Reliability Engineering and System Safety 237 (2023) 109321

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Simplified models of remaining useful life based on stochastic orderings

Luciano Sánchez [a,*], Nahuel Costa [a], Inés Couso [b]

[a] Oviedo University, Computer Science Department, 33202, Gijón, Spain
[b] Oviedo University, Department of Statistics, Operations Research and DM, 33202, Gijón, Spain

ARTICLE INFO

Keywords:
Remaining useful life
Prognostics and health management
Interpretability
Variational inference
Monotonic neural networks

ABSTRACT

A method for designing simple models of the remaining lifetime of a system is proposed. A health state model is learned, the output of which varies consistently with the remaining useful life. The model and the criterion used to measure how well it fits the data are jointly learned. The goal of this joint search is to find the criterion, within a family of stochastic orderings, for which the model has the simplest expression. The performance of the new method is comparable to recent AI-based models, such as recurrent networks, convolutional networks or variational autoencoders, and depends on a much smaller number of parameters than these methods, so it can be applied in systems with reduced computational capacity.

1. Introduction

Recent estimators of the Remaining Useful Life (RUL) of a system employ machine learning techniques such as Convolutional Neural Networks, Recurrent Networks or Variational Autoencoders [1]. The accuracy and interpretability of these techniques is remarkable and, in principle, they are the technology of choice for new designs. However, the computational complexity of machine learning-based lifetime estimation algorithms may limit their practical application, because the control electronics of many devices are not powerful enough to implement state-of-the-art algorithms.

Taking the example of the well-known CMAPSS benchmark [2], a RUL model based on multilayer perceptrons depends on a few hundred parameters [3]. A convolutional neural network may require up to 3K weights [4], and some recurrent networks need up to 50K variables to be stored in memory [5]. The most recent techniques, for example the variational autoencoder, may require up to 700K variables [1] (see Fig. 1). Storage space is not necessarily a problem, but embedded processors in many systems do not have the capacity to handle matrices and vectors of the required size. There are software frameworks for AI in edge or embedded devices [6], but using the latest AI-based lifetime prediction techniques on small to medium-sized systems (think, for instance, of an industrial fan or a light vehicle) is not realistic unless the equipment is connected to the cloud and diagnostics are performed remotely [7].

The problem addressed in this study is to exploit recent machine learning techniques to find a simple model, dependent on a very small number of parameters, that can be used to infer the lifetime of equipment that does not have a permanent internet connection and has access to only moderate computing power. We will not apply these limitations to the process of learning the model (since, except in adaptive models, the learning is not done with the computer embedded in the system) but to the task of autonomously inferring the RUL of the equipment from the signals measured in it.

To achieve this result, we do not look for a precise predictor of the remaining lifetime of the equipment, but we will limit ourselves to synthesise a State of Health (SoH) variable that varies consistently with the lifetime (i.e., whenever the expected lifetime of the equipment decreases, this synthetic health variable also decreases in value). In this way, as will be explained in the next sections, the expression of the predictive model can be made considerably simpler, since it is no longer necessary to model certain nonlinearities in the system; increases in ageing need not necessarily be proportional to reductions in service life.

Model learning will not be based on minimising prediction error, but on maximising a correlation between model predictions and training data [8,9]. However, unlike previous work, the proposed algorithm simultaneously searches for the model and the best function (within a family of stochastic orderings) that defines the correlation between the model and the data. The aim of this joint search is to find the correlation for which the SoH model has the simplest expression. In this study we employ the family proposed in [10], which is representable by a type of neural network to be introduced later.

2. Related work

In this section, we present a literature review on probabilistic health estimators in the context of condition-based maintenance optimisation

* Corresponding author.
E-mail addresses: luciano@uniovi.es (L. Sánchez), costanahuel@uniovi.es (N. Costa), couso@uniovi.es (I. Couso).

https://doi.org/10.1016/j.ress.2023.109321
Received 31 January 2023; Received in revised form 29 March 2023; Accepted 23 April 2023
Available online 3 May 2023

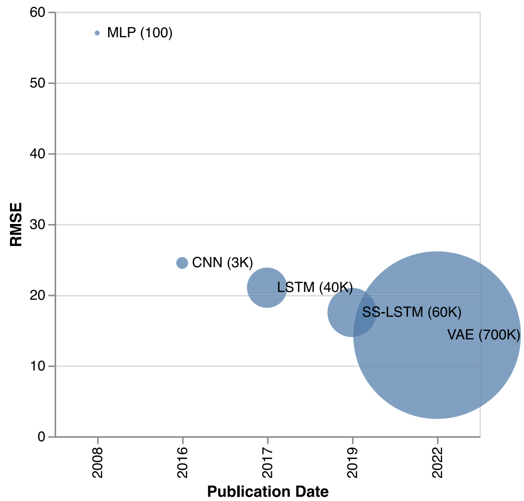

Fig. 1. Historical evolution of the number of parameters of the neural networks used to estimate the RUL of a system versus the accuracy obtained with the corresponding models. The area of the circles is proportional to the number of parameters needed to solve the CMAPSS problem. The Y coordinate is the root mean square error of each model.

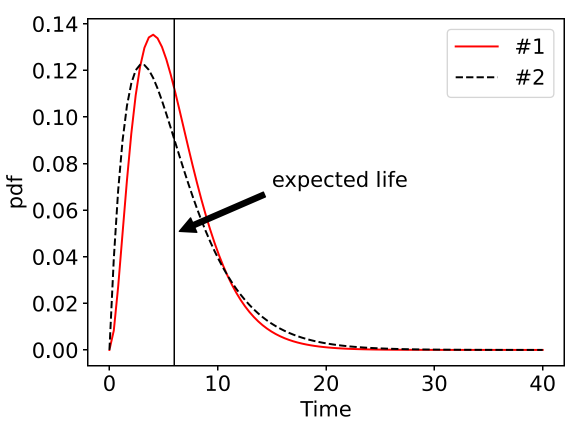

Fig. 2. Probability distributions of the RULs of two systems #1 and #2 whose RULs follow distributions $\gamma(3,2)$ and $\gamma(2,3)$, respectively.

models for stochastically deteriorating systems. The novelty and contributions of the proposed model are also presented, followed by the organisation of the article.

2.1. Literature review

RUL estimation plays an important role in condition-based maintenance (CBM) and in Prognostics and Health Management (PHM) [11]. The practical applications of the technique proposed in this paper mainly concern the first line: simplify the planning of inspections using CBM and define a maintenance policy that optimises system performance based on certain criteria, such as cost, availability or reliability.

The popularity of CBM is largely based on stochastic deterioration models [12]. It is common to divide CBM methods into three categories [13]: discrete-state and steady-state deterioration models, for environments where the system state is observable, and proportional hazard models (PHM) for systems operating in dynamic environments [14]. Discrete-space methods are, in general, related to Markov processes [15] and steady-state methods are based on stochastic processes; in most cases, on Brownian motion [16], Gamma processes [17] or Gaussian and Inverse Gaussian [18].

Deterioration models based on stochastic processes update the probability of system failure with continuously collected information about the state of the system. In a real working environment, however, there may be constraints that limit the use of steady-state models: the theoretical model of the system may have an imperfect fit under real-world conditions [19], or the environment may be time-varying [20], requiring the use of adaptive models [21–23]; in this context, it has also been studied how to combine stochastic methods with machine learning, taking advantage of pre-existing run-to-failure data [24,25], or how to make resilient estimates to imperfect inspection data [26].

The use of intelligent techniques to make up for the lack of information about the physical process or its state has the trade-off, as mentioned in the introduction, that the algorithms required are increasingly complex [27,28]. The balance between the cost of equipment inspections and the cost of the electronic equipment for CBM embedded in the system is important in some applications, and this study proposes to simplify, by means of a new machine learning algorithm, the equipment health estimation algorithm. Some previous work has evaluated simplified models of the RUL [29], but to the best of our knowledge the approach to be followed in this work, which is to design a machine learning algorithm that produces a health model that is dependent on few parameters and that is comonotone with the RUL, has not been studied before. From a methodological point of view, the techniques employed are related to the Ref. [30].

2.2. Novelty and contribution

The contributions of this study are:

- Development of a learning algorithm for a health model dependent on a reduced set of parameters, usable in systems with low computing power and/or without data connection to an external computing centre.
- Methodology for joint learning of a health model and the best ordinal correlation between health and RUL.
- Definition of a parametric family of ordinal correlations through a neural encoding of the generating function of a stochastic order.

