

An evidential similarity-based regression method for the prediction of equipment remaining useful life in presence of incomplete degradation trajectories

Francesco Cannarile^{1,2}, Piero Baraldi¹, Enrico Zio^{1*,3,4}

¹*Energy Department, Politecnico di Milano, Via la Masa 34, 20156 Milano, Italy*

*enrico.zio@polimi.it

² *Aramis Srl, Via pergolesi 5, Milano, Italy*

³ *Chair on System Science and the Energetic Challenge, Fondation EDF, Centrale Paris and Supelec, Paris, France*

⁴*Eminent Scholar, Department of Nuclear Engineering, College of Engineering, Kyung Hee University, Republic of Korea*

Abstract

Data-driven methods for direct prognostic map the relationship between monitored parameters and equipment Remaining Useful Life (RUL). They typically require the availability of a set of run-to-failure degradation trajectories for model training. Yet, in many industrial applications, equipment is often replaced before they fail to avoid catastrophic consequences on production and safety. Then also, incomplete degradation trajectories are available. In this work, we develop a method for predicting equipment RUL, and the related uncertainty based on both complete and incomplete degradation trajectories. The method is based on the combined use of a similarity measure and Evidence Theory (EvT). Application of the method on two case studies shows that it provides accurate RUL predictions, also in comparison with a similarity-based regression method of literature.

Keywords: Evidence Theory (EvT), Incomplete Information, Incomplete Degradation Trajectories, Fault Prognostics, Remaining Useful Life (RUL), Evidential Similarity-based Regression (EvSR).

1. INTRODUCTION

Various data-driven methods have been proposed for predicting the Remaining Useful Life (RUL) of degrading equipment [1], i.e., the time before the piece of equipment stops fulfilling its design functions. Such data-driven methods rely on the availability of collected data in the field during equipment operation [1] and are of interest when an explicit model of the equipment degradation process is not available. Among data-driven methods one can distinguish (i) degradation-based approaches, which aim at modeling the future evolution of the equipment degradation, (ii) direct RUL prediction approaches, which attempt to directly predict the equipment RUL [2].

Degradation-based approaches are based on statistical models that learn the equipment degradation evolution from time series of observations of the degradation state [3]. The predicted degradation state is, then, compared with a failure criterion (failure threshold) representative of the degradation state beyond the which the equipment fails performing its required functions. Examples of modeling techniques used in degradation-based approaches are Wiener Processes (WP) [4], Gamma Processes (GP) [5], Inverse Gaussian Processes (IGP) [6], Semi-Markov Models (SMM) [7], Hidden Semi-Markov Models (HSMM) [8], General Path Models (GPM) [9] and fuzzy transition models [10] based on Mamdani models [11].

Degradation-based approaches provide informative and transparent outcomes, giving the current prediction of the equipment RUL and the prediction of the entire equipment degradation trajectory. However, proper degradation state indicators and failure thresholds must be identified, which can be difficult tasks in some industrial applications, especially when the knowledge of the degradation process is limited and only few and/or irregular degradation trajectories are available.

Direct RUL prediction approaches typically resort to machine learning techniques that map the direct relationship between the observed signals and the equipment RUL, without passing through predicting the equipment degradation state evolution and setting a failure threshold [12]. When few degradation trajectories with no clear patterns of regularity are available for model training, similarity-based (also known as instance-based) learning algorithms [13] have been proved to be effective given that they do not perform explicit generalization. Contrarily, non-linear regression methods such as those based on Artificial Neural Networks (ANN) [14,15], which are characterized by large numbers of parameters to be tuned, tend to underperform due to the scarcity of training examples [16]. Indeed, direct RUL prediction approaches assume the availability of a set of *complete* run-to-failure degradation trajectories for

training. In industrial applications, these can be collected on components which are cheap and not critical from the point of view of production and safety. In some cases, they can even be collected in ad-hoc performed laboratory tests.

Condition monitoring data are typically collected between consecutive maintenance interventions which, in many industrial applications are conservatively periodically scheduled to avoid failures. As a consequence, since no or few failures are observed, the available data refer to degradation trajectories which are right censored or *incomplete*, i.e., they do not contain data collected in the last part of the component life.

Furthermore, performing experimental run-to-failure tests to collect condition monitoring data during the entire equipment life is impracticable for safety-critical and high-value components such as those used by aerospace, nuclear and oil and gas industries characterized by very high reliability and unique or new designed components. In other industrial sectors, such as the automotive, the design and execution of test campaigns to collect run-to-failure degradation trajectories requires significant cost, time and resources. For example, in the automotive electronic [17] a temperature cycling test designed to adequately represent the life of automotive electronic components can easily take 2-5 months [17]. In many practical cases, such stress tests have to be repeated multiple times, e.g., whenever new failure modes are discovered or design changes are introduced.

Also, for both degradation-based and direct-RUL prediction approaches, it is fundamental to provide an assessment of the expected mismatch between the ground truth and the predicted RUL. For this, all sources of uncertainty affecting the RUL prediction must be considered:

- the aleatory uncertainty caused by the variability of the degradation process itself (e.g., due to the micro-structural differences between pieces of the same equipment or to unforeseen future loads, operational settings, and external conditions) [17-18];
- the epistemic uncertainty due to imperfect knowledge of the equipment degradation process model and its parameters [19].

Uncertainty management, is thus, a fundamental task in prognostics [20], in order to allow them confidently planning maintenance actions [21].

In this context, the objective of this work is twofold:

- 1) developing a prognostic method capable of exploiting all the available information, including data from right-censored degradation trajectories and not only from complete, run-to-failure trajectories;
- 2) representing the uncertainty in the RUL prediction.

For this, we propose a novel method called Evidential Similarity-based Regression (EvSR), which combines Similarity-based Regression (SR) with Evidence Theory (EvT). Similarity-based approaches have been successfully applied in fault prognostics. In [22], the authors have developed a fuzzy instance-based prognostic approach which builds local fuzzy models and applied it to engine RUL prediction. For a given test engine, the fuzzy model identifies a cluster of peers characterized by similar instances with comparable operational characteristics. The final RUL estimate is, then, a similarity weighted average of the cluster instances. In [23], a library of degradation models has been created using linear regression and used to construct health indicators. The test component RUL is, then, estimated through a weighted sum of the RULs of the most similar instances. The method has been successfully applied to the PHM 2008 challenge turbofan data. In [13], a fuzzy point-wise similarity measure among degradation trajectories is defined to associate weights to the training trajectories. The test component RUL is, then, obtained as a weighted sum of RULs of the similar instances. Only run-to-failure degradation trajectories have been considered to train the prognostic model in all these works.

The use of EvT (also referred to as Dempster-Shafer theory, Belief Function Theory (BFT)) [24] in fault prognostics, has been proposed in few works which typically rely on degradation-based approaches to forecast the future equipment degradation states. In [25], a trajectory similarity-based approach relying on the use of a K-Nearest Neighbours (KNN) classifier based on the belief function is proposed to jointly predict observations and degradation states of aircraft engines. In [26], an evolving real-time neuro-fuzzy system is used to forecast future observation values and an evidential Markovian classifier based on EvT is trained to classify the turbofan degradation states using the predicted observations. In [27], an Evidential Hidden Markov Model (EvHMM), which combines probabilistic Hidden Markov Model (HMM) with EvT, is developed to estimate degradation states of turbofan engines. Notice that these works only exploit the information content of complete run-to-failure trajectories.

To the best of our knowledge, the only attempt to combine SR with EvT is given in [28]. SR is used to predict the RUL of sea water filters in an energy production plant

and EvT is exploited to provide a left-bounded interval $[RUL^{inf}, +\infty]$, which guarantees that the lower bound of the probability that the RUL of the test equipment is larger than RUL^{inf} is greater than a required confidence level. Yet, in [28] the use of EvT is limited by the facts that a) right-censored trajectories are not included in the training set, b) only a lower bound is provided for quantifying the uncertainty on the RUL prediction, c) EvT is not used to provide a RUL point prediction.

