM08 contest (Saxena et al., 2008a), and provides accurate results on RUL computation with an explainable similarity-based approach on the CMAPSS case study. Other approaches have been developed to address the same case study that outperform the Reference Approach but are inspired in Machine Learning techniques whose reasoning cannot be explained. As the proposed approach is oriented to explainable solutions, Machine Learning techniques are not been considered. The main goal of the comparison between the PA and the RA is to demonstrate that the PA provides benefits at computing RUL of technical systems with an explainable reasoning compared to other approaches in the same conditions. Explainable reasoning is



(a) With one operational mode (FD001)



(b) With six operational modes (FD002)

Fig. 5. KPCA: individual and cumulative contribution of Principal Components (PCs).

still a crutial point for the implementation of predictive maintenance in complex systems such as aircraft. The RA implements a SBM for RUL estimation of aircraft jet engines (see Fig. 4). The main differences between the PA and RA are in the way the HI is generated and the fusion rule for the RUL estimation. The RA performs a manual process for feature selection. A linear regression model is used to merge the selected features into a consolidated degradation reference trend for each training unit. For the fusion rule, the RA uses the contribution of almost all possible individual RUL_i that can be obtained from comparing a testing unit with the reference degradation trends in the SBM library. Some non-relevant individual RUL_i are discarded manually using an expert rule. However, as the authors of the RA point out, the expert rules they implement for the final RUL calculation are highly application dependent. It means that the rule to fuse the individual RUL_i should not be directly used for a different application, some expert knowledge is needed for its adaptation. The RA came out best in the 2008 Prognostic and Health Management challenge (Saxena et al., 2008a). A little adaptation of RA is done in the current study as it turns out difficult to replicate the expert rule the RA used for deleting non-representative reference degradation trends for the fusion rule. In the context of this research, the RA uses all possible individual RUL_i without deleting the non-representative reference degradation trends.

Remaining Useful Life (RUL) estimation using the PA and the RA is assessed using the performance indicators of Section 3.5. A performance gain is used to study the benefits of the proposed approach implementation. A performance gain is a percentage that shows a relative improvement compared to the reference approach. A performance gain of 100% means an ideal improvement in which the error has been reduced to 0 and the estimated RUL is equal to the true RUL in all the tested engines. The performance gain is computed as follows:

$$PG = \left(1 - \frac{PA}{RA}\right) \times 100\% \tag{26}$$

If $PG > 0$, it means the PA performed better than the RA. If $PG < 0$, it means the PA does not improve results compared to the RA.

5.2. KPCA contribution assessment

For the assessment of the KPCA contribution for HI generation, a comparison of the RUL estimations using the PA and the RA is performed. The fusion rule of RA is used for both PA and RA. This is to measure the actual contribution of the KPCA without any influence of the fusion rule. Table 3 summarizes the PHM08 score for RUL estimation using the proposed approach and the reference approach on the testing subset introduced in Section 4.2. The proposed approach improves significantly the performance in RUL estimation as can be seen on the performance gain percentage in Table 3. The results of the proposed approach always outperformed those of the reference approach. The performance was also evaluated using the RMSE and the mean error with similar results as for the PHM08 score.

5.3. Complete approach contribution assessment

The complete proposed approach of a SBM for RUL estimation includes the use of the KPCA for HI generation and a fusion rule for the RUL estimation that can benefit from the acquired knowledge during the SBM operation to improve the RUL estimation. For the proposed fusion rule assessment, two different scenarios are being considered. The first scenario assumes that the testing units do not

Table 3
Performance gain of implementing the KPCA for Health Index generation using testing units at different lifetime percentage.

Lifetime records (%)	Proposed Approach performance with the RA fusion rule (PHM08 score)	Reference approach performance (PHM08 score)	Performance Gain (%)
40	17299	117648	85.30
50	7080	78723	91.01
60	2117	10184	79.21
70	642	1136	43.48
80	503	653	22.97
90	1639	6370	74.27

Table 4
RUL estimation assessment using the first scenario of the proposed fusion rule and the PHM08 score as a performance indicator.

