



**HAL**  
open science

## Remaining useful life estimation in heterogeneous fleets working under variable operating conditions

Sameer Al-Dahidi, Francesco Di Maio, Piero Baraldi, Enrico Zio

### ► To cite this version:

Sameer Al-Dahidi, Francesco Di Maio, Piero Baraldi, Enrico Zio. Remaining useful life estimation in heterogeneous fleets working under variable operating conditions. *Reliability Engineering and System Safety*, 2016, 156, pp.109 - 124. 10.1016/j.ress.2016.07.019 . hal-01786958

**HAL Id: hal-01786958**

**<https://hal.science/hal-01786958>**

Submitted on 23 Mar 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# REMAINING USEFUL LIFE ESTIMATION IN HETEROGENEOUS FLEETS WORKING UNDER VARIABLE OPERATING CONDITIONS

Sameer Al-Dahidi<sup>1</sup>, Francesco Di Maio<sup>1\*</sup>, Piero Baraldi<sup>1</sup>, Enrico Zio<sup>1,2</sup>

<sup>1</sup> *Energy Department, Politecnico di Milano, Via La Masa 34, 20156 Milan, Italy*

<sup>2</sup> *Chair System Science and the Energy Challenge, Fondation Electricité de France (EDF), CentraleSupélec, Université Paris Saclay, Grande Voie des Vignes, 92290 Chatenay-Malabry, France*

**Abstract.** The availability of condition monitoring data for large fleets of similar equipment motivates the development of data-driven prognostic approaches that capitalize on the information contained in such data to estimate equipment Remaining Useful Life (RUL). A main difficulty is that the fleet of equipment typically experiences different operating conditions, which influence both the condition monitoring data and the degradation processes that physically determine the RUL. We propose an approach for RUL estimation from heterogeneous fleet data based on three phases: firstly, the degradation levels (states) of an homogeneous discrete-time finite-state semi-markov model are identified by resorting to an unsupervised ensemble clustering approach. Then, the parameters of the discrete Weibull distributions describing the transitions among the states and their uncertainties are inferred by resorting to the Maximum Likelihood Estimation (MLE) method and to the Fisher Information Matrix (FIM), respectively. Finally, the inferred degradation model is used to estimate the RUL of fleet equipment by direct Monte Carlo (MC) simulation. The proposed approach is applied to two case studies regarding heterogeneous fleets of aluminum electrolytic capacitors and turbofan engines. Results show the effectiveness of the proposed approach in predicting the RUL and its superiority compared to a fuzzy similarity-based approach of literature.

**Keywords:** Failure Prognostics, Remaining Useful Life (RUL), Heterogeneous Fleet, Homogeneous Discrete-Time Finite-State Semi-Markov Model, Aluminum Electrolytic Capacitors, Turbofan Engines.

## Notation and list of acronyms

RUL	Remaining Useful Life	$S$	Number of degradation states (final consensus clusters) of equipment
RNNs	Recurrent Neural Networks	$S_{final}$	Number of degradation states including the failure state of equipment
RVMs	Relevance Vector Machines	$i$	Index of degradation state, $i = 1, \dots, S_{final}$
PF	Particle Filtering	$\bar{X}$	Dataset matrix of the collected measurements
SVMs	Support Vector Machines	$C_{candidate}$	Possible number of clusters in the final consensus clustering $S$ , $C_{candidate} \in [C_{min}, C_{max}]$
NF	Neuro-Fuzzy	RMSE	Root Mean Square Error for prognostics
HDTFSSMM	Homogeneous Discrete-Time Finite-State Semi-Markov Model	AI	Accuracy Index for prognostics
MLE	Maximum Likelihood Estimation	$\alpha - \lambda$ accuracy	$(\alpha - \lambda)$ accuracy index for prognostics
FIM	Fisher Information Matrix	PI	Precision Index for prognostics
MC	Monte Carlo simulation	CR	Coverage Rate for prognostics
$N_{max}$	Number of Monte Carlo simulation trials	$RUL_p(t_l)$	True RUL of $p$ -th equipment at the measurement time $t_l$
FCM	Fuzzy $C$ -Means	$\overline{RUL}_p(t_l)$	Estimated RUL of $p$ -th equipment at the measurement time $t_l$
CSPA	Cluster-based Similarity Partitioning Algorithm	$I_p$	Number of measurements of $p$ -th equipment
$P$	Number of equipment in the fleet	$Z$	Number of signals of each degradation trajectory
$P_{training}$	Number of equipment used for training	$z$	Index of signal
$P_{training}^c$	Number of complete-run-to-failure trajectories used for training	$t_l^{(p)}$	Index of the measurement time of $p$ -th equipment, $l = 1, \dots, I_p$
$P_{training}^{pic}$	Number of incomplete-run-to-failure trajectories (right-censored) used for training	$M$	Number of discrete time steps between two successive measurements
$P_{test}$	Number of equipment used for testing	$ESR^{measured}$	Measurements of the degradation indicator
$p$	Index of equipment, $p = 1, \dots, P_{training}$ and/or $P_{test}$	$T$	Temperature profiles experienced by the capacitors
$\theta = \{q, \beta\}$	Parameters of the discrete Weibull distribution	$ESR^{norm}$	Capacitors degradation indicator
$\hat{\theta} = \{\hat{q}, \hat{\beta}\}$	Estimated parameters of the discrete Weibull distribution	DB	Davies-Bouldin criteria
$H$	Number of base clusterings	$c_{t_l}^p$	Coverage value of $p$ -th equipment at the measurement time $t_l$
$j$	Index of base clustering	$(\alpha - \lambda)_{t_\lambda}^p$	$(\alpha - \lambda)$ value of $p$ -th equipment at the measurement time $t_\lambda$
$C_{opt}^j$	Optimum number of clusters of the $j$ -th base clustering	$w_{t_l}^p$	Width value of $p$ -th equipment at the measurement time $t_l$
PPI	Prognostic Performance Indicator	$m_p$	The monotonicity of the $p$ -th degradation trajectory, $p = 1, \dots, P_{training}$

## 1. Introduction

Prognostics of failures aims at forecasting the Remaining Useful Life (RUL) of an equipment, i.e., the amount of time the equipment can continue performing its functions under its design specifications [1]–[4]. Knowledge of the RUL would allow avoiding system unscheduled shutdowns by defining efficient maintenance strategies that exploit the full RUL for operation. This would increase the system availability and safety, while reducing maintenance costs [2], [4], [5]. For these attractive reasons, there is an increasing interest of industry for failure prognostics [3], [6], [7].

Approaches for RUL estimation can be generally categorized into model-based and data-driven [2], [8]–[13]. Model-based approaches use physics models to describe the degradation behaviour of the equipment [4], [9],

[14], [15]. For example, Li et al. [16], [17], have proposed two prediction models of defect propagation in bearings; Oppenheimer et al. [18], have modelled a rotor shaft crack growth using the Forman law of linear elastic fracture mechanics and used the model for predicting its health condition and, correspondingly, estimating its RUL; Di Maio et al. [19], have explored the combination of exponential regression and Relevance Vector Machines (RVMs) for estimating the RUL of partially degraded thrust ball bearings; Cadini et al. [20], have used Particle Filtering (PF) for estimating the RUL of equipment subject to fatigue crack growth; modelled by Paris-Erdogan law [21], and for defining the optimal policy of condition-based equipment replacement. Despite the fact that these approaches lead to accurate prognostics results, uncertainty arising due to the assumptions and simplifications of the adopted models may pose limitations on their practical deployment [3], [9], [22]–[24].

Contrarily, data-driven prognostics approaches do not use any explicit physical model, but rely exclusively on the availability of process data related to equipment health to build (black-box) models that capture the degradation and failure modes of the equipment [4], [23]. For example, Di Maio et al. [22], have introduced a data-driven fuzzy similarity-based prognostics approach for estimating the RUL of equipment subject to fatigue cycles; Recurrent Neural Networks (RNNs) [25], Neuro-Fuzzy (NF) systems [26] and Support Vector Machines (SVMs) [27] have also been used for prognostics, with success. In spite of the recognized potential of these data-driven approaches, challenges still exist for their practical applications [4], [22]:

1) to build accurate models, data-driven approaches require sufficiently representative run-to-failure data (i.e., time series data up to the threshold value beyond which the equipment loses its functionality) which, in some practical cases, might be expensive or impractical to obtain; for this reason, data-driven approaches are more commonly applied for equipment of relatively short life than for safety-critical and slow-degrading equipment, for which complete run-to-failure trajectories are rarely available [4], [28];

2) these approaches are computationally intense [4];

3) with these models it is difficult to provide a measure of confidence on the RUL predictions, i.e., the uncertainty affecting the predictions [22], [29];

4) these approaches do not provide a clear physical interpretation of the current degradation condition of the equipment under observation, i.e., they behave like black-boxes [30].

To overcome these challenges, it seems worthwhile to consider and make use of the knowledge and data coming from similar equipment, forming what in the industrial context is called a fleet [6], [31], rather than relying solely on the knowledge and data coming from a single equipment. This will improve our knowledge concerning the equipment behaviour, reduce prognostics uncertainty and, thus, improve the efficiency of the fault prognostics task. A fleet of  $P$  pieces of equipment might:

- 1) have identical technical features and usage, and work in the same operating conditions, thus forming an identical fleet, e.g., a fleet of identical diesel engines located in one ship [6]; knowledge derived

from a fleet of this nature has been used for defining thresholds for anomaly detection [32] and for diagnosing faults [33] of equipment identical to the fleet members;

- 2) share some technical features and work in similar operating conditions, but show differences either on some features or on their usage, forming a so-called homogenous fleet, e.g., a fleet of trains working over a common route [34]; knowledge derived from this type of fleet has been used for developing diagnostics approaches for enhancing maintenance planning [34];
- 3) have different and/or similar technical features, but undergo different usage with different operating conditions, forming a so-called heterogeneous fleet, e.g., a fleet of highly standardized steam turbines of pressurized water reactors nuclear power plants [35]; this type of fleet can provide wider knowledge concerning the equipment behaviour [6], [31], [36].

The variability of behaviour of the members of the different types of industrial fleet above mentioned gives rise to a variability in a population of elements, in mathematical terms.

Most of the existing fleet-wide approaches for failure prognostics treat only the information gathered from identical and/or the homogenous fleets rather than from heterogeneous ones [14].

The objective of the present work is to develop a data-driven prognostics approach capable of capitalizing the information coming from an heterogeneous fleet of  $P$  pieces of equipment for the prediction of the progression of the degradation process to failure and, correspondingly, the estimation of the RUL of equipment [15].

The proposed approach is based on an Homogeneous Discrete-Time Finite-State Semi-Markov Model (HDTFSSMM). More specifically, the approach can be structured in three phases: offline identification of the degradation levels by an unsupervised ensemble clustering method, previously proposed by some of the authors [37] (see Appendix A.1), i.e., the identification of the health states of the Markov model that are explained by the different operating conditions experienced by the equipment during its life (whose number is generally “a priori” unknown, making the problem unsupervised) by an unsupervised ensemble clustering method, previously proposed by some of the authors [37] (see Appendix A.1); fleet data Maximum Likelihood Estimation (MLE) of the parameters of the discrete Weibull distributions [38]–[40] that are assumed to describe the transitions among the states and estimation of their uncertainties by the Fisher Information Matrix (FIM) [41]; use of the inferred degradation model in a direct Monte Carlo (MC) simulation to estimate the RUL of a new equipment of the fleet [42].

As we shall show, the strengths of the proposed approach lie in its capability of:

- 1) providing informative and transparent outcomes, by supplying the prediction of the current health state of the equipment and its RUL that is directly related to the operating conditions that influence the equipment degradation process. This can be very useful, since it allows identifying the adjuvant operating condition that may enhance the degradation process towards failure and, thus, scheduling proactively the proper maintenance interventions [43]

- 2) quantifying the uncertainty affecting the RUL predictions of the equipment, due to the variable operating conditions experienced by the fleet equipment, whose information is used for the HDTFSSMM parameters identification. This uncertainty assessment, which describes the expected mismatch between the real and predicted equipment failure times, can be used by the maintenance decision maker to plan maintenance interventions with the required confidence [43];
- 3) benefiting from the availability of also incomplete run-to-failure degradation trajectories (i.e., right-censored data) for enhancing the RUL estimations, rather than relying solely on complete run-to-failure degradation trajectories that might be difficult or even infeasible to be acquired, e.g., for high reliable equipment;
- 4) providing more accurate estimates of the equipment RUL when dealing with multidimensional problems, i.e., condition monitoring of multivariate signals.

The proposed approach is applied to two case studies regarding i) an heterogeneous fleet of aluminum electrolytic capacitors used in electric vehicles powertrains [44] and ii) an heterogeneous fleet of turbofan engines used in aircraft industry [45], both under variable operating conditions, that have been devised as particularly suitable to demonstrate the abovementioned strengths 3) and 4), respectively. The performance of the proposed approach is verified with respect to five metrics (i.e., Root Mean Square Error (RMSE), Accuracy Index (AI), Precision Index (PI), Coverage Rate (CR) and  $\alpha - \lambda$  accuracy [46]) and is compared with a data-driven prognostics approach from literature (i.e., the fuzzy similarity-based approach introduced in [22]).

The remaining of this paper is organized as follows. In Section 2, the proposed data-driven prognostics approach for the estimation of the RUL of equipment belonging to an heterogeneous fleet working under variable operating conditions is illustrated. In Sections 3 and 4, the results of the application of the proposed approach to the two case studies are presented and compared with those obtained by the fuzzy similarity-based approach. Finally, some conclusions are drawn in Section 5.