2.3. Overview

The structure of the paper is as follows: Section 3 introduces a family of stochastic comparisons between random variables that can be represented by monotone neural networks. Section 4 defines the SoH model and its learning algorithm. Section 5 presents testable experimental results on a classical benchmark (the CMAPSS problem mentioned above) and two applications to real data, from the diagnosis of aviation turbofan engines and road tunnel ventilation fans. The paper concludes in Section 6, where future work on this topic is also indicated.

3. Learning by comparing random variables

Knowledge of RUL, in general, is not deterministic: there can be no absolute certainty about when equipment will develop a fault. It is common to characterise our knowledge about the RUL by a probability distribution, whereby it is possible to define concepts such as the expected lifetime or the probability that the RUL is higher than a given value.

For ease of explanation, let us assume we have two systems, whose RULs are characterised by two probability distributions $\gamma(3,2)$ and $\gamma(2,3)$. See Eq. (19) for the mathematical expression of the density function of the γ distribution and Fig. 2 for the graphical representation of both.

The expected lifetime of both systems is the same: 6 units of time (vertical line in Fig. 2). However, the probability that the life of the

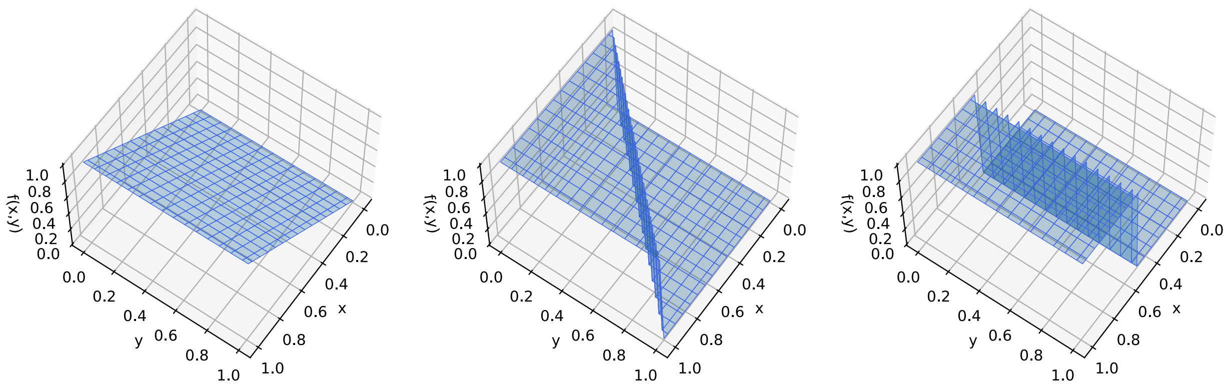

Fig. 3. Functions $f_1(x,y) = x$ (left), $f_2(x,y) = 1_{x>y}$ (centre) and $f_3(x,y) = 1_{x>0.5}$ (right) that are associated with dominance in expectation, statistical precedence and first-order stochastic dominance, respectively.

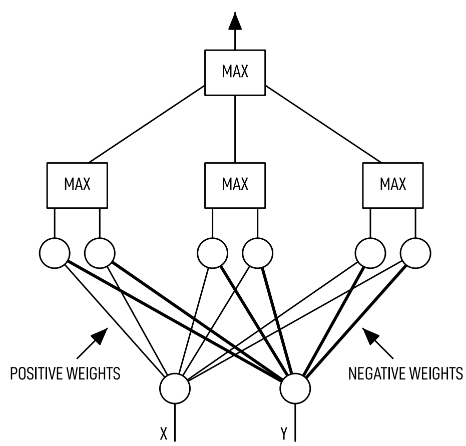

Fig. 4. Sill's monotone neural network [31] can encode an increasing function in one argument and decreasing in the other if the sign of the weights connecting the input "Y" to the neurons of the first layer is reversed.

first system is higher than the life of the second is (assuming that their respective lifetimes are independent from each other)

$$\int_0^\infty dx \int_0^x \gamma(x;3,2) \cdot \gamma(y;2,3)\; dy = 0.525, \tag{1}$$

while the probability that the second system lasts longer than the first is $1 - 0.525 = 0.475$, so the first equipment is more likely to last longer. On the other hand, the probability that the first equipment lasts more than 10 units of time is

$$\int_{10}^\infty \gamma(x;3,2)\; dx = 0.125, \tag{2}$$

while the probability that the second system lasts more than 10 units is 0.155, then the second system is more likely to last more than 10 units of time. In the case that we have several equipments and we have to order them from lowest to highest RUL, it is clear from this example that there is not an order that is inherently better than the others.

Continuing with the same example, let the RULs of the first and second systems be two independent random variables A and B. If the ordering criterion consisted of assigning each system its expected life (this criterion is called "dominance in expectation"),

$$A \prec_E B := E(A) \le E(B) \tag{3}$$

both equipments would have the same rank. If it is decided to compare them according to the probability that one lasts more than the other, i.e.

$$A \prec_S B := P(A \le B) \ge P(B \le A) \tag{4}$$

then $B \prec_S A$ (the criterion used is statistical precedence). Finally, if we decide to select the one with higher probability of lasting more than 10 units of time,

$$A \prec_{10} B := P(A \ge 10) \le P(B \ge 10) \tag{5}$$

then $A \prec_{10} B$ (which is a weak type of first-order statistical dominance).

The three orderings mentioned are stochastic preferences [32]. Recently a study has been developed in which it is justified that many comparisons between random variables in the literature share a simple definition, which is based on a function of two arguments $f(x,y)$, monotonic increasing in x and monotonic decreasing in y [10]. This definition is

$$A \prec B := E(f(A,B)) \le E(f(B,A)). \tag{6}$$

As an example, Fig. 3 shows three functions $f_1(x,y) = x$, $f_2(x,y) = 1_{x>y}$ and $f_3(x,y) = 1_{x>\alpha}$ that are associated with the $\prec_E$, $\prec_S$ and $\prec_\alpha$ criteria introduced in this example. The importance of this definition is that many stochastic preferences and orders can be constructed from a function f with the above-mentioned properties, which allows for joint learning of the SoH model and the criterion used to measure how well the model fits the data.

3.1. Joint learning of the SoH model and a stochastic comparison criteria

The training set comprises a sample of S systems whose RULs are independent random variables $\mathrm{R}^1, \dots, \mathrm{R}^S$. There is a vector $\mathbf{x}^s$ for each system containing a sequence of measurements taken on the sth system over time.

The objective of the learning problem is to obtain:

(1) A model of the SoH, which is a mapping that assigns a random variable SoH $\sim p_{\theta(\mathbf{x})}$ to a system characterised by a vector of measurements $\mathbf{x}$. The probability distribution of this random variable depends on a parameter $\theta(\mathbf{x})$. This parameter is, in turn, a function of $\mathbf{x}$.
(2) A function f, defining a criteria for comparing two random variables.

The need to obtain this second function is explained below. Consider the following matrix, formed by pairwise comparisons of the random variables comprising the training set:

$$P = \begin{pmatrix} 0 & \mathrm{R}^1 \prec \mathrm{R}^2 & \dots & \mathrm{R}^1 \prec \mathrm{R}^S \\ \mathrm{R}^2 \prec \mathrm{R}^1 & 0 & \dots & \mathrm{R}^2 \prec \mathrm{R}^S \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{R}^S \prec \mathrm{R}^1 & \mathrm{R}^S \prec \mathrm{R}^2 & \dots & 0 \end{pmatrix} \tag{7}$$

Let be

$$g(x,y) = f(x,y) - f(y,x) \tag{8}$$

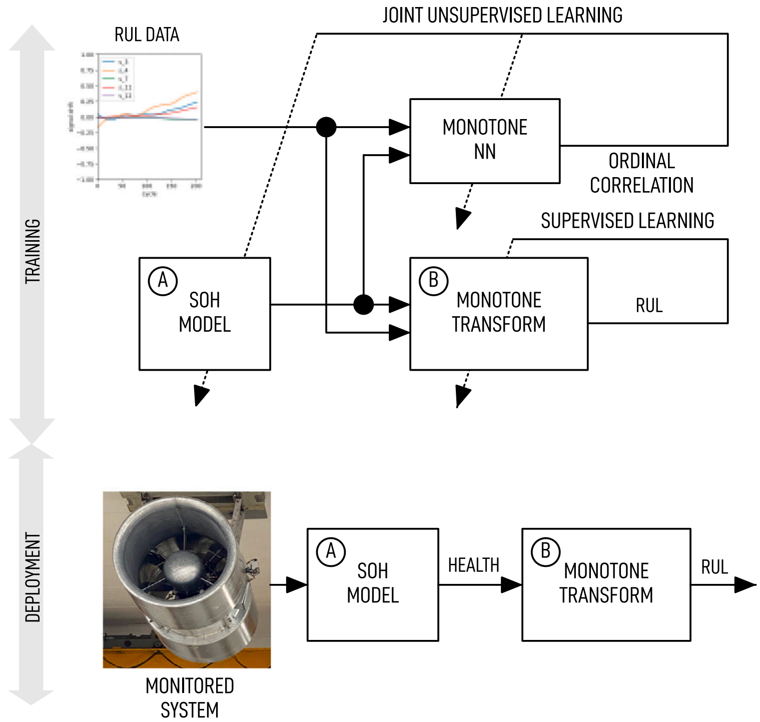

Fig. 5. Block diagram of the proposed method. The learning is performed from time-to-end RUL data, and results in two models: (A) a model of the health state, which produces an output that is comonotonic with the RUL, and (B) a transformation that converts the health state values into RUL values.