In this work, we overcome the limitations a), b) and c) of [28] by proposing a complete framework combining SR with EvT. We adopt the view of Smets' Transferable Belief Model (TBM) [29], which is based on the assumption that beliefs manifest themselves at two mental levels: the "credal" level where beliefs are entertained and the "pignistic" level where beliefs are used to make decisions. At credal level, Smets proposed that beliefs are represented by belief functions [29]. When a decision must be made, the beliefs held at the credal level induce a pignistic probability function at the pignistic level. This probability function is called the pignistic probability function. The proposed EvSR method considers run-to-failure and right-censored degradation trajectories as agents whose state of knowledge on the actual RUL of the testing equipment is encoded in a Basic Belief Assignment (BBA). The construction of these BBAs is based on the estimation of the similarity between test and training trajectories. Then, Dempster's rule of combination is applied to combine the BBAs corresponding to each training trajectory. The resulting combined BBA is, then, used to compute the corresponding pignistic distribution, from which one can derive: 1) the pignistic expectation, which represents the RUL point prediction and can be regarded as our decision point on the unknown RUL and 2) pignistic quantiles, which allow properly defining a two-sided prediction interval to quantify the uncertainty in the RUL prediction. The proposed EvSR approach is verified and compared with a SR-based approach with respect to two real case studies concerning the RUL prediction of turbofan engines and knives used in the packaging industry.

This work, is organized as follows: Section 2 states the assumptions on the available data and the problem. Section 3 recalls the method for RUL prediction based on SR and the basic concept of EvT. Section 4 describes the proposed EvSR for RUL prediction and uncertainty quantification. In Section 5, we discuss the metrics to quantitatively assess the performance and how to tune the hyperparameters of the proposed methodology. The case study and the application of the proposed method are reported in Section 6. Finally, Section 7 concludes the work..

2. PROBLEM STATEMENT

The objective of this work is to estimate the RUL of a test component for which the degradation trajectories of N pieces of equipment similar to the one currently monitored have been observed. Among the available degradation trajectories, $N_F < N$ are run-to-failure trajectories and the remaining $N_H = N - N_F$ are right-censored degradation trajectories (i.e., trajectories corresponding to equipment replaced before failure). We refer to these two classes of trajectories as “*complete*” and “*incomplete*” trajectories, respectively. We consider the case in which it is difficult to obtain complete trajectories, e.g., new or safety critical components for which running to failure can be lengthy, costly or dangerous. Therefore, to be realistic, we assume $N_F \leq N_H$ and a limited number of available degradation trajectories (e.g., $N = N_F + N_H < 30$).

Let $\mathbf{z}^i(\tau_i) = [z_1^i(\tau_i) \dots z_M^i(\tau_i)] \in \mathbb{R}^M$, $i = 1, \dots, N$, $\tau_i = 1, \dots, n_i$, be the vector of M features extracted from signal measurements performed at time τ_i , on the i^{th} component. The time n_i corresponds to the failure time if the i^{th} trajectory is complete ($i = 1, \dots, N_F$) or to the last observation performed on the component before replacement if the trajectory is incomplete ($i = N_F + 1, \dots, N$). We assume that an expert is capable of providing an upper bound $T_{max} > 0$ on the equipment life duration. This implies that for the i^{th} right-censored degradation trajectory the ground truth RUL at time τ_i , $y^i(\tau_i)$, lies in the interval, $I^i(\tau_i) = (y_L^i(\tau_i), y_U^i(\tau_i)]$, where

$$y_L^i(\tau_i) = n_i - \tau_i \quad i = N_F + 1, \dots, N; \tau_i = 1, \dots, n_i \quad (1)$$

$$y_U^i(\tau_i) = T_{max} - \tau_i \quad i = N_F + 1, \dots, N; \tau_i = 1, \dots, n_i \quad (2)$$

Instead, we exactly know the ground truth RUL of the i^{th} complete degradation trajectory

$$y^i(\tau_i) = n_i - \tau_i \quad i = 1, \dots, N_F; \tau_i = 1, \dots, n_i \quad (3)$$

We consider a case in which the failure thresholds for the extracted features are not known. In this setting, fault prognostics is framed as a regression problem: given a training set, U , formed by N_F realizations of complete degradation trajectories $\{\mathbf{z}^i(\tau_i), y^i(\tau_i), \tau_i = 1, \dots, n_i\}, i = 1, \dots, N_F$ and N_H realizations of incomplete degradation trajectories $\{\mathbf{z}^i(\tau_i), I^i(\tau_i), \tau_i = 1, \dots, n_i\}, i = N_F + 1, \dots, N$ of a stochastic process $(\mathbf{Z}(\tau), Y(\tau)) \in \mathbb{R}^M \times (0, +\infty)$, our task is to find a function

$$f: \mathbb{R}^{M \times k} \rightarrow (0, +\infty) \quad (4a)$$

$$f = f(\mathbf{z}_{\tau-k+1:\tau}) \quad (4b)$$

$$\mathbf{z}_{\tau-k+1:\tau} = [\mathbf{z}(\tau - k + 1), \dots, \mathbf{z}(\tau - 1), \mathbf{z}(\tau)] \quad k = 1, \dots, \tau \quad (4c)$$

that associates to a test pattern $\mathbf{z}_{\tau_{test}-k+1:\tau_{test}}^{test} = [\mathbf{z}^{test}(\tau_{test} - k + 1), \dots, \mathbf{z}^{test}(\tau_{test} - 1), \mathbf{z}^{test}(\tau_{test})] \in \mathbb{R}^{M \times k}$ at the current time instant τ_{test} , the corresponding output $y^{test}(\tau_{test})$. The regression curve f in Eq. (4) can depend on all degradation history up to time current time (i.e., $k = \tau$), or only on the last k observations ($k < \tau$).

3. RUL ESTIMATION BASED ON SIMILARITY REGRESSION AND BASIC CONCEPTS OF EVIDENCE THEORY

3.1. RUL estimation based on similarity regression

This Section briefly illustrates the similarity-based regression method for the prediction of the RUL of a test component proposed in [13]. The distance $\delta_i(\tau_{test}, \tau_i)$ between the test pattern $\mathbf{z}_{\tau_{test}-k+1:\tau_{test}}^{test}$ and the training pattern $\mathbf{z}_{\tau_i-k+1:\tau_i}^i$ extracted from the i^{th} complete training trajectories is defined by:

$$\delta_i(\tau_{test}, \tau_i) = \|\mathbf{z}_{\tau_{test}-k+1:\tau_{test}}^{test} - \mathbf{z}_{\tau_i-k+1:\tau_i}^i\|_2 \quad i = 1, \dots, N_F; \tau_i = k, \dots, n_i \quad (5)$$

where $\|\mathbf{z}_1 - \mathbf{z}_2\|_2$ is the Euclidean distance between vectors \mathbf{z}_1 and \mathbf{z}_2 . The similarity measure $\mu_i(\tau_{test}, \tau_i)$ between the test trajectory segment $\mathbf{z}_{\tau_{test}-k+1:\tau_{test}}^{test}$ and the training trajectory segment $\mathbf{z}_{\tau_i-k+1:\tau_i}^i$ is defined by the Gaussian kernel function:

$$\mu_i(\tau_{test}, \tau_i) = \exp\left(-\frac{\delta_i^2(\tau_{test}, \tau_i)}{2\sigma}\right) \quad i = 1, \dots, N_F; \tau_i = k, \dots, n_i; \sigma > 0 \quad (6)$$

where the scale parameter σ is set by the analyst to shape the desired interpretation to proper similarity degree. This similarity has turned out to give robust results in similarity-based regression due to its gradual smoothness [13]. For each complete training trajectory, $i = 1, \dots, N_F$, the segment of length k with largest similarity, $\mathbf{z}_{\tau_i^*-k+1:\tau_i^*}^i$, is first identified

$$\mathbf{z}_{\tau_i^*-k+1:\tau_i^*}^i: \tau_i^* = \operatorname{argmax}_{\tau_i=k, \dots, n_i} \mu_i(\tau_{test}, \tau_i) \quad (7)$$