## 2. Data-driven prognostics approach for heterogeneous fleets

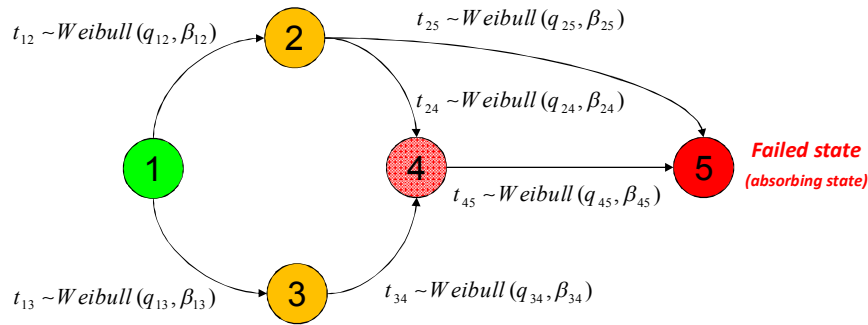
In this Section, we describe the proposed data-driven prognostics approach for estimating the RUL of equipment of an heterogeneous fleet formed by  $P_{training}$  pieces of equipment. The approach capitalizes the available information (i.e., data and knowledge) collected on the members of the fleet, regarding their degradation behaviours under different operating conditions.

The degradation process of the equipment of the fleet is assumed to follow a HDTFSSMM [47]–[50], for which the states are re-generation states [51], [52]:

- when the generic  $i$ -th state is reached, the previous history of the process is assumed not to influence the future evolution (i.e., Homogenous);
- the transitions among states are allowed only at discrete time instants (i.e., Discrete-Time);

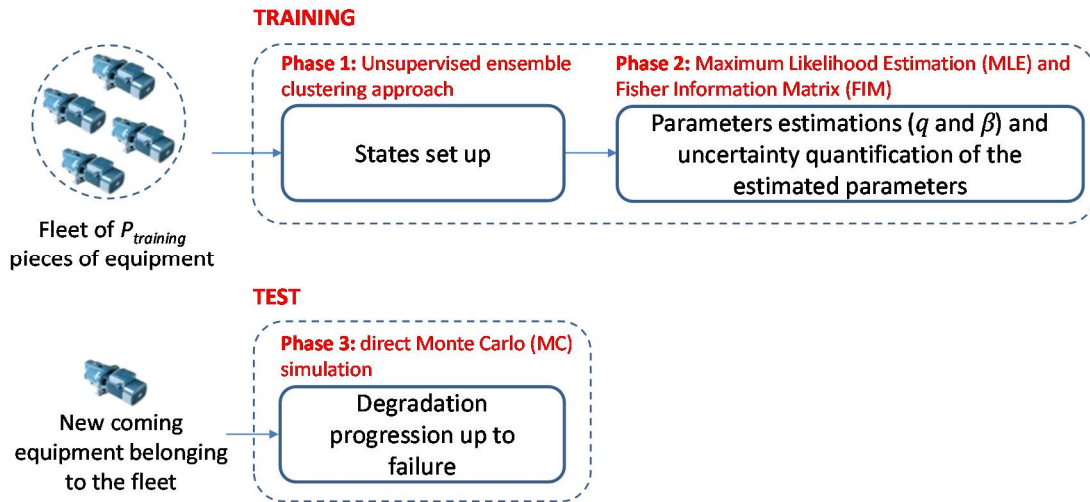
- the equipment are described in a finite set of degradation levels or states influenced and explained by different operating conditions (i.e., Finite-State) and;
- the transition rates depend on the time spent in the current state (sojourn time) (i.e., Semi-Markov), which can follow any arbitrary distribution. In this work, we assume discrete Weibull distributions, as these are the probability distributions most commonly used to describe degradation processes of industrial equipment [49], [53], [54].

For ease of clarity, Figure 1 shows an example of a physical degradation process described by a HDTFSSMM of five states, starting from state 1 and evolving into two alternative paths (states 2 and 3) that depend on the operating conditions experienced by the equipment up to failure in state 5. The transition times from generic state  $i$  to generic state  $j$  are described by Weibull distributions with parameters  $q_{ij}$  and  $\beta_{ij}$ ,  $i, j = 1, 2, \dots, 5$ .



**Figure 1: An example of a HDTFSSM modelling the degradation behaviour of generic equipment.**

The proposed method entails three phases and is sketched in Figure 2.



**Figure 2: Sketch of the proposed data-driven prognostics approach for a fleet of  $P_{training}$  pieces of equipment.**

Let us assume the availability of  $I_p$  measurements for each of the  $p = 1, \dots, P_{training}$  pieces of equipment (hereafter called training trajectories), that are monitored at discrete times throughout the time horizon of observation. The collected trajectories come from pieces of equipment that are undergoing different usages under different operating conditions and, thus, the fleet they belong to can be considered heterogeneous. Among the training trajectories,  $P_{training}^C$  are complete run-to-failure trajectories (i.e., trajectories that last all

the way to the instance when the degradation state reaches the threshold value beyond which the equipment loses its functionality) and  $P_{training}^{ic} = P_{training} - P_{training}^c$  are incomplete run-to-failure trajectories (i.e., trajectories that do not reach the failure threshold – right-censored data). Each  $p$ -th trajectory is a  $z$ -dimensional trajectory of  $Z$  signals representative of the equipment behaviour and of the operating conditions the equipment is subjected to.

**Phase 1:** *Unsupervised ensemble clustering approach for setting up the number of states of the HDTFSSMM.*

For each  $p$ -th equipment, the  $z$ -dimensional measurements are taken at predefined times  $t_1^{(p)}, t_2^{(p)}, \dots, t_l^{(p)}, t_{l_p}^{(p)}$ , where  $l = 1, \dots, I_p$  and  $t_l^{(p)} - t_{l-1}^{(p)} = M$  is the number of discrete time steps between two successive measurements. The measurements taken on all the  $P_{training}$  pieces of equipment are, then, appended to the overall dataset matrix  $\bar{X}$ .

The objective is to partition the collected data in  $\bar{X}$  into  $S$  dissimilar groups (whose number is ‘‘a priori’’ unknown), such that data belonging to the same group characterize the degradation state of the HDTFSSMM that has to be built. To guarantee that data populating the same state are more similar than those belonging to other groups, the unsupervised ensemble clustering approach sketched in Figure 3 [37] is adopted. It ensembles  $H$  base clusterings performed considering different groups of signals. Depending on the application, different criteria can be adopted for grouping the signals, e.g., in [37] the authors have used two groups of signals with the first dedicated to signals representative of the equipment behaviour, and the second to the operating conditions, whereas in other applications characterized by a large number of measured signals, groups of signals formed by highly correlated signals can be used [55].

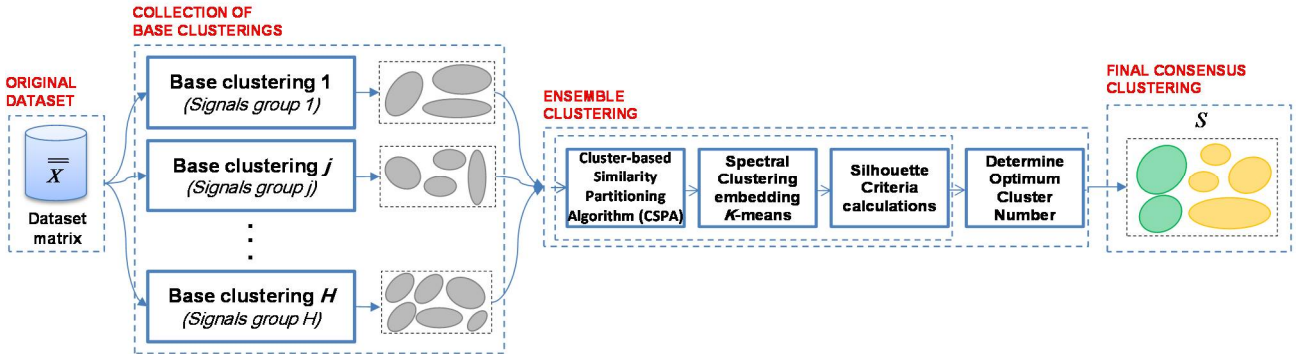


Figure 3: The unsupervised ensemble clustering approach [37].

In brief, the unsupervised ensemble clustering is based on the combination of (refer to Appendix A.1 for more details): 1) a Cluster-based Similarity Partitioning Algorithm (CSPA) [56] to quantify the co-association matrix that describes the similarity among the two base clusterings; 2) Spectral Clustering [57], [58], embedding an unsupervised  $K$ -Means algorithm [59], [60], to find the final consensus clustering based on the available co-association matrix; 3) the Silhouette index [61] to quantify the goodness of the obtained clusters by choosing the optimum number  $S$  of clusters in the final consensus clustering as that with the maximum Silhouette value, i.e., such that clusters are well separated and compacted.



The final consensus clusters  $S$  can be seen as the health states representative of the different degradation levels of the equipment, that are influenced and explained by different operating conditions [37]. The failure state (i.e., an absorbing state) at which the degradation level reaches the threshold value is added to those states to build the transition diagram of the equipment operation with  $S_{final}$  states (i.e.,  $S_{final} = S + 1$  states).

**Phase 2: States transition parameters estimation and their uncertainty quantification.**

Once the topology of the model is fully defined, the parameters governing the transitions among the degradation states are to be estimated. We assume they obey a discrete Weibull distribution with probability mass function, cumulative distribution function and survival function defined as  $f(t) = (q^{(t-1)})^\beta - (q^t)^\beta$ ,  $F(t) = 1 - (q^t)^\beta$  and  $R(t) = (q^t)^\beta$ , respectively, and whose parameters to be identified are  $\underline{\theta} = \{q, \beta\}$  (with  $0 < q < 1$  and  $0 < \beta$ ) [62], [63].

The parameter estimation problem is here formulated as an optimization problem, where the likelihood of having for each  $p$ -th equipment pairs of state-measurement time  $Q^{(p)} = [S_1^{(p)}, t_1^{(p)}, \dots, S_l^{(p)}, t_l^{(p)}, S_{l_p}^{(p)}, t_{l_p}^{(p)}]$  describing the degradation through the  $S_{final}$  degradation states is to be maximized [49].

Resorting to the Maximum Likelihood Estimation (MLE) technique [41], the log-likelihood function to be maximized under the constraints  $0 < q < 1$  and  $0 < \beta$  can be defined as in Eq. (1):

$$L = Pr(Q|\underline{\theta}) = \prod_{p=1}^{P_{training}} Pr(Q^{(p)}|\underline{\theta})$$

$$\log L = \sum_{p=1}^{P_{training}} \log \left( \prod_{i=1}^{S_{final}} \left( \prod_{n=1}^{n_i^{(p)}} \left( f_{ij}(\tau_n, \underline{\theta}) \cdot R_{ig}(\tau_n, \underline{\theta}) \right)^\delta \cdot \left( R_{ij}(\tau_n, \underline{\theta}) \right)^{1-\delta} \right) \right) \quad (1)$$

where  $f_{ij}(\tau_n, \underline{\theta}) = (q_{ij}^{(\tau_n-1)})^{\beta_{ij}} - (q_{ij}^{(\tau_n)})^{\beta_{ij}}$  is the probability that the  $n$ -th transition,  $n = 1, \dots, n_i^{(p)}$ , occurs from state  $i$  to another state  $j$ , with sojourn time  $\tau_n$ ,  $R_{ig}(\tau_n, \underline{\theta}) = (q_{ig}^{\tau_n})^{\beta_{ig}}$  is the probability that the  $n$ -th transition does not occur to another state  $g$  before the end of the sojourn time  $\tau_n$ ,  $\delta$  is equal to 1 if the last observed state is the failure state (i.e., for  $P_{training}^c$  complete run-to-failure trajectories), whereas  $\delta$  is equal to 0 if the last observed state is a non-failure state (i.e., for  $P_{training}^{ic}$  incomplete run-to-failure trajectories) [64], [65].

Confidence intervals on the parameters  $\underline{\hat{\theta}}$  can be obtained by exploiting the asymptotic properties of the Maximum Likelihood estimator  $\underline{\hat{\theta}}$  and resorting to Fisher Information Matrix (FIM) (refer to [41] for more details).

Once the  $\underline{\hat{\theta}}$  are estimated, one must verify that the identified health states and the respective transition parameters represent a monotonic degradation trend, from less to more degraded states, so that the model built can be effectively used for RUL estimation. In this work, we rely on the calculation of the monotonicity Prognostic Performance Indicator (PPI) [75], [76] to verify that the proposed health states result in a monotonic trend:

$$Monotonicity = \frac{1}{P_{training}} \left| \sum_{p=1}^{P_{training}} m_p \right| \quad (2)$$

$$m_p = \frac{\text{no.of } d/dx \geq 0}{I_p - 1} - \frac{\text{no.of } d/dx < 0}{I_p - 1}, p = 1, \dots, P_{\text{training}}$$

where  $m_p$  and  $I_p$  are the monotonicity and the number of measurements of the  $p$ -the degradation trajectory. In practice, the monotonicity PPI of Eq. (2) ranges in  $[0,1]$ , where 1 indicates large monotonicity property and 0 vice versa (i.e., a value of 1 entails that the equipment really degrades from less to more degraded states over time, whereas a value of 0 entails that the equipment undergoes a self-healing and, thus, its health state can be improved over time (refer to [76] for more details)).