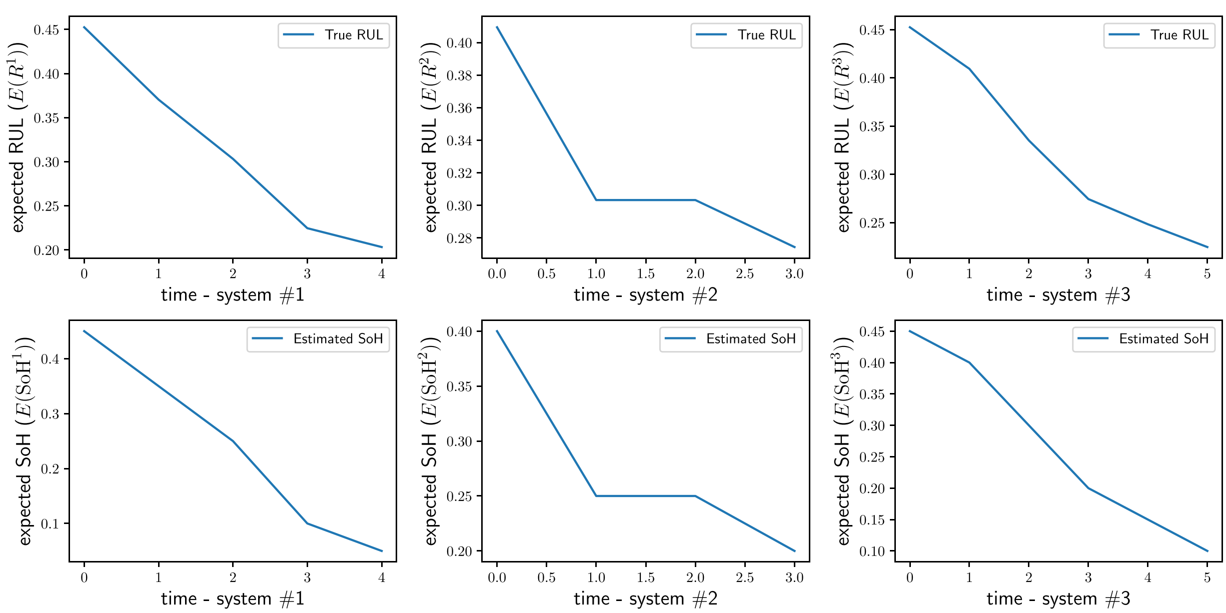

Fig. 6. Top: True expected RULs for the systems in the example in Section 3.3. Bottom: expected SoHs estimated with the proposed method. The estimates are comonotonical with the expected life of the system.

and

$$G = \begin{pmatrix} 0 & E(g(\mathrm{R}^1, \mathrm{R}^2)) & \dots & E(g(\mathrm{R}^1, \mathrm{R}^S)) \\ E(g(\mathrm{R}^2, \mathrm{R}^1)) & 0 & \dots & E(g(\mathrm{R}^2, \mathrm{R}^S)) \\ \vdots & \vdots & \ddots & \vdots \\ E(g(\mathrm{R}^S, \mathrm{R}^1)) & E(g((\mathrm{R}^S, \mathrm{R}^2))) & \dots & 0 \end{pmatrix} \tag{9}$$

G contains essentially the same information as P: the elements G_{ij} of this matrix are positive when the RUL of the ith system is "higher" (in a stochastic sense) than that of the jth, zero if equal or non-comparable and negative if lower.

Let $\mathrm{SoH}^s \sim p_{\theta(\mathbf{x}^s)}$ be the output of the model when the input is $\mathbf{x}^s$. The matrix

$$H = [H_{ij}] = [E(g(\mathrm{SoH}^i, \mathrm{SoH}^j))] \tag{10}$$

encodes the pairwise comparisons between the model's predictions. We propose that the evaluation function of a model counts the coincidences and discrepancies between the signs of the elements of the P and H matrices. The precise method by which these differences will be counted will be defined later, in Section 4.2. As will be seen, the metric is analogous to statistical tests that measure the correlation between ordinal values, and the proposed algorithm is also related to certain

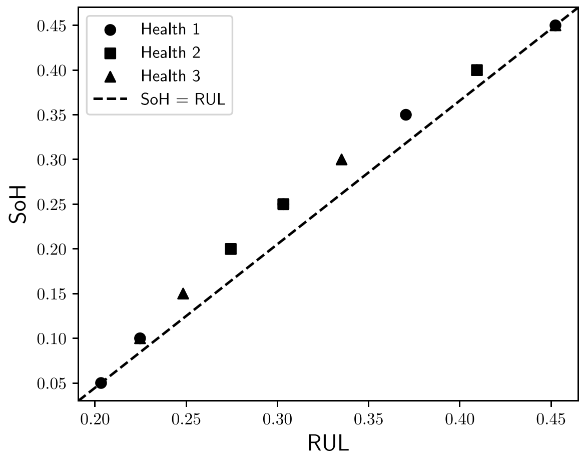

Fig. 7. Representation of the values of $E(\mathrm{R})$ as a function of $E(\mathrm{SoH})$, showing that one can be obtained from the other by a monotonic transformation.

algorithms of the "learning to rank" type [33], although in this case the comparisons are not between numerical values, but between random variables.

3.2. Using monotone neural networks to represent a stochastic order

Neural networks are a convenient way to parameterise the f functions mentioned in the previous section, so that the joint learning of the health state model and the stochastic ordering can be posed as an optimisation problem of a vector of real numbers; a part of the components of this vector will serve to define the function $\theta(\mathbf{x})$ and the remaining part will be the weights of the neural network defining f.

There are several neural architectures for representing monotonic functions with respect to each of their inputs. The case we are concerned with in this study (a monotonic function increasing with respect to one variable and decreasing with respect to the other) is not explicitly covered in any previous work, to our knowledge. Among the different options available, we have chosen the architecture proposed by Sill [31], because it can be easily adapted to the problem at hand. In this neural network, monotonicity is guaranteed because the weights of the input layer are positive (thanks to an exponential transformation). If, using a similar transformation, the weights connecting the second input to the MAX layer (see Fig. 4) are forced to be negative, the resulting function is monotonic increasing with respect to the first input and monotonic decreasing with respect to the second input, as desired.

3.3. Outline of the method

Fig. 5 contains a block diagram of the proposed method. The learning is performed from time-to-end RUL data, and results in two models: (A) a simple health state model, which produces an output that is comonotonic with the RUL, and (B) a transformation that converts the health state values into RUL values. The learning of health model is unsupervised, as the model does not approximate the known RUL values, but their correlation with the SoH. This correlation between RUL and SoH is measured by a function that depends on the monotone neural network described in the preceding section; this network is only used during learning, and is not part of the model that is deployed.

Having stated the main ideas, we illustrate the steps of the proposed method by means of an illustrative example. Let us suppose that have 3 systems with the same properties, each of them characterised by a sequence $\mathbf{x}^s$, $s = 1, 2, 3$ of measurements of a variable, and whose RULs follow uniform probability distributions:

$$\mathrm{R}^s \sim U[0, e^{-\sum_i x_i^s}] \tag{11}$$

Table 1
Table with the measurements of the example in Section 3.3.