LRP (%)	FD001			FD002			FD005		
	PA (PHM08 score)	RA (PHM08 score)	PG	PA (PHM08 score)	RA (PHM08 score)	PG	PA (PHM08 score)	RA (PHM08 score)	PG
40	2 259	39 041	**94.21%**	13 521	15 512	**12.83%**	6 561	12 361	**46.92%**
50	489	21 710	**97.75%**	7 874	11 349	**30.62%**	10 180	16 084	**36.70%**
60	235	5 303	**95.57%**	5 522	9 996	**44.76%**	3 551	20 883	**83.00%**
70	137	1 385	**90.13%**	5 049	5 955	**15.22%**	4 018	19 537	**79.43%**
80	103	1 435	**92.85%**	3 060	3 760	**18.62%**	2 495	18 712	**86.67%**
90	91	3 898	**97.66%**	1 364	2 770	**50.78%**	1 776	12 352	**85.62%**

LRP: Lifetime Records Percentage, PA: Proposed Approach, RA: Reference Approach, PG: Performance Gain

reach the end of life before they are taken out from operation, as it can happen because of a maintenance policy. In this first scenario the accumulated knowledge over the SBM operation can not be used for to improve the fusion rule and the number of Retrieved Reference Degradation Trends (RRDT) to compute the individual RUL_i remains the same as it was initialized (see Fig. 3.4). The second scenario assumes that the testing units do reach their end of life and the true RUL of the testing units become available. The information from the tested units is used to update the number of RRDT for the fusion rule.

For the complete assessment of the proposed approach, three datasets of CMAPSS have been selected: FD001, FD002 and FD005. These datasets have been selected because they present one single failure mode at different operational modes. Tables 4, 5 and 6 shows the assessment of the fusion rule of the proposed approach considering the first scenario, and comparing the results against the reference approach using three performance indicators: PHM08 score, RMSE, and mean error correspondingly. The results of the proposed approach outperformed those of the reference approach independently of the performance indicators for the three selected datasets.

It is important to point out that the proposed approach needs fewer data records to have higher performance in the RUL estimation. For example with the FD001 dataset, it can be noticed for example that the PHM08 score of the PA for a lifetime at 40% is better than the RA with a lifetime at 60%. A normal behavior in SBM for RUL estimation is that they have better results when more records are available to perform the similarity analysis between the testing units and the reference degradation trends in the SBM library. The need of fewer records to achieve better results is an evidence of the benefits of the proposed approach over the reference approach; accurate results of the RUL will be known earlier and this will lead to a better decision making process in a maintenance department. Overall, there is better results with the proposed approach for the three databases (i.e. FD001, FD002, and FD005). The “RMSE” and “Mean error” show as well homogeneous improvements of the PA over the RA on all studied datasets (see Table 5 and Table 6). The minimum performance gain is around 13%, for the FD002 database.

In interesting behavoir can be observed for FD005 in Table 4, when the lifetime records percentage increases, the PHM08 score sometimes also increases which means a RUL estimation error increment. For example, the performance of the proposed approach is lower at 50% of the lifetime records than at 40% using the PHM08 score. This is not an expected behavior as prognostics tend to be more precise with more data to be assessed. Having a closer look on the RUL estimation for the engines in the FD005 dataset, the estimated value was often higher than the true RUL which is highly penalized by the PHM08 score. This can provide an explanation of lower performances with more percentage of lifetime records for RUL estimation in Table 4. This behavior is not visible using the RMSE and the mean error as performance indicators (see Table 5 and Table 6) as they are based on an average value. The performance indicator choice has a high impact on the approach evaluation.

Table 7 summarizes the results of the proposed approach using the second scenario for the fusion RUL and the PHM08 score as performance indicator. The overall improvement on the RUL estimation is at least 38% on the three datasets used. This is not only a improvement compared to the reference approach but also an improvement compared to the first scenario of the proposed approach. Similar results where also observed using the RMSE and the mean error as performance indicator. Comparing the results in Tables 4 and 7, it can be seen the second scenario outperforms the benefits of the of the first scenario of the proposed approach. It means that the more available data to estimate the number of RRDT for the fusion rule, the better results are obtained.