**Phase 3:** *Direct Monte Carlo (MC) simulation of the degradation progression for the online estimation of the RUL.*

The RUL of new coming equipment belonging to the fleet at time  $t$  is performed by resorting to a direct MC simulation [66], viz:

1. At the actual time  $t$ , classify the actual health state of the equipment into one of the  $S_{\text{final}}$  states found in phase 1, on the basis of the online measured signals values representative of the equipment behaviour and of the operating conditions the equipment is subjected to;
2. Suppose the actual state is recognized at state  $i$  and the time of the first entrance in state  $i$  is  $t_i$ . Find the residual sojourn time  $\tau_i$  provided that the equipment has already sojourned in state  $i$  for  $\tilde{T} = t - t_i$  units of time, by sampling the times  $\tau_{i,j}$  from  $u_{i,j} = F(\tau_{i,j} | \hat{\theta}_{i,j})$  for each  $j = 1, \dots, S_{\text{final}}$  [66]:
  - 2.1. Sample  $u_{i,j} \sim U(0,1)[1 - F(\tilde{T})] + F(\tilde{T})$ , where  $U(0,1)$  is a random number uniformly distributed between 0 and 1;
  - 2.2. Calculate  $\tau_{i,j} = F^{-1}(u_{i,j}) - (\tilde{T})$ , where  $F(\tau_{i,j} | \hat{\theta}_{i,j}) = 1 - (q_{i,j}^{\tau_{i,j}})^{\beta_{i,j}}$ .

The sojourn time  $\tau_i$  is  $\tau_i = \min(\tau_{i,j})$ ; whereas the arrival state  $j$  is the one that corresponds to  $\arg[\min(\tau_{i,j})]$ : the equipment enters state  $j$  at the respective time  $\tau_{i,j}$ .

3. Sample the sojourn times in the proceeding states from the corresponding cumulative distributions up to failure state as done in 2);
4. Repeat 2) and 3) for  $N_{\text{max}}$  times, where  $N_{\text{max}}$  is the number of MC simulations trials;
5. The samples statistics allow estimating the probability distribution of the RUL and calculate, for example, its mean value and the associated uncertainty bounds, e.g., the 10<sup>th</sup> and 90<sup>th</sup> percentiles.

It seems worth mentioning that we resort to a direct MC simulation, instead of developing analytical methods for RUL prediction, because it is usually impossible to construct closed-form analytical solutions for the RUL probability distributions of complex systems: even though analytical methods are computationally cheaper, they are not readily suitable to account for all types of uncertainty in prognostics [67], especially for the uncertainties on the parameters affecting the state transition times distributions.

### 3. Case study I: Aluminum electrolytic capacitors in fully electrical vehicles

In this Section, the potentiality of the proposed approach is demonstrated on an artificial case study properly designed to mimic the realistic degradation behavior of an heterogeneous fleet of  $P_{training} = 100$  aluminum electrolytic capacitors used in automotive industry [44]. In [44], some of the authors propose a model-based prognostics approach based on PF for the estimation of the RUL of aluminum electrolytic capacitors, assuming that the degradation model of the capacitor is available and the values of its parameters are known. The data-driven approach here proposed, instead, avoids resorting to a physical model because in practice this may not be available. Also, the data-driven approach proposed is capable of exploiting incomplete degradation trajectories (i.e., right-censored data).

#### 3.1 Simulation setup and data description

The main degradation mechanism of electrolytic capacitors is the vaporization of the electrolyte, whose degradation speed is largely influenced by the component working temperature [68]. According to [44], the Normalized Equivalent Series Resistance ( $ESR^{norm}$ ) is considered as a degradation indicator and its evolution has been simulated for a fleet of capacitors using a physics-based model [44]. The simulated degradation trajectories begin with an initial  $ESR^{norm}$  value equal to 100% and end at the failure of the capacitor, assumed to be the first time instant at which  $ESR^{norm}$  exceeds the failure threshold of 200% [69]. During the capacitor life, the following  $Z = 2$  signals are measured:

- 1)  $ESR^{measured}$ , the direct measurements of the degradation indicator, i.e., the measurements of the Normalized Equivalent Series Resistance ( $ESR^{norm}$ );
- 2) the temperature experienced by the capacitor ( $T$ ), which represents the operating condition most influencing the degradation process of the capacitor. According to design experts [70], [71], temperature variations experienced by the capacitors during life are mainly caused by i) the seasonality of the environmental external temperature and by ii) the aging (barely up to 10% of its initial temperature value). Therefore, the simulated temperature profiles follow an arbitrary sinusoidal function that justifies seasonality, by adding to this a shift sigmoidal function accounting for aging.

The heterogeneity among the capacitors that belong to the fleet is guaranteed by considering arbitrary parameters of the sinusoidal and the sigmoidal functions.

For clarification purposes, Figure 4 shows the simulated data of two capacitors (capacitor 1 and capacitor 2 – dark and light shade of color, respectively): Figure 4 (top) shows the capacitors degradation process ( $ESR^{norm}$ ), Figure 4 (left bottom) shows the  $ESR$  measurements ( $ESR^{measured}$ ), whereas Figure 4 (right bottom) shows the temperature profiles experienced by the capacitors ( $T$ ). It is worth noticing that the higher the temperature (capacitor 2 – light shade of color) is, the faster the vaporization process due to the increase of the self-heating effects and, hence, faster is the failure process as shown in Figure 4 (top) [44], [72].

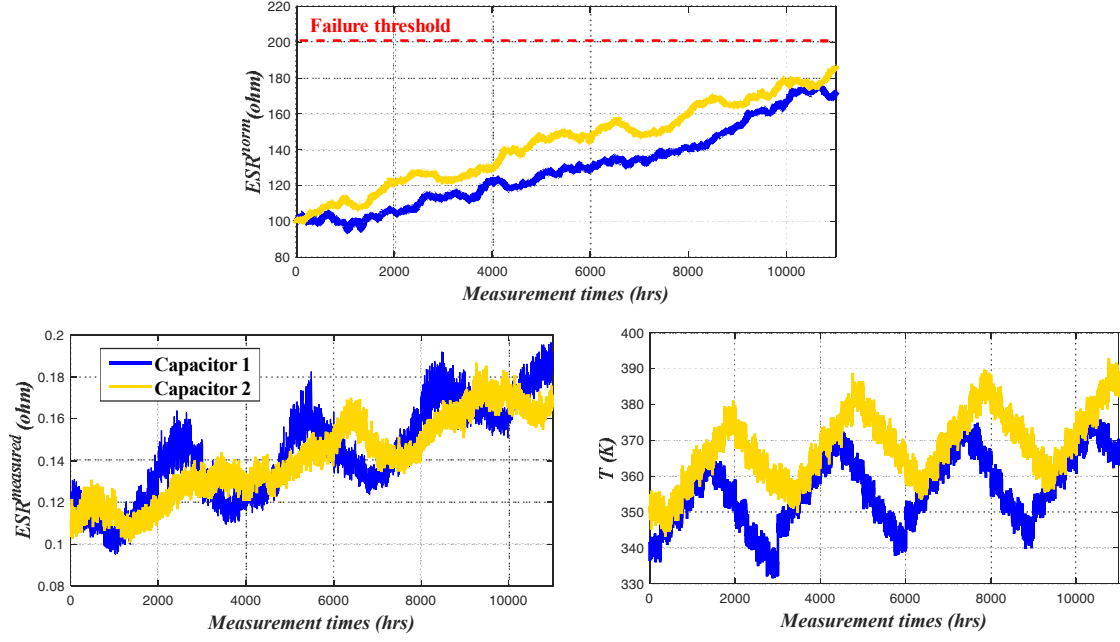


Figure 4: The true degradation process ( $ESR^{norm}$ ) (top), the  $ESR$  measurements ( $ESR^{measured}$ ) (left bottom) and the temperature profiles experienced by the capacitors ( $T$ ) (right bottom).

As already said, real data from complete run-to failure trajectories are expensive or impractical to collect, especially for highly reliable equipment: to simulate this, from the heterogeneous fleet of  $P_{training} = 100$  capacitors,  $P_{training}^c = 20$  are assumed to show degradation trajectories that last all the way to the failure threshold, whereas  $P_{training}^{ic} = 80$  show incomplete run-to failure trajectories. In Figure 5, examples of the  $ESR^{norm}$  signals of the complete and incomplete run-to-fail degradation trajectories (in dark and light shade of color, respectively) are shown.

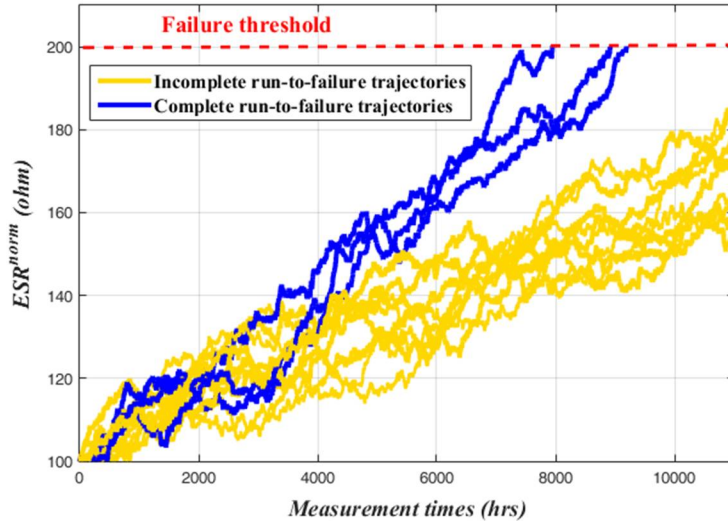


Figure 5: Examples of simulated complete and incomplete run-to-failure degradation trajectories.

All the measurements of both the complete and incomplete run-to-failure degradation trajectories are stored in the dataset matrix  $\bar{X}$  that is used in Sections 3.2 and 3.3 for the purpose of comparing the performance of the

proposed approach with the fuzzy similarity-based approach. For computational convenience, 1000 time steps between two successive measurements (i.e.,  $M = 1000$ ) are considered.

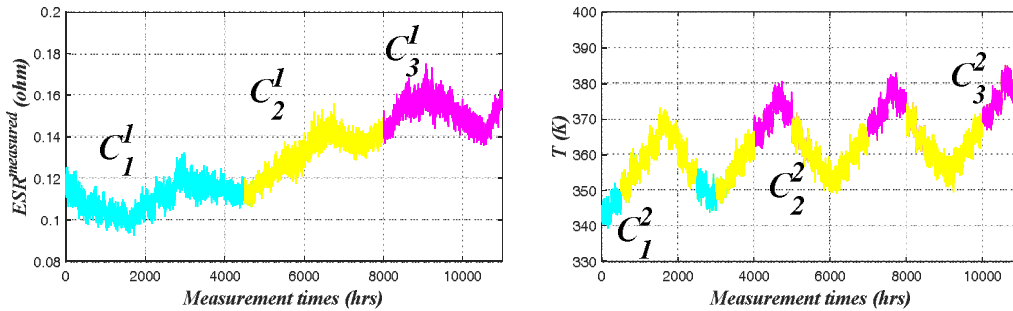
### 3.2 Application of the proposed data-driven prognostics approach

The application of the proposed data-driven prognostics approach is here described step-by-step, following the scheme presented in Section 2.

**Phase 1:** The states of the HDTFSSMM describing the capacitor  $ESR$  degradation progression under different temperature conditions are identified by resorting to the unsupervised ensemble clustering approach [37] with  $H = 2$  base clusterings ( $j = 1$  for the capacitor degradation behaviour and  $j = 2$  for the temperature conditions that influence the capacitor degradation process).

An unsupervised Fuzzy  $C$ -Means (FCM) algorithm [73] is used for each base clustering: by resorting to Davies-Bouldin (DB) validity criterion [74] the correct number of clusters  $C_{opt}^1 = 3$  and  $C_{opt}^2 = 3$  is identified for  $j = 1$  and  $j = 2$  base clusterings, respectively. In practice, this result means that if we would like to cluster the data (i.e., to define the states) from the independent perspective of degradation and operational conditions, we should use 3 states for the former and (incidentally) 3 for the latter.

For ease of clarity, Figure 6 (left) shows the three different levels of the degradation ( $ESR^{measured}$ ) of a capacitor (namely low degrading ( $C_1^1 - L_{ESR}^{measured}$ ), medium degrading ( $C_2^1 - M_{ESR}^{measured}$ ) and high degrading ( $C_3^1 - H_{ESR}^{measured}$ )), obtained by the  $j = 1$  base clustering. Similarly, Figure 6 (right) shows the three different temperatures experienced by the capacitor (i.e., low temperature ( $C_1^2 - L_T$ ), medium temperature ( $C_2^2 - M_T$ ) and high temperature ( $C_3^2 - H_T$ )), obtained by the  $j = 2$  base clustering.



**Figure 6: The base clustering results of  $ESR^{measured}$  (left) and the  $T$  (right) of one capacitor.**

The optimum number of states that guarantees a comprehensive representation of the possible degradation under variable operational conditions is given by the number  $S$  of final consensus clusters that is selected according to the Silhouette values [61] for different numbers of clusters  $C_{candidate}$  that span in the interval  $[3,9]$ , where the lower bound (3) is the minimum between  $C_{opt}^1$  and  $C_{opt}^2$ , and the upper bound (9) is the number of the largest combination of the two base clusters (i.e.,  $3 \times 3$ ) [37]. The optimum number  $S$  of states is found to be  $S = 9$ , at which the Silhouette measure is maximized (i.e., 0.71) (star in Figure 7 (top)) [61].

This result leads to the transition diagram with  $S_{final} = 10$  states shown in Figure 7 (bottom), where the 10<sup>th</sup> state is the failure state (i.e., an absorbing state) at which the degradation level reaches the threshold value. The equipment may transit through all states before failure. In fact, at each state, three types of transitions are possible: i) transitions to the immediate neighbour states of different degradation levels, e.g., transitions between states 1 and 4, ii) transitions to the immediate neighbour states of different temperature conditions, e.g., transitions between states 2 and 3, and iii) transitions to the immediate neighbour states of different degradation level and temperature condition, e.g., transitions between states 2 and 6.