System	Measurements	RUL	
# 1	$x^1 = [0.1, 0.2, 0.2, 0.3, 0.1]$	$\mathrm{R}^1 \sim U[0, 0.406]$	$E(\mathrm{R}^1) = 0.203$
# 2	$x^2 = [0.2, 0.3, 0.0, 0.1]$	$\mathrm{R}^2 \sim U[0, 0.549]$	$E(\mathrm{R}^2) = 0.274$
# 3	$x^3 = [0.1, 0.1, 0.2, 0.2, 0.1, 0.1]$	$\mathrm{R}^3 \sim U[0, 0.449]$	$E(\mathrm{R}^3) = 0.224$

The data for a conventional algorithm would be the three measurement sequences and the expectation of their RULs (see Table 1). On the contrary, the proposed algorithm is only provided with an ordering between the RULs. In this particular case, let us assume that $\mathrm{R}^1 \prec \mathrm{R}^3 \prec \mathrm{R}^2$, which is encoded in the following matrix of pairwise comparisons:

$$P = \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} \tag{12}$$

The conventional RUL prediction algorithm would ideally conclude with the numerical estimator

$$E(\mathrm{R}) = \frac{1}{2} e^{-\sum_i x_i} \tag{13}$$

whereas the proposed algorithm would end up with two outcomes: (a) a potentially simpler definition of the probability distribution of the SoH, for instance

$$\mathrm{SoH} \sim U[0, 1 - \sum_i x_i] \tag{14}$$

and (b) a monotone neural network-based definition of the function defining the best stochastic order, for instance

$$f(x, y) = \begin{cases} 1 & x \geq y \\ 0 & x < y \end{cases} \tag{15}$$

because, for this particular choice of f, the signs of the elements of the matrix

$$H = \begin{pmatrix} 0 & 1.875 & 0.75 \\ -1.875 & 0 & -0.75 \\ -0.75 & 0.75 & 0 \end{pmatrix} \tag{16}$$

coincide with the signs of P. Note that the elements of the matrix H have been calculated as indicated in Eq. (10), and in this particular case

$$A \sim U[0, v_1] \quad B \sim U[0, v_2]$$
$$E(f(A, B)) = \int_0^{v_1} dx \int_x^{v_2} \frac{1}{v_1} \frac{1}{v_2} dy = -\frac{v_1}{2v_2} \tag{17}$$

$$E(g(A, B)) = \frac{v_2}{2v_1} - \frac{v_1}{2v_2} \tag{18}$$

Fig. 6 shows, graphically, the expected RUL values for each of the three systems in this example and the estimates of the SoH using Eq. (14). Note that the expectation of the SoH is different than the expectation of the true RUL, but both expectations are comonotonic, thus monotonic regression can be used to find a transformation that converts $E(\mathrm{R})$ to $E(\mathrm{SoH})$, as shown in Fig. 7.

The SoH estimator (recall Eq. (14)) can be implemented in different ways. In this work we propose a simple expression that produces reasonable results in many practical problems, as discussed in the next section.

4. A simple SoH model

As seen in the previous example, the proposed method does not seek an accurate model of the RUL but a model that decreases in time consistently with the RUL, together with a stochastic ranking criteria under which the model estimates are in the same order as the training set data. In this section we propose a model that depends on a small number of parameters but, as will be seen in the experimental results section, it offers a good practical behaviour. In any case, the proposed

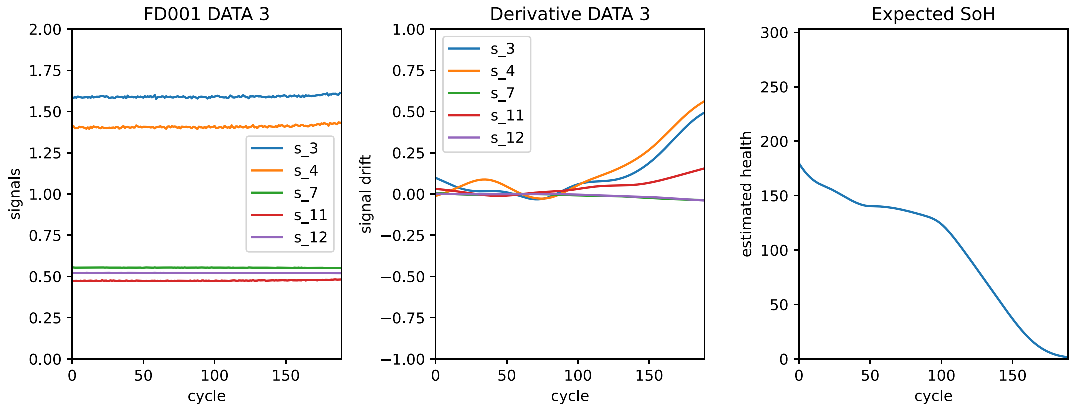

Fig. 8. Left: Signals $\mathbf{x}$ of the engine $s = 3$ of the CMAPSS/FD001 problem, with number of cycles $M_3 = 179$ and measured up to failure, and therefore with label $\hat{\mathrm{L}}^3 = 0$. Centre: signal derivative $\dot{\mathbf{x}}(t)$ ($\times 10^3$). Right: Estimated health for each time instant according to Eq. (21).

method is applicable to any lifetime estimator; this model is given as a reference, but can be replaced by any other function appropriate to the problem.

Let SoH_t be the SoH of a system at time instant t. Let we assume that $\mathrm{SoH}_t \sim \gamma(k, \theta_t)$ (which is a common assumption in this kind of problems [17]). The density function of this distribution is

$$f(\tau; \theta_t, k) = \frac{\tau^{k-1} e^{-\tau/\theta_t}}{\theta_t^k \Gamma(k)}, \tag{19}$$

so that the expected SoH at time $t = 0$ is $E(\mathrm{SoH}_0) = k\theta_0$.

At each time instant t_i, $i = 1, \dots, N$ the system undergoes a degradation $D_i \in [0, 1]$, such that

$$f(\tau; \theta_{i+1}, k) = f(\tau; \theta_i D_i, k), \tag{20}$$

thus the SoH at time t_i also follows a gamma distribution,

$$\mathrm{SoH}_i \sim \gamma(k, \theta_0 \prod_{j=0}^{i-1} D_j) = \gamma(k, \theta_0 e^{\sum_{j=0}^{i-1} \log D_j}). \tag{21}$$

The term D_i measures the reduction in the expected health of the system between instants t_i and t_{i+1}. A value of $D_i = 1$ means that no event affecting the life of the equipment has occurred during that period, and a value of (for example) $D_i = 0.5$ means that the expected health has been halved at that instant of time.

We also propose that the logarithm of the instantaneous degradation D_i depends only on the derivatives of the sequence of measurements of the system at time t_i. Given two parameter vectors α and β, we define the logarithm of the deterioration as

$$\log D_i = -\|\mathrm{relu}(\alpha \cdot \dot{\mathbf{x}}_i - \beta)\| \tag{22}$$

where $\mathrm{relu}(x) = \max(x, 0)$.

This model is based on a low number of parameters, compared to recent AI-based methods (the number of parameters is twice the number of signals) and has a simple physical interpretation: at each instant of time the system does not deteriorate ($\log D_i = 0$) unless the derivatives of the signals (affected by a scaling factor α, with sign) exceed a threshold β. If this threshold is exceeded, the higher the (scaled) derivative of the signal, the greater the deterioration (see Fig. 8 for an example).

Note also that with this model it is not necessary to assume that the useful life is piecewise linear [34], and that it is immediate to use the deterioration profile to identify at what time instant the damage has occurred and how large it has been (as already seen in Fig. 8).

4.1. Fitting the model to a data sample

Let $\mathrm{R}_i^s \sim \gamma(k, \theta_i^s)$, $s = 1, \dots, S$, be the remaining life of a system s at time t_i. The training set consists, as mentioned, of S sequences of measurements of the signals $\mathbf{x}^s = [x_i^s]_{i=1\dots M_s}$ (recall the left part of Fig. 8), accompanied by a label L^s each, measuring the health of each system in the last period of the sequence M_s. As we are only going to use the signs of the pairwise comparisons of these measurements, it is not necessary to have a numerical estimate of the expected lifetime of the equipment; ordinal labels such as “GOOD”, “NORMAL” or “BAD” are enough.

Let us assume that

- The labels are ordered: $\mathrm{L}^1 \preceq \mathrm{L}^2 \preceq \dots \preceq \mathrm{L}^S$
- The equipment is not completely deteriorated at the initial instant, but that the life expectancy of the equipment at that time is not known.
- The initial SoH linearly depends on the expected number of time instants $M_s + E(\mathrm{SoH}^s) = M_s + k\theta_{M_s}^s$ that elapse from the beginning of the sequence of measurements until the end of life of the equipment, that is:

$$k\theta_0^s = \eta(M_s + k\theta_{M_s}^s) + \delta \tag{23}$$

and

$$\theta_{M_s}^s = \theta_0^s e^{\sum_{j=0}^{M_s-1} \log D_j} \tag{24}$$

therefore,

$$k\theta_{M_s}^s = \frac{(\eta M_s + \delta) e^{\sum_{j=0}^{M_s-1} \log D_j}}{(1 - \eta e^{\sum_{j=0}^{M_s-1} \log D_j})} \tag{25}$$

The learning objective is to estimate the set of parameters α, β, η, δ and k for which the matrix H of precedences between the SoH of the equipment and the matrix P of precedences between the labels (recall Eqs. (9) and (10)) are concordant, as explained in the next section.