5.4. Proposed approach assessment using a confidence interval

Table 8 summarizes the PHM08 score at a confidence of 95% for the three databases. As can be expected, the performance indicator improves (PHM08 score decreases) since providing a RUL interval instead of a single value increases the chances to estimate closer values to the true RUL. For the three databases, a confidence threshold of 95% seems to improve the RUL estimation compared to an approach without confidence bounds, but the results have to be analyzed carefully. By decreasing the confidence threshold, the score decreases as well. Fig. 6 shows the *score* versus the evolution of the confidence threshold for the RUL estimation on FD001 sliced at 40% can be observed. The blue line represents the *score* for the proposed approach without the use of confidence interval. It results in a constant curve. The red line represents the *score* of the proposed approach with the use of various values of confidence thresholds. With a confidence interval of 100%, the proposed approach with confidence bounds provides the same result than without confidence bounds (single value RUL estimation represented by a blue

Table 5
RUL estimation assessment using the first scenario of the proposed fusion rule and the RSME as a performance indicator.

LRP (%)	FD001			FD002			FD005		
	PA (RMSE)	RA (RMSE)	PG	PA (RMSE)	RA (RMSE)	PG	PA (RMSE)	RA (RMSE)	PG
40	42.87	58.01	**26.10%**	39.07	44.89	**12.96%**	36.11	42.33	**14.71%**
50	33.23	54.84	**39.40%**	34.86	42.84	**18.63%**	35.29	41.64	**15.25%**
60	25.96	48.60	**46.58%**	31.81	40.56	**21.59%**	29.24	41.77	**30.01%**
70	20.96	37.63	**44.29%**	26.41	35.08	**24.73%**	29.27	40.68	**28.05%**
80	17.00	30.23	**43.78%**	23.15	28.48	**18.72%**	25.13	39.74	**36.77%**
90	14.30	41.57	**65.59%**	20.68	28.44	**27.26%**	22.95	41.04	**44.09%**

LRP: Lifetime Records Percentage, RMSE: Root Mean Square Error, PA: Proposed Approach, RA: Reference Approach, PG: Performance Gain

Table 6
RUL estimation assessment using the first scenario of the proposed fusion rule and the mean error percentage as a performance indicator.

LRP (%)	FD001			FD002			FD005		
	PA (mean error %)	RA (mean error %)	PG	PA (mean error %)	RA (mean error %)	PG	PA (mean error %)	RA (mean error %)	PG
40	28.47	32.96	**13.60%**	27.71	31.95	**13.26%**	26.83	31.57	**15.02%**
50	27.72	38.07	**27.19%**	31.09	37.28	**16.60%**	30.71	37.84	**18.84%**
60	25.10	45.00	**44.22%**	36.64	44.53	**17.71%**	32.28	47.75	**32.39%**
70	28.05	52.70	**46.79%**	40.09	52.99	**24.34%**	42.64	62.78	**32.08%**
80	33.96	64.03	**46.97%**	52.33	64.06	**18.32%**	53.88	91.30	**40.99%**
90	55.35	183.29	**69.80%**	94.16	128.31	**26.62%**	97.84	201.75	**51.50%**

LRP: Lifetime Records Percentage, PA: Proposed Approach, RA: Reference Approach, PG: Performance Gain

Table 7
RUL estimation assessment using the second scenario of the proposed fusion rule and the PHM08 score as a performance indicator.

LRP (%)	FD001			FD002			FD005		
	PA (PHM08 score)	RA (PHM08 score)	PG	PA (PHM08 score)	RA (PHM08 score)	PG	PA (PHM08 score)	RA (PHM08 score)	PG
40	1 140	39 041	**97.08%**	7 135	15 512	**54.00%**	5 428	12 361	**56.09%**
50	300	21 710	**98.62%**	6 123	11 349	**46.05%**	3 688	16 084	**77.07%**
60	129	5 303	**97.56%**	4 124	9 996	**58.74%**	2 267	20 883	**89.14%**
70	31	1 385	**97.78%**	3 614	5 955	**39.31%**	1 259	19 537	**93.55%**
80	55	1 435	**96.18%**	2 332	3 760	**37.97%**	309	18 712	**98.35%**
90	61	3 898	**98.43%**	784	2 770	**71.72%**	956	12 352	**92.26%**

LRP: Lifetime Records Percentage, PA: Proposed Approach, RA: Reference Approach, PG: Performance Gain

Table 8
RUL estimation comparison between the PA and PA-CI for FD001, FD002, and FD005 databases using a PHM08 score as a performance indicator.