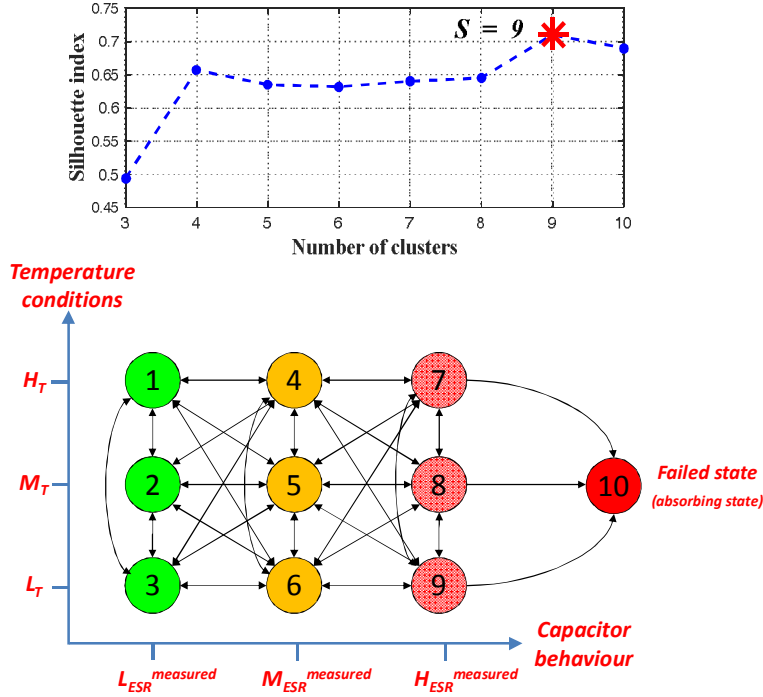


Figure 7: Silhouette values vs. cluster numbers (top) and the built transition diagram (bottom).

In practice, this means that the different degradation behaviors (namely, low, medium and high degrading) of the fleet capacitors, influenced and explained by the different temperature conditions experienced by the capacitors during their lives (namely, low, medium and high temperatures), can be retrieved for: i) estimating the actual health state of a new capacitor, ii) tracking its degradation progression, iii) informing the diagnosis with the adjuvant temperature conditions that may enhance the degradation process towards failure, iv) estimating its RUL and, correspondingly, v) scheduling proactively a proper maintenance intervention.

**Phase 2:** Resorting to the MLE and to the FIM, the estimated parameters  $\hat{\theta}$  governing the transitions among the degradation states, and their 95% confidence intervals  $\hat{\theta}$ , are obtained, respectively. For clarification purposes, Figure 8 shows the estimated parameters  $\hat{\theta}$  of the transitions that might occur from state 5 ( $q$  and  $\beta$ , top and bottom Figure, respectively).

It can be seen that the transitions from state 5 to states 1, 2, 3 and 10 seem not possible (from the available fleet data) and, hence, the estimated  $q$  and  $\beta$  values are  $\sim 1$  and 0, respectively (circles), whereas the transitions to states 4, 6, 7, 8 and 9 are possible and the uncertainty (error) on their estimated parameters values (dots) is

clearly influenced by the number of transitions that have actually occurred and recorded in the dataset (small number of transitions entails large uncertainty on the estimations, whereas large number of transitions entails small uncertainty on the estimations), but can be kept under control.

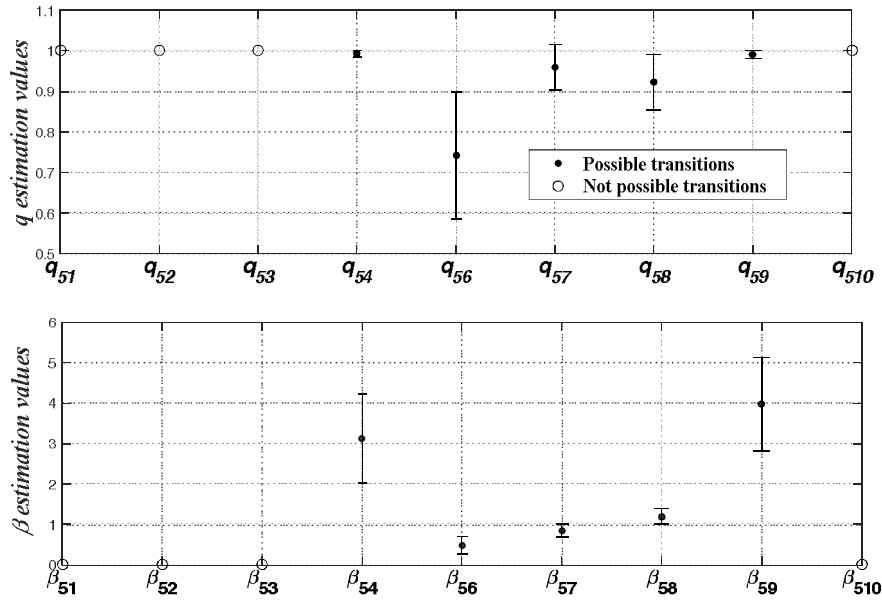


Figure 8: Estimated parameters values and their confidence intervals of the transitions occurred from state 5.

To verify that the developed model is representative of a monotonic degradation process, the monotonicity PPI of Eq. (2) is calculated for a sequence of 1000 degradation paths that were simulated by a direct MC with the estimated  $\hat{\theta}$  and, then, compared with the monotonicity of the  $P_{training} ESR^{norm}$  degradation trajectories (Figure 5).

Without loss of generality, Figure 9 shows one simulated capacitor degradation path: if the temperature conditions are neglected, it is clear that it monotonically degrades through the three degradation levels ( $L_{ESR}^{measured}$ ,  $M_{ESR}^{measured}$ ,  $H_{ESR}^{measured}$ ).

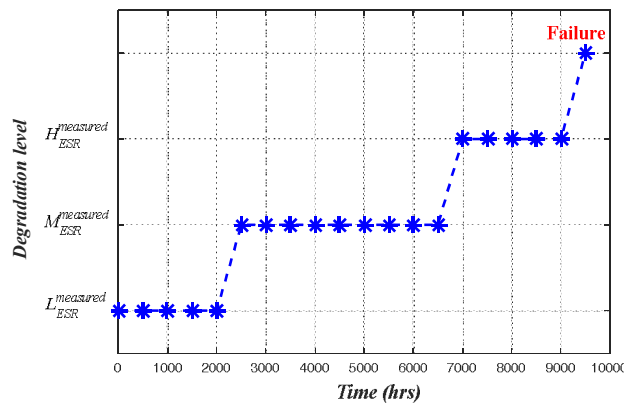
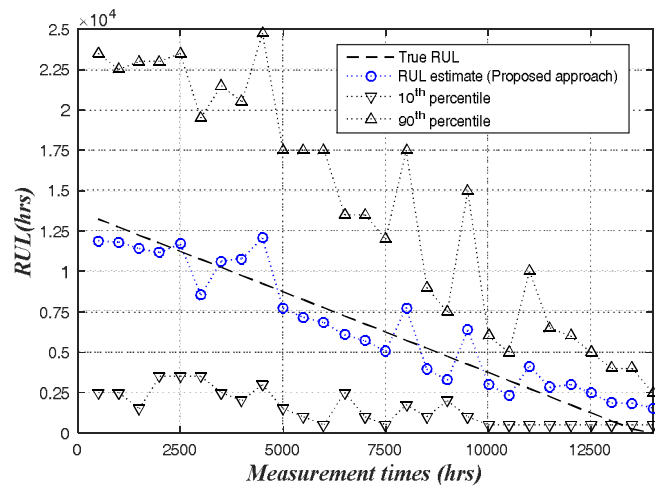


Figure 9: Sample realization of the degradation process of a capacitor.

The monotonicity PPI turns out to be equal to 0.71 and 0.52 for the simulated degradation paths and the real degradation trajectories, respectively. We can, thus, conclude that the built model degradation is representative of a monotonic degradation process and, thus, it can be effectively used for the purpose of RUL estimation.

**Phase 3:** The online estimation of the RUL of  $P_{test} = 50$  new capacitors (complete run-to-failure trajectories) at each time  $t$  is performed by resorting to the direct MC simulation scheme. In order to obtain a good compromise between accuracy and computational time,  $N_{max} = 1000$  MC trials have been used in the simulation.

The prognostics approach provides a prediction of the RUL in the form of a probability density function whose mean values (circles) and the associated 10<sup>th</sup> and 90<sup>th</sup> percentiles (triangles) for one capacitor are shown in Figure 10.



**Figure 10: RUL prediction and corresponding 10<sup>th</sup> and 90<sup>th</sup> percentiles of one capacitor.**

It can be seen that: 1) the estimated RUL is reasonably close to the actual remaining life, and 2) as the capacitor ages, the prediction intervals for the remaining life become narrower, that is, the prediction uncertainty decreases. This reduction is due to the reduction of the time horizon, which makes the RUL prediction task less affected by the variability of the fleet transitions and of the temperature conditions experienced by the capacitors at the early stages of their lives.

In conclusion, one of the strengths of the proposed approach lies in its capability of associating the RUL predictions with a measure of confidence, contrarily to many state-of-the-art data-driven approaches for prognostics like the fuzzy similarity-based approach (as we shall see in Section 3.3), for accommodating the variable operating conditions experienced by the equipment during its life. This assessment of the uncertainty can be used by the maintenance decision maker to confidently plan maintenance interventions [77].

Furthermore, the capability of the proposed approach of using incomplete run-to-failure degradation trajectories for enhancing the RUL predictions is shown in the following Section, by comparison with the fuzzy similarity-based approach.

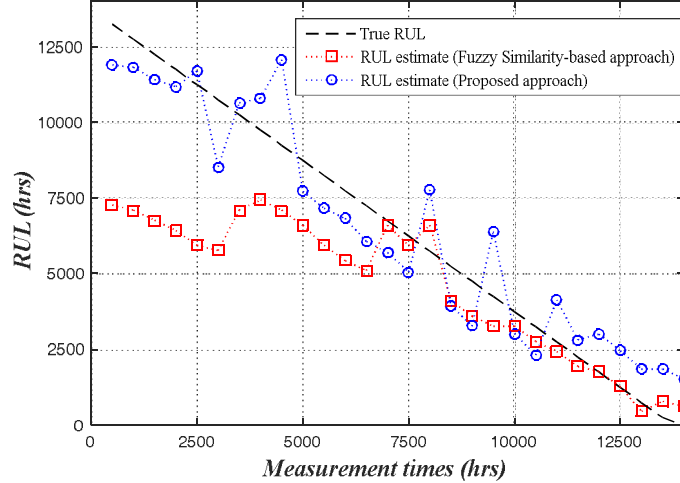


### 3.3 Comparison with a data-driven fuzzy similarity-based approach

In this Section, a data-driven prognostics approach from literature, the fuzzy similarity-based approach [22], has been applied to the aluminum electrolyte capacitors case study and the obtained results are compared with those obtained by the proposed approach.

Briefly, the basic idea of the fuzzy similarity-based approach is to evaluate the similarity between the test degradation trajectory and the  $P_{training}^c = 20$  complete run-to-failure training trajectories, and to use the RULs of these latter to estimate the RUL of the former, considering how similar they are [30], [78], [79]. The similarity is quantified by resorting to the definition of an ‘‘approximately zero’’ fuzzy set taken as a bell-shaped function whose parameters can be set by following a trial and error procedure on a validation set of complete run-to-failure trajectories. For more details on the fuzzy similarity-based approach, the interested reader may refer to [78].

The estimates of the RUL obtained resorting to the fuzzy similarity-based approach for one capacitor are plotted in Figure 11 in squares (together with those obtained by the proposed approach of Section 3.2 in circles). The predictions provided by the two approaches are comparable: the benefit in accuracy of the proposed approach with respect to the fuzzy similarity-based approach (in particular at the early stage of its life) is justified by the use of the complete dataset for training the proposed approach, whereas only the complete run-to failure trajectories are used to train the fuzzy similarity-based approach.