Then, we associate to each training trajectory the normalized weight ω_i^* :

$$\omega_i^* = \frac{\omega_i}{\sum_{e=1}^{N_F} \omega_e} \quad (8)$$

with

$$\omega_i = \mu_i(\tau_{test}, \tau_i^*) \quad (9)$$

and the RUL:

$$y^i(\tau_i^*) = n_i - \tau_i^* \quad (10)$$

corresponding to that of the last pattern of the most similar segment. The RUL $y^{test}(\tau_{test})$ of test pattern $\mathbf{z}_{\tau_{test}-k+1:\tau_{test}}^{test}$ is estimated as a similarity-weighted sum:

$$\hat{y}^{test}(\tau_{test}) = \sum_{i=1}^{N_F} \omega_i^* y^i(\tau_i^*) \quad (11)$$

The ideas behind weighting the individual $y^i(\tau_i^*)$ is that: (i) all trajectories in the training set bring useful information for estimating the RUL of the test trajectory; (ii) those segments of the training trajectories which are most similar to the most recent segment of the test trajectory should be more informative in the prediction of the future behavior of the test trajectory. With respect to the assessment of the uncertainty affecting the RUL prediction, the Similarity-based Conditional Cumulative Distribution (SCCD) has been defined as [30]:

$$F(y_{test}(\tau_{test}) | \mathbf{z}_{\tau_{test}-k+1:\tau_{test}}^{test}) = \sum_{i=1}^{N_F} 1_{y_{test}(\tau_{test}) \in (0, y^i(\tau_i^*)]}(y_{test}(\tau_{test})) \omega_i^* \quad (12)$$

where $1(\cdot)$ denotes the indicator function of a set. Then, the $(1 - \epsilon) \times 100$ prediction interval, $\hat{I}_\epsilon(\tau_{test}) = [\hat{y}_{\epsilon,L}(\tau_{test}), \hat{y}_{\epsilon,U}(\tau_{test})]$, $\epsilon \in (0, 1)$, can be obtained as

$$\hat{y}_{\epsilon,L}(\tau_{test}) = \inf \left\{ y_{test}(\tau_{test}) \in (0, +\infty) \mid F(y_{test}(\tau_{test}) | \mathbf{z}_{\tau_{test}-k+1:\tau_{test}}^{test}) > \frac{\epsilon}{2} \right\} \quad (13)$$

$$\hat{y}_{\epsilon,U}(\tau_{test}) = \inf \left\{ y_{test}(\tau_{test}) \in (0, +\infty) \mid F(y_{test}(\tau_{test}) | \mathbf{z}_{\tau_{test}-k+1:\tau_{test}}^{test}) > 1 - \frac{\epsilon}{2} \right\} \quad (14)$$

3.2. Basic concepts of Dempster-Shafer evidence theory

To overtake the limitation of the methodology discussed in Subsection 3.1, which exploits only the information provided by complete degradation trajectories, we propose a novel approach for RUL prediction, which will be referred to as Evidential Similarity Regression (EvSR), and combines Similarity-based Regression (SR) with Evidence Theory (EvT) (also known as Dempster-Shafer theory or Belief Function Theory

(BFT)) [24]. Adopting the view of Smets' Transferable Belief Model (TBM) [29], we assume that beliefs manifest themselves at two mental levels: the "credal" level where beliefs are entertained and the "pignistic" level where beliefs are used to make decisions. At credal level, beliefs are represented by belief functions [29], whereas, when a decision must be made at the pignistic level, the beliefs held at the credal level induce a probability function which is called the pignistic probability function. In what follows, we briefly introduce the necessary notions of EvT.

3.2.1. Mass Functions

Let X denote a variable taking values in a domain Θ called the frame of discernment. A discrete mass function m with focal sets A_1, \dots, A_G is a function from 2^Θ to $[0, 1]$ verifying $m(A_g) = m_g > 0$ for all $g \in \{1, \dots, G\}$, $m(A) = 0$ for all $A \notin \{A_1, \dots, A_G\}$, and:

$$\sum_{g=1}^G m_g = 1 \quad (15)$$

Mass function m is called Basic Belief Assignment (BBA) and each number m_g is interpreted as a degree of belief attached to the proposition $X \in A_g$, and no more specific proposition, based on some evidence. The triplet $(\Theta, \{A_g\}_{g=1:G}, m)$ is called Dempster-Shafer Belief Structure (DSBS) [31].

Given a normal mass function m , i.e., such that $m(\emptyset) = 0$, with focal sets A_1, \dots, A_G , the corresponding belief and plausibility functions are defined, respectively, as:

$$bel(A) = \sum_{\{g: A_g \subseteq A\}} m_g \quad (16)$$

$$pl(A) = \sum_{\{g: A_i \cap A \neq \emptyset\}} m_g \quad (17)$$

for all $A \subseteq \Theta$. These two functions are linked by the relation $Pl(A) = 1 - Bel(\bar{A})$ for all $A \subseteq \Theta$. $Bel(A)$ can be interpreted as the degree to which the evidence supports A , and $Pl(A)$ as the degree to which the evidence is not contradictory to A .

According to Dempster interpretation [40], the mass function m defines a set of probability measures $\mathcal{P}(m)$ on the measurable space $(\Theta, \mathcal{B}(\Theta))$, where $\mathcal{B}(\Theta)$ denotes the Borel σ -algebra on Θ , such that:

$$(18)$$

$$bel(A) \leq \mathbb{P}(A) \leq pl(A), \forall \mathbb{P} \in \mathcal{P}(m)$$

for any measurable subset A of Θ . A probability measure \mathbb{P} satisfying the above inequalities is said to be compatible with m [33]. Therefore, the mass function provides information about the probability of a set in an imprecise manner by giving the interval in which the probability lies [34].

The mass functions, m_1 and m_2 , induced by two distinct pieces of evidence, can be combined using Dempster's combination rule [33-35]:

$$m_{1,2}(A) = m_1 \oplus m_2(A) = \frac{1}{1 - \chi} \sum_{B \cap C = A} m_1(B)m_2(C) \quad (19)$$

for all $A \subseteq \Theta, A \neq \emptyset, m_1 \oplus m_2(\emptyset) = 0$, where:

$$\chi = \sum_{B \cap C = \emptyset} m_1(B)m_2(C) \quad (20)$$

is the degree of conflict between m_1 and m_2 . If $\chi = 1$, there is a logical contradiction between two pieces of evidence and they cannot be combined. Dempster's rule is commutative, associative, and it admits as neutral element the vacuous mass function defined as $m(\emptyset) = 0$.

It may occur that one doubts the reliability of a source inducing the BBA m . In this case, the discounting operation can be used to reduce by a factor $\gamma \in [0,1]$ the belief assigned by m to the evidence conveyed by the information [35]:

$$\tilde{m}(A) = (1 - \gamma)m(A) \quad \forall A \subseteq \Theta \quad (21a)$$

$$\tilde{m}(A) = (1 - \gamma)m(A) \quad \forall A \subseteq \Theta \quad (21b)$$

The mass assigned to the frame of discernment Θ in Eq. (7) represents the ignorance about the value of X , because it indicates the absence of evidence that the value of X belongs to any subset A of Θ .