4.2. Learning algorithm

There are different ways of counting the number of discrepancies between P and H. We will use the subtraction between the number of matrix elements with matching sign and the number of elements with discordant sign, divided by the number of matrix elements, which is the well-known Kendall’s τ_A statistic mentioned before. We will consider, as in Ref. [35], that there are non-comparable elements. Since we intend to use descent algorithms to solve the optimisation problem, we will make the smooth approximation $\mathrm{sign}(x) \approx \tanh(\kappa x)$ (for a sufficiently high value of a constant κ) thus

$$\tau_A(P, H) = \frac{1}{S^2} \left(\sum_{i=1}^{S} \sum_{j=1}^{S} \tanh\left(\frac{\kappa P_{ij}}{\max(|P|)}\right) \cdot \tanh\left(\frac{\kappa H_{ij}}{\max(|H|)}\right) + C_0 - D_0 \right) \tag{26}$$

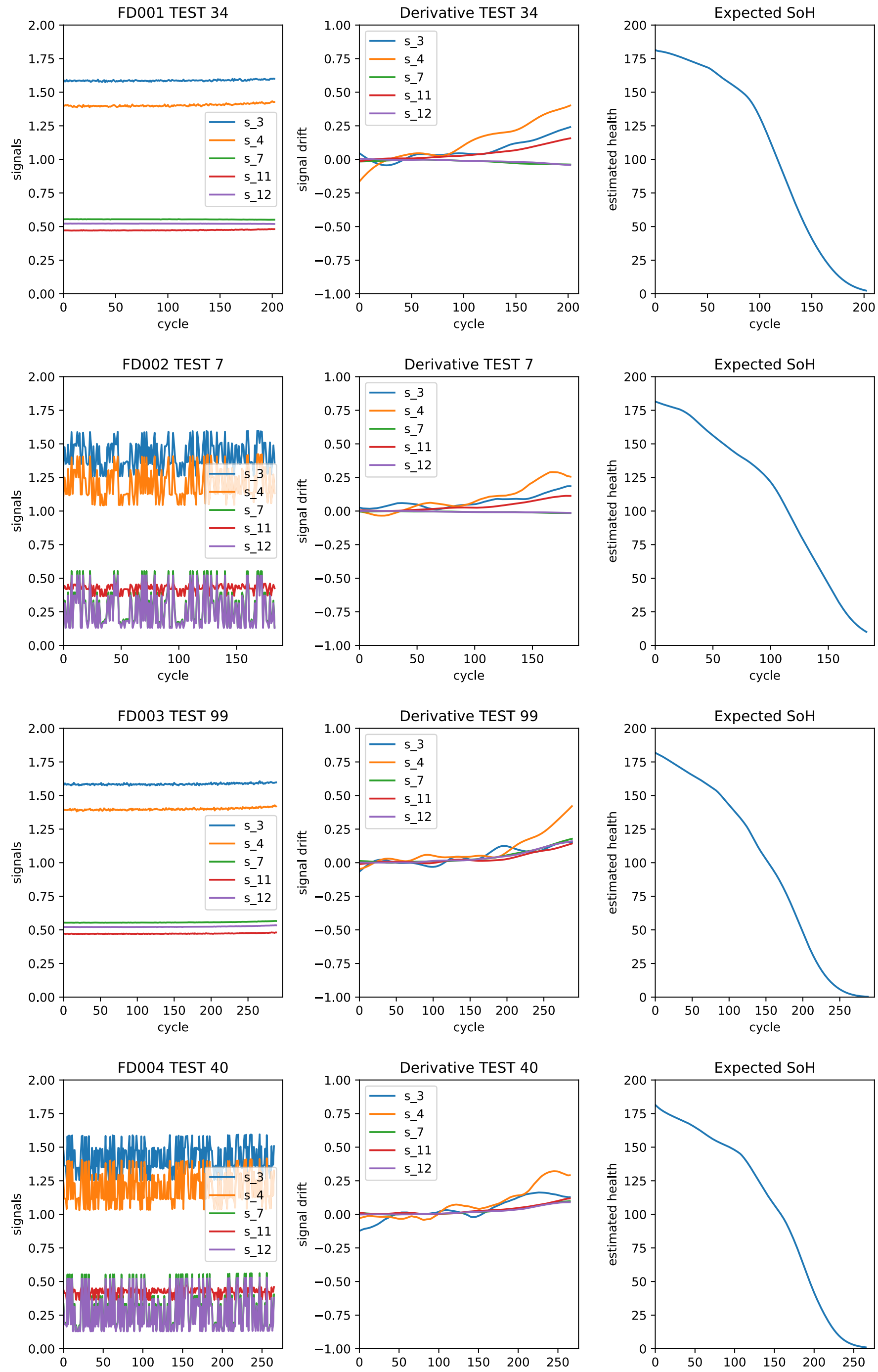

Fig. 9. From top to bottom and from left to right: measured signals, derivatives of the signals and synthesised health signal for engines FD001-34, FD002-7, FD003-99 and FD004-40 from the test sets of these problems. Fig. 10 shows the estimation of the RUL for the same problems using the RVE method.

where C_0 and D_0 measure the discrepancies in the pairs of values where any of the graph elements is zero:

$$
\begin{aligned}
C_0 &= \sum_{i=1}^{S}\sum_{j=1}^{S} 1_{P_{ij}=H_{ij}=0} \\
D_0 &= \sum_{i=1}^{S}\sum_{j=1}^{S} |H_{ij}| \cdot 1_{P_{ij}=0} + \sum_{i=1}^{S}\sum_{j=1}^{S} |P_{ij}| \cdot 1_{H_{ij}=0}
\end{aligned}
\tag{27}
$$

The purpose of the learning problem is to maximise $\tau_A(P,H)$ with respect to the parameters $\alpha,\beta,\eta,\delta,k$ and the weights of the neural network codifying f. In the experiments presented in the next section, this optimisation has been performed by alternating two steps: in the first one, the parameters defining the monotone neural network f are left frozen and τ_A is maximised with respect to the parameters $\alpha,\beta,\eta,\delta,k$. In the second, the opposite is done: the ageing model is frozen and the parameters of the neural network f are optimised. The process is repeated until one of the two optimisations cannot improve the initial error value.

4.3. Conversion of model output to numerical estimations

Recall that the maximisation of $\tau_A(P,H)$ does not make use of the expected values of the RULs in the training set, only the signs of their pairwise comparisons. In this sense, the prediction provided by Eq. (21)

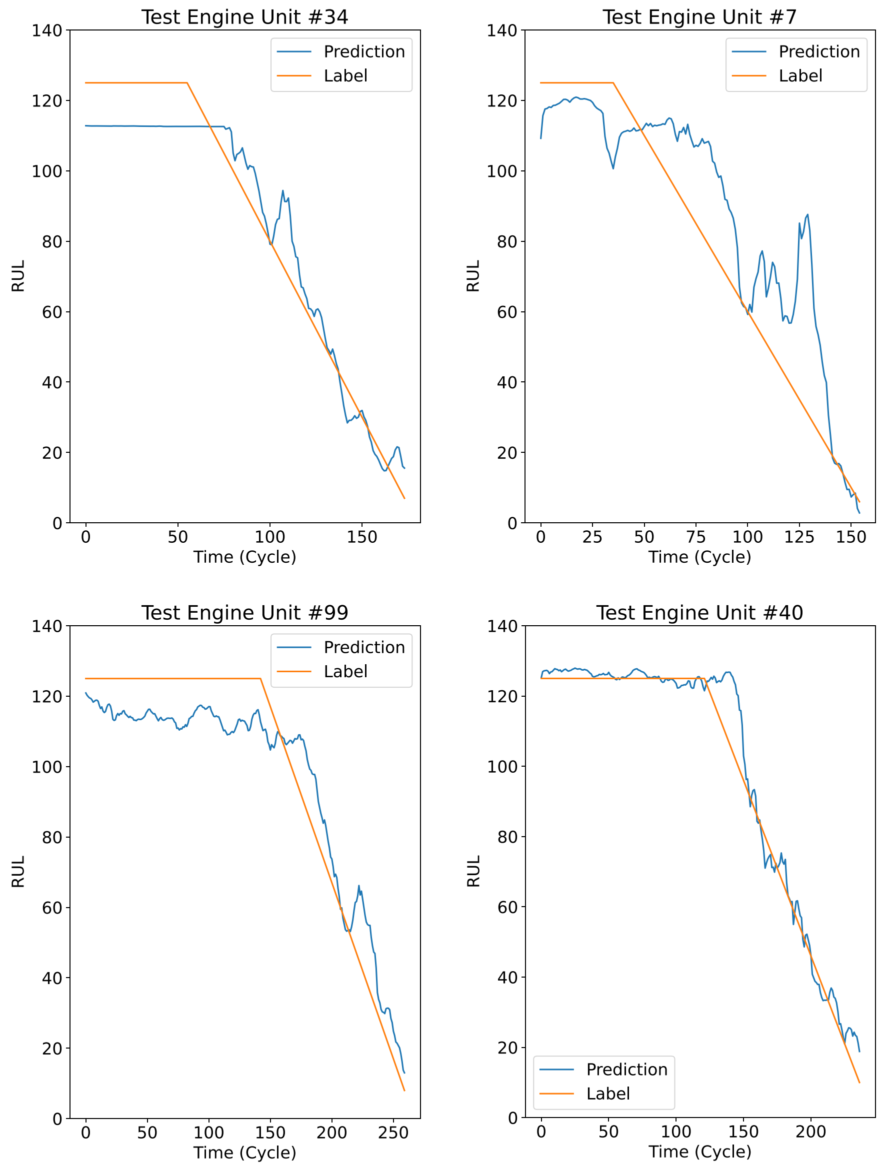

Fig. 10. Estimation of the RUL of engines FD001-34, FD002-7, FD003-99 and FD004-40 using the RVE method.

