POLITECNICO DI MILANO

Dipartimento di Energia

Dottorato di Ricerca in

Scienze e Tecnologie Energetiche e Nucleari

DEVELOPMENT OF ADVANCED COMPUTATIONAL METHODS FOR PROGNOSTICS AND HEALTH MANAGEMENT IN ENERGY COMPONENTS AND SYSTEMS

Doctoral dissertation of:
Francesca MANGILI

Supervisor:
Prof. Enrico ZIO

Co-supervisors:
Dott. Piero BARALDI
Dott. Giulio GOLA
Dott. Bent Helge NYSTAD

Tutor:
Dott. Francesco DI MAIO

Coordinatore del Corso di Dottorato:
Prof. Carlo BOTTANI

2013 - XXV CICLO

Acknowledgments/Ringraziamenti

First of all, I would like to thank Fridtjov Øwre, Øivind Berg, Davide Roverso and Mario Hoffmann for the financial support provided by the Halden Reactor Project to this PhD work and the many opportunities offered me during these three years.

Desidero poi ringraziare il relatore di questa tesi, Prof. Enrico Zio, per avermi sapientemente guidata e motivata durante questo dottorato e aver mostrato di credere in me.

Sono fortemente in debito con Piero Baraldi, per la pazienza, la disponibilità con cui ha sempre seguito il mio lavoro, e per tutto quello che mi ha insegnato.

Un sincero ringraziamento va anche a Giulio Gola e Bent Nystad per il loro prezioso aiuto e le molte utili discussioni che abbiamo avuto durante i miei soggiorni in Norvegia.

Grazie anche agli amici del LASAR, per aver dato colore a tutto il tempo speso in dipartimento.

Non può mancare un grande grazie alla mia famiglia, che mi sostiene sempre, nonostante da qualche anno faccia fatica a capire cosa faccio di lavoro.

Infine, grazie a Stefano per tantissimi motivi, tra cui l'aver dato significato a ogni cosa e aver reso piacevoli anche i momenti più faticosi semplicemente con la sua presenza.

Contents

PART I

PART II

Report A[*] P. Baraldi, F. Mangili, E. Zio, G. Gola, B. Nystad. *An ensemble-based approach for process parameter estimation and health state assessment in offshore oil platforms*.. LASAR[†] and Halden Reactor Project Work Report, 2013.

Paper I P. Baraldi, F. Mangili, E. Zio, *A Kalman Filter-based Ensemble Approach for Turbine Creep Prognostics*. IEEE Transactions on Reliability 2012, Vol. 61(4), pp. 966-977.

Paper II P. Baraldi, F. Cadini, F. Mangili, E. Zio, *Model-Based and Data-Driven Prognostics under Different Available Information*. Accepted for publication on Probabilistic Engineering Mechanics, 2013.

Paper III P. Baraldi, F. Mangili, E. Zio, *Investigation of uncertainty treatment capability of model-based and data-driven prognostic methods using simulated data*. Reliability Engineering and System Safety 2013, Vol 112C, pp. 94-108.

Paper IV P. Baraldi, F. Di Maio, F. Mangili, E. Zio. *Equipment Uncertainty in Remaining Useful Life Prediction: a Belief Function Theory Treatment*. Ready for submission, 2013.

Paper V P. Baraldi, F. Mangili, E. Zio. *An approach to Remaining Useful Life prediction based on Gaussian Process Regression for degradation modeling*. Ready for submission, 2013.

Paper VI P. Baraldi, F. Mangil, E. Zio, G. Gola, B.H. Nystad. *An Approach to Combining Different Representation of Uncertainty in Data-Driven Prognostics Using Belief Function Theory*. Ready for submission, 2013.

---

[*] This report is being submitted as official Halden Work Report (HWR).
[†] LASAR (Laboratorio di Analisi di Segnale ed Analisi di Rischio): Laboratory of Signal Analysis and Risk Analysis of the Department of Nuclear Engineering of the Polytechnic of Milan.

Acronyms & Notation

AHP	Analytic Hierarchy Process
Cov	Coverage
BBA	Basic Belief Assignment
BDF	Belief Density Function
BFT	Belief Function Theory
CDF	Cumulative Density Function
KF	Kalman Filter
KR	Kernel Regression
GPR	Gaussian Process Regression
MAE	Mean Absolute Error
PDF	Probability Density Function
PF	Particle Filtering
rMAE	Mean relative absolute error
RUL	Remaining Useful Life
SBR	Similarity Based Regression
SIR	Sampling Importance Resampling
SRCC	Spearman's Rank Correlation Coefficient

ΔP	Pressure drop
$\dot{M}$	Mass flow rate
T	Sea water temperature
δ	Choke valve opening
q	Reference index of the observable parameters
Q	Number of observable parameters
r	Reference index of similar pieces of equipment
R	Number of similar pieces of equipment
$z_q^r(t_j)$	Value of parameter q measured for equipment r at time t_j
$\mathbf{z}_j^r$	Vector of parameters values $[z_1(t_j),...,z_q(t_j),...,z_Q(t_j)]$
$z_{\theta,j}^r$	Indication of the degradation state of equipment r at time t_j
Θ_j^r	Random variable representing the degradation state of equipment r at time t_j
θ_j^r	Degradation state of equipment r at time t_j
$\hat{\theta}_j^r$:	Predicted degradation state of equipment r at time t_j
θ_{th} :	Failure threshold
t_j	Discretized time instants
t_J	Present time
t_F	Failure time
RUL_j^r :	Random variable representing the RUL of equipment r at time t_j
rul_j^r :	RUL of equipment r at time t_j

$\hat{rul}_j^r$:	Prediction of the RUL of equipment r at time t_j
$CI_J(\alpha)$	Prediction interval $[rul_J^{\inf}(\alpha), rul_J^{\sup}(\alpha)]$ containing the RUL value with probability $\geq \alpha$
e_j^r	RUL prediction error
$p_X(x)$	Probability density function of the random variable X
$E[X \mid \Im]$	Expected value of X conditioned on the information $\Im$
σ_X^2	Variance of the random variable X
$\sigma_{a(b,c)}^2$	Component *a*, *b* or *c* of the RUL prediction error variance
g	Physics-based degradation model
u	Observation model
η	Data-driven degradation model
f	Prognostic model
s_j^r:	Measure of similarity between trajectories
m_X	BBA for the uncertain variable X
Ω_X	Domain of X
$\mathrm{Bel}_m(Y)$	Belief assigned by the BBA m_X to a set Y
$\mathrm{Pl}_m(Y)$	Belief assigned by the BBA m_X to a set Y
γ, λ	Trust and similarity parameters in the SBR approach
$\mu_\Theta(t)$	Prior mean of the Gaussian process in the GPR approach
$C_\Theta(t,t')$	Prior covariance function in the GPR approach
b_X	BDF for the uncertain variable X
A	Transition matrix in the KF ensemble approach
H	Observation matrix in the KF ensemble approach
R	Observation noise covariance matrix in the KF ensemble approach
q	Process noise variance in the KF ensemble approach
ρ_j	Validation residuals in the bootstrapped ensemble approach
χ	Model of the process and noise prediction error variance components in the bootstrapped ensemble approach
ε	Turbine blade creep strain

1 Introduction

For industry, unforeseen equipment failure is costly, both for repair and lost revenue. The unexpected discovery in March 2002 of extensive corrosion in the reactor vessel head—one of the vital barriers preventing a radioactive release—at the Davis-Bessel nuclear power plant in Ohio led to a more than 2 years outage and to estimated repair costs exceeding $600 millions [1][3]. The failure of rotating equipment causes to thermal power industry millions of dollars of unrecovered costs in lost energy production, plus additional millions of dollars paid in replacement equipment costs paid by casualty insurers [2]. Many other examples of costly unanticipated failure could easily be found, e.g., in the fields of oil & gas, aerospace, information technology for business-critical applications, etc. The maintenance strategies that are set up to face this problem traditionally fall into two categories: (i) preventive maintenance and (ii) corrective maintenance. The former is performed on a time-based schedule regardless of the actual condition of the equipment and may result in unnecessary maintenance; the latter avoids any unnecessary maintenance by only repairing equipment which has already failed, but it may take long times and result in significant lost revenue [3]. In recent times, Prognostics and Health Management (PHM) systems have gained more and more attention among researchers and in industry. The PHM community has envisioned the possibility of optimizing the time at which performing maintenance actions by knowing the actual state of health of the equipment and predicting its future evolution [4-5]. A successful application of PHM has the potential of increasing systems availability and safety, by keeping maintenance costs reasonably low [6-7]. In general terms, the tasks of PHM are [8-9]: (1) monitoring the equipment and detecting the presence of anomalies (detection); (2) assessing the health state of the system (diagnostics); (3) predicting the equipment Remaining Useful Life (RUL), i.e., the amount of time it will continue to perform its function according to design specifications (prognostics); (4) supporting decisions on what actions to take to assure safety and efficient production with the minimum costs (maintenance planning). Compared to fault detection and diagnosis techniques (itemized tasks 1 and 2), which have been extensively investigated in the last decades, prognostics (itemized task 3) is a relatively new research field which is still in the development phase [10]. On the other side, the possibility of effective maintenance planning (itemized task 4) grounds on the availability of reliable prognostic predictions.

In this context, the goal of the PhD project here presented is to develop an efficient approach for the prediction of the remaining useful life of industrial equipment, as the necessary requirement for PHM to be effectively employed in industry for improved maintenance.

The work has been performed within a cooperation between the Laboratorio di Analisi di Segnale ed Analisi di Rischio (LASAR, Laboratory of Signal Analysis and Risk Analysis) of the Department of Nuclear Engineering of the Polytechnic of Milan (http://lasar.cesnef.polimi.it/) and the division of Computerised Operation Support Systems (COSS) of the Institute For Energy Technologies (IFE), OECD Halden Reactor Project in Halden, Norway (http://www.ife.no).

[3] References from [1] to [81] are literature references whereas references from [82] to [95] have been produced within this PhD work.

1.1 Review of prognostic approaches

In general, prognostic methods can be classified in model-based and data-driven [11]. Among data-driven methods one can distinguish between those based on degradation modeling and those directly predicting the RUL [12].

Model-based methods use an explicit mathematical model of the degradation process to predict the future evolution of the degradation state and, thus, the RUL of the equipment [13]. Examples of degradation models are the non-linear stochastic model of fatigue crack dynamics [14-15]. In practice, even when the model of the degradation process is known, the RUL estimate may be difficult to obtain, since the degradation state of the equipment may not be directly observable and/or the measurements may be affected by noise and disturbances. In these cases, model-based estimation methods aim at inferring the dynamic degradation state and provide a reliable quantification of the estimation uncertainty on the basis of the sequence of available noisy measurements [16]. Many approaches rely on Bayesian methods for updating the prediction upon collection of new data and information [17-18]: the exact Kalman filter has been largely used in case of linear state space models and independent, additive Gaussian noises, whereas analytical or numerical approximations of the exact Bayesian solution (such as the Extended Kalman filter, the Gaussian-sum filters or the grid-based filters) have been applied in most realistic cases where the dynamics of degradation is non-linear and/or the associated noises are non-Gaussian [19-20]. Recently, numerical approximations based on the Monte Carlo sampling technique (known as particle filters) have gained popularity for their flexibility and ease of design [21-22].

On the other side, data-driven methods are used when an explicit model of the degradation process is not available, but sufficient historical data have been collected. Being a rapidly developing research filed, the research on data-driven prognostics has embraced a vast number of techniques and algorithms [7] that come from multiple research fields, such as reliability engineering [23-24], Bayesian inference [16,25], regression analysis [26-28], artificial intelligence [7,29-30], etc. All these methods are based on statistical models that can ‘learn’ a prognostic model from the data. In this respect, artificial neural networks are often used [26, 30, 31]; other examples are Autoregressive Moving Average techniques [32], Relevance Vector Machines [26, 33], similarity-based regression methods [12,29]. Recently, ensemble approaches, based on the aggregation of multiple model outcomes, have been introduced due to the superior robustness and accuracy with respect to single models [34] and the possibility of estimating the uncertainty of the predictions [35]. The interested reader can found a more extensive review of prognostic methods for remaining useful life estimation [7] and [36].

These methods can be use to learn from data the trend of the degradation evolution, or directly the relation between the observable parameter and the equipment RUL [7]: we talk about ‘degradation-based prognostics’ in the former case [32,37-38], and ‘direct RUL prediction’ in the latter [30,39]. The degradation-based approach relies on degradation state prediction followed by a failure criteria evaluation, which in general consists in the evaluation of the time at which degradation will exceed the failure threshold, i.e., the maximum value of degradation allowed for the equipment to continue performing its functions according to the given specifications. Data-driven approaches learn the degradation model from the time series of the observed degradation states through regression/trend analysis or stochastic process modeling [12,37]. In the direct RUL prediction, instead, the equipment RUL is directly derived from a data-driven

model that learns from the data the relations between the current observation and the equipment RUL, without passing through degradation state estimation and failure criterion definition [29-30].

Three main open issues in prognostics are the pre-treatment of noisy or unreliable field data, the choice of a suitable prognostic approach given the type and quality of information available, the quantification of the prediction uncertainty.

1.2 Contribution of this thesis work

With reference to the three open issues presented above, the following objectives have guided the work performed during this PhD research activity [81-90]:

1. developing methods for the verification and pre-treatment of degradation-related field data, in order to derive reliable prognostic information from them;
2. identifying the most representative situations of information available, and developing properly tailored prognostic approaches to tackle each of them;
3. analyzing the sources of uncertainty affecting the RUL prediction and providing methods for their quantification;
4. evaluating the proposed prognostic methods in terms of the assumptions they require, their accuracy in predicting the equipment RUL and their ability of providing measures of confidence in the RUL predictions produced.