LRP (%)	FD001			FD002			FD005		
	PA-CI (PHM08 score)	PA (PHM08 score)	PG	PA-CI (PHM08 score)	PA (PHM08 score)	PG	PA-CI (PHM08 score)	PA (PHM08 score)	PG
40	1 932	2 259	**14.48%**	6 132	13 521	**54.65%**	3 188	6 561	**51.41%**
50	412	489	**15.73%**	3 956	7 874	**49.76%**	4 455	10 180	**56.24%**
60	197	235	**16.27%**	2 769	5 522	**49.86%**	1 762	3 551	**50.37%**
70	112	137	**17.81%**	2 573	5 049	**49.03%**	1 780	4 018	**55.69%**
80	84	103	**17.97%**	1 494	3 060	**51.18%**	1 265	2 495	**49.29%**
90	76	91	**16.62%**	785	1 364	**42.43%**	903	1 776	**49.14%**

LRP: Lifetime Records Percentage, PA-CI: Proposed Approach with Confidence Interval, PA: Proposed Approach, PG: Performance Gain

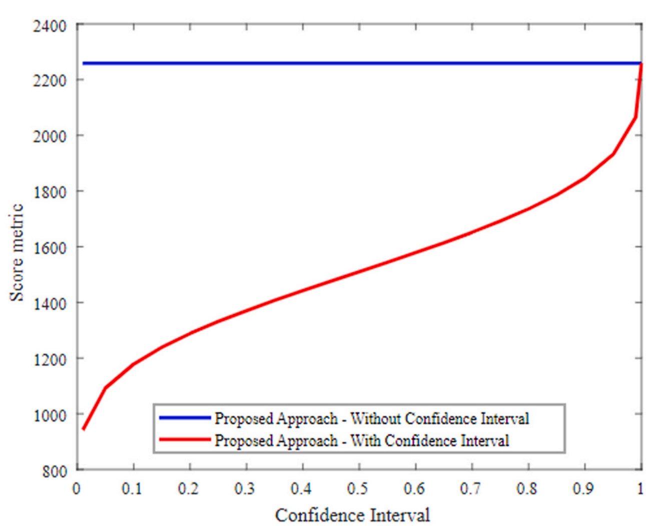


Fig. 6. Confidence threshold influence for RUL estimation on FD001 sliced at 40%.

line in Fig. 6). With a confidence threshold of 0.95, it leads to a confidence interval of 95% on the RUL estimation. The *score* is then evaluated using Fig. 21 and Fig. 22, which resulting into a single value (i.e. no confidence interval on results that are from a confidence interval). Even if the performance indicator seems to be improving while decreasing the confidence threshold, it does not mean it is advisable. Fig. 6 shows the perverted effect of the confidence threshold on the RUL estimation and the *score*. The lower the confidence threshold is, the bigger the RUL estimation interval is and the bigger the uncertainty is when computing the RUL.

A different methodology is needed in order to evaluate the performance of the RUL estimation using confidence interval. For this aim, a Percentage of Disparity (PoD) that can evaluate the confidence interval for a given confidence threshold is proposed. The PoD is computed through the quotient between the relative variation and the mean estimation (PoD = RelativeRUL/MeanRUL) and it is an indicator of how dispersed the values are in the confidence interval. Table 9 presents the average range of the relative RUL variation (i.e. difference between the maximum and the minimum RUL estimation, respectively Fig. 18 and Fig. 19), the mean RUL estimation and a Percentage of Disparity (PoD). It can be observed that the PoD and the relative RUL range is different for each database. The mean PoD for FD001, FD002 and FD005 are respectively 5%, 21% and 20%. To analyzed these results a maintenance policy that allows no more than 10% of error on the RUL estimation is assumed. For FD005, a PoD of 20% means the confidence threshold of 95% is not a good choice as it does not meet the maintenance policy. Some modifications would be needed in the confidence threshold to reach the desired error value. The confidence interval can be seen as an allowed error rate for the RUL estimation. A big interval leads to big uncertainty on the RUL. The PoD highlighted this problem with the database FD005. This interval width is directly linked to the confidence threshold. Decreasing the confidence threshold leads to the increase of the RUL estimation interval.