is a health signal, which varies consistently with the remaining lifetime but the expected value of SoH is not a RUL estimator.

As seen in the example in Section 3.3, to convert the predictions produced by Eq. (21) into RUL predictions one must solve a monotonic regression problem where the input variables are the expectations of the SoH for the elements of the training set and the output variables are the expected RULs in the training set. Note that this step does not significantly increase the complexity of the diagnostic algorithm, since the results of this monotonic regression problem can easily be stored in a look-up table of only a few hundred values, and this table can be stored in the non-volatile memory of the system computer. Moreover, in many practical applications this last step is not necessary, as an alarm can be triggered by a threshold in SoH and health does not need to be defined in time units.

5. Experimental results

The proposed method is validated in three ways. In the first part, the CMAPSS problem is used to measure how much accuracy is lost when different recent machine learning techniques are replaced by a simplified model. In the second part, the new solution is evaluated on a real turbo-fan engine diagnosis problem and compared to previous solutions of the same problem. Finally, the third part shows a real case where this method is implemented for CBM of a fan for road tunnels.

5.1. Measuring the simplicity-accuracy balance

To assess the balance between simplicity and accuracy of the proposed procedure we use the CMAPSS benchmark. This problem is representative of the problems we want to solve because it is small in size, then it is amenable to a simple method, and at the same time it is probably the benchmark to which the most extensive catalogue of AI-based RUL learning methods has been applied. A selection of intelligent techniques, from the multilayer perceptron (which has a complexity comparable to our method) to variational autoencoders (VAEs), is shown in Table 2 (taken from Ref. [1]). The methods are ordered from lowest to highest complexity and, in the last row, the results of the proposed method are shown both in terms of the root mean square error and the score proposed in [2] by the authors of the benchmark problem.

In view of the comparison table, co-learning a simple comonotone model and a stochastic comparison criteria produces results comparable

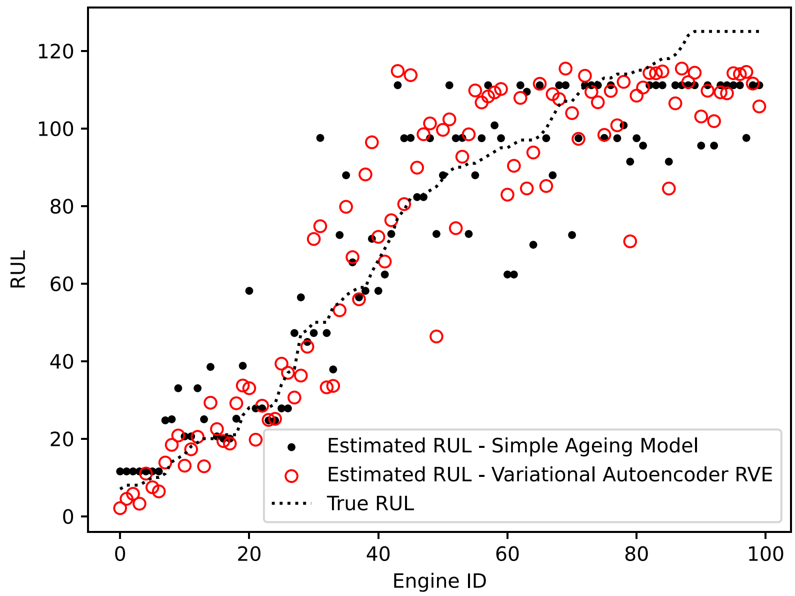

Fig. 11. FD001 problem: Graphical comparison between RUL predictions made with a variational autoencoder and with the proposed method. The dotted line corresponds to the actual RUL values (ordered from lowest to highest). The empty circles are the RVE prediction and the filled circles are the proposed method.

Table 2
Comparison of results on a selection of intelligent techniques applied to the CMAPSS problem.
Source: Upper part of the table reproduced from Ref. [1].

	FD001		FD002		FD003		FD004	
	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
MLP [3]	37.56	18 000	80.03	7 800 000	37.39	17 400	77.37	5 620 000
SVR [3]	20.96	1380	42.00	590 000	21.05	1600	45.35	371 000
RVR [3]	23.80	1500	31.30	17 400	22.37	1430	34.34	26 500
CNN [3]	18.45	1299	30.29	13 600	19.82	1600	29.16	7890
Deep LSTM [3]	16.14	338	24.49	4450	16.18	852	28.17	5550
Semi-Supervised [5]	12.56	231	22.73	3366	12.10	251	22.66	2840
DCNN [4]	12.61	273	22.36	10 412	12.64	284	23.31	12 466
MS-DCNN [4]	**11.44**	**196**	19.35	3747	**11.67**	**241**	22.22	4844
VAE+RNN [1]	15.81	326	24.12	4183	14.88	722	26.54	5634
RVE [1]	13.42	323	**14.92**	**1379**	12.51	256	**16.37**	**1846**
Simple Ageing Model	16.79	437	19.51	1555	17.48	588	22.82	3960

to those of a CNN or a Deep LSTM [3]. In particular, in the most complex problems (FD002 and FD004, with six operating points) the proposed model is not substantially different from techniques three orders of magnitude more complex. Fig. 9 shows the health signals synthesised by this method, which can be compared with the estimates of a VAE on the same problem in Fig. 10. The predictions are similar over a wide range of values and, from a practical point of view the gain of the more complex methods hardly justifies their additional complexity for this particular problem, as can be seen in Fig. 11.

We would also like to point out that, unlike most methods in the literature, the proposed method generalises better than the alternatives to unseen data because an RUL estimator trained with any of the problems can be applied, without a substantial loss of efficiency, to any of the other three, as shown in Table 3. To our knowledge, none of the methods published to date have this capability.

5.2. Mechanical deterioration of turbofan aircraft engines

In this section we address the problem of predicting the mechanical deterioration of two-shaft high bypass ratio turbofan engines with real data taken from the aircraft industry. The two main sections of any engine prone to failure are the high pressure compressor (HPC) and the turbine (HPT). HPC deterioration is mainly due to increased tip clearance or loss of material in a blade or in the span. Turbine deterioration can be due to combustion problems or damage to the blade. HPT deterioration is slower and more difficult to diagnose by analysing the signals monitored in the engine.

Table 3
Evaluation of a model trained on one problem on test data from a different problem. Results on each problem's own test set are marked in bold. The cross-tabulated results show that a simplified RUL estimator can have remarkable generalisation capability.

	FD001		FD002		FD003		FD004	
	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
Trained on FD001	**16.79**	**437**	19.29	1688	22.91	897	23.26	4232
Trained on FD002	17.65	517	**19.51**	**1555**	24.63	1259	24.94	4607
Trained on FD003	18.85	676	22.16	5297	**17.48**	**588**	22.74	8724
Trained on FD004	17.12	490	20.04	1764	19.03	992	**22.82**	**3960**

Forty-three aircraft with a slow deterioration pattern and no readily detectable anomalies were selected. Each of these aircraft has been given two labels to quantify the health of its HPC and HPT. Each label is one of the words (from best to worst health) “good”, “good to normal”, “normal”, “normal to high”, “high” or “bad”.

A comparative analysis of the following five techniques has been carried out:

- A cycle-count deterioration model. This is a baseline method: each engine has an expected life of 5000 cycles at the start of the series, and the RUL is the result of subtracting the number of cycles performed by the engine from this initial value.

Table 4
Mean absolute errors of RUL prediction of aircraft HPC and HPT systems with different techniques.