The necessity of pre-treating data directly or indirectly related to the equipment degradation (objective 1) arises in real industrial applications where reliable prognostic information has to be drawn from field data which are typically affected by noise, sensor faults, extrapolation errors, etc. [8-9]. In fact, all these sources of uncertainty can be potentially amplified when projected in the future to predict the equipment RUL. For this reason, it is very important to reduce the uncertainty injected into the prognostic model by the input and training data. Providing a thorough analysis and a general solution to this problem is very difficult, since the problems that can arise in this context and the relative solutions depend strongly on the specific application. However, the importance of this problem has been standing out during the development of this thesis work when dealing with real industrial data, and several tools for their pre-treatment have been developed in the context of a case study related to the erosion of choke valves used in the oil & gas industry [91-94]. In particular, we developed a monotonicity-based index for evaluating the quality of the available degradation indications and a clustering analysis for validating the reliability of some uncertain observable parameters [91]. Finally, an ensemble of kernel regression models has been developed for improving the estimates of the uncertain parameters that are found unreliable [92-94].

The second objective listed above deals with the lack of general guidelines for tackling the different prognostic problems that one may encounter depending on the forms of information and data that are available for the assessment of the evolution to failure of degrading equipment. Indeed, depending on the situation, different prognostic methods may be applied [9-10]. Three practical situations with different sources of information available for the prognostic task have been identified in this work [82]. These situations and the prognostic methods proposed in this work to tackle each of them are summarized in Figure 1. In all situations it is assumed to have available a sequence of observation related to the degradation of the

equipment of interest. In the first case considered in this work, hereafter referred to as case 1, it has been assumed to have available a stochastic model of the degradation process and thus, a model-based particle filtering (PF) method [16,25], has been set-up to predict the distribution of the system RUL and online-update it when new observations are collected [82-83]. The second situation considered, hereafter referred to as cases 2, has available a number of observations related to the degradation evolution and the failure times of a set of similar systems operating under similar conditions. Two sub-cases can be distinguished: a first one (case 2A) where the relation between observations and degradation is not known and thus one has to resort to direct RUL prediction, and a second one (case 2B) where this relation is known and can be used to develop degradation-based prognostics. In case 2A a prognostic method using similarity-based regression (SBR) [12,29,40] has been developed [88-89], in case 2B Gaussian Process Regression (GPR) [41-42] has been used to model the evolution of the degradation process [86-87]. Furthermore, since, in a situation where more prognostic methods can be applied, it is often hard to chose among them, a combined approach which aggregates the different outcomes of the SBR and the GPR methods has also been developed [92]. Finally, a third situation considered, hereafter referred to as case 3, assumes to have available a sequence of degradation observations only for the equipment of interest. In this situation, it is necessary to resort to degradation-based prognostics, since no failure data are available from similar pieces of equipment which have failed in the past. Although the situation is similar to that of case 2B, more robust methods have to be used in this case, since much fewer data are available. For this, we resorted to the ensemble technique and proposed two approaches based on a Kalman Filter (KF)-aggregated ensemble [30, 81] and a bootstrapped ensemble of degradation models [35,82-85].

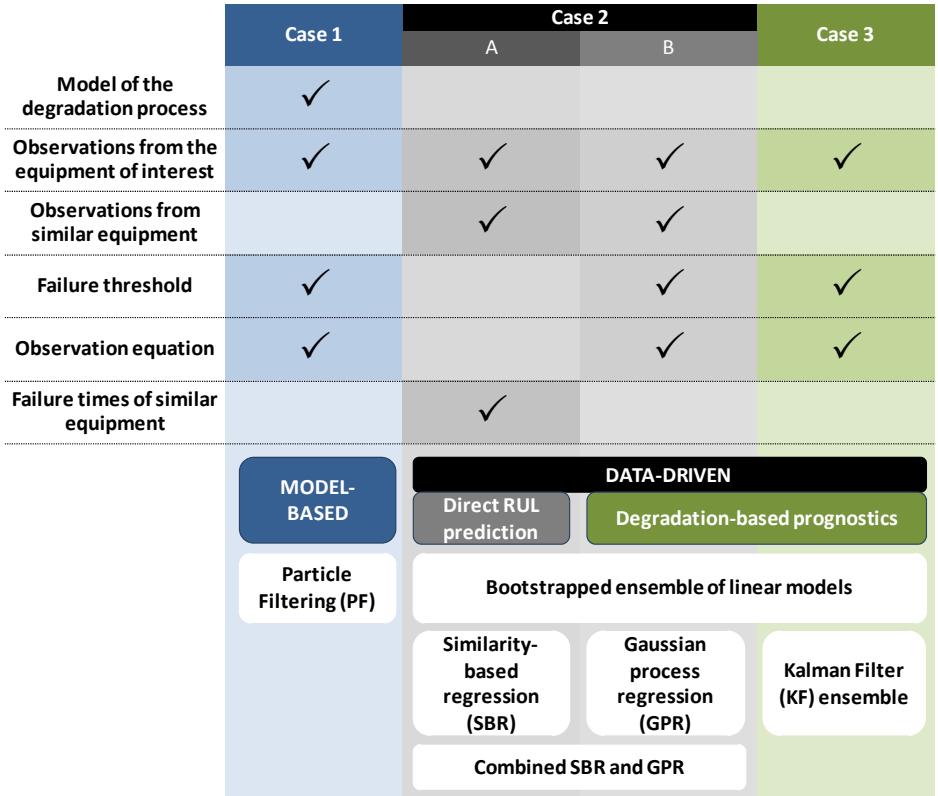

Figure 1: pictorial view of the situations of information available considered and the corresponding prognostics methods developed to tackle them.

To perform their task, prognostic models have to project the current equipment condition in time, in the absence of future measurements about its degradation state and operational conditions. This leads to propagating large uncertainties which need to be assessed and managed (objective 3), by associating uncertainty estimates to the RUL predictions provided, so to measure the expected mismatch between the real and predicted equipment failure times, which can be used by the maintenance planner to confidently schedule maintenance actions, according to the desired risk tolerance. In this context, the contribution of this work has been the development of prognostic methods capable of correctly quantifying their prediction error and supplying a reliable prediction interval (instead of a point prediction) for the value of the RUL [82-90]. Also, this work has contributed to the way of investigating the capabilities of different prognostic approaches to deal with the uncertainty in the RUL prediction by considering the variance of the RUL prediction error as a suitable measure of prognostic uncertainty, decomposing it into three components representing the three main sources of uncertainty identified above, studying the contribution of the different sources of uncertainty and validating different prognostic approaches with respect to their capability of providing accurate estimates of the RUL prediction error variance [82].

Two case studies with simulated data have been developed to validate the prognostic approaches poposed (objective 4): the linear growth of creep damage in the blades of a helium gas turbine of a Gas Turbine Modular Helium Reactor (GT-MHR) [81-85] and the non-linear growth of creep damage in ferritic steel equipment (simulated data) [86-88]. A further case study taken from the nuclear industry and concerning the clogging of filters in a Boiling Water Reactor (BWR) condenser has also been considered to demonstrate the applicability of the prognostic methods on real data [87-90].

1.3 Structure of the thesis

The thesis comprises two parts. Part I, subdivided in six Sections, addresses the objectives undertaken, illustrates the methods developed and employed in this PhD work, and discusses some of the results obtained in the case studies. Part II is a collection of a work report and six selected papers, written as a result of this PhD work, to which the reader is referred for further details.

The first objective considered in this work, i.e., the pre-treatment of noisy and unreliable field data for their use in prognostics, is introduced in Section 2 and more extensively discussed in Report A of Part II, in the context of choke valve erosion assessment. Section 3 introduces the concepts behind remaining useful life prediction and discusses the two important topics of prognostic model selection based on the available information (objective 2), and treatment of prognostic uncertainty (objective 3), which are further detailed in Papers II and III of Part II. The central Section 4 focuses on the description of the prognostic methods developed during this PhD to tackle the three situations of information available summarized in Figure 1 and provide accurate RUL prediction with reliable measures of the prediction uncertainty. In Section 4 methods are organized based on the specific situation of information they have been designed for. Section 5 discusses some results obtained in the validation of these methods on case studies about the linear growth of creep damage in turbine blades and the clogging of sea water filters. Both Sections 4 and 5 make extensive reference to the six selected papers of Part II, which treat more in details the prognostic approaches developed and show the results of their application to some case study among those considered. Tables 1, 2

and 3 summarize the structure of this thesis with respect to the objectives undertaken, the prognostic methods presented and, respectively, the case studies considered.

Table 1: structure of the work with respect to the objectives undertaken

Objective	**PART I Section**	**PART II Paper/Report**
1. Pre-treatment of prognostic data	2	A
2. Model selection based on the available information	3.1, 4	II
3. Treatment of prognostic uncertainty	3.2, 4	III
4. Evaluation of the proposed prognostic methods	5	I-VI

Table 2: structure of the work with respect to the prognostic methods presented

Method	**PART I Section**	**PART II Paper/Report**
Kalman filter ensemble	4.3.1	I
Particle Filtering	4.1.1, 5.1	II-III
Bootstrapped Ensemble	4.3.2, 5.1	II-III
Similarity-based regression (SBR)	4.2.1, 5.2	IV
Gaussian process regression (GPR)	4.2.2, 5.2	V
Combined SBR and GPR methods	4.2.3, 5.2	VI

Table 3: structure of the work with respect to the case studies considered

Case study	**PART I Section**	**PART II Paper/Report**
Linear creep growth in turbine blades (simulated data)	5.1	I-III
Non-linear creep growth in ferritic steel (simulated data)	-	IV-V
Clogging of filters in a BWR condenser (real data)	5.2	IV-VI
Erosion of choke valves in extraction wells (real data)	2	A

2 Pre-treatment of prognostic data

In the industrial application of the prognostic approaches developed and investigated in this work, we may very likely have to face additional problems related to the limited and unreliable information available about the equipment degradation state or the future operating conditions. During this thesis work, this problem has been encountered in the context of a case study with real data taken from eroding choke valves located topside at wells in offshore platforms on the Norwegian Continental Shelf [43,91-94].

In oil and gas industries, choke valves are used to control flow rates and protect the equipment from unusual pressure fluctuations. They are based on a throttle mechanism made of two circular disks, one fixed and one rotating, each with a pair of circular openings to create variable flow areas (see Figure 2).

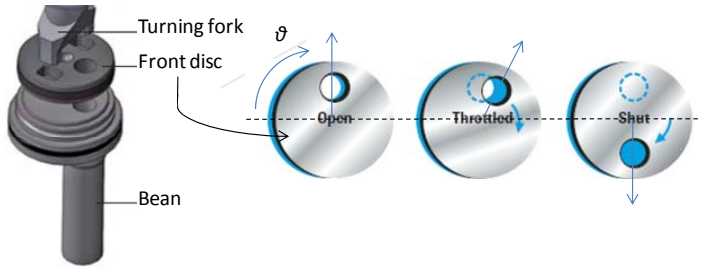

Figure 2: Typical scheme of a choke valve: by rotating the disk the flow is throttled (picture taken from www.vonkchokes.nl).

Combined mechanisms of erosion-corrosion cause choke lifetimes of less than one year [44]. To avoid failures, effective erosion management strategies must be implemented, requiring the development of reliable models to estimate the erosion and predict the lifetime of choke valves [45-46]. A common indicator of the valve flow capacity is the flow coefficient C_V. For a specific valve opening δ, erosion produces a gradual increase of the valve area available for flow transit, thus determining an increase of C_V. For this reason, knowing the value of the flow coefficient is fundamental for assessing the erosion state of the choke, and, consequently, for defining a failure criterion, evaluating correctly the life duration of valves, and, further on, modeling the degradation process.

During operation, C_V is not directly measured but computed as a function of the pressure drop through the choke, ΔP, the oil, water and gas flow rates, $\dot{M}_o$, $\dot{M}_w$ and $\dot{M}_g$, respectively, and the valve opening δ [47]. Thus, for a correct assessment of the choke erosion state, it is fundamental to have reliable measurements or estimates of the parameters ΔP, δ, $\dot{M}_o$, $\dot{M}_w$ and $\dot{M}_g$, used to compute the flow coefficient C_V. Nevertheless, only the pressure drop ΔP, and the valve opening δ are measured during daily standard inspections, whereas measures of water, oil and gas flows rates are taken downstream of the choke with a multiphase flow separator only during well tests performed about once per month. On daily bases, the values of $\dot{M}_o$, $\dot{M}_w$ and $\dot{M}_g$ for a single well are allocated based on the measured total production and on the values of physical parameters, such as pressures and temperatures related to the specific well. In practice, the resulting indication of the choke valve state is very noisy and lacks the physical monotonicity of the erosion process; allocated values of oil, gas and water flow rates are conjectured to be the cause of the large inaccuracies and uncertainties in the calculation of the actual valve flow coefficient. To

verify this and improve the accuracy and reliability of the available information, different tools for the pre-treatment of degradation-related data have been developed during this thesis.

First, a monotonicity index has been proposed to evaluate the reliability of a sequence of degradation indications [91]. Then, a method for verifying reliability of some uncertain parameters in the observation vector $\mathbf{z}_j$ has been developed, by means of the Fuzzy C Means (FCM) clustering algorithm [48-49,91]. Based on expert knowledge, the Q-dimensional space of the observable parameters is partitioned in two subspaces of reliable and uncertain parameters. FCM is applied to the projections of the dataset of available observations $\mathbf{D} = \{\mathbf{z}^r_{1:n^r}\}_{r=1}^{R}$ into the subspace of the reliable parameters (e.g., pressure drop and choke opening in the choke valve case study) and the subspace of the uncertain parameters (e.g., oil, water and gas flow rates in the choke valve case study). The two partitions are compared to investigate the coherence of the information conveyed by the parameters. Finally, since the values of oil, gas and water flow rates in the choke valve case study are found unreliable, a method for improving their estimates has been developed based on the relations among all observable parameters. To this aim, an ensemble of Kernel Regression (KR) models, a distance-based regression algorithm [50-51], is devised to avoid the need of selecting the optimal model and to increase the robustness and reduce the uncertainty of the estimate [52-53, 91-94]. Diversity is injected in the ensemble by differentiating the training procedure for each KR model. Different procedures for the aggregation of the KR model outcomes have been compared; in particular, an original procedure, based on the weighted average of the single model outcomes with weights calculated using the Analytic Hierarchy Process (AHP) [54,91] has been developed for those situation where performance-based aggregation algorithms cannot be applied due to the unavailability of reliable observations for the value of the uncertain parameters.