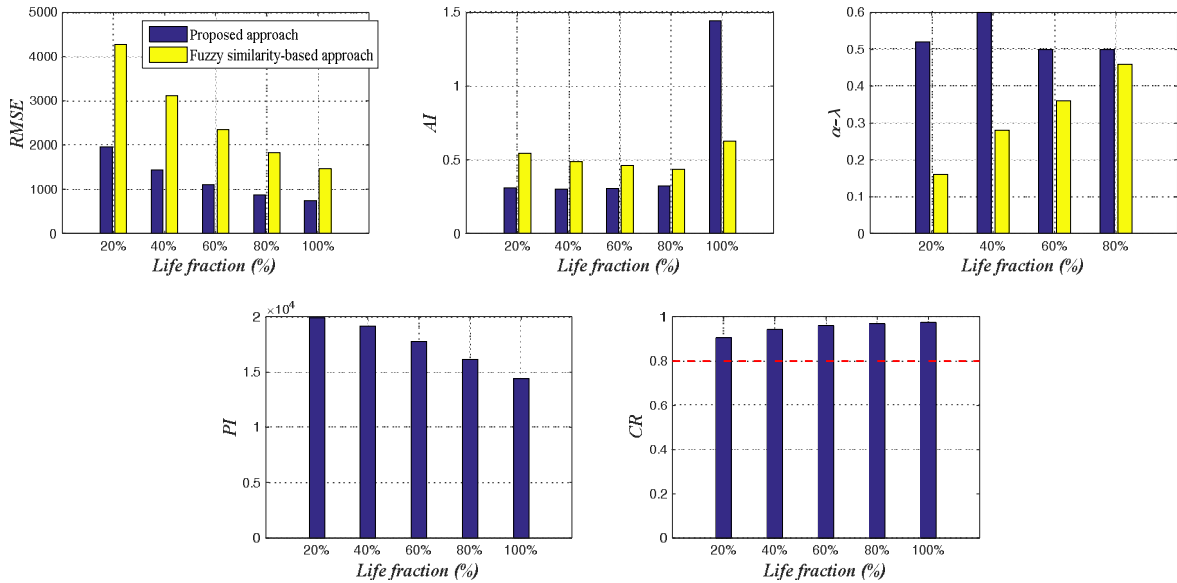


**Figure 11: Comparison of the RUL predictions for one capacitor provided by the proposed approach and the fuzzy similarity-based approach.**

To further evaluate the proposed approach with respect to the fuzzy similarity-based approach, the approaches are applied to the  $P_{test} = 50$  test trajectories in correspondence of five life fractions after the beginning of the capacitors lives, i.e., 20%, 40%, 60%, 80% and 100% (i.e., the entire life of the capacitor). In particular, the following performance metrics are considered [46] (refer to Appendix A.2 for more details on their definitions):

- 1) Root Mean Square Error (RMSE): it is defined as the average error of the RUL prediction. Small RMSE values indicate more accurate predictions;
- 2) Accuracy Index (AI): it is defined as the relative error of the RUL prediction. Small AI values indicate more accurate predictions;
- 3)  $\alpha - \lambda$  accuracy: it determines whether a prediction falls within specified  $\alpha$ -bounds, e.g.,  $\alpha=10\%$  accuracy, of the true RUL at a specific time  $t_\lambda$  to failure, e.g.,  $\lambda=20\%$ ,  $40\%$ ,  $60\%$  and  $80\%$ . Large values of this index indicate more accurate predictions;
- 4) Precision Index (PI): it computes the relative width of the prediction interval. Small values of PI indicate more precise predictions;
- 5) Coverage Rate (CR): it is the probability that the true RUL value lies within the RUL prediction interval. Larger CR value means that the predicted interval has high chance to cover the true RUL.

Figure 12 shows the average of the five performance metrics over the  $P_{test} = 50$  test trajectories, when the whole dataset ( $P_{training} = 100$  trajectories) is used to train the proposed approach (dark shade of color) and when the complete run-to-failure trajectories ( $P_{training}^c = 20$  trajectories) are used to train the fuzzy similarity-based approach (light shade of color).



**Figure 12: Comparison of the five performance metrics for the proposed approach and the fuzzy similarity-based approach.**

It can be seen that the proposed approach provides a more satisfactory performance in terms of accuracy of the RUL predictions, i.e., lower RMSE and higher  $\alpha - \lambda$  metrics (Figure 12, top). This is due to the fact that the proposed method exploits information taken from both the complete run-to-failure trajectories and the incomplete run-to-failure trajectories, whereas the fuzzy similarity method only uses the first source of information. Furthermore, the complete run-to-failure trajectories used for training the fuzzy similarity-based approach are characterized by short lives (refer to Figure 5) and, thus, the fuzzy similarity method tends, on

average, to underestimate the component RUL at the beginning of its degradation trajectory. When the component is close to failure, the performance of the two methods becomes comparable since the incomplete run-to-failure trajectories are not informative of the component behavior just before the failure. In particular, the AI of the fuzzy similarity-based method is more satisfactory than that of the proposed method.

In practice, this result suggests a scheme for RUL prediction, in which the proposed approach can be used to provide the RUL estimate for the first 80% portion of the capacitor life ( $\sim 1.6 \times 10^4$  hrs average capacitor life) and, then, the fuzzy similarity-based approach can be utilized to provide the RUL estimates in the remaining 20%.

Moreover, Figure 12 (bottom) shows that the fuzzy similarity-based approach does not quantify the uncertainty on the RUL predictions. On the contrary, the width of the prediction interval provided by the proposed approach decreases with the equipment life, but it is large, which explains the satisfactory coverage value, i.e., larger than the desired value of 0.8 (horizontal dotted line). In fact, the large width of the prediction interval is due to the large variability of the fleet transitions and of the temperature conditions experienced by the capacitors during the early stages of their lives. With respect to the capability of the fuzzy similarity-based method to estimate a confidence interval, it is worth mentioning that in [43] the considered fuzzy similarity-based approach has been combined with a Belief Function Theory (also called Dempster–Shafer or evidence theory [80]) for providing RUL predictions with the associated uncertainty. Other similarity-based approaches for prognostics have been developed and their capabilities of quantifying the uncertainty affecting the RUL predictions have been demonstrated [30]. The investigation of these approaches is outside the scope of the present work.

To verify the capability of the proposed approach to provide accurate RUL predictions, in terms of RMSE, AI and  $\alpha - \lambda$  accuracy, with proper representation of the uncertainty, i.e., PI and CR, even in case of limited number of complete run-to-failure trajectories, a sensitivity test has been designed.

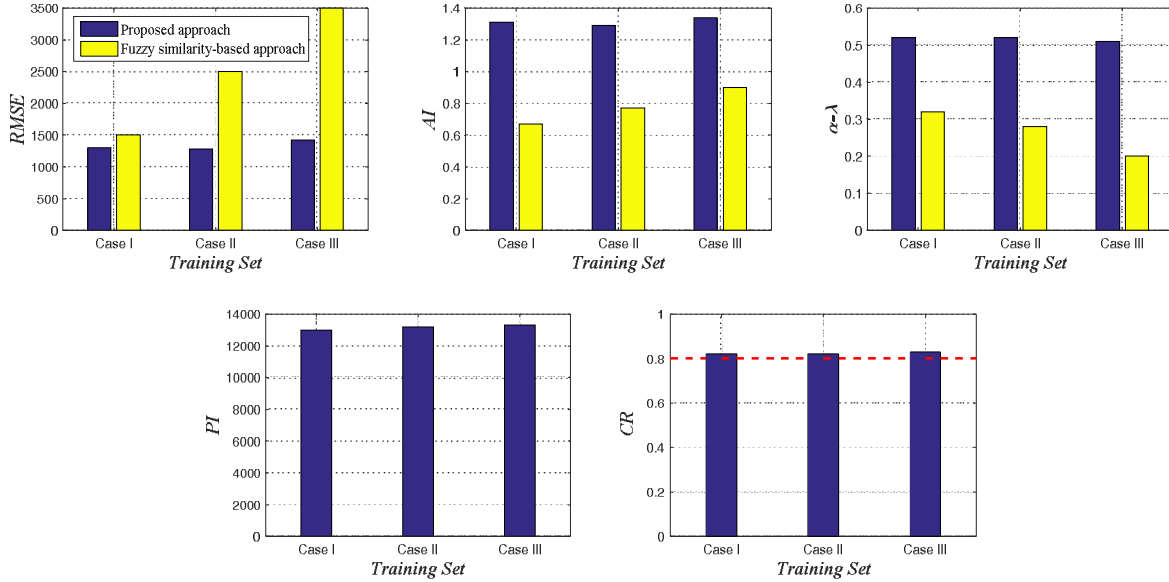
With this aim, the  $P_{training} = 100$  trajectories have been divided into different combinations of complete ( $P_{training}^c$ ) and incomplete ( $P_{training}^{ic}$ ) run-to-failure trajectories, for training the proposed approach and the fuzzy similarity-based approach, as reported in Table 1:

	$P_{training}^c$	$P_{training}^{ic}$
<b>Case I (Subsection 3.1)</b>	20	80
<b>Case II</b>	10	90
<b>Case III</b>	5	95

**Table 1: Different combinations of complete and incomplete run-to-failure training trajectories.**

Figure 13 shows the average values of the considered five performance metrics over the  $P_{test} = 50$  test trajectories. It can be seen that: 1) the proposed approach consistently outperforms the fuzzy similarity-based approach, in terms of RMSE and  $\alpha - \lambda$  accuracy, which confirms the benefits from utilizing the incomplete run-to-failure trajectories  $P_{training}^{ic}$ . The benefits are most remarkable when complete run-to-failure trajectories

are very limited, i.e., Case III, and becomes marginal, e.g., for the RMSE, when abundant complete run-to-failure trajectories are available, i.e., Case I (Figure 13 (top)). This result shows the capability of the proposed approach of accurately predicting the RUL even in cases of a limited number of complete run-to-failure trajectories. Moreover, 2) the prediction interval width tends to be constant and assures the desired coverage level of 0.8, in all three cases (Figure 13 (bottom)).



**Figure 13: Comparison of the five performance metrics for the proposed approach and the fuzzy similarity-based approach.**

As last remark, Figure 13 (middle top) shows that the AI of the proposed approach tends to be constant over the three cases. Conversely, the AI provided by the fuzzy similarity-based approach tends to increase as the number of complete run-to-failure trajectories decreases. This is due to the fact that the fuzzy similarity-based approach relies solely on the availability of complete run-to-failure trajectories.

#### 4. Case Study II: Aircraft engine turbo machinery

In this Section, the capability of the proposed approach of accurately estimating the equipment RUL is shown and compared with the fuzzy similarity-based approach, on a case of multi-dimensional data regarding an heterogeneous fleet of  $P = 260$  complete run-to-failure trajectories generated under different operational conditions of turbofan engine degradation and made available in NASA Ames Prognostics Data Repository [45].

##### 4.1 Data description

Each degradation trajectory is 24-dimensional (i.e.,  $Z = 24$  signals): 21 signals representative of the engine degradation behaviour and 3 signals representative of the engine operating conditions, which are contaminated with noise. The degradation trajectories are complete run-to-failure simulations that have been carried out

using C-MAPSS, a well-known simulation program for transient operation of modern commercial turbofan engines [45].

It is worth mentioning that among the 21 engine degradation behaviour signals, few signals show binary or constant values during the engines lives and, hence, they are excluded from the analysis focusing only on 15 signals indexed by 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 20, and 21 [30]. For more details on the engine run-to-failure simulation, interested readers may refer to [81].

Moreover, it is worth mentioning that the degradation trajectories have been simulated under variable operational conditions, including altitude, Mach number and throttle resolver angle (thrust setting) that have a substantial effect on the engines degradation process [82]. Additionally, the amount and rate of damage accumulation for each engine instance can be different, causing variable engine life, depending on various factors carried out within the C-MAPSS simulations [81], [82]. For these reasons, the population of the engines can be considered as an heterogeneous fleet.

To proceed with the analysis, as described in Section 2, the dataset is divided into: 1)  $P_{training} = 50$  trajectories used to train the proposed approach and the fuzzy similarity-based approach, 2)  $P_{test} = 150$  trajectories used to verify the performance of the proposed approach with respect to the fuzzy similarity-based approach and 3) the remaining trajectories used to optimize the values of the fuzzy similarity-based approach parameters, i.e., the parameters of the bell-shaped function [22].

For proper utilization of the dataset for prognostics, the following pre-processes have been done [83]:

- 1) For each run-to-failure trajectory, we only consider the data upon degradation onset and neglect the non-degraded data, as proposed in [83];
- 2) an exponential filter approach, whose parameter value is set by following a trial and error procedure is adopted to smooth local fluctuations degradation trajectory due to disturbance or noise, as proposed in [82], [83];
- 3) the engines degradation behaviour signals are collected under variable operational conditions and exhibit no prominent trend along the engines' lives, and hence, the data are normalized by resorting to the multi-regime health assessment method proposed in [82].

Figure 14 shows few examples of the 15-dimensional trajectories of the pre-processed signals data.

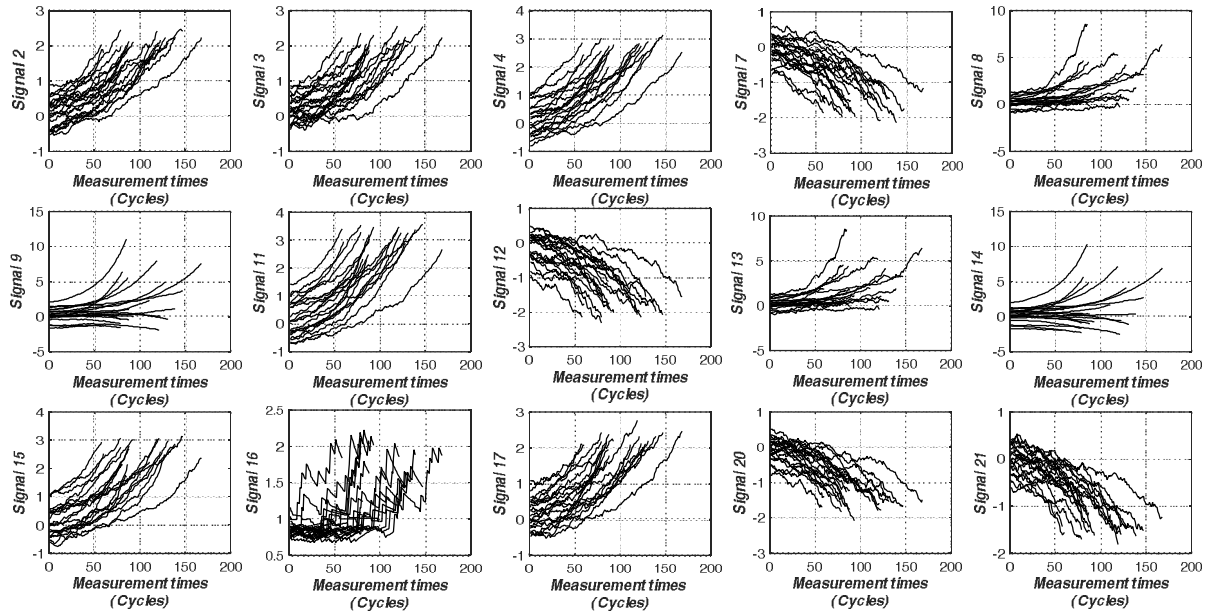


Figure 14: Examples of pre-processed signals data.