Table 9
RUL estimation assessment using a Relative RUL estimation method and a Mean RUL estimation method for FD001, FD002, and FD005 databases and using a percentage of disparity as a performance indicator.

LRP (%)	FD001			FD002			FD005		
	M-RUL (PHM08 score)	R-RUL (PHM08 score)	PoD	M-RUL (PHM08 score)	R-RUL (PHM08 score)	PoD	M-RUL (PHM08 score)	R-RUL (PHM08 score)	PoD
40	129	4	**2.84%**	127	15	**11.57%**	135	16	**11.77%**
50	110	4	**3.40%**	112	15	**13.66%**	116	15	**12.86%**
60	88	4	**4.23%**	95	17	**18.00%**	97	16	**16.26%**
70	67	4	**5.64%**	75	17	**22.13%**	79	18	**22.59%**
80	48	4	**7.29%**	57	16	**28.89%**	57	15	**25.95%**
90	31	3	**8.13%**	39	13	**32.73%**	39	11	**29.82%**

LRP: Lifetime Records Percentage, M-RUL: Mean RUL estimation score, R-RUL: Relative RUL score, PoD: Percentage of Disparity

6. Conclusion

This study proposed two different improvements for similarity-based models for RUL estimation. As a first contribution a health index building method using the Kernel Principal Component Analysis (KPCA) was proposed. It allowed the creation of the degradation reference trends that are stored in the SBM library in a generic way, by using automatic and unsupervised techniques. As a second contribution a novel fusion rule for RUL estimation was proposed. It is a self-adaptive method that can benefit from the accumulated knowledge over the SBM operation.

For the assessment of both contributions different tests where performed using the CMAPSS case study and comparing the results against a reference approach that also implements a SBM for RUL estimation but in contrast, it uses several experts rules that are not easily adaptable on different applications. The validation of the proposed SBM aimed at proving that it can outperform other approaches with explainable reasoning. This is an important point for the implementation of predictive maintenance in critical systems such as aircraft.

In a first row of test, the contribution of the KPCA for the HI generation was successfully validated. The results were better than those of the reference approach. In a second row of test the complete approach was validated: the implementation of the KPCA and the proposed fusion rule. Different performance indicators where used for the approach validation: the PHM08 score, the RSME and the mean error. It provides a better idea of how the SBM performs with the proposed approach. For the fusion rule two different scenarios where explored. The first scenario does not consider the accumulated knowledge over the SBM operation. The second scenario benefits from the accumulated knowledge to update the fusion rule. Both scenarios shown better results compared to the reference approach that uses experts rules for the RUL estimation.

As a complementary analysis, the proposed approach was also studied if the RUL estimation should be in confidence interval. The performance indicators used for single value RUL estimation do not provide a good overview of the RUL estimation using confidence bounds. Another performance evaluation was proposed using the percentage of disparity in the confidence interval. Performance of the RUL estimation would rely on maintenance policies and the confidence threshold that should be modified to meet the expected results.

Further work focuses on enhancing the proposed approach so to consider the presence of several failure modes with the same case study, which is currently a limitation of the presented approach. In addition, the proposed SBM for RUL estimation can also be implemented in other case studies. Spontaneous important health index improvements, such as may happen in batteries (Zhou et al., 2016), are not managed by the current approach, and will be addressed in the future.

CRediT authorship contribution statement

Sébastien Schwartz: Conceptualization, Methodology, Software, Writing - Original Draft, **Juan Jose Montero Jimenez:** Writing - Review & Editing, **Rob Vingerhoeds:** Supervision, Writing - Review & Editing, **Michel Salaün:** Supervision, Writing - Review & Editing

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bektas, O., Jones, J.A., Sankararaman, S., Roychoudhury, I., Goebel, K., 2019. A neural network framework for similarity-based prognostics. MethodsX 6, 383–390. https://doi.org/10.1016/j.mex.2019.02.015

Berri, P.C., Dalla Vedova, M., Mainini, L., 2021. Computational framework for real-time diagnostics and prognostics of aircraft actuation systems. Comput. Ind. 132, 103523. https://doi.org/10.1016/j.compind.2021.103523

Bleakie, A., Djurdjanovic, D., 2013. Analytical approach to similarity-based prediction of manufacturing system performance. Comput. Ind. 64 (6), 625–633. https://doi.org/10.1016/j.compind.2013.02.013

Cai, H., Jia, X., Feng, J., Li, W., Pahren, L., Lee, J., 2020. A similarity based methodology for machine prognostics by using kernel two sample test. ISA Trans. 103, 112–121. https://doi.org/10.1016/j.isatra.2020.03.007

Carpenter, J., Bithell, J., 2000. Stat. Med. 19 (9), 1141–1164 10.1002/(SICI)1097-0258(20000515)19:9 < 1141::AID-SIM479 > 3.0.CO;2-F.