In the choke valve case study, we have noticed that, depending on the well, either the physics-based model for flow rates allocation or the data-driven KR models ensemble can provide the most accurate estimates of the flow rates to be used for computing the flow coefficient and assessing the valve erosion [92]. Physics-based approaches rely on the assumption of deep understanding of the system behaviours and detailed knowledge of geometry, material properties and other characteristics of the system. On the other side, data-driven approaches are only accurate when applied to the same, or similar, operating condition for which data have been collected. Outside the training space of the empirical models, physics-based models should be preferred, because their results come from universal first principles. The combination of physics-based and data-driven models, usually termed hybrid models, aims to balance out their different errors and to augment the robustness and interpretability of physics-based models with the sensitivity of data-driven methods [9,34,45]. In this work, a hybrid approach has been applied to improve the accuracy of the estimates of the uncertain parameters provided by the physic-based model, by appending to it the ensemble of Kernel Regression (KR) models [93-94]. In order to exploit the KR models only when actually needed, a local performance-based approach is adopted for the aggregation of the different model outcomes [55].

The tools developed for data pre-treatment are briefly sketched in this section; a detailed description of the algorithms and all the results of their application to the choke valves case study, are given in Report A of Part II.

2.1 Monotonicity index

In general, a sequence of degradation indications $z_{\theta,j}^{r}$ is expected to be monotonic, since degradation cannot decrease in time unless maintenance actions are performed or the degradation process includes some self-healing mechanism. A quantitative index of monotonicity is the Spearman's rank correlation used in statistics to assess how well the relationship between two variables can be described using a monotonic function [56]. The Spearman's rank correlation coefficient $SRCC^{r}$ between the degradation indication $z_{\theta,j}^{r}$ and time t_j^r, at which the observations of the r-th trajectory are taken, is computed as:

$$SRCC^{r} = 1 - \frac{6\sum_{j=1}^{n^r}\left(R(z_{\theta,j}^{r}) - j\right)}{n^r\left(n^{r\,2} - 1\right)} \tag{1}$$

where $R(z_{\theta,j}^{r})$ and j are the ranks (i.e., the relative positions) of observation $\mathbf{z}_j^r$ when all observations are ordered with respect to the values of $z_{\theta,j}^{r}$ and t_j^r, respectively. Values of $SRCC^{r}$ close to 1 are expected for a monotonic quantity.

2.2 Cluster analysis

Let $\mathbf{D}$ be a generic set of N observations $\mathbf{z}_j = [\mathbf{z}_j^{\mathrm{R}}; \mathbf{z}_j^{\mathrm{U}}]$, of Q parameters which can be divided in a vector $\mathbf{z}_j^{\mathrm{R}}$ of q^{R} reliable parameters, and another vector $\mathbf{z}_j^{\mathrm{U}}$ of q^{U} uncertain parameters, whose observed values are doubted to be reliable. In general, the distinction between reliable and uncertain parameters can be achieved based on expert judgment, data analysis or by resorting to data validations techniques which allow detecting anomalous behaviors in datasets. In the choke valve case study, we have $\mathbf{z}_j^{\mathrm{R}} = [\Delta P, \delta]$ and $\mathbf{z}_j^{\mathrm{U}} = [\dot{m}_o, \dot{m}_w, \dot{m}_g]$ according to expert judgment [91].

In this work, a procedure has been proposed for verifying whether the information provided by the uncertain parameters in $\mathbf{z}_j^{\mathrm{U}}$ is coherent with that of the reliable parameters in $\mathbf{z}_j^{\mathrm{R}}$. This is done by considering the relative positions of the observations in the q^{R}-dimensional subspace of the reliable parameters, and in the q^{U}-dimensional subspace of the uncertain parameters [91]. An effective technique to find a structure in a collection of unlabeled objects is unsupervised clustering, consisting in the organization (partition) of the observations into non-overlapping, non-empty groups (clusters) so that observations of the same cluster are *similar* between them and *dissimilar* to the observations belonging to other clusters [57]. For the validation of the uncertain parameters, two different partitions (Γ^{R} and Γ^{U}) of the dataset $\mathbf{D}$ into C clusters are considered: Γ^{R} is obtained using the unsupervised Fuzzy C-Means (FCM) clustering technique in the reliable parameters space, whereas Γ^{U} obtained by applying the same technique in the uncertain parameters space. The cluster structure thereby identified by Γ^{R} is assumed as reference in the comparison with the partition Γ^{U}, since it is built using only the reliable information in $\mathbf{D}$. The clusters of Γ^{R} and Γ^{U} are bi-univocally associated by minimizing the distance between the two partitions [58] and crisp partitions Ω^{R} and Ω^{U} are obtained from the fuzzy partitions Γ^{R} and Γ^{U}, respectively, by assigning a observation $\mathbf{z}_j$ to a given cluster if its degree of membership to the cluster, exceeds a predefined threshold, which represents the

required degree of confidence for the assignment. The crisp partitions Ω^{R} and Ω^{U} are compared by considering the difference between the sets of observations assigned to associated clusters. A large difference in the assignment of the observations to the clusters in the two partitions is taken as a symptom that the information conveyed by the uncertain parameters may be misleading.

A further analysis based on a supervised clustering technique is also proposed, where a clustering algorithm based on Mahalanobis metrics [58] is used to obtain a partition Γ^{M} of the entire Q-dimensional dataset as close as possible to that obtained based only on the reliable parameters (Γ^{R}). Based on the Mahalanobis metrics providing the partition Γ^{M}, a measure of the importance in the clustering of the different parameters is calculated and used to verify the coherence of the information conveyed by the uncertain parameters with that conveyed by the reliable ones.

2.3 Improving process parameter estimates

The analysis presented in Section 2.1. and 2.2 performed on the choke valve case study, led to the conclusion that, in some circumstances, which may correspond to specific operating conditions, the physics-based model used to provide the allocated values of the mass flow rates $\dot{M}_o$, $\dot{M}_w$ and $\dot{M}_g$ does not work properly, and thus the allocations are noisy and uncertain (see Report A of Part II). For this reason, a hybrid ensemble approach has been developed for improving the estimates of process parameters provided by partially uncertain models or measurements [93-94].

The ensemble, sketched in Figure 3, is fed with the observation $\mathbf{z}_j$, and provides improved estimates $\hat{\mathbf{z}}_j^{U}$ of the uncertain parameters vector $\mathbf{z}_j^{U}$.

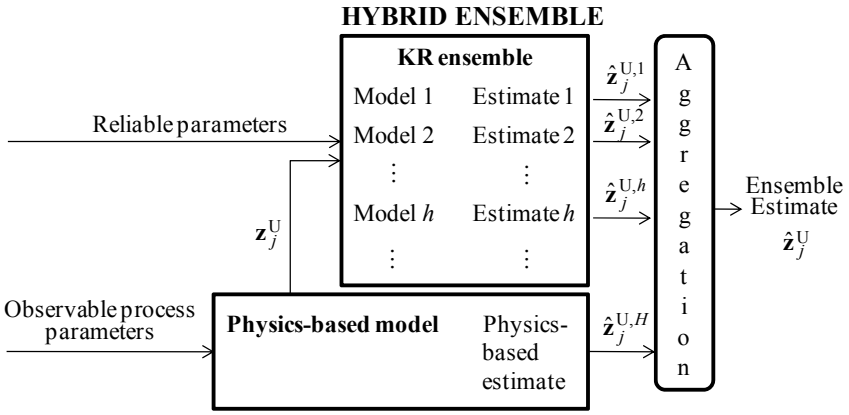

Figure 3: Pictorial view of the hybrid ensemble approach.

The ensemble includes the physics-based model used to estimate the uncertain parameters $\mathbf{z}_j^{U}$ and H data-driven Kernel Regression (KR) models [59], which learn the relationships between the parameters from the dataset of all available observations $\mathbf{D}$ or, if possible, from a subset of it, containing only reliable data. For example, in the choke valve case study, a subset of the entire available dataset $\mathbf{D}$ containing only the reliable observations taken during well tests should be used for more robust training of the KR models. KR models return as outputs the estimates of the uncertain parameters they have been fed with in input. Such estimates

are obtained as weighted averages of the training observations, with weights decreasing in function of the distance between training and test observations: by use of a Gaussian kernel function [59], training observations closer to the test one are weighed more in the averaging because they are assumed to be more similar to it. Diversity between the KR models of the ensemble is obtained by varying the parameters used to compute the distance between training and test observations.

Performance-based aggregation techniques have proved to be effective for the aggregation of the outcomes of the individual models of the ensemble. However, they require the availability of a validation dataset of reliable observations for the value of the parameters under estimations. In the context of the estimation of uncertain parameters, such validation dataset may not be available. In this work, an original procedure for the aggregation of the different ensemble model outcomes in the absence of a validation dataset has been developed, based on the AHP technique (see also Section 3.3 of Report A) [91-92]. AHP is used to assign performance weights to the models of the ensemble. AHP is a multi-criteria decision method that uses hierarchic structures to represent a decision problem [54] and provides ranking of the different models outcomes using relative performance measurements, without resorting to an absolute measurement of the model performance. It consists of two main steps: 1) structuring a hierarchy; 2) assigning priorities to the elements of each hierarchy level by comparative judgments of the elements based on a pre-defined scale. A three level hierarchy structure is used: the ensemble models (level 3) undergoes pairwise comparisons with respect to two criteria Z_1 and Z_2 (level 2) towards the goal (level 1) of obtaining high model accuracy. The first criterion Z_1, chosen to evaluate the relative importance of model h with respect to model h' in the estimate of a test observation $\mathbf{z}_j$, is the relative similarity of the two models outcomes $\hat{\mathbf{z}}_j^{\mathrm{U},h}$ and $\hat{\mathbf{z}}_j^{\mathrm{U},h'}$ to the remaining models outcome $\hat{\mathbf{z}}_j^{\mathrm{U},h''}$, $h'' \neq h, h'$. Assuming that the model outcomes of the models left out of the pairwise comparison are distributed around the correct value, this criterion assigns larger weights to the model whose outcome is more similar to those of the models left out. The similarity of two estimates $\hat{\mathbf{z}}_j^{\mathrm{U},h}$ and $\hat{\mathbf{z}}_j^{\mathrm{U},h'}$ has been estimated by the inverse of their Euclidean distance $d_j^{h,h'}$; the relative importance of a model h with respect to model h' when model h'' is taken as reference is defined by the ratio $d_j^{h',h''} / d_j^{h,h''}$ and the overall relative importance of model h with respect to model h' is given by the product of the relative importance with respect to all remaining models $h'' \neq h, h'$. A second criterion Z_2 for evaluating the performance of a model takes into account the RMSE in reconstructing the reliable parameters in $\mathbf{z}_j^{\mathrm{R},h}$, i.e. the root mean square difference between the reconstructed and measured values. This second criterion is based on the idea that robust and accurate models should be able to correctly reconstruct the reliable parameters despite the noise on the uncertain parameters. Based on the comparisons performed with respect to criteria Z_1 and Z_2, the AHP provides a ranking of the different model outcomes which is used to weight them before they are averaged to provide a single aggregated estimate of the uncertain parameters.

In the choke valve case study, reliable observations for the mass flow rates are collected during well tests, so that it has been possible to apply a local performance-based techniques for aggregating the ensemble model outcomes [93-94]. This technique relies on the idea that each model can perform well in some regions of the parameter space and poorly in others; then, the aggregation has been guided by the local performance of each model, i.e., its reconstruction accuracy on training observations similar (and for this reason also called neighbors) to that under test (see also Section 4.1 of Report A) [55]. This allows exploiting the data-driven KR models only when they outperforms the physics-based model, thus avoiding affecting its accuracy when it actually outperforms the KR models. For each parameter q^{U} to estimate, the local performance

aggregation approach adopted assigns to the generic model h in the ensemble a weight proportional to the model performance evaluated considering the inverse of its mean square estimation error over the N training observations $\mathbf{z}_{1:N}^{neig}$ closest to the test observation $\mathbf{z}_j$. The estimation error made by model h in providing the estimates $\hat{z}_{q,1:N}^{\text{neig},h}$ of the q-th parameter is obtained by comparing it to the corresponding well test measurement. The final estimate $\hat{z}_{q,j}$ of parameter q is obtained as the average of the multiple model estimates $\hat{z}_{q,j}^{h}$ weighted with the performance-based weights.

In Figure 4 the values of oil, water and gas flow rates measured during the well tests in the choke valve case study are compared with the estimates obtained by the physics-based model and by the locally-aggregated hybrid ensemble. Results show that the hybrid ensemble and the physics-based model estimates are in general very similar except for the oil flow in well 1 where the ensemble estimates significantly outperform the physics-based model estimates. This results from the fact that the locally-aggregated hybrid ensemble can correctly mediate between the physics-based and data-driven approaches, since the local-performance aggregation technique is able to automatically distinguish those cases in which the physics-based model works properly from those in which it is convenient to integrate the data-driven models.