3.2.2. Pignistic probability distribution

Since the use of probabilities in a decision context is strongly supported by rationality arguments [29], a belief function has to be transformed into a probability function for decision making at pignistic level. One example of probability measure compatible with

m (i.e., such that Eq. (18), holds) is the pignistic probability measure \mathbb{P}^{bet} with corresponding pignistic probability distribution p_{bet} [33-34]:

$$p_{bet}(x) = \sum_{A \subseteq \Theta, A \neq \emptyset} m(A) \frac{1_A(x)}{|A|} \quad (22)$$

where $1_A(\cdot)$ denotes the characteristic function of A and $|A|$ its cardinality. For each focal element A , the pignistic probability distribution is bound between the belief and plausibility functions:

$$bel(A) \leq p_{bet}(A) \leq pl(A) \quad (23)$$

3.2.3. Evidence theory for continuous bounded frame of discernment

The frame of discernment Θ is assumed to be finite in Subsections 3.2.1 and 3.2.2. If Θ is a continuous bounded interval on the real line and the focal are a finite number of intervals on the real line, all the expressions defined above are unchanged [35]. For the sake of simplicity, we set $\Theta = [x_L, x_U]$ and we represent our state of knowledge regarding a quantity X by a mass function $m: 2^{[x_L, x_U]} \rightarrow [0,1]$ with g interval focal sets $A_g = [a_g, b_g] \subseteq [x_L, x_U]$, $g \in \{1, \dots, G\}$. In particular, considering measurable subset A of $[x_L, x_U]$ of the form $A = (x_L, x]$, $x \in [x_L, x_U]$, Eq. (23) becomes:

$$bel((x_L, x]) \leq \mathbb{P}((x_L, x]) = F(x) \leq pl((x_L, x]) \quad (24)$$

where $F(x)$ is the Cumulative Distribution Function (CDF) corresponding to the probability measure $\mathbb{P} \in \mathcal{P}(m)$ compatible with m [31, 33]. In this case, the cardinality of a given set A in Eq. (22) is replaced by its Lebesgue measure:

$$|A| = \int A(x) dx \quad (25)$$

and the pignistic probability function p_{bet} is [31]:

$$p_X^{bet}(x) = \sum_{g=1}^G m([a_g, b_g]) \frac{I_{[a_g, b_g]}(x)}{|b_g - a_g|} \quad x \in [x_L, x_U] \quad (26)$$

3.2.4. Pignistic expectation and quantiles

The pignistic probability function defined in Eq. (26) allows defining some summary statistics, such as the pignistic expectation \mathbb{E}_X^{bet} and the pignistic quantiles $q_X^{bet}(\epsilon)$ of order $(1 - \epsilon)$, $\epsilon \in (0,1)$. The pignistic expectation \mathbb{E}_X^{bet} :

$$\mathbb{E}_X^{bet} = \int_{x_L}^{x_U} xp_X^{bet}(x)dx = \sum_{g=1}^G m([a_g, b_g]) \frac{(a_g + b_g)}{2} \quad (27)$$

can be regarded as the decision point on the unknown quantity X . The pignistic quantiles $q_X^{bet}(\epsilon)$ are defined as

$$q_X^{bet}(\epsilon) = \inf\{x \in [x_L, x_U] | F_X^{bet}(x) \geq 1 - \epsilon\} \quad \forall \epsilon \in (0,1) \quad (28)$$

where F_X^{bet} is the pignistic cumulative distribution:

$$F_X^{bet}(x) = \int_{\tilde{x} < x} p_X^{bet}(\tilde{x})d\tilde{x} \quad x \in [x_L, x_U] \quad (29)$$

By choosing $x_{\epsilon,L} = q_X^{bet}(\epsilon)$ and $x_{\epsilon,U} = q_X^{bet}(1 - \epsilon)$, the interval $I_X^{bet}(\epsilon) = [x_{\epsilon,L}, x_{\epsilon,U}]$, satisfies:

$$\mathbb{P}_X^{bet}(X \in I_X^{bet}(\epsilon)) = 1 - \epsilon \quad (30)$$

i.e., $I_X^{bet}(\epsilon)$ is a $(1 - \epsilon) \times 100\%$ prediction interval for X and it can be regarded as the decision interval on the unknown quantity X .

4. EVIDENTIAL SIMILARITY-BASED REGRESSION

The key idea of the proposed method is to consider complete and incomplete training trajectories as agents whose state of knowledge regarding the unknown RUL of a test trajectory at current time τ_{test} , $X = Y^{test}(\tau_{test})$, is represented by a mass function $\tilde{m}_{\tau_{test}}^i, i = 1, \dots, N$. This latter is defined by the similarity measure between the trajectory segment containing the last k measurements extracted from the test trajectory $\mathbf{z}_{\tau_{test}-k+1:\tau_{test}}^{test}$ and the segment $\mathbf{z}_{\tau_i-k+1:\tau_i}^i$ extracted from the i^{th} training trajectory. Then, the weights ω_i , associated to each training trajectory, $i = 1, \dots, N$, are assigned by applying Eqs. (7) and (9), and the discount factor $\gamma_i \in [0,1)$ (Eq. 21):

$$\tilde{m}_{\tau_{test}}^i(I^i(\tau_i^*)) = \gamma_i \omega_i \quad (31a)$$

$$\tilde{m}_{\tau_{test}}^i(\Theta_{\tau_{test}}) = 1 - \gamma_i \omega_i \quad (31b)$$

where $\Theta_{\tau_{test}}$ is the frame of discernment, i.e., the domain $[0, T_{max} - \tau_{curr}] \subset \mathbb{R}$ of the unknown quantity $RUL_{\tau_{tes}}$ with T_{max} representing the maximum life duration of the equipment provided by an expert (see Section 2). Parameter γ_i represents the analyst prior opinion about the maximum information that can be derived from the i^{th} training

trajectory with respect to the estimation of the test trajectory RUL [28]. Notice that $\gamma_i < 1$ implies that a part of belief is assigned to the *ignorance* represented by $\Theta_{\tau_{test}}$, even in the unrealistic case of a reference trajectory exactly identical to the test one [28]. When no prior knowledge is available on parameter γ_i , this can be set through cross-validation.

The definition of the quantity $I^i(\tau_i^*)$ in Eq. (31), which indicates the uncertain interval of the component RUL, depends on whether the index i refers to a complete or a right-censored training trajectory. In case of complete degradation trajectories, $i \in \{1, \dots, N_F\}$,

$$I^i(\tau_i^*) = [n_i - \tau_i^*, n_i - \tau_i^* + 1] \quad i = 1, \dots, N_F \quad (32)$$

since we only know that at time $\tau_i^* - 1$ the component was working and at time τ_i^* it was failed, whereas in the case of incomplete degradation trajectories, $i \in \{N_F + 1, \dots, N\}$,

$$I^i(\tau_i^*) = (n_i - \tau_i^*, T_{max} - \tau_i^*] \quad i = N_F + 1, \dots, N \quad (33)$$

since we know that the component failure would have been after time τ_i^* , which corresponds to the time at which the component was preventively replaced, but before time T_{max} , which, according to the expert is an upper bound of the component life duration.

Once the mass functions $\tilde{m}_{\tau_{test}}^i$ associated to each training trajectories have been computed, they are combined into the BBA $\tilde{m}_{\tau_{test}}$ by applying Eqs. (19-20). The BBA $\tilde{m}_{\tau_{test}}$ represents our state of knowledge on the unknown component RUL at time τ_{test} , $Y^{test}(\tau_{test})$. From $\tilde{m}_{\tau_{test}}$ the pignistic probability density function and the pignistic cumulative distribution function can be derived using Eqs. (26) and (27), respectively.

The bounds of a $(1 - \epsilon) \cdot 100\%$ prediction interval $\hat{I}_\epsilon(\tau_{test}) = [\hat{y}_{\epsilon,L}(\tau_{test}), \hat{y}_{\epsilon,U}(\tau_{test})]$ such that $\mathbb{P}(Y^{test}(\tau_{test}) \in \hat{I}_\epsilon(\tau_{test})) = 1 - \epsilon, \epsilon \in (0,1)$ are obtained by applying Eqs. (29-30) and the pignistic expectation $\hat{y}^{test}(\tau_{test})$ by Eq. (28). We consider $\hat{y}^{test}(\tau_{test})$ and $\hat{I}_\epsilon(\tau_{test})$ as our “decision” point and interval, respectively.

5. PERFORMANCE METRICS AND HYPERPARAMETER TUNING

In this Section, we briefly describe the metrics used to quantitatively assess and compare the performance of prognostic models SR and EvSR and how to tune the models hyperparameters. Considering a generic test trajectory $(\mathbf{z}(\tau), y(\tau)), \tau = 1, \dots, n$, we

refer to $\hat{y}(\tau)$ and $\hat{I}_\epsilon(\tau) = [\hat{y}_{\epsilon,L}(\tau), \hat{y}_{\epsilon,U}(\tau)]$ as a generic point estimate and an $(1 - \epsilon) \times 100\%$ prediction interval of the test component RUL at time $\tau, y(\tau)$, respectively.