All the measurements of the complete run-to-failure degradation trajectories are stored in the dataset matrix  $\bar{X}$ , that is used in Sections 4.2 and 4.3 for the purpose of comparing the performance of the proposed approach with the fuzzy similarity-based approach. For computational convenience, 5 time steps between two successive measurements (i.e.,  $M = 5$ ) are considered.

#### 4.2 Application of the proposed data-driven prognostics approach

Figure 14 shows that the available signals have different trends (decreasing or increasing). This has suggested us to identify the number of states of the HDTFSSM model by grouping the signals based on their correlation, i.e., groups containing correlated signals together [55]. In this regard,  $H = 3$  groups are identified, namely group 1 (signals 8, 9, 13, 14), group 2 (signals 2, 3, 4, 11, 15, 16, 17) and group 3 (signals 7, 12, 20, 21). For clarification purposes, Figure 15 shows examples of one signal from each group: Figure 15 (left) shows signal 8 of group 1 characterized by a slightly increasing trend, whereas Figure 15 (middle) shows signal 2 of group 2 characterized by a rapid increasing trend and Figure 15 (right) shows signal 7 of group 3 characterized by a rapid decreasing trend.

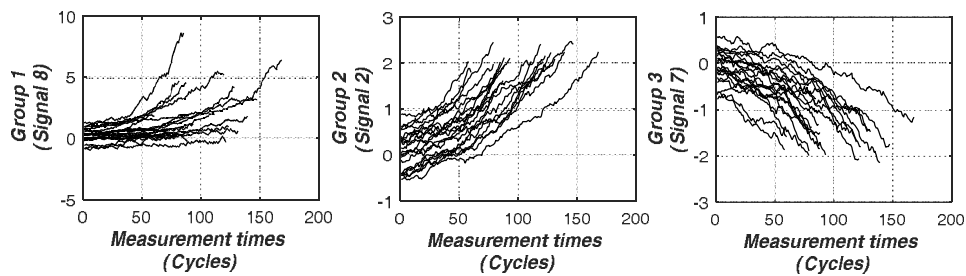


Figure 15: Examples of one signal of the three groups obtained by the correlation criteria.

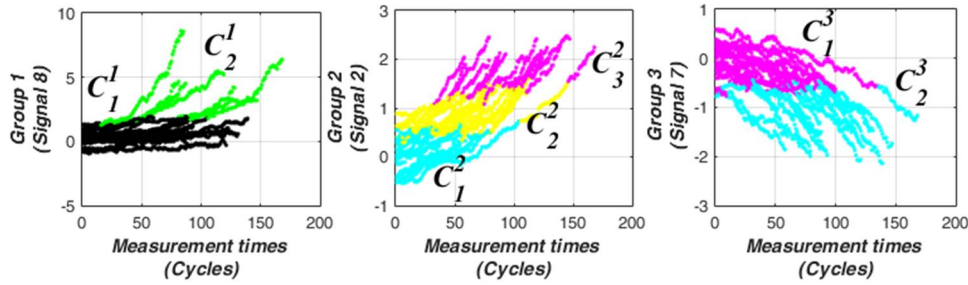
**Phase 1:** The states of the HDTFSSMM describing the engine degradation progression are identified by resorting to the unsupervised ensemble clustering approach [37], with  $H = 3$  base clusterings ( $j = 1$  for the engine degradation behaviour obtained by the signals of group 1,  $j = 2$  for the engine degradation behaviour obtained by the signals of group 2 and  $j = 3$  for the engine degradation behaviour obtained by the signals of group 3).

Table 2 reports the number of base clusters obtained by the unsupervised FCM [73] and DB validity criterion [84] along with the optimum number of clusters obtained in the final consensus clustering  $S$ .

$j=1$ Base clustering	$j=2$ Base clustering	$j=3$ Base clustering	Consensus clustering
$C_{opt}^1 = 2$	$C_{opt}^2 = 3$	$C_{opt}^3 = 2$	$S = 8$

**Table 2: Number of base clusters along with the optimum number of clusters in the final consensus clustering.**

For ease of clarity, Figure 16 (left) shows the two different clusters obtained by the  $j = 1$  base clustering for signal 8 (group 1), whereas Figure 16 (middle) shows the three different clusters obtained by the  $j = 2$  base clustering for signal 2 (group 2) and Figure 16 (right) shows the two different clusters obtained by the  $j = 3$  base clustering for signal 7 (group 3).



**Figure 16: Examples of base clustering one signal from the three groups obtained by the correlation criteria.**

The optimum number of states that guarantees a comprehensive representation of the possible degradation is given by the number  $S$  of final consensus clusters that is selected according to the Silhouette values [61] for different numbers of clusters  $C_{candidate}$  that span in the interval [2,12]. The optimum number  $S$  of states is found to be  $S = 8$ , at which the Silhouette measure is maximized [61].

This result leads to the transition diagram with  $S_{final} = 9$  states, where the 9<sup>th</sup> state is the failure state (i.e., an absorbing state) at which the engine is considered failed.

**Phase 2:** Resorting to the MLE and to the FIM [41], the estimated parameters  $\hat{\theta}$  governing the transitions among the degradation states and the 95% confidence intervals on the estimated parameters  $\hat{\theta}$  are obtained, respectively. Figure 17 shows the estimated parameters  $\hat{\theta}$  of the possible (dots) and the non-possible transitions (circles) from state 1 ( $q$  and  $\beta$ , top and bottom Figure, respectively).

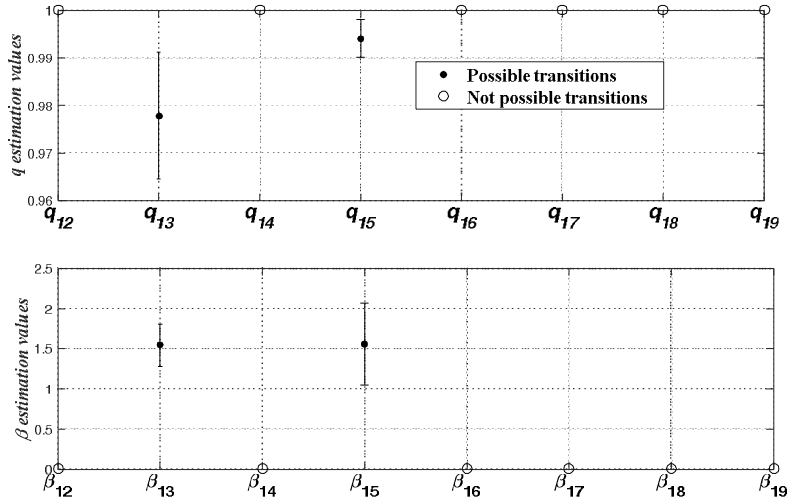


Figure 17: Estimated parameters values and their confidence intervals of the transitions occurred from state 1.

To verify that the developed model is representative of a monotonic degradation process, the monotonicity PPI of Eq. (2) is calculated for a sequence of 1000 degradation paths that were simulated by a direct MC with the estimated  $\hat{\theta}$  and, then, compared with the monotonicity of the  $P_{training}$  degradation trajectories of the  $H = 3$  groups (Figure 15).

Without loss of generality, Figure 18 shows one simulated engine degradation path: it is clear that it monotonically degrades through different degradation levels.

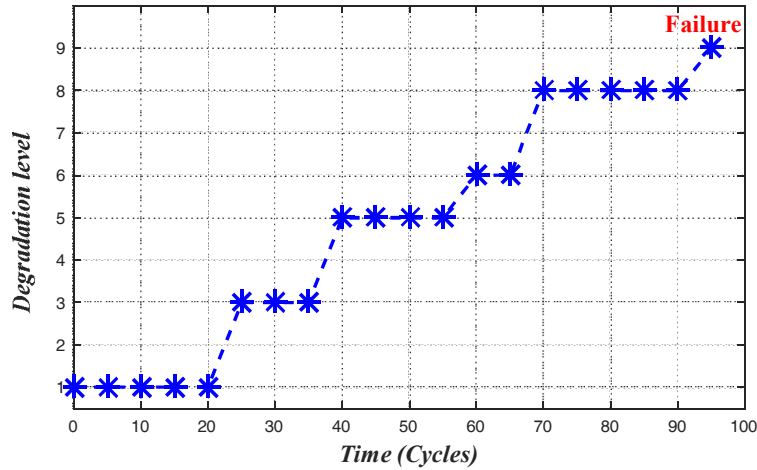


Figure 18: Sample realization of the degradation process of an engine.

The monotonicity PPI turns out to be equal to 0.996, 0.4, 0.61 and 0.5 for the simulated degradation paths, group 1, group 2 and group 3, respectively. We can, thus, conclude that the built model degradation is representative of a monotonic degradation process and, thus, it can be effectively used for the purpose of RUL estimation.

**Phase 3:** The online estimation of the RUL for  $P_{test} = 150$  new engines is performed by resorting to the direct MC simulation. In order to obtain a good compromise between accuracy and computational time,  $N_{max} = 1000$



MC trials have been used in the simulation. Figure 19 shows the mean RUL predictions (squares) and the associated 10<sup>th</sup> and 90<sup>th</sup> percentiles (triangles) for one engine.

It can be seen that 1) the estimated RUL is reasonably close to the actual remaining life, and 2) as the engine ages, the prediction intervals shrink toward the true RUL, that is, the prediction uncertainty decreases.

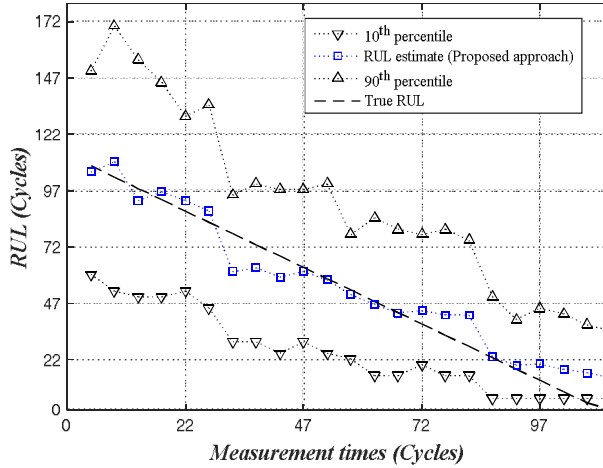


Figure 189: RUL prediction and corresponding 10<sup>th</sup> and 90<sup>th</sup> percentiles of one engine.

### 4.3 Comparison with the fuzzy similarity-based approach

In this Section, the fuzzy similarity-based approach [22] has been applied to the turbofan engines case study and the obtained results are compared with those obtained by the proposed approach.

Table 3 reports the average values of the considered five performance metrics over the  $P_{test} = 150$  test trajectories. The comparison results show that the proposed approach is more satisfactory, since it provides more accurate RUL predictions (lower RMSE and AI and larger  $\alpha - \lambda$  accuracy values) and it assures the desired coverage level of 0.8. In fact, increasing the number of the signals impacts the similarity evaluation between the test trajectory and the training trajectories and, thus, influences the capability of the fuzzy similarity-based approach of providing accurate estimations of the RUL.

	Proposed approach	Fuzzy similarity-based approach
RMSE	18.87	21.96
CR	0.82	NA
PI	57.69	NA
$\alpha$ - $\lambda$ accuracy ( $\alpha=10\%$ , $\lambda=50\%$ )	0.41	0.15
AI	0.36	0.49

Table 3: Average value of the performance metrics for the  $P_{test}=150$  test trajectories obtained by the proposed approach and the fuzzy similarity-based approach.

## 5. Conclusions

The availability of condition monitoring data of heterogeneous fleets of equipment installed worldwide and experiencing different operational conditions motivates the development of new data-driven approaches to capitalize the information coming from the fleet for improving the RUL estimation.

In this work, we propose an approach of fleet prognostics that consists in building an homogeneous discrete-time finite-state semi-markov model, whose states are the degradation levels that the equipment can experience throughout life and that are identified by resorting to an unsupervised ensemble clustering approach. The parameters of the discrete Weibull distributions governing the transitions among the states and their uncertainties are inferred by resorting to the Maximum Likelihood Estimation (MLE) and to the Fisher Information Matrix (FIM), respectively. The inferred degradation model is embedded within a direct Monte Carlo (MC) simulation scheme to estimate the RUL of new coming equipment.

The proposed approach has been shown superior to the fuzzy similarity-based approach, when applied to two case studies regarding i) an heterogeneous fleet of aluminum electrolytic capacitors used in electric vehicles powertrains and ii) an heterogeneous fleet of turbofan engines used in aircraft industry, both working under variable operating conditions. In the former case, the capability of the proposed approach in benefiting from the availability of also incomplete run-to-failure degradation trajectories for enhancing the RUL estimation, rather than relying solely on the complete run-to-failure trajectories is verified, whereas in the latter case the capability of the proposed approach in accurately estimating the RUL when dealing with multi-dimensional data is verified.

Moreover, the proposed approach has been able to provide informative and transparent outcomes by supplying a prediction of the current health state of the equipment and its RUL, and to quantify the uncertainty affecting the equipment RUL predictions due to the variable operating conditions experienced by the fleet equipment during their lives.

## Acknowledgements

The participation of Sameer Al-Dahidi and Piero Baraldi is supported by the European Union Project INNOVATION through Human Factors in risk analysis and management (INNHF, [www.innhf.eu](http://www.innhf.eu)) funded by the 7<sup>th</sup> framework program FP7-PEOPLE-2011-Initial Training Network: Marie-Curie Action. The participation of Enrico Zio to this research is partially supported by the China NSFC under grant number 71231001.