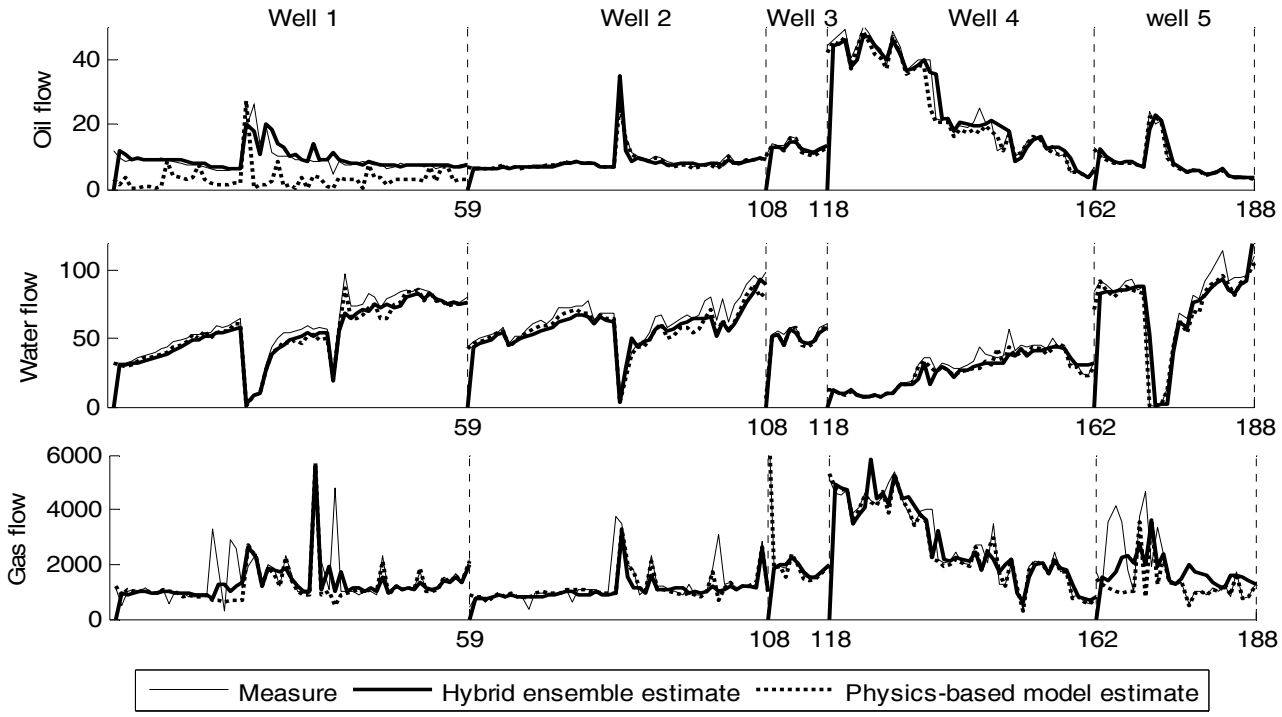

Figure 4: comparison of the measured valued of oil, water and gas flow rates with the estimates obtained by the physics-based model and the hybrid ensemble.

3 Remaining Useful life prediction

The aim of prognostics is to estimate the Remaining Useful Life of degrading equipment. Let us discretize, for ease of exposition, the continuum time variable t into a sequence of time instants t_j, $j = 1,2,...$ assumed to be equally spaced. We indicate by θ_j the degradation level of a piece of equipment at time t_j and by $\mathbf{z}_j = [z_1(t_j),...,z_q(t_j),...,z_Q(t_j)]$ a vector of observations of Q relevant parameters $z_q(t_j)$ related to the equipment degradation, measured at time t_j. In this work, we assume that the equipment is subject to a single degradation mechanism described as a random process; we do not consider the effects that other competing degradation mechanisms can have on the equipment degradation. It is also assumed that the equipment degradation cannot exceed a maximum acceptable level, hereafter referred to as *failure threshold*, θ_{th}, which is fixed and identical among similar equipment. Notice that in prognostics the failure threshold does not necessarily indicate complete failure of the system, but, for safety margins, it is often set to a conservative value of the degradation beyond which the risk of complete failure exceeds tolerance limits or the performance of the system does not fulfill the requirements [60]. Setting at $t_0 = 0$ the beginning of the working life of equipment *r*, its life duration is given by the time t_F at which its degradation crosses the failure threshold, and its RUL at the present time t_J is equal to

$$rul_J = t_F - t_J . \tag{2}$$

3.1 Information and data for prognostics

Different forms of information and data may be available for the assessment of the evolution to failure of degrading equipment. Depending on the situation, different prognostic approaches may be applied [9-10]. The main sources of information upon which prognostics can be based have been classified in the following six categories summarized in Table 4 and further detailed here:

1. Physical model of the degradation mechanism (source *A*, Table 4), e.g., described by a first-order Markov process:

$$\theta_j = g(\theta_{j-1}, \boldsymbol{\gamma}_{j-1}) \; ; \; \theta_0 \sim p_{\Theta_0}(\theta_0) \tag{3}$$

where $p_{\Theta_0}(\theta_0)$ is the initial distribution of the degradation at time t_0, g is a possibly non-linear transition function describing the value of a one-time-step degradation increment, and $\boldsymbol{\gamma}_j$, $j = 1,2,...$ is a sequence of mutually independent vectors of state noises. Model g can contain parameters referring to system inherent characteristics (material, physical, chemical, geometrical, etc.), which may vary from one individual system to another of the same type: this variability is described by probability distribution functions.

Table 4: main sources of information for prognostics

Source	Description	Mathematical representation
A	Stochastic model of the degradation process	eq.(3)
B	Sequence of observations related to the degradation of the test equipment collected at $t_{1:J}$	$\mathbf{z}_{1:J}^{test}$
C	Historical sequences of observations related to the degradation of a set of *R* failed systems collected at $n(r)$ time instants t_j, $r = 1,2,...,R$, $j = 1,2,...,n(r)$	$\{\mathbf{z}_{1:n^r}^{r}\}_{r=1}^{R}$
D	Value of the failure threshold	θ_{th}
E	Measurement equation	eq.(4)
F	Durations of lives of the set of *S* failed systems *s*.	$\{t_F^r\}_{r=1}^{R}$

2. A set of observations, collected at different time instants, during the life of the equipment whose RUL we want to predict (source of information B, Table 4), hereafter referred to as test equipment, or of a population of R identical or similar pieces of equipment (source of information C, Table 4), hereafter referred to as reference equipment. Among the observable process parameters in $\mathbf{z}_j^{test}$ there can be a direct measure of the degradation state of the system (e.g., depth of a crack fracture, elongation of a creeping component, etc.) or they can be only indirectly related to it (e.g., the time of travel or the intensity of ultrasonic waves for non-destructive inspections). In the following, we indicate by $\mathbf{z}_{1:n^r}^{r}$ the observations collected at times $t_{1:n^r}^{r}$ during the degradation of the *r*-th reference piece of equipment, and by $\mathbf{z}_{1:J}^{test}$ those collected for the test equipment until the present time $t_{1:J}^{test}$. The superscript *test* will be omitted when not necessary to the clarity of explanation.
3. The value of the failure threshold θ_{th} (source *D*, Table 4).
4. The observation equation (source *E*, Table 4), i.e., the physical model describing the relation between the observation $\mathbf{z}_j$ and the actual degradation state θ_j of the system:

$$\mathbf{z}_j = u(\theta_j, \boldsymbol{\gamma}'_j) \tag{4}$$

where u is a known function, in general non-linear, and $\boldsymbol{\gamma}'_j$ is a vector of measurement noises.

5. The life durations $\{t_F^r\}_{r=1}^{R}$ of the R reference pieces of equipment which have failed in the past (source *F*, Table 4); notice that, the actual value of the RUL of the *r*-th failed equipment can be computed at any time $t_j^r < t_F^r$ according to eq. (2)

Three cases are considered in this work (Figure 1), in which the sequence of observations $\mathbf{z}_{1:J}$ collected during the life of the test equipment (source of information *B*) is available in combination with other different sources of information.

In case 1, the physical model of the evolution of the degradation state is known, as well as the distribution and evolution in time of all its characteristic and external parameters (source *A*). Other sources of information available are the value of the failure threshold θ_{th} (source *D*) and the observation equation

(source *E*) linking the observations with the degradation state. This situation is typical for well known degradation mechanisms, such as the crack or creep growth processes, which have been widely studied in laboratory.

In case 2, a set of observations $\{\mathbf{z}_{1:n^r}^r\}_{r=1}^R$, from the R reference pieces of equipment (source *C*) is assumed to be available. Each sequence of observations $\mathbf{z}_{1:n^r}^r$ is hereafter called training trajectory. This situation is typical for short-life systems, for which many trajectories to failure can be observed. In this situation, we have distinguished two sub-cases A and B: in case 2A, the duration of equipment lives $\{t_F^r\}_{r=1}^R$ (source *F*) is known; in case 2B, instead, we know the measurement equation (source *E*) and the failure threshold θ_{th} (source *D*). Notice that, in case 2B, life duration data could be available instead of the failure threshold, since the value of this latter can be derived, in general, from the degradation state of the reference equipment at the time of its failure, which can be in turn derived from the observations $\{\mathbf{z}_F^r\}_{r=1}^R$, when the observation equation is available and the training trajectories have been collected until failure of the reference equipment.

Finally, in case 3, the information available is the observation equation (source *E*) and the failure threshold θ_{th} (source *D*). This situation can occur in case of very reliable systems, e.g., those used in the nuclear industry, which have very long life durations and are usually renewed before failure happens.

3.2 Uncertainty treatment

A prognostic model plays the central role of receiving in input different available pieces of information concerning the state of health of the equipment of interest and predicting its RUL by projecting the current system condition in time [61-62]. Since the prediction of the Remaining Useful Life (RUL) of degrading equipment is performed in the absence of future measurements concerning the equipment degradation and operational conditions, the prognostic task is necessarily affected by large uncertainty.

In this work, the sources of uncertainty affecting the RUL prediction are classified in three categories:

a. Randomness in the future degradation of the equipment. This intrinsic uncertainty in the degradation process has several causes such as the unknown future load profile, and operation and environmental conditions.

b. Modeling error, i.e., inaccuracy of the prognostic model used to perform the prediction. In model-based prognostic approaches, this source of uncertainty takes into account the assumptions and simplifications made on the form and structure of the model, and the uncertainty on the model parameters. In data-driven approaches, it relates, for example, to the incomplete coverage of the data set used to train the empirical model.

c. Uncertainty in current and past equipment degradation data, which are used by the prognostic model to elaborate the RUL prediction. These data are usually acquired by sensors with some measurement noise or derived from diagnostic systems assessing the equipment health state with some degree of uncertainty.

The model and observation uncertainties (sources *b* and *c*) can be interpreted as the epistemic uncertainty related to our limited knowledge about the degradation process and the equipment degradation state, whereas the process uncertainty (source *a*) is related to the stochastic uncertainty intrinsic to the degradation process.

The challenge of managing uncertainties associated with prognostics has been recently addressed in [60-61]. Uncertainty management in prognostics entails to identify, classify and analyze uncertainty sources with the aim of associating to the RUL predictions provided by a prognostic model an estimate of its uncertainty [9,60,63-64], i.e., a measure of the expected degree of mismatch between the real and predicted equipment failure time. This information can be used by the maintenance planner to confidently plan maintenance actions, according to the desired risk tolerance [62].

As degradation evolves randomly in time, due, e.g., to the scatter in the microstructural and manufacturing characteristics of equipment, the loading and external conditions variability, etc., the damage state of equipment r at any future time t_j, is better represented by a random variable $\Theta(t_j)$ rather than by a deterministic quantity [65]. As a consequence, also the equipment RUL at the present time t_J should be represented by a random variable RUL_J [83]. Thus, the prognostic output of interest with respect to an equipment with current degradation state θ_J is the probability density function (PDF) $p_{RUL_J}(rul \mid \theta_J)$. The uncertainty described by such distribution regards the future stochastic evolution of the equipment degradation and, thus, it is irreducible.

A realization rul_J of the random variable RUL_J can be written as:

$$rul_J = E[RUL_J \mid \theta_J] + a_J \tag{5}$$

where $E[RUL_J \mid \theta_J]$ is the expected value of the RUL of a piece of equipment with degradation θ_J at time t_J, and a_J is a random variable with zero mean and variance σ_a^2 which represents the uncertainty on the future evolution of degradation (source of uncertainty *a* in the itemized list above).

In practice, the exact value of the current degradation state θ_J as well as the true distribution of the equipment RUL are not know, in general, and have to be estimated from the available information. Thus, the objective of applying prognostics to equipment r, having available the information set $\Im$ is to estimate the PDF $p_{RUL_J}(rul \mid \Im)$.

Let us assume to build a prognostic model f that, given in input the observations $\mathbf{z}_{1:J}$, generates at time t_J the prediction $\hat{rul}_J$, i.e. $\hat{rul}_J = f(\mathbf{z}_{1:J} \mid \Im)$, based on the information set $\Im$. The realization rul_J of the random variable RUL_J can be re-written as:

$$rul_J = \hat{rul}_J + e_J \tag{6}$$

where $e_J = \hat{rul}_J - rul_J$ is a random variable with zero mean and variance $Var[e_J \mid \Im]$ representing the error made by model f in predicting the equipment RUL. Then, in practice, since it is often hard to accurately estimate the PDF $p_{RUL_J}(rul \mid \Im)$, the prognostic information can be well summarized by:

- the expected value $E[RUL_J \mid \Im]$ of RUL_J conditioned on the set of available information $\Im$, an estimate of which, hereafter referred to as the RUL prediction, is provided by the prognostic model $\hat{rul}_J = f(\mathbf{z}_{1:J} \mid \Im)$;

- the variance of the RUL prediction error $\sigma_{e_J}^2 = Var[e_J \mid \Im]$, representing a measure of the accuracy with which the estimate $\hat{rul}_J$ predicts the actual RUL value.