Chen, Z., Cao, S., Mao, Z., 2017. Remaining useful life estimation of aircraft engines using a modified similarity and supporting vector machine (SVM) approach. Energies 11 (1), 28. https://doi.org/10.3390/en11010028

Climente-Alarcon, V., Nair, D., Sundaria, R., Antonino-Daviu, J.A., Arkkio, A., 2017. Combined model for simulating the effect of transients on a damaged rotor cage. IEEE Trans. Ind. Appl. 53 (4), 3528–3537. https://doi.org/10.1109/TIA.2017.2691001

Datta, A., Ghosh, S., Ghosh, A., 2018. PCA, kernel PCA and dimensionality reduction in hyperspectral images. In: Naik, G.R. (Ed.), Advances in Principal Component Analysis. Springer, Singapore, Singapore, pp. 19–46. https://doi.org/10.1007/978-981-10-6704-4_2

Feng, D., Xiao, M., Liu, Y., Song, H., Yang, Z., Zhang, L., 2016. A kernel principal component analysis-based degradation model and remaining useful life estimation for the turbofan engine. Adv. Mech. Eng. 8 (5). https://doi.org/10.1177/1687814016650169

Fournier, Q., Aloise, D., 2019. Empirical comparison between autoencoders and traditional dimensionality reduction methods. 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE). IEEE, pp. 211–214. https://doi.org/10.1109/AIKE.2019.00044

Gretton, A., Borgwardt, K.M., Rasch, M.J., Sch, B., 2013. A kernel two-sample test. J. Mach. Learning Res. 13, 723–773.

Gu, M., Chen, Y., 2016. A framework of multi-index modeling for similarity-based remaining useful life estimation. 2016 3rd International Conference on Information Science and Control Engineering (ICISCE). IEEE, Beijing, China, pp. 31–37. https://doi.org/10.1109/ICISCE.2016.18

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182. https://doi.org/10.1162/153244303322753616

Hassani.N, S.M., Jin, X., Ni, J., 2019. Physics-based Gaussian process for the health monitoring for a rolling bearing. Acta Astronaut. 154, 133–139. https://doi.org/10.1016/j.actaastro.2018.10.029

Huang, C.-G., Huang, H.-Z., Peng, W., Huang, T., 2019. Improved trajectory similarity-based approach for turbofan engine prognostics. J. Mech. Sci. Technol. 33 (10), 4877–4890. https://doi.org/10.1007/s12206-019-0928-3

Jiang, Q., Yan, X., 2013. Weighted kernel principal component analysis based on probability density estimation and moving window and its application in nonlinear chemical process monitoring. Chemom. Intell. Lab. Syst. 127, 121–131. https://doi.org/10.1016/j.chemolab.2013.06.013

Jiang, Y., Zhu, H., Ding, C., Pfeiffer, O., 2019. A novel ensemble fuzzy model for degradation prognostics of rolling element bearings1. J. Intell. Fuzzy Syst. 37 (4), 4449–4455. https://doi.org/10.3233/JIFS-179277

Khelif, R., Malinowski, S., Chebel-Morello, B., Zerhouni, N., 2014. RUL prediction based on a new similarity-instance based approach. 2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE). IEEE, Istanbul, Turkey, pp. 2463–2468. https://doi.org/10.1109/ISIE.2014.6865006

Kumar, U., Galar, D., 2018. Maintenance in the era of industry 4.0: issues and challenges. In: Kapur, P., Kumar, U., Verma, A.K. (Eds.), Quality, IT and Business Operations. Springer, Singapore, Singapore, pp. 231–250. https://doi.org/10.1007/978-981-10-5577-5_19