5.1. Performance metrics for complete test degradation trajectories

- 1) The Root Mean Square Error (RMSE) metric [28] is used to assess the quality of the point RUL prediction $\hat{y}(\tau)$ provided by a prognostic model. It is defined as

$$RMSE = \sqrt{\frac{1}{n} \sum_{\tau=1}^n (\hat{y}(\tau) - y(\tau))^2} \quad (34)$$

The RMSE measures the accuracy of the estimated RUL $\hat{y}(\tau)$ and is desired to be as small as possible.

- 2) The Interval Score IS_ϵ metric [35] is used to assess the quality of the RUL prediction interval $\hat{I}_\epsilon(\tau)$ provided by a prognostic model. It is defined by:

$$IS_\epsilon = \sum_{\tau=1}^n (IS_{\epsilon,A}(\tau) + IS_{\epsilon,B}(\tau)) \quad (35)$$

with

$$IS_{\epsilon,A}(\tau) = \frac{2}{1-\epsilon} (\hat{y}_{\epsilon,L}(\tau) - y(\tau)) 1_{\{y(\tau) < \hat{y}_{\epsilon,L}(\tau)\}}(y(\tau)) + \frac{2}{1-\epsilon} (y(\tau) - \hat{y}_{\epsilon,U}(\tau)) 1_{\{y(\tau) > \hat{y}_{\epsilon,U}(\tau)\}}(y(\tau)) \quad (36a)$$

$$IS_{\epsilon,B}(\tau) = \hat{y}_{\epsilon,U}(\tau) - \hat{y}_{\epsilon,L}(\tau) \quad (36b)$$

where $1(\cdot)$ denotes the indicator function of a set. The metric considers a trade-off between the desiderata of producing prediction intervals which contain the target ground truth RUL (first addend $IS_{\epsilon,A}(\tau)$, in Eq. (36)) and which are narrow (second addend $IS_{\epsilon,B}(\tau)$ in Eq. (36)) [35]. Smaller is the IS_ϵ , more satisfactory the prediction intervals are.

5.2. Performance metrics for incomplete degradation trajectories

Since metrics 1) and 2) require the knowledge of the ground truth RUL of the test trajectory, they cannot be applied in case of incomplete test degradation trajectories. Therefore, we introduce two additional prognostics metrics that exploit the only information available about $y(\tau)$ for incomplete degradation trajectories, i.e., $y(\tau) \in I(\tau) = (y_L(\tau), y_U(\tau)] = (n - \tau, T_{max} - \tau]$.

- 3) The Prediction Interval Coverage PIC_ϵ [36] measures the accuracy of the prediction interval $\hat{I}_\epsilon(\tau)$ when only the interval $I(\tau)$ to which the ground truth RUL belongs is known. It is defined by:

$$PIC_\epsilon = \frac{1}{n} \sum_{\tau=1}^n PIC_{\epsilon}(\tau) \quad (37)$$

with

$$PIC_{\epsilon}(\tau) = \begin{cases} 1 & I(\tau) \subseteq \hat{I}_\epsilon(\tau) \\ \frac{\text{diam}(I(\tau) \cap \hat{I}_\epsilon(\tau))}{\text{diam}(I(\tau))} & I(\tau) \cap \hat{I}_\epsilon(\tau) \neq \emptyset \\ 0 & I(\tau) \cap \hat{I}_\epsilon(\tau) = \emptyset \end{cases} \quad (38)$$

where $\text{diam}(I(\tau))$ and $\text{diam}(I(\tau) \cap \hat{I}_\epsilon(\tau))$ refer to the width of intervals $I(\tau)$ and $(I(\tau) \cap \hat{I}_\epsilon(\tau))$, respectively. Eq. (39) implies that if the target $I(\tau)$ is entirely contained within the estimated prediction interval $\hat{I}_\epsilon(\tau)$, then $PIC_\epsilon(\tau)$ is equal to 1. If the estimated prediction interval does not contain the entire target $I(\tau)$, but the intersection of the two is nonempty, then $PIC_\epsilon(\tau)$ is equal to the ratio between the width of the interval $I(\tau) \cap \hat{I}_\epsilon(\tau)$ and the width of the interval $I(\tau)$, $\text{diam}(I(\tau))$. Finally, if the intersection between the estimated prediction interval and the real target $I(\tau)$ is empty, then, the coverage $PIC_\epsilon(\tau)$ is 0. This metric is desired to be close to 1.

- 4) Inside metric D :

$$D = \frac{1}{n} \sum_{\tau=1}^n 1_{\hat{y}(\tau) \in I(\tau)}(\hat{y}(\tau)) \quad (39)$$

The inside metric D measures the percentage of times that the point estimate of the predicted RUL is greater than the period of time in which the component has not failed, i.e., the condition $\hat{y}(\tau) \in I(\tau)$ holds. The metric is desired to be closer to 1.

5.3. Hyperparameter tuning

The proposed EvSR method is characterized by the presence of three parameters, which should be properly tuned:

- the Gaussian kernel scale β used in Eq. (9);
- the segment length $k \in \{1, 2, \dots\}$ used in Eqs. (4-7);
- the discount factors γ_i used in Eq. (31), $i = 1, \dots, N$.

We assume that γ_i , has the same value γ_F for all the complete training trajectories, i.e., $\gamma_i = \gamma_F, i = 1, \dots, N_F$ and the same value γ_H for all incomplete training trajectories, i.e., $\gamma_i = \gamma_H, i = N_F + 1, \dots, N$. Notice that a SR approach requires the setting of parameters β and k . Hyperparameter tuning is performed by optimizing the prediction accuracy metrics for complete and incomplete degradation trajectories. To this purpose, the RMSE and the PIC are computed on a validation set of complete and incomplete degradation trajectories, respectively, considering different values of the parameters to be set. Then, a trade-off solution among those belonging to the Pareto optimal set is chosen by resorting to the TOPSIS method [37], giving equal importance to the two considered objectives.

6. CASE STUDIES

The proposed EvSR prognostic method has been verified with respect to data extracted from a well-known benchmark and an industrial case study.

6.1. Case Study 1: RUL prediction of turbofan engines

We consider simulated degradation trajectories of a heterogeneous fleet of $N = 218$ turbofan engines operating under variable operational conditions. The data are taken from the NASA Ames Prognostics CoE Data Repository [38] and have been preprocessed as in [15], obtaining $M = 6$ relevant features describing the component degradation evolution. Figure 1 shows the evolution of the six extracted features during the life of a turbofan engine.

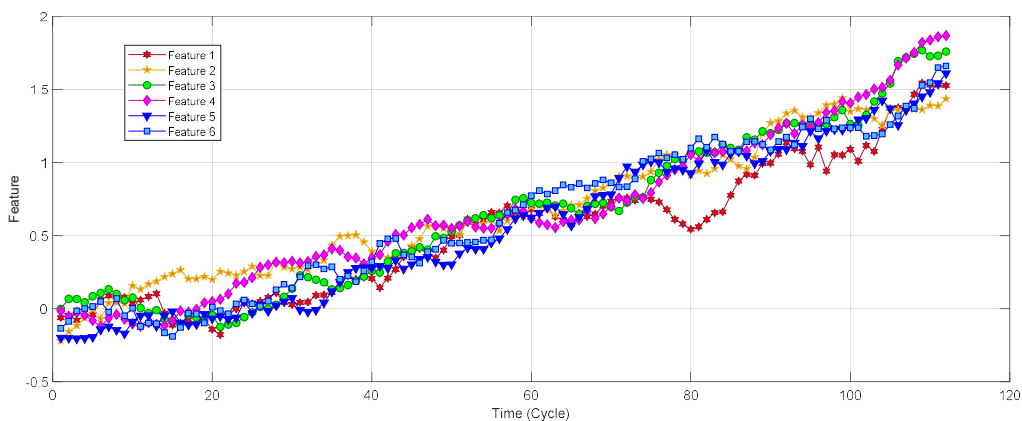


Figure 1. Time evolution of the extracted 6 relevant features during the entire life of a turbofan engine.