The variance $\sigma_{e_J}^2$ includes in addition to the process uncertainty (source of uncertainty *a*), the model uncertainty (source of uncertainty *b*) and the observation uncertainties (sources of uncertainty *c*). The model uncertainty can be due to modeling errors already included in the available information $\Im$ or introduced while building the prognostic model that generates the RUL prediction from the available data; the observation uncertainty is due to measurement noises and other sources of error affecting the observations $\mathbf{z}_j$, j=1,2,…, r=1,2,… available. In Paper III of Part II it is shown that, under some independency assumption, the prediction error variance $\sigma_{e_J}^2$, can be decomposed into three components representing the three sources of uncertainty *a*, *b*, and *c*:

$$\sigma_{e_J}^2 = E[(\hat{rul}_J - rul_J)^2] = \sigma_a^2 + \sigma_b^2 + \sigma_c^2 \tag{7}$$

In practice, when using a different framework than the probability theory to represent uncertain quantities, e.g., the belief function theory presented in Section 4.2.1 below, it may not be possible to produce an estimate of $\sigma_{e_J}^2$. Then, our goal in performing prognostics, has been, more generally, to provide the estimate $\hat{rul}_J$ of the expected value $E[RUL_J \mid \Im]$ of RUL_J, and a credible prediction interval $CI_J(\alpha) = [rul_J^{\inf}(\alpha), rul_J^{\sup}(\alpha)]$ to which the value of RUL_J can be assigned with a confidence of at least α. Within a probabilistic framework, the prediction interval can easily be derived from the estimated prediction error variance $\sigma_{e_J}^2$ under the hypothesis that the prediction error has a Gaussian distribution.

A more detailed discussion about uncertainty treatment in prognostics can be found in Paper III of Part II.

4 Modeling approaches for RUL prediction

This Section illustrates the different modeling approaches undertaken during this PhD work to cope with the three prognostic cases outlined in Section 3.1 and Figure 1. Each approach is described with reference to a well defined situation of information available; this provides a structured procedure for guiding the choice of the prognostic approach to be adopted depending on the information available in the specific situation encountered.

However, in real industrial applications, it is common to face hybrid situations characterized by the availability, at the same time, of more sources of information than those considered for each particular case of Figure 1. In such situation, it is possible to apply multiple prognostic approaches among those presented, each of which makes use of different pieces of available information to provide the RUL prediction. In such situations, the identification of the correct prognostic approach to be applied may not be trivial, also because some sources of information can be partially inaccurate or affected by large uncertainty. In this context, one can chose between two main strategies: (i) identifying the best performing approach or (ii) properly combining the outcomes of the different approaches. Concerning the first strategy (i), an analysis of the sensitivity of the performance of different prognostic approaches to the quality and quantity of the information available has been carried out in the context of the linear creep growth case study presented in Section 5.1, by considering a case in which all the sources of information listed in Table 4 are available at the same time, so that all types of prognostic approaches discussed can be applied. Concerning the second strategy (ii), many aggregation techniques able to correctly weight the different outcomes of multiple models have been proposed in literature [34], some of which have been adopted and further developed also during this PhD (Sections 2.4 and 4.3.1); however, these strategies do not consider, in general, the problem of incorporating different types of prediction uncertainty measures. A solution to this problem has been proposed during this work, for combining predictions whose uncertainty is represented in the two different frameworks of probabilistic and evidential reasoning (Section 4.2.4).

4.1 Model-based prognostics (Case 1)

In the context of available information shaped by case 1, a model based approach is in general a good choice if one is confident about the accuracy of the available model of the degradation process. In this case, at time t_J, the current degradation state θ_J is not directly known, but some information about its value is conveyed by the stochastic dynamic model of the degradation process in eq. (3) and the observation equation in eq. (4), the sequence of J observations $\mathbf{z}_{1:J}$ related to the equipment degradation state, which are assumed to be known. Thus, instead of estimating $p_{RUL_J}(rul|\theta_J)$ we are forced to restrict our objective to estimating the probability density function (PDF) $p_{RUL_J}(rul|\mathbf{z}_{1:J})$, conditioned on the observations $\mathbf{z}_{1:J}$ [82-84].

In this setting, it is desired to infer the unknown PDF $p_{\Theta_j}(\theta|\mathbf{z}_{1:J})$ of the degradation Θ_j of the test equipment at the future times $t_j > t_J$ on the basis of all the observations $\mathbf{z}_{1:J}$. The RUL cumulative probability distribution $P_{RUL_J}(rul|\mathbf{z}_{1:J})$ is then computed from $p_{\Theta_j}(\theta|\mathbf{z}_{1:J})$ as the probability that the failure threshold θ_{th}, whose value is also known, is exceeded before time $t_J + rul$. A Bayesian approach is typically used to estimate the filtered posterior PDF $p_{\Theta_j}(\theta|\mathbf{z}_{1:J})$ by a recursive computational procedure divided into successive prediction and update stages [18,22]. Unfortunately, except for a few cases, including

linear Gaussian state space models (Kalman filter) and hidden finite-state space Markov chains (Wohnam filter), it is not possible to evaluate analytically the filtered posterior PDF $p_{\Theta_J}(\theta \mid \mathbf{z}_{1:J})$, since it would require the evaluation of complex high-dimensional integrals. For this reason, approximation methods are necessary [17-18]. Due to its flexibility and ease of design, a numerical approximation of the Bayesian estimate based on the Monte Carlo sampling technique and called particle filtering, has been set-up in this work to predict the distribution of the system RUL and online-update it when new observations are collected [18,22,82]. Here, only a synthetic description of particle filtering is given; for a deeper discussion about the Bayesian approach to prognostic, and for the analytical details of the PF method the reader is refer to Paper II of Part II.

4.1.1 Particle Filtering

The Sampling Importance Resampling (SIR) version of PF has been adopted in this work. The proposed approach improves the one previously proposed [16] by taking into account the uncertainty on the parameters of the model of the degradation process and addressing the particle degeneration problem by means of the resampling algorithm.

The SIR PF method is based on sampling a large number K of trajectories $\{\theta_{0:F}^k\}_{k=1}^K$ (called particles), by recursively sampling the state θ_j^k from the transition PDF $p_{\Theta_j}(\theta_{j+1}^k \mid \theta_j^k)$, which can be derived from the physical model in eq. (3), until the failure threshold θ_{th} is exceeded and the life duration t_F^k of the k–th particle is recorded. The value rul_J^k of the particle RUL at time step t_J can then be computed from eq. (2). When an observation $\mathbf{z}_J$ is collected for the test trajectory, each particle is assigned a weight w_J^k proportional to the likelihood of its degradation state d_J^k at the time t_J, i.e., the probability $p_{\mathbf{z}}(\mathbf{z}_J \mid d_J^k)$ of observing $\mathbf{z}_J$ given that the equipment degradation is d_J^k [16]. The PDF $P_{RUL_J}(rul \mid \mathbf{z}_{1:J})$ is then approximated by an histogram of the K weighted values rul_J^k, $k = 1,...,K$, of the particle RULs at time t_J; the weighted average and the weighted standard deviation of the rul_J^k represent the prediction $\hat{rul}_J$ of the expected value $E[RUL_J \mid \mathbf{z}_{1:J}]$ of RUL_J and the estimate $\hat{\sigma}_{e_J}^2$ of the prediction error variance $\sigma_{e_J}^2$, respectively. The prediction interval $CI_J(\alpha) = [rul_J^{\inf}(\alpha), rul_J^{\sup}(\alpha)]$, i.e., the interval expected to contain the true RUL value rul_J with a probability of α, is obtained by setting $rul_J^{\inf}(\alpha)$ and $rul_J^{\sup}(\alpha)$ equal to the $(1-\alpha)/2$ and $(1+\alpha)/2$ percentiles, respectively, of the RUL distribution estimated by the PF.

Unfortunately, the procedure illustrated suffers from the so called degeneracy phenomenon: after few samplings, the weight variance increases and most of the *K* particle weights become negligible so that the corresponding trajectories do not contribute to the estimate of the PDF of interest [18,22]. As a result, the approximation of the target distribution $P_{RUL_J}(rul \mid \mathbf{z}_{1:J})$ becomes very poor and significant computational resources are spent trying to update particles with minimum relevance. The solution implemented to overcome this problem is the bootstrap resampling algorithm [17]. When degeneracy occurs, e.g. after few iterations of the weight updating procedure, *K* samples are drawn with replacement from the swarm of *K* particles; the k-th particle is sampled with a probability proportional to its weight value w_J^k and the sequence of degradation state $\theta_{1:J}^k$ until time t_J is retained for the resampled particle k' and recursively augmented with degradation states $\theta_{j+1}^{k'}$ sampled one more time from the transition PDF. The *K* resampled particles are then assigned the same weight $1/K$. Then, the filtering procedure continues with the original trajectories $\theta_{1:F}^k$ and the associated weights w_J^k replaced by new trajectories $\theta_{1:F}^{k'}$ with weights $w_J^{k'} = 1/K$.

As for the uncertainty in the RUL prediction, in this approach the randomness of the degradation process (source of uncertainty A) is described by the model, whereas the observation equation accounts for the observation noise (source of uncertainty C). Thus, these two causes of uncertainty are accounted for in the RUL prediction through the procedure of particle sampling and weights updating, respectively. On the contrary, the contribution of model uncertainty (source of uncertainty B) to the RUL prediction uncertainty is not directly considered, since it is assumed that the degradation dynamics model and the observation equation are exactly known. The effects of this uncertainty on the RUL prediction will be further discussed in Section 5.1.2. Notice, however, that if the uncertainty on the model parameters can be quantified and a probability distribution assigned to the value of the uncertain model parameter, the PF approach can be adjusted to handle also this source of uncertainty [83].

4.2 Data-driven prognostics based on current and historical degradation trajectories (Case 2)

In case 2, the information $\Im$ available at time t_i includes the dataset $\mathbf{D} = \{\mathbf{z}^r_{1:n^r}\}_{r=1}^{R}$ of the degradation-related observations collected for R training trajectories, and the sequence of observations $\mathbf{z}^{test}_{1:J}$ related to the degradation state of the test equipment. In this context of available information, we may not be able to estimate the probability distribution $p_{RUL_J}(rul \mid \mathbf{D})$ of RUL_J. In practice, our objective is limited to providing the estimate $\hat{rul}_J = f(\mathbf{z}^{test}_{1:J} \mid \mathbf{D})$ of the RUL expected value $E[RUL^r_J \mid \mathbf{D}, \mathbf{z}^{test}_{1:J}]$, based on the historical trajectories observed, and a credible prediction interval $CI_J(\alpha) = [rul^{\inf}_J(\alpha), rul^{\sup}_J(\alpha)]$ to which the value of RUL^r_J can be assigned with a confidence of, at least, α.

In case 2A, the relation between a sequence of observations $\mathbf{z}_{j:j'}$ and the equipment RUL at time $t_{j'}$, can be directly learned from the available data by artificial intelligence techniques, e.g., artificial neural networks [30], kernel methods [66], etc. This is possible because the equipment RUL, $rul^r_{j'}$, can be derived in correspondence of any sequence of observation $\mathbf{z}^r_{j:j'}$ from the knowledge of the equipment life durations t^r_F through eq. (2). In Section 4.2.1, a similarity-based regression (SBR) method, using a representation of the uncertain variable RUL^{test}_J based on belief function theory (BFT), is presented to derive an empirical model that receives in input a subset of the test sequence of observations $\mathbf{z}^{test}_{1:J}$, and produces as output the RUL prediction $\hat{rul}^{test}_J$ and the prediction interval $CI^{test}_J(\alpha) = [rul^{\inf}_J(\alpha), rul^{\sup}_J(\alpha)]$.

In case 2B, since the relation between the observation and the degradation state of the equipment is known, one can derive an indication $z^r_{\theta,j}$ about the degradation state θ^r_j of the *r*-th equipment at any time t^r_j. Then, a degradation-based approach can be applied in this case. The idea, is to use regression methods, e.g., linear and non-linear regression models, general degradation path models [7,36], etc., to build a model $\hat{\theta}_j = \eta(t \mid \mathbf{D}, \mathbf{z}^{test}_{1:J})$ of the evolution in time of equipment degradation. The prediction $\hat{rul}_J$ of the equipment RUL at time t_J is then obtained from the relation $\eta(t_J + \hat{rul}_J) = \theta_{th}$. A flexible method based on Gaussian process regression (GPR) and presented in section 4.2.2, has been developed in this thesis work to model non-linear degradation processes. The RUL prediction $\hat{rul}^{test}_J$ and the prediction interval $CI^{test}_J(\alpha) = [rul^{\inf}_J(\alpha), rul^{\sup}_J(\alpha)]$ are based on a probabilistic representation of the uncertain variables Θ^{test}_J and RUL^{test}_J.

Finally, a strategy for aggregating the prediction provided by the SBR and the GPR approaches with their different representation of the uncertain variable RUL_J^{test}, has been developed to provide more accurate and robust prediction in a case where both methods can be applied, since all sources of information considered for sub-cases A and B are available at the same time. This strategy, presented in Section 4.2.4, resorts to the definition of the BFT on the continuous real axis $\Re$ [67] to allow transferring the RUL_J^{test} PDF provided by the GPR approach in the BFT framework in which the SBR prediction is represented.

4.2.1 Similarity-based prognostics

In this Section, we present a direct RUL prediction approach to prognostic which uses the set of training trajectories $\{\mathbf{z}^r_{1:n^r}\}_{r=1}^R$ to perform a data-driven analysis of similarity with the test trajectory $\mathbf{z}^{test}_{1:J}$ and provide the prediction $\hat{rul}_J^{test}$ of the test equipment RUL [12,29,40,88-89], and the prediction interval $CI_J^{test}(\alpha) = [rul_J^{\inf}(\alpha),\ rul_J^{\sup}(\alpha)]$ [88-89]. In this Section, only a brief overview of the proposed prognostic method is given. For the analytical details of the procedure, the author is referred to Paper IV of Part II.