## References

- [1] E. Zio, "Prognostics and Health Management of Industrial Equipment," *Diagnostics Progn. Eng. Syst. Methods Tech.*, pp. 333–356, 2012.
- [2] D. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot, "Hidden Markov models for failure diagnostic and prognostic," *Progn. Heal. Manag. Conf.*, pp. 1–8, 2011.
- [3] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.
- [4] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges

- and opportunities,” *Mech. Syst. Signal Process.*, vol. 23, no. 3, pp. 724–739, 2009.
- [5] M. Bevilacqua and M. Braglia, “Analytic hierarchy process applied to maintenance strategy selection,” *Reliab. Eng. Syst. Saf.*, vol. 70, no. 1, pp. 71–83, 2000.
- [6] G. Medina-Oliva, A. Voisin, M. Monnin, and J.-B. Leger, “Predictive diagnosis based on a fleet-wide ontology approach,” *Knowledge-Based Syst.*, vol. 68, pp. 40–57, 2014.
- [7] K. Le Son, M. Fouladirad, A. Barros, E. Levrat, and B. Iung, “Remaining useful life estimation based on stochastic deterioration models: A comparative study,” *Reliab. Eng. Syst. Saf.*, vol. 112, pp. 165–175, Apr. 2013.
- [8] C. Sankavaram, B. Pattipati, A. Kodali, K. Pattipati, M. Azam, S. Kumar, and M. Pecht, “Model-based and data-driven prognosis of automotive and electronic systems,” in *2009 IEEE International Conference on Automation Science and Engineering*, 2009, pp. 96–101.
- [9] L. H. Chiang, E. Russell, and R. D. Braatz, *Fault detection and diagnosis in industrial systems*, vol. 12, no. 10–11. 2001.
- [10] M. Schwabacher and K. Goebel, “A survey of artificial intelligence for prognostics,” *Assoc. Adv. Artif. Intell. AAAI Fall Symp. 2007*, pp. 107–114, 2007.
- [11] X. S. Si, W. Wang, C. H. Hu, and D. H. Zhou, “Remaining useful life estimation - A review on the statistical data driven approaches,” *Eur. J. Oper. Res.*, vol. 213, no. 1, pp. 1–14, 2011.
- [12] J. Z. Sikorska, M. Hodkiewicz, and L. Ma, “Prognostic modelling options for remaining useful life estimation by industry,” *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1803–1836, 2011.
- [13] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, “Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications,” *Mech. Syst. Signal Process.*, vol. 42, no. 1–2, pp. 314–334, 2014.
- [14] J. Liu, W. Wang, F. Ma, Y. B. Yang, and C. S. Yang, “A data-model-fusion prognostic framework for dynamic system state forecasting,” *Eng. Appl. Artif. Intell.*, vol. 25, no. 4, pp. 814–823, 2012.
- [15] P. Baraldi, F. Cadini, F. Mangili, and E. Zio, “Model-based and data-driven prognostics under different available information,” *Probabilistic Eng. Mech.*, vol. 32, no. 0, pp. 66–79, 2013.
- [16] Y. Li, T. R. Kurfess, and S. Y. Liang, “Stochastic prognostics for rolling element bearings,” *Mech. Syst. Signal Process.*, vol. 14, no. 5, pp. 747–762, 2000.
- [17] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk, and S. Liang, “Adaptive prognostics for rolling element bearing condition,” *Mech. Syst. Signal Process.*, vol. 13, no. 1, pp. 103–113, 1999.
- [18] C. Oppenheimer and K. Loparo, “Physically based diagnosis and prognosis of cracked rotor shafts,” *AeroSense 2002*, vol. 4733, pp. 122–132, 2002.
- [19] F. Di Maio, K. L. Tsui, and E. Zio, “Combining Relevance Vector Machines and exponential regression for bearing residual life estimation,” *Mech. Syst. Signal Process.*, vol. 31, pp. 405–427, Aug. 2012.
- [20] F. Cadini, E. Zio, and D. Avram, “Model-based Monte Carlo state estimation for condition-based component replacement,” *Reliab. Eng. Syst. Saf.*, vol. 94, no. 3, pp. 752–758, Mar. 2009.
- [21] P. C. Paris and F. Erdogan, “A Critical Analysis of Crack Propagation Laws,” *Journal of Basic Engineering*, vol. 85, no. 4, p. 528, 1963.
- [22] F. Di Maio and E. Zio, “Failure Prognostics By A Data-Driven Similarity-Based Approach,” *Int. J. Reliab. Qual. Saf. Eng.*, vol. 20, no. 01, p. 1350001, 2014.
- [23] V. T. Tran and B.-S. Yang, “Machine Fault Diagnosis and Prognosis : The State of The Art,” *Int. J. Fluid Mach. Syst.*, vol. 2, no. 1, pp. 61–71, 2009.
- [24] W. Li and H. Pham, “An Inspection-Maintenance Model for Systems With Multiple Competing Processes,” *IEEE Trans. Reliab.*, vol. 54, no. 2, pp. 318–327, Jun. 2005.
- [25] P. Campolucci, A. Uncini, F. Piazza, and B. D. Rao, “On-line Learning Algorithms for Locally Recurrent Neural Networks,” *IEEE Trans. Neural Netw.*, vol. 10, no. 2, pp. 253–271, 1999.
- [26] W. Q. Wang, M. F. Golnaraghi, and F. Ismail, “Prognosis of machine health condition using neuro-fuzzy systems,” *Mech. Syst. Signal Process.*, vol. 18, no. 4, pp. 813–831, Jul. 2004.
- [27] V. A. Sotiris and M. Pecht, “Support Vector Prognostics Analysis of Electronic Products and Systems,” in *Artificial Intelligence for Prognostics*, 2007, pp. 120–127.
- [28] H. Zhang, C. Hu, X. Kong, and W. Zhang, “A Model For Residual Life Prediction Based on Brownian Motion in Framework of Similarity,” *Asian J. Control*, vol. 18, no. 4, pp. 1–11, 2016.
- [29] E. Zio, F. Di Maio, and M. Stasi, “A data-driven approach for predicting failure scenarios in nuclear systems,” *Ann. Nucl. Energy*, vol. 37, no. 4, pp. 482–491, 2010.
- [30] T. Wang, J. Yu, D. Siegel, and J. Lee, “A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems,” in *PHM. International Conference*, 2008, pp. 1–6.
- [31] M. Monnin, B. Abichou, A. Voisin, and C. Mozzati, “Fleet historical cases for predictive maintenance,” *The International Conference Surveillance 6, Compiègne*.
- [32] R. Patrick, M. J. Smith, C. S. Byington, G. J. Vachtsevanos, K. Tom, and C. Ly, “Integrated software platform for fleet data analysis, enhanced diagnostics, and safe transition to prognostics for helicopter component CBM,” *Proceedings of Annual Conference of the Prognostics and Health Management Society*.

- [33] A. Reymonet, J. Thomas, and N. Aussenac-Gilles, "Ontology based information retrieval: An application to automotive diagnosis," *Proceedings of International Workshop on Principles of Diagnosis*, pp. 9–14, 2009.
- [34] P. Umiliacchi, D. Lane, and F. Romano, "Predictive Maintenance of railway subsystems using an Ontology based modelling approach," in *Proceedings of 9th world Conference on Railway Research, Lille, France*.
- [35] S. Al-Dahidi, F. Di Maio, P. Baraldi, E. Zio, and R. Seraoui, "A Framework for Reconciling the Consensus Clusterings of a Fleet of Nuclear Power Plants Turbines for Fault Diagnosis," *Appl. Soft Comput.*, p. Under Review, 2015.
- [36] V. Agarwal, N. J. Lybeck, R. Bickford, and R. Rusaw, "Development of Asset Fault Signatures for Prognostic and Health Management in the Nuclear Industry," *PHM, IEEE Conf.*, 2014.
- [37] S. Al-Dahidi, F. Di Maio, P. Baraldi, E. Zio, and R. Seraoui, "A Novel Ensemble Clustering for Operational Transients Classification with Application to a Nuclear Power Plant Turbine," *Int. J. Progn. Heal. Manag.*, vol. 6, no. SP3, pp. 1–21, 2015.
- [38] M. A. Jazi, C. D. Lai, and M. H. Alamatsaz, "A discrete inverse Weibull distribution and estimation of its parameters," *Stat. Methodol.*, vol. 7, no. 2, pp. 121–132, 2010.
- [39] T. Nakagawa and S. Osaki, "A New Discrete Weibull Distribution," *Ieee Transactions On Reliability*, vol. R-24, no. 2, pp. 196–197, 1975.
- [40] H. Rinne, *The Weibull distribution: a handbook*. CRC Press, 2008.
- [41] M. Kendall and Stuart A., *The Advanced Theory of Statistics*, vol. 2. Charles Griffin and Company limited, London & High Wycombe, 1968.
- [42] F. Cadini, E. Zio, and D. Avram, "Monte Carlo-based filtering for fatigue crack growth estimation," *Probabilistic Eng. Mech.*, vol. 24, no. 3, pp. 367–373, Jul. 2009.
- [43] P. Baraldi, F. Mangili, and E. Zio, "A belief function theory based approach to combining different representation of uncertainty in prognostics," *Inf. Sci. (Ny)*, vol. 303, pp. 134–149, May 2015.
- [44] M. Rigamonti, P. Baraldi, E. Zio, D. Astigarraga, and A. Galarza, "Particle Filter-Based Prognostics for an Electrolytic Capacitor Working in Variable Operating Conditions," *IEEE Trans. Power Electron.*, vol. 31, no. 2, pp. 1567–1575, 2016.
- [45] A. Saxena and K. Goebel, "C-MAPSS Data Set." NASA Ames Prognostics Data Repository, Moffett Field, CA., 2008.
- [46] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, "Metrics for Offline Evaluation of Prognostic Performance," *Int. J. Progn. Heal. Manag.*, no. 1, pp. 1–20, 2010.
- [47] M. Dong and Y. Peng, "Equipment PHM using non-stationary segmental hidden semi-Markov model," *Robot. Comput. Integr. Manuf.*, vol. 27, pp. 581–590, 2011.
- [48] C. T. Chen, "Dynamic preventive maintenance strategy for an aging and deteriorating production system," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 6287–6293, 2011.
- [49] R. Moghaddass and M. J. Zuo, "A parameter estimation method for a condition-monitored device under multi-state deterioration," *Reliab. Eng. Syst. Saf.*, vol. 106, pp. 94–103, 2012.
- [50] M. H. Shu, B. M. Hsu, and K. C. Kapur., "Dynamic performance measures for tools with multi-state wear processes and their applications for tool design and selection," *Int. J. Prod. Res.*, vol. 48, no. 16, pp. pp. 4725–4744, 2010.
- [51] S. M. Ross, *Introduction to probability models*. Academic Press Ltd, 2014.
- [52] H. Thorisson, "The coupling of regenerative processes," *Adv. Appl. Probab.*, vol. 15, no. 3, pp. 531–561, 1983.
- [53] R. Moghaddass and M. J. Zuo, "An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process," *Reliab. Eng. Syst. Saf.*, vol. 124, pp. 92–104, 2014.
- [54] S. Al-Dahidi, F. Di Maio, P. Baraldi, and E. Zio, "Supporting Maintenance Decision with Empirical Models Based on Fleet-Wide Data," in *The 49th ESReDA Seminar on: Innovation through Human Factors in Risk Assessment & Maintenance, Belgium*, 2015, pp. 1–12.
- [55] P. Baraldi, R. Canesi, E. Zio, R. Seraoui, and R. Chevalier, "Genetic algorithm-based wrapper approach for grouping condition monitoring signals of nuclear power plant components," *Integr. Comput. Aided. Eng.*, vol. 18, no. 3, pp. 221–234, 2011.
- [56] A. Strehl and J. Ghosh, "Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions," *Journal of Machine Learning Research*, vol. 3, pp. 583–617, 2002.
- [57] U. von Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, Aug. 2007.
- [58] P. Baraldi, F. Di Maio, and E. Zio, "Unsupervised Clustering for Fault Diagnosis in Nuclear Power Plant Components," *Int. J. Comput. Intell. Syst.*, vol. 6, no. 4, pp. 764–777, Aug. 2013.
- [59] M. Su and C. Chou, "A modified version of the K-means algorithm with a distance based on cluster symmetry," *Pattern Anal. Mach. Intell. ....*, vol. 23, no. 6, pp. 674–680, 2001.
- [60] X. Z. Fern and W. Lin, "Cluster ensemble selection," *Statistical Analysis and Data Mining*, vol. 1, no. 3, pp. 787–797, 2008.
- [61] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. C, pp. 53–65, Nov. 1987.