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., Liu, H., 2018. Feature selection: a data perspective. ACM Comput. Surv. 50 (6), 1–45. https://doi.org/10.1145/3136625

Li, J., Li, X., He, D., 2019. A directed acyclic graph network combined with CNN and LSTM for remaining useful life prediction. IEEE Access 7, 75464–75475. https://doi.org/10.1109/ACCESS.2019.2919566

Li, X., Ding, Q., Sun, J.-Q., 2018. Remaining useful life estimation in prognostics using deep convolution neural networks. Reliab. Eng. Syst. Saf. 172, 1–11. https://doi.org/10.1016/j.ress.2017.11.021

Liang, Z., Gao, J., Jiang, H., Gao, X., Gao, Z., Wang, R., 2019. A degradation degree considered method for remaining useful life prediction based on similarity. Comput. Sci. Eng. 21 (1), 50–64. https://doi.org/10.1109/MCSE.2018.110145829

Montero Jimenez, J.J., Schwartz, S., Vingerhoeds, R., Grabot, B., Salaün, M., 2020. Towards multi-model approaches to predictive maintenance: a systematic literature survey on diagnostics and prognostics. J. Manuf. Syst. 56, 539–557. https://doi.org/10.1016/j.jmsy.2020.07.008

NASA, Prognostic Center of Excellence Datasets, ⟨https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/⟩ (2007).

Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space, The London, Edinburgh, and Dublin. Philos. Mag. J. Sci. 2 (11), 559–572. https://doi.org/10.1080/14786440109462720

Ramasso, E., Saxena, A., 2014. Performance benchmarking and analysis of prognostic methods for CMAPSS datasets. Int. J. Progn. Health Manag. 15.

Saxena, A., Goebel, K., Simon, D., Eklund, N., 2008a. Damage propagation modeling for aircraft engine run-to-failure simulation. 2008 International Conference on Prognostics and Health Management. IEEE, pp. 1–9. https://doi.org/10.1109/PHM.2008.4711414

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., Schwabacher, M., 2008b. Metrics for evaluating performance of prognostic techniques. 2008 International Conference on Prognostics and Health Management. IEEE, Denver, CO, USA, pp. 1–17. https://doi.org/10.1109/PHM.2008.4711436

Schölkopf, B., Smola, A., Müller, K.-R., 1997. Kernel principal component analysis. In: Goos, G., Hartmanis, J., van Leeuwen, J., Gerstner, W., Germond, A., Hasler, M., Nicoud, J.-D. (Eds.), Artificial Neural Networks -? ICANN'97, vol. 1327. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 583–588. https://doi.org/10.1007/BFb0020217

Wang, T., Yu, Jianbo, Siegel, D., Lee, J., 2008. A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems. 2008 International Conference on Prognostics and Health Management. IEEE, Denver, CO, USA, pp. 1–6. https://doi.org/10.1109/PHM.2008.4711421

Yu, W., Kim, I.Y., Mechefske, C., 2019. Remaining useful life estimation using a bidirectional recurrent neural network based autoencoder scheme. Mech. Syst. Signal Proces. 129, 764–780. https://doi.org/10.1016/j.ymssp.2019.05.005

Yu, W., Kim, I.Y., Mechefske, C., 2020. An improved similarity-based prognostic algorithm for RUL estimation using an RNN autoencoder scheme. Reliab. Eng. Syst. Saf. 199, 106926. https://doi.org/10.1016/j.ress.2020.106926

Zhang, W., Liu, J., Gao, M., Pan, C., Huusom, J., 2019. A fault early warning method for auxiliary equipment based on multivariate state estimation technique and sliding window similarity. Comput. Ind. 107, 67–80. https://doi.org/10.1016/j.compind.2019.01.003

Zhao, Z., Liang, Bin, Wang, X., Lu, W., 2017. Remaining useful life prediction of aircraft engine based on degradation pattern learning. Reliab. Eng. Syst. Saf. 164, 74–83. https://doi.org/10.1016/j.ress.2017.02.007

Zhou, Y., Huang, M., Chen, Y., Tao, Y., 2016. A novel health indicator for on-line lithium-ion batteries remaining useful life prediction. J. Power Sources 321, 1–10. https://doi.org/10.1016/j.jpowsour.2016.04.119