The first requirement of the approach is to define a measure to evaluate the similarity between trajectories. This is done considering the pointwise difference d_j^r between *n*-long sequences of observations $\mathbf{z}^r_{j-n+1:j}$ of the training trajectories and the sequence of the *n* latest observations $\mathbf{z}^{test}_{J-n+1:J}$ of the test trajectory. The similarity s_j^r of the training trajectory segment $\mathbf{z}^r_{j-n+1:j}$ to the test trajectory is defined as bell-shaped function of the distance measure d_j^r:

$$s_j^r = \exp\left(-\frac{(d_j^r)^2}{\lambda}\right) \tag{8}$$

The arbitrary parameter λ can be set by the analyst to shape the desired interpretation of similarity: the smaller is the value of λ the stronger the definition of similarity. For the prediction of the test equipment RUL, a RUL prediction $\hat{rul}^r_{j^*}$ is assigned to each training trajectory $r = 1:R$, by considering the difference between the trajectory failure time t_F^r and the last time instant $t^r_{j^*}$ of the trajectory segment $\mathbf{z}^r_{j^*-n+1:j^*}$ which has the maximum similarity $s^r_{j^*}$ with the test trajectory:

$$\hat{rul}^r_{j^*} = t_F^r - t^r_{j^*} \tag{9}$$

Then, the prediction $\hat{rul}_J^{SB}$ of the test equipment RUL at time t_J^{test} is given by the similarity weighted sum of the predictions $\hat{rul}^r_{j^*}$:

$$\hat{rul}_J^{SB} = \frac{\sum_{r=1}^{R} s^r_{j^*} \hat{rul}^r_{j^*}}{\sum_{r=1}^{R} s^r_{j^*}} \tag{10}$$

As for the estimate of the prediction uncertainty, the distribution of the RUL predictions $r\hat{u}l^r_{j*}$ associated to the different training trajectories provides a measure of the prediction uncertainty. However, our knowledge about the actual equipment RUL is limited and partial, e.g., because the number of historical trajectories available is small, and the predictions $r\hat{u}l^r_{j*}$ are the result of the arbitrary, although reasonable, measure of similarity adopted rather than the outcome of a real stochastic process. Then, in this work, we have considered more appropriate to resort to belief function theory (also called Dempster-Shafer or evidence theory [68-69]), rather than to probability theory, for representing the uncertain variable RUL_J based on the available information.

The similarity measure s^r_j defined in eq. (8) is interpreted as a measure about the relevance of the source of information represented by trajectory r. Then, a basic belief assignment (BBA) $m^r_{RUL^{test}_J}$ is associated to its RUL prediction as follows: [40]:

$$\begin{aligned} m^r_{RUL^{test}_J}(\{r\hat{u}l^r_{j*}\}) &= \gamma \cdot s^r_{j*} \\ m^r_{RUL^{test}_J}(\Omega_{RUL^{test}_J}) &= 1-\gamma \cdot s^r_{j*} \end{aligned} \tag{11}$$

where $\Omega_{RUL^{test}_J} = [0, t^{\max}_F - t^{test}_J]$ is the frame of discernment, i.e., the domain of RUL^{test}_J, and the arbitrary parameter $\gamma \in [0,1]$ defines the degree of trust given to the reference trajectories: if $\gamma < 1$ a part of belief will always be assigned to the entire RUL domain Ω_{RUL_J}, even in the case a reference trajectory were exactly identical to the test one. Notice that the mass assigned to Ω_{RUL_J} represents our *ignorance* about RUL^{test}_J, since it is the amount of evidence about its value that cannot be assigned to any subset of the RUL domain.

The distinct sources of information inducing the BBAs, $m^r_{RUL^{test}_J}$ can be combined according to the Dempster's rule of combination [40,70] to obtain the BBA $m_{RUL^{test}_J}$. The information conveyed by this BBA can be represented by the belief $\mathrm{Bel}_m(CI_J)$ and the plausibility $\mathrm{Pl}_m(CI_J)$ of the hypothesis $RUL^{test}_J \in CI_J = [rul^{\inf}_J, rul^{\sup}_J]$ defined, respectively, as

$$\mathrm{Bel}_m(CI_J) = \sum_{CI'_J \subseteq CI_J} m_{RUL_J}(CI'_J) \tag{12}$$

and

$$\mathrm{Pl}_m(CI_J) = \sum_{CI'_J \cap CI_J \neq \varnothing} m_{RUL_J}(CI'_J) \tag{13}$$

The belief associated to an interval CI_J represents the amount of belief that directly supports the hypothesis $RUL^{test}_J \in [rul^{\inf}_J, rul^{\sup}_J]$, whereas the plausibility represent the maximum belief that could be committed to this hypothesis if further information became available. Then, belief and plausibility can be seen as lower and upper bounds on the probability that the hypothesis $RUL^{test}_J \in [rul^{\inf}_J, rul^{\sup}_J]$ is true. In this work, the left bounded intervals $CI_J(\alpha) = [rul^{\inf}_J(\alpha), +\infty]$ to which a predefined belief α is assigned, have been considered as a prediction interval for the RUL value, since the probability of $RUL^{test}_J \in CI_J(\alpha)$ is larger

than α, and thus the probability of a failure happening before $rul_J^{\inf}$ is lower than $1-\alpha$. This way the prediction interval $CI_J(\alpha)$ provides to a maintenance planner able to specify a maximum acceptable failure probability of $1-\alpha$ the latest time $rul_J^{\inf}(\alpha)$ at which, according to the available information, a probability to have a failure lower than $1-\alpha$ is guaranteed.

Notice that the prediction interval thus obtained accounts for all three sources of uncertainty listed in Section 3.2. Indeed, the evidence conveyed by the predictions $\hat{rul}_{j*}^r$ is related to the process uncertainty, since each trajectory represents a different degradation process, and to the model uncertainty, since each prediction $\hat{rul}_{j*}^r$ can be interpreted as the outcome of a different similarity-based prognostic model; moreover, the discounting strategy accounts for the presence of noise on the observations, since noise reduces the similarity between trajectories and thus the belief assigned to their predictions, and, as a consequence, increases the amplitude of the prediction interval, i.e. the estimated prediction uncertainty.

4.2.2 Degradation-based: Gaussian Processes

When direct indications $z_{\theta,1:n^r}^r$ about the equipment degradation state are available for a number R of training trajectories, the most natural data-driven technique for RUL estimation is to model the equipment degradation evolution and compare the extrapolated future degradation trajectory to a failure criterion. In this Section, a prognostic approach that uses Gaussian process regression [41-42] to explicitly model the evolution of equipment degradation as a stochastic process based on the available sequences of degradation indications $z_{\theta,1:n^r}^r$ is proposed [86-87]. GPR is a powerful and flexible method for performing nonparametric probabilistic inference over functions [41] and can be effectively used for modeling degradation as a stochastic process [86-87]. To do that, it is necessary to assume that the distribution of the degradation states is Gaussian with different mean $E[\Theta(t_j)]$ and variance $Var[\Theta(t_j)]$ at every time instant t_j. This hypothesis is done in GPR to allow analytical calculations that make inference simpler. Although one cannot always prove that degradation states are normally distributed, in the absence of outstanding evidence to support a different assumption, we suggest preferring this one to others that would make the inference from data more difficult.

For mapping the function $\Theta(t)$ given the input t, the GPR defines a *prior* for it in the form of a distribution over functions specified by a Gaussian Process (GP). A GP is a collection of random variables any finite number of which has a joint Gaussian distribution. A real GP $\Theta(t)$ is completely specified by its mean function $\mu_\Theta(t) = E[\Theta(t)]$ and covariance function $C_\Theta(t,t') = E[(\Theta(t) - \mu_\Theta(t'))(\Theta(t') - \mu_\Theta(t'))]$. This *prior* is taken to represent our prior beliefs over the kind of functions we expect to observe. Typically the *prior* mean and covariance functions adopted will have some free parameters, called hyper-parameters. Although the choice of the covariance function must be specified by the user, various methods have been proposed for determining the corresponding hyper-parameters from training data [41]. In this work, the hyper-parameters have been optimized by maximizing the marginal likelihood of the dataset set of input/output pair $\mathbf{D}_{t/z}^{train} = \{(t_j^r; z_{\theta,j}^r)_{j=1:n^r;r=1:R}\}$ drawn from the training trajectories.

Given the prior information about the GP, the set of hyper-parameters, and the observation dataset $\mathbf{D}_{t/z}^{+}$, including datasets $\mathbf{D}_{t/z}^{train}$ and $\mathbf{D}_{t/z}^{test} = \{(t_j^{test}; z_{\theta,j}^{test})_{j=1:J}\}$, drawn, respectively, from the training and test trajectories, GPR provides a method for predicting the conditional probability density function (PDF) $p_{\Theta_j^{test}}(\theta \mid \mathbf{D}_{t/z}^{+})$ of the future degradation states Θ_j^{test}, $t_j^{test} > t_J^{test}$ of the test trajectory. For this, a restriction

on the *prior* distribution to contain only those functions that agree with the observed data is imposed; the resulting posterior distribution of the output $\Theta(t_j^{test} \mid \mathbf{D}_{t/z}^{+})$ in correspondence of the input t_j^{test} is Gaussian with mean $E[\Theta(t_j^{test}) \mid \mathbf{D}_{t/z}^{+}]$ and variance $Var[\Theta(t_j^{test}) \mid \mathbf{D}_{t/z}^{+}]$ given by [41]:

$$\begin{aligned} E[\Theta(t_j^{test}) \mid \mathbf{D}_{t/z}^{+}] &= \mu_{\Theta}(t_j) + \mathbf{C}_{\Theta}(\mathbf{t}^{train}, t_j^{test})[\mathbf{C}_{\Theta}(\mathbf{t}^{train}, \mathbf{t}^{train})]^{-1}(\mathbf{z}^{train} - \mu_{\Theta}(\mathbf{t}^{train})) \\ Var[\Theta(t_j^{test}) \mid \mathbf{D}_{t/z}^{+}] &= C_{\Theta}(t_j^{test}, t_j^{test}) - \mathbf{C}_{\Theta}(t_j^{test}, \mathbf{t}^{train})[\mathbf{C}_{\Theta}(\mathbf{t}^{train}, \mathbf{t}^{train})]^{-1}\mathbf{C}_{\Theta}(\mathbf{t}^{train}, t_j^{test}) \end{aligned} \tag{14}$$

where $\mathbf{t}^{train}$ and $\mathbf{z}^{train}$ are the vectors of all the inputs and, respectively, outputs in $\mathbf{D}_{\tau/z}^{+}$ and $\mathbf{C}_{\Theta}$ indicates the covariance matrix containing the values of $C_{\Theta}(t_j^r, t_{j'}^{r'})$ evaluated for all possible pairs of inputs in $\mathbf{t}^{train}$.

Since the dataset $\mathbf{D}_{t/z}^{+}$, used for conditioning the *prior* GP, is drawn partly from the reference equipment and partly from the test equipment, it is possible to make the GPR capable of learning both the structure underlying the degradation processes which is common to all similar equipment, and the specific variation around this structure that characterizes the degradation of the test equipment. This result has been obtained by using a covariance function of the following form [71,86-87]:

$$C_{\Theta}(t_j^r, t_{j'}^{r'}) = C_{\Theta}^{1}(t_j^r, t_{j'}^{r'}) + C_{\Theta}^{\backslash 2}(t_j^r, t_{j'}^{r'})\delta(r, r') + \sigma_z^2\delta(r, r')\delta(j, j') \tag{15}$$

where σ_z^2 is the variance of the white Gaussian noise affecting the degradation indications $z_{\delta,j}^r$ and the reference index assigned to the test trajectory is $r = R+1$. The first term of the kernel corresponds to the covariance associated to the common structure underlying all degradation trajectories; the second represents the covariance owing to the variation of each trajectory around the common structure of all degradation trajectories. This term assumes a finite value only when the inputs t_j^r and $t_{j'}^{r'}$ refers to the same trajectory, since we have assumed the variation specific to each trajectory to be uncorrelated across trajectories. Finally, the third term accounts for the observation noise associated to the indications $z_{\delta,j}^r$ of the degradation state δ_j^r.

Given the value of the failure threshold, assumed here to be known, and the conditional distribution of the degradation state $p_{\Theta_j^{test}}(\theta \mid \mathbf{D}_{t/z}^{+})$, the RUL cumulative distribution function (CDF) $P_{RUL_J^{test}}(rul \mid \mathbf{D}_{t/z}^{+})$ is computed as the probability that the degradation Θ_j^{test} at the future time $t_j^{test} = \tau_J^{test} + rul$ will exceed the failure threshold θ_{th} (Paper V, Part II). The prediction $\hat{rul}_J$ is then given by the mean of the distribution described by $P_{RUL_J^{test}}(rul \mid \mathbf{D}_{t/z}^{+})$ and the prediction bounds $rul_J^{\inf}(\alpha)$ and $rul_J^{\sup}(\alpha)$, by, respectively, its $(1-\alpha)/2$ and $(1+\alpha)/2$ percentiles; finally, its variance can be taken as an estimate of the prediction error variance $\sigma_{e_J}^2$.

As for the different sources of uncertainty, the process and noise uncertainties, learned from the training trajectories, are represented by the degradation state variance $Var[\Theta(t_j^{test}) \mid \mathbf{D}_{t/z}^{+}]$ and projected to the prediction of the RUL CDF. The model uncertainty is related, instead, to the choice of the mean and covariance functions of the *prior* GP and the optimization of their hyper-parameters. Having used a single *prior* GP for the regression, we doubt that the resulting prediction interval $CI_J(\alpha)$ can capture the modeling uncertainty. For a more comprehensive treatment of the prognostic uncertainty, an ensemble approach can be adopted (see, for example, Section 4.3.2 of this thesis or reference [28]) which allows accounting for model

uncertainty, by considering the distribution of the outcomes of different models trained on random replicates of the training dataset or under different *prior* assumptions.