The 218 trajectories have been randomly divided into 11 folds: 10 folds formed by 20 degradation trajectories and one-fold containing the remaining 18 degradation trajectories. A condition in which the failure time of some turbofan engines is unknown

has been artificially created by applying a right-censoring mechanism to 50% of the trajectories in each fold. To this aim, we have randomly selected the trajectories and randomly sampled from a uniform distribution with support $[0.5n, n]$ (where n is the ground truth time to failure of a randomly selected turbofan engine), the time at which the observation of the component life is interrupted. This particular setting has been designed for emulating the real industrial applications, where few complete degradation trajectories (e.g., $N < 30$) are available and many of them are right-censored ($N_H \geq N_F$). Further within each fold, Principal Component Analysis (PCA) has been first applied to reduce the dimensionality of the feature vector. The number of principal components to retain has been set by identifying in the scree plot the “elbow” point where the slope changes from steep to shallow, in this way, we have found that independently from the considered fold, only the first principal component has been selected, i.e., the dimension of the feature vector has been reduced from $M = 6$ to $M = 1$. The performance of SR and EvSR methods are evaluated within each fold resorting to a twice nested Leave-One-Out Cross Validation (LOOCV): the outer loop has been used to compute the metrics, whereas the inner loop has been used to set the hyperparameters, as discussed in Subsection 5.3. Table 1 reports the considered values of the parameters $\beta, \gamma_H, \gamma_F$ and k .

Hyperparameter	Considered values
β	$\{2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7, 2^8, 2^9, 2^{10}\}$
γ_F, γ_H	$\{0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95\}$
k	$\{1, 2, 3, 4, 5\}$

Table 1. Considered hyperparameter values.

Since no prior knowledge is available on the maximum life duration of turbofan engines, the variable T_{max} has been set equal to 266.2 (cycles): this value corresponds to the 110% of the longest life duration among the 218 available degradation trajectories (i.e., $T_{max} = 242 * \frac{11}{10} = 266.2$). Notice that the training of the SR method uses only the complete degradation trajectories, i.e., 9 for folds 1-10 and 8 for folds 11, whereas, the EvSR uses all the available degradation trajectories, i.e., 19 for folds 1-10 and 17 for folds 11.

6.2. Results

Table 2 compares the performances of the developed EvSR method with those of the SR method considering complete degradation test trajectories. Table 2 reports the average of the performance metrics over the 11 folds and its standard deviation. With

respect to the EvSR, we have considered the cases in which the model is trained with only complete degradation trajectories (EvSR1) and with complete and incomplete degradation trajectories (EvSR2).

Metric (complete test trajectories)	SR (training set: only complete degradation trajectories)	EvSR1 (training set: only complete degradation trajectories)	EvSR2 (training set: complete and right-censored degradation trajectories)
RMSE (to be minimized)	37.14 ± 6.32	35.19 ± 5.75	33.63 ± 4.39
$IS_{0.90}$ (to be minimized)	186.31 ± 55.67	166.55 ± 44.02	135.52 ± 18.94

Table 2. Prognostic performances evaluated considering complete test degradation trajectories.

EvSR2 performs significantly better than SR from the points of view of prediction accuracy, being the RMSE of EvSR2 9.46% lower than that of SR and of prediction interval quality, being EvSR2 $IS(0.90)$ 27.25% lower than that of SR. Since the performance of EvSR1 is almost comparable with that of SR, we can conclude that the inclusion of the right-censored degradation trajectories in the training set for model training, allows increasing the EvSR2 performance, which distinguishes EvSR1 from EvSR2. Figure 2 shows an example of RUL point predictions obtained by the considered methods applied on a complete degradation trajectory. The lower RMSE of EvSR2 is mainly due to more accurate predictions at the beginning of the life of the considered test trajectory.

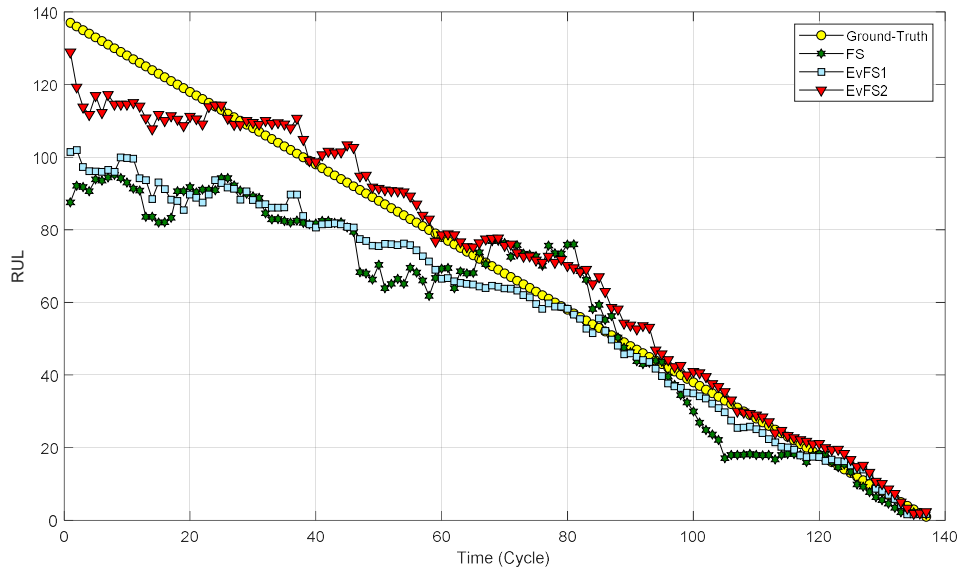


Figure 2. Predicted RUL by SR (hexagram), EvSR1(square) and EvSR2 (triangle) on a complete test degradation trajectory.

Figure 3 shows an example of RUL prediction intervals obtained on a different test trajectory. Also in this case, the inclusion of right-censored degradation trajectories in the training set allows improving interval predictions quality, especially at the beginning of the component life.

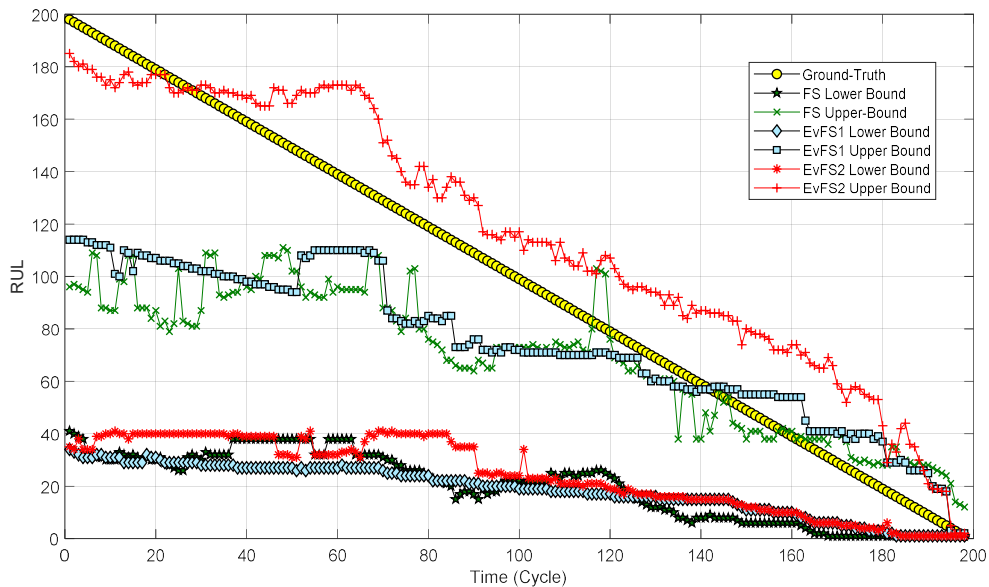


Figure 3. Interval predictions provided by SR, EvSR1 and EvSR2 on a complete test degradation trajectory.