Method	HPC	HPT
Cycle counting	1737	1693
Neural network	1387	1526
Random Forest	1422	1582
Fuzzy Ageing Model	1380	1697
Simple Ageing Model	**1283**	1504

- A random forest regression model, in which the input features were obtained from the aircraft data using the method explained in [36].
- An ageing model based on fuzzy rules [37].
- The model proposed in this study.

Fig. 12 shows two boxplots with the test errors of fifty repetitions of the five models applied to the prediction of HPC and HPT degradation. In Table 5 the average values of the test errors of these models are shown; to obtain the numerical equivalence, the RUL values of 5000, 3000, 1000, 1000, 250, 10 and 0 cycles have been assigned to the labels “good”, “good to normal”, “normal”, “normal to high”, “high” or “bad”, respectively. The results are consistent with experience: the prediction of turbine deterioration is not substantially different from the trivial “cycle counting” method, but compressor deterioration is more predictable. The good generalisability of the method proposed in this study, already shown in the previous experimentation, allows obtaining results that, on average, are better than the alternatives . The prediction error of the compressor RUL is statistically lower than the best alternative (Wilcoxon test p-value less than 1%). The RUL of the turbine is also lower on average, but the difference is not statistically representative (see Table 4).

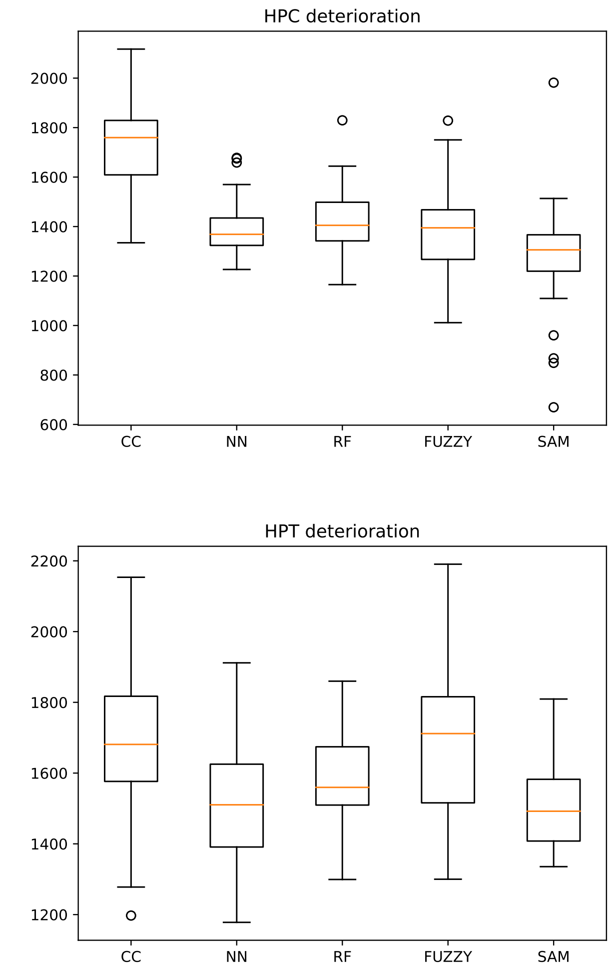

Fig. 12. Left part: Compressor deterioration (HPC): boxplots with the error dispersions of the results in the test set after ten repetitions of the algorithms “cycle counting” (CC), “neural network” (NN), “Random Forest” (RF), “Fuzzy Rule-Based System” (FUZZY) and “Simple Ageing Model” (SAM) methods. Right hand side: Turbine deterioration (HPT).

5.3. Condition monitoring of road tunnel fans

The operation of a fan can be affected by various mechanical anomalies, including imbalances, poor lubrication, vibrations caused by operation close to the stall area, vibrations caused by a loosened support, sudden starts or stops. To these factors must be added electrical and electronic problems (electric motor or inverter failures) and couplings or parasitic vibrations that can occur in installations where several fans operate in coordination.

The left-hand side of Fig. 13 shows a 30 kW axial fan used for ventilation in road tunnels. The right-hand side of the same figure shows a detail of the accelerometers installed to measure vibrations during operation. This type of equipment has a capture card with an industrial microprocessor, which controls the operating point (defined by the flow rate and pressure) and monitors eight health signals: the temperature in the three AC motor windings, the temperature in three bearings and the vibrations in two of the bearings.

In order to implement the fan monitoring system, tests have been carried out in which non-destructive failures of the fan (removal of lubricant from the bearings, loosening of brackets, etc.) have been provoked. The data measured in these experiments were used to identify a model of the fan, which in turn was used to extrapolate the deteriorations up to the end of the fan’s lifetime, for eight different types of deterioration and four operating points. Dataset 1P-1F has one operating point and one deterioration type (front bearing). The 4P-8F dataset has four operating points and eight different types of failure. Datasets 1P-8F and 4P-1F are intermediate between the two. The data files comprise the eight health variables mentioned above and the operating point, and have been made publicly available on [38].

The results of the experimentation are shown in Table 5. This table contains three methods: the “cycle counting” method assumes that the fan lifetime is the average lifetime of the elements in the training set, minus the number of cycles performed in the test set. The RVE method [1] is used as a measure of the state of the art in RUL prediction with intelligent data-driven methods. Finally, the result of the method developed in this study is presented. As can be seen in the table, the proposed method produces a coherent health signal in all cases. In the three simplest problems it is of equal or better quality than the more complex method, and is compatible with the limitations of the microprocessor embedded in the fan.

Table 5
Mean absolute errors of RUL prediction, fan deterioration problems.

Method	1P-1F	4P-1F	1P-8F	4P-8F
Cycle counting	66.12	55.97	46.62	48.95
VAE	35.48	37.29	30.31	**20.50**
Simple Ageing Model	**31.56**	**28.78**	**29.80**	26.96

6. Conclusions and future work

Life estimation is usually part of predictive maintenance routines and therefore it is not critical that RUL models are run in real time; in fact, two of the practical problems presented in this paper pertain to the aircraft industry, where it is common for this type of analysis to be performed by specialised companies that are only supplied with data on a monthly basis, or less. In such a case, it is not justified to simplify the model if this leads to a loss of accuracy. However, other equipment

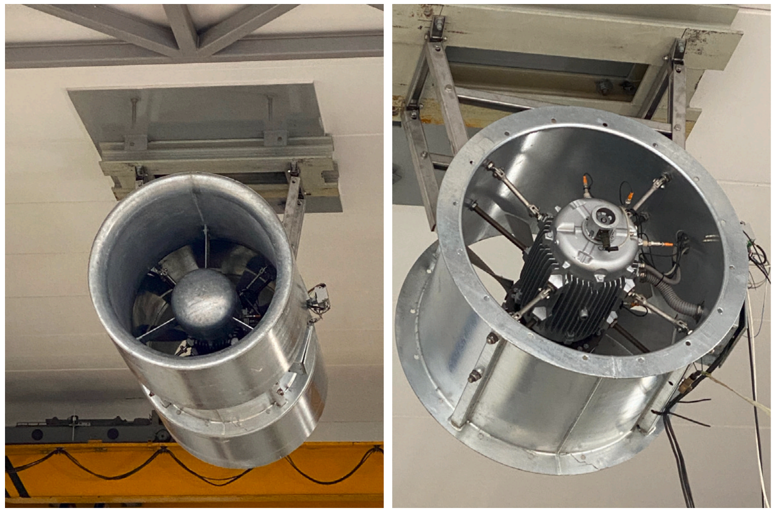

Fig. 13. 30 kW axial fan for road tunnel ventilation. Left side: fan with silencers. Right side: detail of the fan engine and the accelerometers used to measure vibrations.

(industrial fans, light transport vehicles, etc.) are not connected to the Internet and do not have sufficient computing power to apply recent AI methods. The proposed method is oriented towards these applications. A locally synthesised status signal can trigger an alarm before a failure occurs, and it is important to know how far in advance a failure can be anticipated with a local maintenance system versus a connected system.

From a conceptual point of view, the co-learning of a probabilistic model and a criterion for comparing random values, which has been used here to restrict the search for a model to a reduced space, has a wide range of applications. In future work we wish to explore other uses of this technique and, for example, use it to learn models from records where there are missing or censored values, for which a representation based on probability distributions is natural. On the other hand, it has been assumed that deterioration is a stochastic process in which the probability distribution of time to failure is Gamma. The proposed ageing model, where the logarithm of the ratio of life expectancies between two instants is modelled by a linear threshold function, can be extended to other families of probabilities, such as the inverse Gaussian, which can potentially produce better results in some practical problems.