4.2.3 Aggregated degradation-based and direct RUL

Degradation–based prognostics (Section 4.2.2) provides a more informative and transparent output than direct RUL prediction (Section 4.2.1), since it supplies a prediction not only of the current equipment RUL, but of the entire degradation trajectory the equipment will follow which can be checked against, e.g., expert intuition to verify its consistency. Moreover, contrarily to direct RUL prediction, this approach can be applied even when no historical failure data are available. On the other side, direct RUL prediction does not require the identification of a degradation indicator and of a failure threshold which may introduce further uncertainty and sources of errors. Thus, when both degradation-based and direct RUL prediction approaches can be applied, it is hard to choose between them. For this reason, we suggest to resort to an ensemble approach to develop both prognostic models and aggregate their outcomes so to removes the necessity of choosing the best performing one and potentially increasing the robustness and accuracy of the prediction.

In this Section, we consider this ensemble approach for combining the SBR and the GPR models presented in Sections 4.2.1 and 4.2.2. The main obstacle in the aggregation of the RUL predictions provided by the SBR and the GPR approaches is the combination of their two different representations of the uncertain variable RUL_J. In this context, an important contribution of this thesis work has been the development of a technique for aggregating uncertainty measures generated in the two different frameworks of evidential and probabilistic reasoning. For this, the definition of the BFT on the continuous real axis $\Re$ has been used [67], where masses generalizes into densities so to allow representing also probability density functions.

To extend the belief function theory to real numbers, it is assumed that masses are only allocated to closed intervals $[\underline{rul}, \overline{rul}]$. A convenient graphical representation of these intervals is the half-plane highlighted in grey in Figure 5, which contains all the sets of pairs $(\underline{rul}, \overline{rul}) \in \Re^2 : 0 \le \underline{rul} \le \overline{rul}$ representing the interval $[\underline{rul}, \overline{rul}]$.

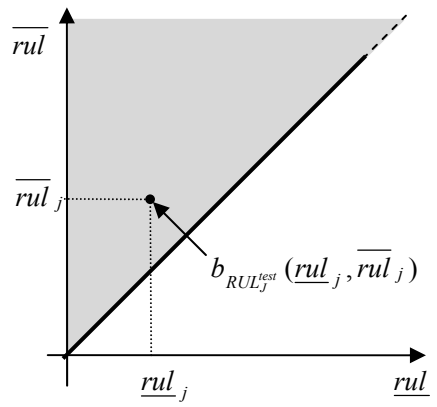

Figure 5: graphical representation of intervals $[\underline{rul}, \overline{rul}]$.

A belief density function (BDF) $b_{RUL_J^{test}}(\underline{rul}, \overline{rul})$ is defined on this half-plane; the total belief $\mathrm{Bel}_b([\underline{rul}_j, \overline{rul}_j])$ assigned by the BDF b to an interval is the integral of $b_{RUL_J^{test}}$ over the triangle $(\underline{rul}, \overline{rul}) \in \Re^2 : \underline{rul} \geq \underline{rul}_j, \overline{rul} \leq \overline{rul}_j$ highlighted in grey in Figure 6 (left), whereas its plausibility $\mathrm{Pl}_b([\underline{rul}_j, \overline{rul}_j])$ is the integral of $b_{RUL_J^{test}}$ over the half-plane $(\underline{rul}, \overline{rul}) \in \Re^2 : 0 \leq \underline{rul} \leq \overline{rul}_j$, $\overline{rul} \geq \underline{rul}_j$ highlighted in grey in Figure 6 (right).

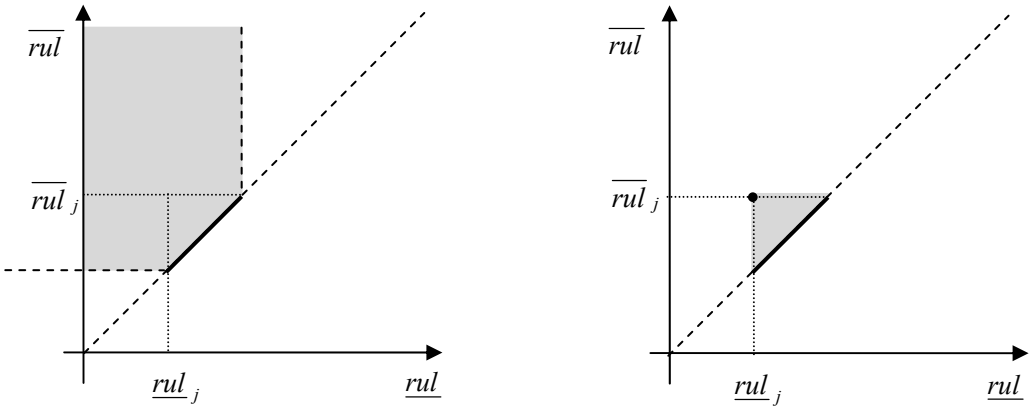

Figure 6: graphical representation of the belief (left) and plausibility (right) associated to the interval $[\underline{rul}_j, \overline{rul}_j]$.

Within the framework of continuous BFT, the BBA $m_{RUL_J^{test}}(\{\hat{rul}_{j*}^r\}_{r=1}^R, \Omega_{RUL_J^{test}})$ defined in the SBR approach can be represented by the BDF $b_{RUL_J^{test}}^{SBR}$ assigning the mass $m_{RUL_J^{test}}(\Omega_{RUL_J^{test}})$ to the interval $[0, rul_J^{\max}]$ and the masses $m_{RUL_J^{test}}(\{\hat{rul}_{j*}^r\})$ to the degenerated intervals of null dimension $[\hat{rul}_{j*}^r, \hat{rul}_{j*}^r]$, j=1:R, lying on the boundary $\underline{rul} = \overline{rul}$ of the half-plane of all possible RUL intervals. Concerning the GPR prediction, the straightforward interpretation of a PDF within the framework of continuous BFT would assign finite density $b_{RUL_J^{test}}([rul, rul]) = p_{RUL_J^{test}}(rul | \mathbf{D}_{\tau/z}^+)$ only to the points on the boundary $\underline{rul} = \overline{rul} = rul$. However, since our knowledge of the equipment RUL is partial and based on uncertain assumptions, a better strategy for representing the GPR prediction in the BFT framework, is to consider it as a pignistic PDF , i.e., the probability that a rational person will assign to an interval $[\underline{rul}, \overline{rul}]$ based on the evidence induced on $\Re$ by the underlying BDF $b_{RUL_J^{test}}$ which value is unknown [70]. According to the pignistic transformation [70], many BDF $b_{RUL_J^{test}}$ can induce the same pignistic probability $Betb$. In this work, we evoke the principle of minimal commitment, that formalizes the idea that one should never give more support than justified to any subset of $\Omega_{RUL_J^{test}}$, for selecting the least committed belief function in a set of equally justified belief functions induced by the pignistic PDF $Betb(rul) = p_{RUL_J^{test}}(rul | \mathbf{D}_{\tau/z}^+)$. This way, we can also relax the hypothesis underlying the RUL PDF predicted by the GPR approach, which is based on the uncertain assumption of normally distributed degradation states and on a rather arbitrary choice of the *prior* mean and covariance functions [72]. A method for computing the least committed BDF $b_{RUL_J^{test}}^{GPR}$ is given in Paper VI of Part II for unimodal pignistic PDFs.

Having represented in the framework of continuous BFT both the BBA resulting from SBR and the PDF generated by GPR, the outcomes of the two approaches can be combined using the Dempster's rule to obtain the combine BDF $b_{RUL_J^{test}}$. The analytical expression of $b_{RUL_J^{test}}^{SBR}$, $b_{RUL_J^{test}}^{GPR}$ and $b_{RUL_J^{test}}$ are given in Paper VI of

Part II. Given the combined BDF $b_{RUL_J^{test}}(\underline{rul}, \overline{rul})$, a credible left-bounded interval $CI_J(\alpha) = [rul_J^{\inf}, +\infty]$ for the value of the RUL can be estimated as for the SBR approach, by considering the interval to which the BDF $b_{RUL_J^{test}}$ assigns the desired belief α.

4.3 Data-driven prognostics based on current degradation trajectory only (Case 3)

This Section tackles the situation of available information described in case 3, which is characterized by the availability of only the sequence of degradation-related observations $\mathbf{z}_{1:J}$ collected during the degradation of the equipment of interest. Also, it is assumed that the value of the failure threshold and the observation equation are known, so that it is possible to define a failure criterion and to derive from the observations the indication $z_{\delta,1:J}$ about the degradation state of the equipment at the different observation time instants $t_{1:J}$. Then, a degradation-based approach can be applied in this case; on the contrary, direct RUL prediction is unfeasible since no historical life duration data are available.

As already mentioned for case 2, empirical modeling of the degradation process can be achieved by training the most suited parametric or non-parametric degradation model, e.g., linear and non-linear regression models, general degradation path models, Gaussian processes etc. [7,36] with the available data. However, in this case, model $\hat{\theta}_j = \eta(t \mid z_{\theta,1:J})$ has to be build based on a single censored sequence of degradation indications $z_{\theta,1:J}$. Then, the number of observation available is, in general, much smaller and the observations available describe only partially the degradation process, since the equipment of interest has not yet failed. In this context of very limited information available, it has been of primary importance to develop robust prognostic approaches. In this work, the ensembles technique has been exploited to satisfy such requirement. Indeed, an ensemble of prognostic models can improve the reliability of the single model since the weaknesses of the individual models are compensated by the other models in the ensemble [34]. Also, by measuring the performance of the individual models and giving more importance to the predictions of the best performing ones, more accurate RUL predictions can be obtained [35]. Moreover, an ensemble can naturally accommodate new information by adding new models tailored to it, without discarding the old models in the ensemble. Updating the overall ensemble model in this incremental way requires reduced computational efforts in comparison to re-building a model from scratch based on all available (old and new) data [73]. A procedure for ensemble-based incremental learning has been developed in this work and is presented in Paper I of Part II.

The implementation of an ensemble of prognostic models requires (a) building the diverse models and (b) combining their RUL predictions. Concerning task (a), diversity among the individual models of the ensemble can be obtained by using different modeling techniques, differentiating the training parameters or the models structure, selecting different features of the training data, or using different datasets to train individual models [34]. Here, diversity between the models has been obtained by using bootstrapped replicates $\mathbf{D}^h$ of the available data set $\mathbf{D} = \{z_{\theta,1:J}\}$ to train the different models $\hat{\theta}_j = \eta^h(t \mid \mathbf{D}^h)$ of the ensemble. The vital element of this technique is the instability of the prediction method: if perturbing the training set can cause significant changes in the predictor constructed, then a bootstrapped ensemble can improve accuracy [74]. During this thesis work an original technique for the generation of the replicated

training datasets, which further increases the diversity between the models of the ensemble, has been developed and is described in Paper I of Part II.

Concerning task (b), when compared to the failure threshold, the degradation estimated by each model provides a different prediction $\hat{rul}_J^h$ of the equipment RUL. Common techniques for the aggregation of multiple model outcomes are based on statistics, e.g., the simple mean and the median, or on models performance, e.g., the globally or locally weighted average [55]. In general, the aggregation requires associating a weight w^h to the prediction $\hat{rul}_J^h$ made by each ensemble model η^h and combining the predictions by means of a weighted average. In case of large noise affecting the degradation indications $z_{\theta,1:J}$, the traditional aggregation techniques may lead to noisy predictions of the equipment RUL since all the ensemble models are built considering bootstrapped replicates of the same noisy data. A possibility to improve the accuracy and robustness of the predictions $\hat{rul}_J$ in such cases is to properly filter them. In this respect, Section 4.3.1 proposes an effective strategy, inspired by the discrete KF [30, 81], for both aggregating the multiple model outcomes and filtering the ensemble predictions.

The limit of the KF-based ensemble is that it does not automatically provides a measure of the RUL prediction uncertainty. To overcome this limitation, a bootstrapped ensemble approach is used to quantify the prediction uncertainty based on two steps [35,82-85]: in a first step, the model uncertainty is estimated based on the distribution of the RUL predictions of an ensemble of bootstrap models; in the second step, the prediction and noise uncertainties are evaluated based on the RUL prediction error made by the ensemble of models when applied to a validation dataset. This second steps can be straightforward applied in a situation such as the one considered in case 2A, where input/output pairs $(z_{\theta,j}; rul_j)$ for which rul_j is known are available for the validation step, but is not suited for case 3, where no life duration data are available. An original solution for the application of the bootstrap ensemble approach to uncertainty estimate in the absence of life duration data has been developed during this thesis and is presented in Section 4.2.2.

4.3.1 Kalman Filter ensemble

Let us consider an ensemble of H degradation models $\eta^h(t|\mathbf{D}^h)$ trained on an equivalent number of bootstrapped replicates $\mathbf{D}^h$ of the available data set $\mathbf{D}$. At the current time step t_J each model provides a different RUL prediction $\hat{rul}_J^h$.