Table 3 compares the performances of the developed EvSR2 algorithm with those of SR and EvSR1, for right-censored degradation test trajectories.

Metric (right-censored test trajectories)	SR (training set: only complete degradation trajectories)	EvSR1 (training set: only complete degradation trajectories)	EvSR2 (training set: complete and right-censored degradation trajectories)
D (to be maximized)	0.58 ± 0.15	0.62 ± 0.16	0.62 ± 0.10
$PIC_{0.90}$ (to be maximized)	0.46 ± 0.11	0.47 ± 0.09	0.51 ± 0.05

Table 3. Prognostic performance for right-censored test degradation trajectories.

Although the results are more uncertain, being the test set less informative than that used for the computation of the RMSE and $IS_{0.90}$ metrics (Table 2), they lead to similar conclusions.

6.3. Case Study 2: Prediction of the RUL of knives used in the packaging industry

We consider a heterogeneous fleet of $N = 23$ knives used in the manufacturing industry to cut packaging material. The prognostic task is complicated by the fact that few run-to-failure degradation trajectories are available and a failure threshold is not available. These data contain $N_F = 10$ and $N_H = 13$ complete and right-censored degradation trajectories, respectively. To deal with this type of data, we have applied a pre-processing procedure followed by a feature extraction step, which has allowed the identification of $M = 2$ relevant features describing the component degradation evolution. More details on the dataset and feature extraction procedure can be found in [15, 39]. Figure 4 shows the evolution of the two extracted features during a run-to-failure trajectory of a knife.

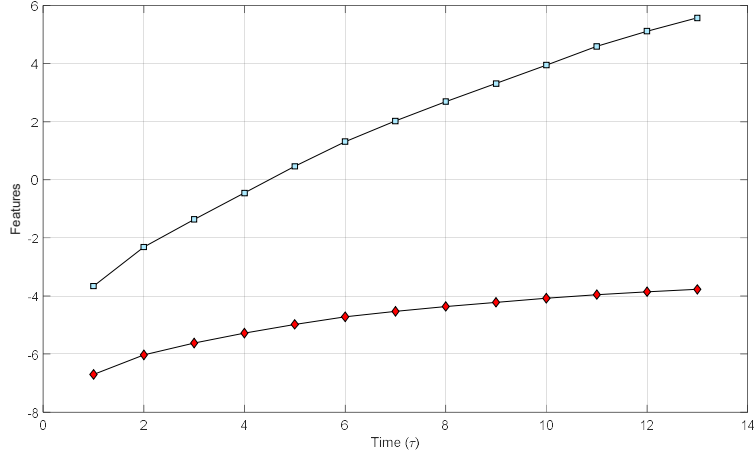


Figure 4. Time evolution of the $M = 2$ relevant features from one knife.

The performance of SR and EvSR are evaluated resorting to a twice nested Leave-One-Out Cross Validation (LOOCV) in which the outer loop has been used to compute the metrics, whereas the inner loop to set the hyperparameters, as discussed in Subsection 5.3. Table 4 reports the considered values of the parameters β , γ_H and γ_F .

Hyperparameter	Considered values
β	$\{2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5\}$
γ_F, γ_H	$\{0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95\}$
k	$\{1, 2\}$

Table 4. Considered hyperparameter values for grid search.

Since no prior knowledge is available on the maximum life duration of knives, the variable T_{max} has been set equal to 25.3 (time units): this value corresponds to the 110% of the longest life duration among the 23 available degradation trajectories (i.e., $T_{max} = 23 * \frac{11}{10} = 25.3$). Notice that the training set is formed by 9 trajectories for the SR and EvSR1 and by 20 for the EvSR2 method.

6.4. Results

Tables 5 and 6 compare the performances of the developed EvSR methods with those of the SR method for complete degradation test trajectories. The averages of the performance metrics over the LOOCVs and its standard deviation are reported.

Metric (complete test trajectories)	SR	EvSR1	EvSR2

	(training set: only complete degradation trajectories)	(training set: only complete degradation trajectories)	(training set: complete and right-censored degradation trajectories)
RMSE (to be minimized)	1.74±1.20	1.79± 1.05	1.56±0.80
$IS_{0.90}$ (to be minimized)	14.85±8.18	14.38 ± 4.17	13.94 ± 2.09

Table 5. Prognostic performance evaluated considering complete test degradation trajectories.

Metric (right-censored test trajectories)	SR (training set: only complete degradation trajectories)	EvSR1 (training set: only complete degradation trajectories)	EvSR2 (training set: complete and right-censored degradation trajectories)
D (to be maximized)	0.44± 0.41	0.47±0.33	0.47± 0.33
$IPC_{0.90}$ (to be maximized)	0.20 ± 0.15	0.72 ± 0.13	0.81 ± 0.13

Table 6. Performance metrics for right-censored test degradation trajectories.

We can conclude that the better performance provided by the EvSR2 model with respect to that of the SR are due to the use of incomplete degradation trajectories in the training set. This is confirmed by the fact that the EvSR2 outperforms the EvSR1 considering the RMSE, $IS_{0.90}$ and $IPC_{0.90}$ metrics.

7. CONCLUSIONS

In this work, we have developed a novel method for estimating the RUL of industrial equipment based on SR measure and EvT. This method allows exploiting all available information, including that contained in the right-censored degradation trajectories. Furthermore, the method is able to properly represent the uncertainty in the RUL prediction by means of prognostic quantiles. The method has been shown able to outperform a similarity-regression based method on two case studies. The developed method is expected to be useful for the prediction of the RUL of components which undergo periodic maintenance, are safety critical or high-value, and, therefore, for which few run-to-failure degradation trajectories are available. Furthermore, the method can be easily extended to more general applications beyond PHM, for which time-to-event analysis is required. In these cases, the proposed method can manage left-censored or interval censored time series data.

One limitation of the developed method is the computational complexity of the Dempster's rule of combination, which grows exponentially in the number of evidential sources [40]. Thus, the use of the proposed method when many sources of evidence are available and real-time RUL predictions are required can be critical. This problem can be partially overcome by reducing the number of evidences to be aggregate, for example, limiting the combinations to the K -nearest training points or by using some approximation method based on Monte Carlo simulations [41]. Another issue of the proposed method concerns the fact that it requires some prior knowledge of the maximum life duration of the equipment under investigation. Since this quantity is typically known with uncertainty during the equipment design phase, future works will be devoted to properly take into account this additional source of uncertainty in the developed methodology.

ACKNOWLEDGEMENT

The participation of Enrico Zio to this research is partially supported by the China NSFC under grant number 71231001.

REFERENCES

- [1] Pecht, M., "Prognostics and health management of electronics", New York, NY, USA: Wiley-Interscience, 2008.
- [2] Mosallam, A., Medjaher, K., Zerhouni, N., "Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction", *Journal of Intelligent Manufacturing*, 27 (5), pp. 1037-1048, 2016.
- [3] Gorjian, N., Ma, L., Mittinty, M., Yarlagadda, P., Sun, Y., "A review on degradation models in reliability analysis", *Engineering Asset Lifecycle Management - Proceedings of the 4th World Congress on Engineering Asset Management, WCEAM 2009*, pp. 369-384, 2009.
- [4] Zhang, Z., Si, X., Hu, C., Lei, Y., "Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods", *European Journal of Operational Research*, 271 (3), pp. 775-796, 2018.
- [5] Le Son, K., Fouladirad, M., Barros, A., "Remaining useful lifetime estimation and noisy gamma deterioration process", *Reliability Engineering and System Safety*, 149, pp. 76-87, 2016.
- [6] Guo, J., Wang, C., Cabrera, J., Elsayed, E.A., "Improved inverse Gaussian process and bootstrap: Degradation and reliability metrics", *Reliability Engineering and System Safety*, 178, pp. 269-277, 2018.
- [7] Xiao, Q., Fang, Y., Liu, Q., Zhou, S., "Online machine health prognostics based on modified duration-dependent hidden semi-Markov model and high-order particle filtering", *International Journal of Advanced Manufacturing Technology*, 94 (1-4), pp. 1283-1297, 2018.