- [62] T. Nakagawa and S. Osaki, "The Discrete Weibull Distribution Applications of the Sojourn - Time Problem to Reliability," vol. 5, no. 5, pp. 300–301, 1975.
- [63] A. A. S. Al-wakeel, "On discrete weibull distribution," pp. 1–9.
- [64] E. Mathieu, Y. Foucher, P. Dellamonica, and J. P. Daures, "Parametric and Non Homogeneous Semi-Markov Process for HIV Control," *Methodol Comput Appl Probab*, vol. 9, pp. 389–397, 2007.
- [65] J. Chen, G. M. Hoversten, D. Vasco, Y. Rubin, and Z. Hou, "A Bayesian model for gas saturation estimation using marine seismic AVA and CSEM data," *Geophysics*, vol. 72, no. 2, p. WA85, 2007.
- [66] M. Compare, P. Baraldi, F. Cannarile, F. Di Maio, and E. Zio, "K-Nearest Neighbour classification and homogeneous finite-state, continuous time, semi-Markov modelling for condition-based diagnostics of industrial components," *Probabilistic Eng. Mech.*, p. Under review, 2015.
- [67] S. Sankararaman and K. Goebel, "Why is the Remaining Useful Life Prediction Uncertain?," in *Annual Conference of the Prognostics and Health Management Society 2013*, 2013, pp. 1–13.
- [68] C. Kulkarni, G. Biswas, X. Koutsoukos, J. Celaya, and K. Goebel, "Integrated diagnostic/prognostic experimental setup for capacitor degradation and health monitoring," in *AUTOTESTCON (Proceedings)*, 2010, pp. 351–357.
- [69] P. Venet, H. Darnand, and G. Grellet, "Detection of faults of filter capacitors in a converter. Application to predictive maintenance," *Telecommunications Energy Conference, INTELEC '93. 15th International*, vol. 2, pp. 229–234 vol.2, 1993.
- [70] A. Lahyani, P. Venet, G. Grellet, and P.-J. Viverge, "Failure prediction of electrolytic capacitors during operation of a switchmode power supply," *Power Electronics, IEEE Transactions on*, vol. 13, no. 6, pp. 1199–1207, 1998.
- [71] M. L. Gasperi, "Life prediction model for aluminum electrolytic capacitors," *Ind. Appl. Conf. 1996. Thirty-First IAS Annu. Meet. IAS '96., Conf. Rec. 1996 IEEE*, vol. 3, no. 1, pp. 1347–1351 vol.3, 1996.
- [72] E. Wolfgang, "Examples for failures in power electronics systems." ECPE Tutorial on Reliability of Power Electronic Systems, Nuremberg, Germany.
- [73] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, vol. 25, no. 3, 1983.
- [74] D. L. Davies and D. W. Bouldin, "A cluster separation measure.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 1, no. 2, pp. 224–227, 1979.
- [75] J. Coble and J. W. Hines, "Identifying optimal prognostic parameters from data: a genetic algorithms approach," in *Proceedings of the Annual Conference of the Prognostics and Health Management Society*, 2009, pp. 1–11.
- [76] J. Coble and J. W. Hines, "Merging Data Sources to Predict Remaining Useful Life – An Automated Method to Identify Prognostic Parameters," University of Tennessee, 2010.
- [77] P. Baraldi, F. Di Maio, F. Mangili, and E. Zio, "A Belief Function Theory Method for Prognostics in Clogging Filters," in *Chemical Engineering Transactions*, 2013, vol. 33, pp. 847–852.
- [78] E. Zio and F. Di Maio, "A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system," *Reliab. Eng. Syst. Saf.*, vol. 95, no. 1, pp. 49–57, 2010.
- [79] L. Angstenberger, *Fuzzy Pattern Recognition*. Kluwer Academic Publishers, 2001.
- [80] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. July 1928, pp. 379–423, 1948.
- [81] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *2008 International Conference on Prognostics and Health Management*, 2008, pp. 1–9.
- [82] T. Wang, "Trajectory similarity based prediction for remaining useful life estimation," University of Cincinnati, 2010.
- [83] M. Rigamonti, P. Baraldi, F. Di Maio, and E. Zio, "Development of a Complete Framework for the Prediction of the RUL of a Turbofan Engine Working in Different Operating Conditions," in *Third European Conference of the PHM Society - PHME16*.
- [84] D. L. Davies and D. W. Bouldin, "A cluster separation measure.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, no. 2, pp. 224–227, 1979.

## Appendices

### *Appendix A.1: The Unsupervised Ensemble Clustering*

The unsupervised ensemble clustering approach is an ensemble approach composed by multiple base clusterings that can be different because of the different clustering algorithm used or because of the different data features extracted from the original collected dataset upon which (some) clustering algorithm is applied. The unsupervised ensemble clustering approach avoids the need of having an "a priori" knowledge of the

number of clusters  $M$  in the final consensus clustering [37]. The flowchart for the method is sketched in Figure 20. The method goes along the following steps:

**Step 1: Adjacency matrix computation.** An adjacency binary similarity matrix  $\bar{A}$ , is built by aggregating the similarities  $\mu$  of the  $H$  base clusterings [56], where for each  $j$ -th base clustering, the similarity  $\mu = 1$ , if two data belong to the same cluster, whereas the similarity  $\mu = 0$ , if they belong to different clusters.

**Step 2: Similarity matrix computation.** From the adjacency binary similarity matrix  $\bar{A}$ , the overall similarity matrix  $\bar{W}$  is computed as the entry-wise average of the  $H$  base clusterings, i.e.  $\bar{W} = (1/H)\bar{A}\bar{A}^T$  [56]. In this way, each entry of the similarity matrix has a value in  $[0,1]$ , which is proportional to how likely a pair of data is, when grouped together.

**Step 3: Spectral Clustering.** Once the overall similarity matrix  $\bar{W}$  is computed, Spectral Clustering is used to reveal the hidden structure of  $\bar{W}$ . The basic idea of Spectral Clustering is to extract the relevant information of the matrix  $\bar{W}$ , by considering the eigenvectors associated to the ascended eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_{C_{candidate}}, \dots, \lambda_N$  of the normalized laplacian matrix  $\bar{L}_{rs}$  of  $\bar{W}$ , to perform dimensionality reduction before clustering in fewer dimensions [57], [58]. The eigenvectors  $\bar{u}_1, \bar{u}_2, \dots, \bar{u}_{C_{candidate}}, \dots, \bar{u}_N$  of the eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_{C_{candidate}}, \dots, \lambda_N$  are calculated and stored in a matrix  $\bar{U}$  with a size  $N \times N$ , where  $C_{candidate} = [C_{min}, C_{max}]$  and  $C_{min}$  and  $C_{max}$  are the minimum and maximum numbers of clusters considered for the final consensus clustering, respectively.

**Step 4: Clustering algorithm.** For each candidate number of clusters  $C_{candidate}$ , the reduced matrix of  $\bar{U}$  with a size  $N \times C_{candidate}$  is partitioned into  $C_{candidate}$  clusters by using a single clustering algorithm and the final consensus clustering  $S_{C_{candidate}}$  is obtained. In this work, we resort to the  $K$ -means algorithm, one of the most used clustering methods, to partition  $\bar{U}$  into  $K = C_{candidate}$  clusters [59], [60].

**Step 5: Final consensus clustering selection.** For each  $C_{candidate}$ , the obtained consensus clustering  $S_{C_{candidate}}$  is evaluated by computing its Silhouette validity index  $SV_{C_{candidate}}$  [61]. The most appropriate consensus clustering  $S_{C^*}$  is the one for which the Silhouette reaches a maximum, for which clusters are well separated and compacted.

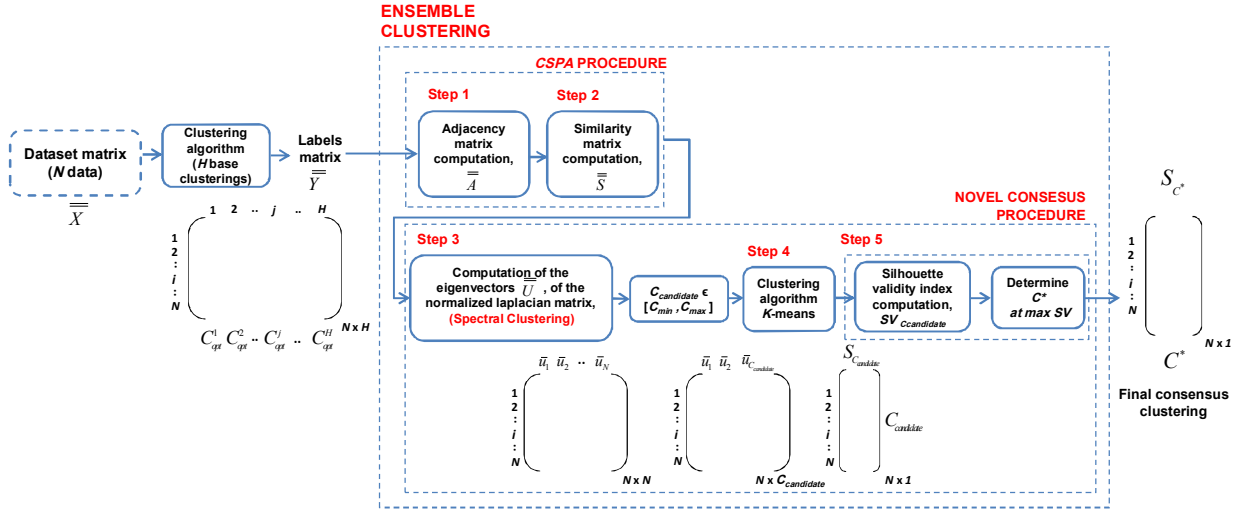


Figure 20: Flowchart of the ensemble clustering approach.

## Appendix A.2 Failure prognostics performance metrics

A failure prognostics approach, aims at estimating the equipment RUL, can be assessed from different perspective with various performance metrics. For instance, the following five metrics are adopted in this work, which are namely, the Root Mean Square Error (RMSE), Accuracy Index (AI), Precision Index (PI), Coverage Rate (CR) and  $\alpha - \lambda$  accuracy index [46]. In brief:

- the Root Mean Square Error (RMSE) index is defined as the average error of the RUL prediction within  $P_{test}$  pieces of equipment (Eq. (A.1)):

$$RMSE_p = \sqrt{\frac{\sum_{l=1}^{I_p} (\widehat{RUL}_p(t_l) - RUL_p(t_l))^2}{I_p}}, RMSE = \frac{\sum_{p=1}^{P_{test}} RMSE_p}{P_{test}} \quad (A.1)$$

where  $\widehat{RUL}_p(t_l)$  and  $RUL_p(t_l)$  are the estimated and the true remaining useful life of the  $p$ -th equipment at the measurement time  $t_l$ , where  $l = 1, \dots, I_p$ .  $RMSE_p$  and  $RMSE$  are the average root mean square error of the  $p$ -th equipment and of the  $P_{test}$  pieces of equipment, respectively. Small RMSE values indicate more accurate predictions.

- The Accuracy Index (AI) is defined as the relative error of the RUL prediction (Eq. (A.2)):

$$AI_p = \sum_{l=1}^{I_p-1} \frac{|\widehat{RUL}_p(t_l) - RUL_p(t_l)|}{RUL_p(t_l)}, AI = \frac{\sum_{p=1}^{P_{test}} AI_p}{P_{test}} \quad (A.2)$$

where  $AI_p$  and  $AI$  are the average accuracy index of the  $p$ -th equipment and of the  $P_{test}$  pieces of equipment, respectively. Small AI values indicate more accurate predictions.

- The Precision Index (PI) considers the width of the prediction interval, which is defined by (Eq. (A.3)):

$$PI_p = \frac{\sum_{l=1}^{I_p} w_{t_l}^p}{I_p}, PI = \frac{\sum_{p=1}^{P_{test}} PI_p}{P_{test}} \quad (A.3)$$

$$w_{t_l}^p = \sup(\widehat{RUL}_p(t_l)) - \inf(\widehat{RUL}_p(t_l))$$

where  $w_{t_l}^p$  is the width value of the  $p$ -th equipment at the measurement time  $t_l$ , where  $l = 1, \dots, I_p$ ,  $PI_p$  and  $PI$  are the average width value of the  $p$ -th equipment and of the  $P_{test}$  pieces of equipment. Small values of  $PI$  indicate more precise predictions.

- The Coverage Rate (CR) is the probability of the prediction interval cover the true value of RUL, larger value of it means the predicted interval has high chance to cover the true RUL (Eq. (A.4)):

$$CR_p = \frac{\sum_{l=1}^{I_p} c_{t_l}^p}{I_p}, CR = \frac{\sum_{p=1}^{P_{test}} CR_p}{P_{test}}$$

$$c_{t_l}^p = \begin{cases} 1 & \text{inf}(\widehat{RUL}_p(t_l)) < RUL_p(t_l) < \text{sup}(\widehat{RUL}_p(t_l)) \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.4})$$

where  $\text{inf}(\widehat{RUL}_p(t_l))$  and  $\text{sup}(\widehat{RUL}_p(t_l))$  are the upper and lower bounds of the 80% confidence interval of RUL prediction interval at the measurement time  $t_l$ , where  $l = 1, \dots, I_p$ ,  $c_{t_l}^p$  is the binary coverage value at the measurement time  $t_l$  of the  $p$ -th equipment,  $CR_p$  and  $CR$  are the average coverage rate of the  $p$ -th equipment and of the  $P_{test}$  pieces of equipment.

- The  $\alpha - \lambda$  accuracy is defined as a binary metric that evaluates whether the prediction accuracy at time  $t_\lambda$  (e.g.,  $\lambda = 0$  for the prediction starting from the beginning of life of the equipment) falls within specified  $\alpha$ -bounds (e.g., 10%). Large values of this index indicate more accurate predictions (Eq. (A.5)).

$$(\alpha - \lambda)_{t_\lambda} = \frac{\sum_{p=1}^{P_{test}} (\alpha - \lambda)_{t_\lambda}^p}{P_{test}}$$

$$(\alpha - \lambda)_{t_\lambda}^p = \begin{cases} 1 & (1 - \alpha) \cdot RUL_p(t_\lambda) < \widehat{RUL}_p(t_\lambda) < (1 + \alpha) \cdot RUL_p(t_\lambda) \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.5})$$

where  $(\alpha - \lambda)_{t_\lambda}^p$  is the  $\alpha - \lambda$  value at the measurement time  $t_\lambda$  of the  $p$ -th equipment,  $(\alpha - \lambda)_{t_\lambda}$  is the average  $\alpha - \lambda$  accuracy value of the  $P_{test}$  pieces of equipment at the measurement time  $t_\lambda$ . Large values of  $\alpha - \lambda$  indicate more precise predictions.