CRediT authorship contribution statement

Luciano Sánchez: Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Investigation, Data curation, Conceptualization. **Nahuel Costa:** Writing – review & editing, Software, Investigation. **Inés Couso:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization.

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.

Data availability

Link to Mendely Data provided in the paper.

Acknowledgements

This work has been partially supported by the Ministry of Economy, Industry and Competitiveness ("Ministerio de Economía, Industria y Competitividad") from Spain/FEDER under grant PID2020-112726-RB-I00 and by Principado de Asturias, Spain, grant SV-PA-21-AYUD/2021/50994.

References

[1] Costa Nahuel, Sánchez Luciano. Variational encoding approach for interpretable assessment of remaining useful life estimation. Reliab Eng Syst Saf 2022;222:108353.
[2] Saxena Abhinav, Goebel Kai, Simon Don, Eklund Neil. Damage propagation modeling for aircraft engine run-to-failure simulation. In: 2008 international conference on prognostics and health management. IEEE; 2008, p. 1–9.
[3] Zheng Shuai, Ristovski Kosta, Farahat Ahmed, Gupta Chetan. Long short-term memory network for remaining useful life estimation. In: 2017 IEEE international conference on prognostics and health management. ICPHM, IEEE; 2017, p. 88–95.
[4] Li Han, Zhao Wei, Zhang Yuxi, Zio Enrico. Remaining useful life prediction using multi-scale deep convolutional neural network. Appl Soft Comput 2020;89:106113.
[5] Ellefsen André Listou, Bjørlykhaug Emil, Æsøy Vilmar, Ushakov Sergey, Zhang Houxiang. Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture. Reliab Eng Syst Saf 2019;183:240–51.
[6] Jeffrey Nicholas, Tan Qing, Villar José R. Simulators and testbeds for IIoT development and validation. In: 2022 IEEE international IOT, electronics and mechatronics conference. IEMTRONICS, IEEE; 2022, p. 1–5.
[7] Dutta Lachit, Bharali Swapna. TinyML meets IOT: A comprehensive survey. Internet of Things 2021;16:100461.
[8] Fahandar Mohsen Ahmadi, Hüllermeier Eyke, Couso Inés. Statistical inference for incomplete ranking data: the case of rank-dependent coarsening. In: International conference on machine learning. PMLR; 2017, p. 1078–87.
[9] Hüllermeier Eyke, Fürnkranz Johannes, Cheng Weiwei, Brinker Klaus. Label ranking by learning pairwise preferences. Artificial Intelligence 2008;172(16–17):1897–916.
[10] Couso Inés, Sánchez Luciano. Machine learning models, epistemic set-valued data and generalized loss functions: an encompassing approach. Inform Sci 2016;358:129–50.
[11] Hu Yang, Miao Xuewen, Si Yong, Pan Ershun, Zio Enrico. Prognostics and health management: A review from the perspectives of design, development and decision. Reliab Eng Syst Saf 2022;217:108063.
[12] Si Xiao-Sheng, Wang Wenbin, Hu Chang-Hua, Zhou Dong-Hua. Remaining useful life estimation–a review on the statistical data driven approaches. European J Oper Res 2011;213(1):1–14.
[13] Alaswad Suzan, Xiang Yisha. A review on condition-based maintenance optimization models for stochastically deteriorating system. Reliab Eng Syst Saf 2017;157:54–63.
[14] Azar Kamyar, Hajiakhondi-Meybodi Zohreh, Naderkhani Farnoosh. Semi-supervised clustering-based method for fault diagnosis and prognosis: A case study. Reliab Eng Syst Saf 2022;222:108405.
[15] Nguyen Van-Thai, Do Phuc, Vosin Alexandre, Iung Benoit. Artificial-intelligence-based maintenance decision-making and optimization for multi-state component systems. Reliab Eng Syst Saf 2022;228:108757.
[16] Wang Yukun, Li Xiaopeng, Chen Junyan, Liu Yiliu. A condition-based maintenance policy for multi-component systems subject to stochastic and economic dependencies. Reliab Eng Syst Saf 2022;219:108174.
[17] Liu Xingheng, Matias José, Jäschke Johannes, Vatn Jørn. Gibbs sampler for noisy transformed Gamma process: Inference and remaining useful life estimation. Reliab Eng Syst Saf 2022;217:108084.

[18] Lee Juseong, Mitici Mihaela. Multi-objective design of aircraft maintenance using Gaussian process learning and adaptive sampling. Reliab Eng Syst Saf 2022;218:108123.
[19] Zhao Yunfei, Smidts Carol. Reinforcement learning for adaptive maintenance policy optimization under imperfect knowledge of the system degradation model and partial observability of system states. Reliab Eng Syst Saf 2022;224:108541.
[20] Jahani Salman, Zhou Shiyu, Veeramani Dharmaraj. Stochastic prognostics under multiple time-varying environmental factors. Reliab Eng Syst Saf 2021;215:107877.
[21] Ayhan Bulent, Kwan Chiman, Liang Steven Y. Adaptive remaining useful life prediction algorithm for bearings. In: 2018 IEEE international conference on prognostics and health management. ICPHM, IEEE; 2018, p. 1–8.
[22] Kwan Chiman, Zhang Xiaodong, Xu Roger, Haynes Leonard. A novel approach to fault diagnostics and prognostics. In: 2003 IEEE international conference on robotics and automation (cat. no. 03CH37422), Vol. 1. IEEE; 2003, p. 604–9.
[23] Zhang Xiaodong, Xu Roger, Kwan Chiman, Liang Steven Y, Xie Qiulin, Haynes Leonard. An integrated approach to bearing fault diagnostics and prognostics. In: Proceedings of the 2005 American control conference. IEEE; 2005, p. 2750–5.
[24] Pang Zhenan, Si Xiaosheng, Hu Changhua, Du Dangbo, Pei Hong. A Bayesian inference for remaining useful life estimation by fusing accelerated degradation data and condition monitoring data. Reliab Eng Syst Saf 2021;208:107341.
[25] Prakash Om, Samantaray Arun Kumar. Prognosis of dynamical system components with varying degradation patterns using model–data–fusion. Reliab Eng Syst Saf 2021;213:107683.
[26] Zhang Nan, Cai Kaiquan, Zhang Jun, Wang Tian. A condition-based maintenance policy considering failure dependence and imperfect inspection for a two-component system. Reliab Eng Syst Saf 2022;217:108069.
[27] Ferreira Carlos, Gonçalves Gil. Remaining useful life prediction and challenges: A literature review on the use of machine learning methods. J Manuf Syst 2022;63:550–62.
[28] Tamssaouet Ferhat, Nguyen Khanh Tp, Medjaher Kamal, Orchard Marcos Eduardo. System-level failure prognostics. Proc Inst Mech Eng O 2023.
[29] Jiang Chen, Vega Manuel A, Todd Michael D, Hu Zhen. Model correction and updating of a stochastic degradation model for failure prognostics of miter gates. Reliab Eng Syst Saf 2022;218:108203.
[30] Wang Han, Liao Haitao, Ma Xiaobing, Bao Rui. Remaining useful life prediction and optimal maintenance time determination for a single unit using isotonic regression and gamma process model. Reliab Eng Syst Saf 2021;210:107504.
[31] Sill Joseph. Monotonic networks. In: Proceedings of the 1997 conference on advances in neural information processing systems, Vol. 10. 1998, p. 661–7.
[32] Shaked Moshe, Shanthikumar J George. Stochastic orders. Springer; 2007.
[33] Fürnkranz Johannes, Hüllermeier Eyke. Preference learning and ranking by pairwise comparison. In: Preference learning. Springer; 2010, p. 65–82.
[34] Heimes Felix O. Recurrent neural networks for remaining useful life estimation. In: 2008 international conference on prognostics and health management. IEEE; 2008, p. 1–6.
[35] Couso Inés, Strauss Olivier, Saulnier Hugo. Kendall's rank correlation on quantized data: An interval-valued approach. Fuzzy Sets and Systems 2018;343:50–64.
[36] Martínez Alvaro, Sánchez Luciano, Couso Inés. Engine health monitoring for engine fleets using fuzzy radviz. In: 2013 IEEE international conference on fuzzy systems (FUZZ-IEEE). IEEE; 2013, p. 1–8.
[37] Martínez Alvaro, Sánchez Luciano, Couso Inés. Aeroengine prognosis through genetic distal learning applied to uncertain engine health monitoring data. In: 2014 IEEE international conference on fuzzy systems (FUZZ-IEEE). IEEE; 2014, p. 1945–52.
[38] Sánchez Luciano, Costa Nahuel, Couso Inés. Condition monitoring of axial fans for road tunnels. Mendeley Data 2023;V1. http://dx.doi.org/10.17632/mzjvw6kbt7.1.