- [8] Cannarile, F., Compare, M., Rossi, E., Zio, E., "A fuzzy expectation maximization based method for estimating the parameters of a multi-state degradation model from imprecise maintenance outcomes", *Annals of Nuclear Energy*, 110, pp. 739-752, 2017.
- [9] Yang, L., Zhao, Y., Ma, X., "Multi-level maintenance strategy of deteriorating systems subject to two-stage inspection", *Computers and Industrial Engineering*, 118, pp. 160-169, 2018.
- [10] Kothamasu, R., Huang, S.H., "Adaptive Mamdani fuzzy model for condition-based maintenance", *Fuzzy Sets and Systems*, 158 (24), pp. 2715-2733, 2007.
- [11] Cerrada, M., Li, C., Sánchez, R.-V., Pacheco, F., Cabrera, D., Valente de Oliveira, J., "A fuzzy transition-based approach for fault severity prediction in helical gearboxes", *Fuzzy Sets and Systems*, 337, pp. 52-73, 2018.
- [12] Gugulothu, N., Vishnu, T.V., Malhotra, P., Vig, L., Agarwal, P., Shroff, G., "Predicting remaining useful life using time series embeddings based on recurrent neural networks", *International Journal of Prognostics and Health Management*, 9 (1), 13 p., 2018.
- [13] Zio, E., Di Maio, F., "A fuzzy similarity-based method for failure detection and recovery time estimation", *International Journal of Performability Engineering*, 6 (5), pp. 407-424, 2010.
- [14] Yang, Z., Baraldi, P., Zio, E., "A comparison between extreme learning machine and artificial neural network for remaining useful life prediction", *Proceedings of 2016 Prognostics and System Health Management Conference, PHM-Chengdu 2016*, 2017.
- [15] Rigamonti, M., Baraldi, P., Zio, E., Roychoudhury, I., Goebel, K., Poll, S., "Ensemble of optimized echo state networks for remaining useful life prediction", *Neurocomputing*, 281 (15), pp. 121-138, 2018.
- [16] Caruana, R., Lawrence, S., Giles, L., "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping", *Advances in Neural Information Processing Systems*, 2011.
- [17] Kleyner, A., Vasan, A., Pecht, M., "A new application for failure prognostics- reduction of automotive electronics reliability test duration", In *Annual Conference of the Prognostics and Health Management Society 2017. Annual Conference of the Prognostics and Health Management Society 2017*, St. Petersburg, Florida.
- [18] Durga Rao, K., Kushwaha, H.S., Verma, A.K., Srividya, A., "Quantification of epistemic and aleatory uncertainties in level-1 probabilistic safety assessment studies", *Reliability Engineering and System Safety*, 92 (7), pp. 947-956, 2007.
- [19] Aven, T., Baraldi, P., Flage, R., Zio, E., "Uncertainty in Risk Assessment: The Representation and Treatment of Uncertainties by Probabilistic and Non-Probabilistic Methods", *Uncertainty in Risk Assessment: The Representation and Treatment of Uncertainties by Probabilistic and Non-Probabilistic Methods*, pp. 1-186, 2014.
- [20] Yan, J., Koç, M., Lee, J., "A prognostic algorithm for machine performance assessment and its application", *Production Planning and Control*, 15 (8), pp. 796-801, 2004.
- [21] Tang, L., Kacprzyński, G. J., Goebel, K., Vachtsevanos, G., "Methodologies for uncertainty management in prognostics", *Proceedings of the IEEE aerospace conference*, 2009 Mar 7-14 2009; Big Sky, MT, 2009.
- [22] Xue, F., Bonissone, P., Varma, A., Yan, W., Eklund, N., Goebel, K., "An instance-based method for remaining useful life estimation for aircraft engines", *Journal of Failure Analysis and Prevention*, 8 (2), pp. 199-206, 2008.

- [23] Wang, T., Yu, J., Siegel, D., & Lee, J., “A similarity-based prognostics approach for remaining useful life estimation of engineered systems”, International conference on prognostics and health management, 6-9 Oct 2008, Denver (CO), 2008.
- [24] Shafer, G., “A mathematical theory of evidence”, Princeton, NJ: Princeton University Press, 1976.
- [25] Ramasso, E., Rombaut, M., Zerhouni, N., “Joint prediction of continuous and discrete states in time-series based on belief functions”, IEEE Transactions on Cybernetics, 43 (1), pp. 37-50, 2013.
- [26] Ramasso, E., Gouriveau, R., “Remaining useful life estimation by classification of predictions based on a neuro-fuzzy system and theory of belief functions”, IEEE Transactions on Reliability, 63 (2), pp. 555-566, 2014.
- [27] Serir, L., Ramasso, E., Zerhouni, N., “Time-Sliced Temporal Evidential Networks: The case of Evidential HMM with application to dynamical system analysis”, 2011 IEEE International Conference on Prognostics and Health Management, PHM 2011 - Conference Proceedings, 2011.
- [28] Baraldi, P., Mangili, F., Zio, E., “A belief function theory-based approach to combining different representation of uncertainty in prognostics”, Information Sciences, 303, pp. 134-149, 2015.
- [29] Smets, P., “The transferable belief model for quantified belief representation”, In D. M. Gabbay, & P. Smets (Eds.), Handbook of defeasible reasoning and uncertainty management systems, 1, pp. 267–301. Dordrecht: Kluwer Academic Publishers, 1998.
- [30] Hansen, B., “Nonparametric Estimation of Smooth Conditional Distributions”, *Technical report*, University of Wisconsin, 2004.
- [31] Durante, F., Fernández Sánchez, J., Trutschnig, W., “On the interrelation between Dempster-Shafer Belief Structures and their Belief Cumulative Distribution Functions”, Knowledge-Based Systems, 52, pp. 107-113, 2013.
- [32] Petit-Renaud, S., Dencœux, T., “Nonparametric regression analysis of uncertain and imprecise data using belief functions”, International Journal of Approximate Reasoning, 35 (1), pp. 1-28, 2004.
- [33] Yager, R.R., Alajlan, N., “Dempster-Shafer belief structures for decision making under uncertainty”, Knowledge-Based Systems, 80, pp. 58-66, 2015.
- [34] Nassreddine, G., Abdallah, F., Denoux, T., “State estimation using interval analysis and belief-function theory: Application to dynamic vehicle localization”, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 40 (5), pp. 1205-1218, 2010.
- [35] Gneiting, T., Raftery, A.E., “Strictly proper scoring rules, prediction, and estimation”, Journal of the American Statistical Association, 102 (477), pp. 359-378, 2007.
- [36] Ak, R., Vitelli, V., Zio, E., “An Interval-Valued Neural Network Approach for Uncertainty Quantification in Short-Term Wind Speed Prediction”, IEEE Transactions on Neural Networks and Learning Systems, 26 (11), art. no. 7050336, pp. 2787-2800, 2015.
- [37] Hwang C.-L., Yoon, K., “Multiple attributes decision making methods and applications”, Springer, Berlin, Heidelberg, 1981.

- [38] Saxena, A., Goebel, K., Simon, D., Eklund, N., “Damage propagation modeling for aircraft engine run-to-failure simulation”, 2008 International Conference on Prognostics and Health Management, PHM 2008, 2008.
- [39] Cannarile, F., Baraldi, P., Compare, M., Borghi, D., Capelli, L., Cocconcelli, M., Lahrac, A., Zio E., “An unsupervised clustering method for assessing the degradation state of cutting tools in the packaging industry”, Safety and reliability: theory and application, in: Proceedings of the European Safety and Reliability Conference, ESREL, 2017.
- [40] Orponen. P., “Dempster’s rule of combination is \neq p-complete”, Artificial Intelligence, 44, pp.245–253,1990.
- [41] Wilson, N, “Algorithms for dempster-shafer theory”. In D.M. Gabbay and P. Smets, editors, Handbook of Defeasible Reasoning and Uncertainty Management Systems, volume 5: Algorithms, pp. 421–475. Kluwer Academic Publishers, 2000.