In general, a Kalman Filter [75] is a recursive method which uses a state dynamics model, a measurement equation and a set of observations to estimate the state of the system at time t. In this work, a methods for its application to the aggregation of the outcomes $\hat{rul}_J^h$, $h = 1:H$ of an ensemble of degradation-based prognostic models η^h has been proposed. In this context, one aims to estimate the expected value $E[RUL(t_J)|\mathbf{z}_{1:J}]$ of the equipment RUL at time t_J. Then, the state is defined as a vector of two elements: the RUL expected value $E[RUL(t_J)|\mathbf{z}_{1:J}]$ and the rate of modification in time of the RUL, i.e., $\left[RUL(t_j) - RUL(t_j + \Delta t)\right]/\Delta t$, which is set constantly equal to -1, since the RUL of a component decreases of one time unit for every time unit passed. This way, assuming $\Delta t = 1$, the state evolution in time can be described through the linear dynamic model:

$$\begin{bmatrix} E[RUL(t_j)\,|\,\mathbf{z}_{1:j}] \\ -1 \end{bmatrix} = \begin{bmatrix} E[RUL(t_j-1)\,|\,\mathbf{z}_{1:j-1}] - 1 + v(t_j) \\ -1 \end{bmatrix} = A \begin{bmatrix} E[RUL(t_j-1)\,|\,\mathbf{z}_{1:j-1}] \\ -1 \end{bmatrix} + \begin{bmatrix} \upsilon(t_j) \\ 0 \end{bmatrix} \quad (16)$$

where $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is the transition matrix and $\upsilon(t_j)$ is an additive white Gaussian process noise with variance q; in practice, since the variance of the process noise is not known, q is set as an arbitrary constant which affects the smoothness of the resultant state estimate time series (lower values of q imply higher amount of smoothing). The process noise $\upsilon(t_j)$ is added to account for the fact that the mean value of the RUL distribution is affected by the stochasticity of the evolution of the degradation level between time $t_j - 1$ and t_j and, as a consequence, does not deterministically decrease of 1 time unit at every time step.

The *observation* vector of the KF is assumed to be formed by the RUL predictions $\hat{rul}_j^h$ of the H models in the ensemble. Then, the measurement equation is:

$$\begin{bmatrix} \hat{rul}_j^1 \\ \hat{rul}_j^2 \\ \vdots \\ \hat{rul}_j^H \end{bmatrix}^T = \begin{bmatrix} E[RUL(t_j)\,|\,\mathbf{z}_{1:j}] + e^1(t_j) \\ E[RUL(t_j)\,|\,\mathbf{z}_{1:j}] + e^2(t_j) \\ \vdots \\ E[RUL(t_j)\,|\,\mathbf{z}_{1:j}] + e^h(t_j) \end{bmatrix} = \mathbf{H} \begin{bmatrix} E[RUL(t_j)\,|\,\mathbf{z}_{1:j}] \\ -1 \end{bmatrix} + \mathbf{e}(t_j) \quad (17)$$

Where $\mathbf{H} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix}^T$ is the observation matrix and $\mathbf{e}(t)$ is the observation noise with covariance matrix $\mathbf{R}$. Notice that although the vector of the RUL predictions is called *observation* vector according to the KF terminology, in this case it is formed by the model outcomes which are not measured quantities. As a consequence, in this application the observation noise covariance matrix $\mathbf{R}$ should contain the prediction error covariance of the ensemble models. Since, however, the prediction error covariances $\sigma_{e_j}^2$ are unknown, assuming independence between the errors of the different models η^h, matrix $\mathbf{R}$ is taken diagonal and equal to:

$$\mathbf{R} = diag(mse_{1\to J}^1, ..., mse_{1\to J}^h, ..., mse_{1\to J}^H) \quad (18)$$

with the performance $mse_{1\to N_p}^h$ of model η^h computed by considering the mean square average over all the indications $z_{\theta,1:J}$ collected at time $t_j \le t_J$ of the relative difference between the time $\hat{t}_j^h$ at which the observed degradation state $z_{\theta,}$ is reached according to model η^h, and time t_j. Thus, the weight assigned to each model is proportional to its performance in estimating the degradation trajectory up to the last available measurement time t_J and matrix $\mathbf{R}$ can be updated each time a new measurement becomes available. Using eq.(18) for the estimation of $\mathbf{R}$ and having set the value of parameter q, the posterior prediction of RUL_J and the corresponding error covariance matrix are obtained through the traditional predict and an update phases of the Kalman Filter (see Paper I of Part II for the analytical details).

A merit of this approach is that it does not imply any constraint on the linearity of the degradation models $\eta^h(t\,|\,\mathbf{D}^h)$, since only the outputs of the predictive models are used in the KF scheme; thus, the non-linearity of the degradation process can be described by using non-linear degradation models, with no additional

complication for the KF-based aggregation procedure. Moreover, the approach removes the necessity of filtering the degradation measurements which is a challenging problem for non-linear degradation processes. The drawback of this approach is that the posterior error covariance matrix provided by the KF can no more be interpreted as an estimate of the prediction error covariance. Thus, the proposed KF approach which constitutes an effective technique for the aggregation of the outcomes of the ensemble models, does not provide an estimation of the uncertainty affecting the RUL.

4.3.2 Bootstrapped ensemble

In their basic form, many regression methods that can be used for building a degradation model $\hat{\theta}_j = \eta(t_j \mid \mathbf{D})$ provide in output a point prediction of the degradation states without any information on the uncertainty of the estimate from which only a point prediction $\hat{rul}_J = f(z_{\theta,J})$ of the RUL can be derived by comparison with the failure threshold. In this Section we propose a method for estimating the uncertainty of the RUL prediction provided by a generic degradation model η.

Under the hypothesis that the model $f(z_{\theta,J})$ is as an unbiased estimator of $E[RUL_J \mid \mathbf{z}_{1:j}]$, i.e., $E[f(z_{\theta,J})] = E[RUL_J \mid \mathbf{z}_{1:j}]$, an estimate of the model error variance, σ_B^2, is obtained from an ensemble of models $\eta^h(z_{\theta,J} \mid \mathbf{D}^h)$, $h = 1:H$ trained using bootstrapped replicates $\mathbf{D}^h$ of a training dataset $\mathbf{D}^{trn}$, drawn from $\mathbf{D}$. Given a generic input $z_{\theta,J}$, the models of the ensemble generate H different predictions $\hat{rul}_J^h$; their variance is taken as the estimate $\hat{\sigma}_B^2$ of the model error variance σ_B^2 [35,82-85], whereas their average is taken as the best prediction $\hat{rul}_J$ of the equipment RUL.

With respect to the estimate of the remaining part of the RUL prediction variance, which is caused by the randomness of the degradation process and the observation noise (sources of uncertainty A and C), i.e. $\sigma_{A+C}^2 = \sigma_A^2 + \sigma_C^2$, an independent validation dataset $\mathbf{D}^{val}$ has to be used. The ensemble of empirical models $\eta^h(t_J \mid z_{\theta,1:J})$ is applied to the observations $z_{\theta,j}^{val}$ in the validation dataset to obtain the RUL predictions $\hat{rul}_j^{val}$, which should be used to calculate, for each validation observation $z_{\theta,j}^{val}$, the prediction residuals ρ_j:

$$\rho_j^2 = (\hat{rul}_j^{val} - rul_j^{val})^2 - \hat{\sigma}_B^2(z_{\theta,j}^{val}) \tag{19}$$

The set of input/output pairs obtained by associating to all the observations $z_{\theta,j}^{val}$, in $\mathbf{D}^{val}$ the corresponding residuals ρ_j is used for training an empirical model $\hat{\sigma}_{A+C}^2(z_{\theta,j}) = \chi(z_{\theta,j})$ of the residual variance approximating the unknown relation between the input $z_{\theta,j}$ and the variance of the residuals [35].

In practice, given the set of available information described for case 3, the residuals ρ_j cannot be computed in correspondence of any of the observation $z_{\theta,j}$, since there are no available pairs $(z_{\theta,j}, rul_j)$ for which rul_j is known. To overcome this problem, we consider a model $\Delta\hat{t}_{j,j'} = \tilde{\eta}(\boldsymbol{\theta}_{j,j'})$, which receives in input a vector of two degradation states $\boldsymbol{\theta}_{j,j'} = [\theta_j, \theta_{j'}]$ and returns in output the estimate $\Delta\hat{t}_{j,j'}$ of the time interval needed to reach the degradation state θ_j starting from $\theta_{j'}$ [82]. In general, model $\tilde{\eta}(\boldsymbol{\theta}_{j,j'})$ can be derived from model $\eta(t_j)$. The prediction $\hat{rul}_j$ is obtained from this model by setting $\theta_j = z_{\theta,j}$ and $\theta_{j'} = \theta_{th}$; in this view, the RUL prediction at time t_j corresponds to the estimate of the time interval $\Delta t_{J,th}$ needed to increase the degradation state from the currently observed degradation state $z_{\theta,J}$ to the failure threshold θ_{th}.

The dataset of input/output pairs $\mathbf{D}_{z/\Delta t}$, associating to all couples of degradation indications $\mathbf{z}_{\boldsymbol{\theta}\,j,j'} = [z_{\theta,j}, z_{\theta,j'}]$ the time interval $\Delta t_{j,j'} = t_{j'} - t_j$ needed to obtain the degradation increment $z_{\theta,j'} - z_{\theta,j}$, is drawn from the available observations and partitioned into a training and a validation dataset, $\mathbf{D}^{trn}_{z/\Delta t}$ and $\mathbf{D}^{val}_{z/\Delta t}$: $\mathbf{D}^{trn}_{z/\Delta t}$ is used to train an ensemble of regression model $\tilde{\eta}^h(\mathbf{z}_{\boldsymbol{\theta}\,j,j'} | \mathbf{D}^{trn,h}_{z/\Delta t})$, whereas $\mathbf{D}^{val}_{z/\Delta t}$ is used to test the ensemble of models $\tilde{\eta}^h$, and collect examples of its prediction error $\Delta\hat{t}^{val}_{j,j'} - \Delta t^{val}_{j,j'}$. From it, one can compute the prediction residuals $\rho_{j,j'}$ associated to the input $\mathbf{z}_{\boldsymbol{\theta}\,j,j'}$, to be used for training the empirical model $\chi(\mathbf{z}_{\boldsymbol{\theta}\,j,j'}) = \hat{\sigma}^2_{A+C}$ for the sum of the process and noise variance components. The value of the variance component σ^2_{A+C} for the RUL prediction error is then obtained by feeding to model χ the input $\mathbf{z}_{\boldsymbol{\theta}\,J,th} = [z_{\theta,j}, \theta_{th}]$ [82]. The prediction error variance is then computed as the sum of the estimates $\hat{\sigma}^2_B$ and $\hat{\sigma}^2_{A+C}$, i.e., $\sigma^2_{e_J} = \mathrm{var}[\hat{rul}^h_J] + \chi(\mathbf{z}_{\boldsymbol{\theta}\,J,th})$. Finally, assuming that the prediction error has a Gaussian distribution, the bounds $rul^{\inf}_J(\alpha)$ and $rul^{\sup}_J(\alpha)$ of the prediction interval can be computed according to the theory of the bootstrap method [35] as:

$$rul^{\inf}_J(\alpha) = \hat{rul}_J - c^{\alpha}_{conf}\,\hat{\sigma}_{e_J} \text{ and } rul^{\sup}_J(\alpha) = \hat{rul}_J + c^{\alpha}_{conf}\,\hat{\sigma}_{e_J} \tag{20}$$

where c^{α}_{conf} is the $1-\alpha/2$ percentile of a Student's *t*-distribution with number of degrees of freedom equal to the number *H* of bootstrap models.

Notice that the training data $(\mathbf{z}_{\boldsymbol{\theta}\,j,j'}; \rho_{j,j'})$, used to build model χ, cover a range of values for the input $\mathbf{z}_{\boldsymbol{\theta}\,j,j'}$ in general different from that of the input $\mathbf{z}_{\boldsymbol{\theta}\,J,th}$ to which the model is applied to obtain the estimate $\hat{\sigma}^2_{A+C}$. This can represents a limit to the quality of this estimate, since in general the performance of empirical models are good when applied to input regions well described by training data, and decrease moving away from these regions.

5 Numerical applications

5.1 Creep growth in turbine blades

The three different cases presented in Section 4 have been considered with reference to the prognostics of a turbine blade in which creep damage is developing [81]. The application has focused on the turbine of a generation IV high temperature gas reactor, which is characterized by rather extreme turbine operational conditions such as working temperatures exceeding 900°C. The predominant damage mechanism affecting turbines operating at such elevated temperatures is creep deformation [76]. A turbine undergoing this degradation process can experience the loss of its blades, one of the most feared failure modes of turbomachinery since it is accompanied by abrupt changes in the power conversion equipment and in the reactor flow conditions and can result in turbine missiles, i.e., irregularly shaped projectiles travelling at high velocities which can impact on barriers in nuclear power plants causing severe damages to the facilities [77]. Figure 7 shows an example of high-pressure turbine deblading occurred in a German power plant [77]. This and the high cost of turbine blade replacement are strong reasons for performing prognostics on creeping turbine blades.

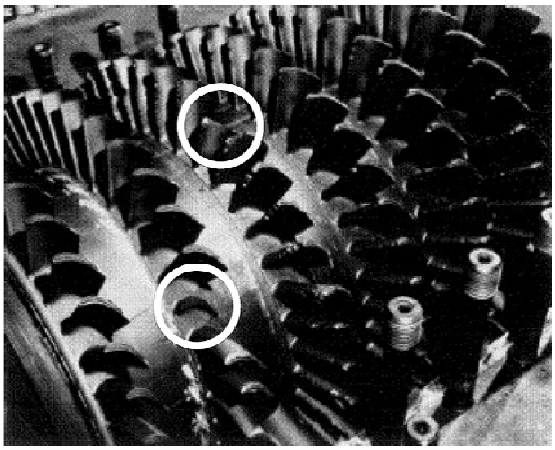

Figure 7: Deblading in a high pressure turbine [77].

Creep is an irreversible deformation process affecting materials exposed to a load below their elastic limit for a protracted length of time and at high temperatures. In the high pressure turbine first stage, blades creep is a major problem due to the high operational temperatures, and is often the life-limiting process [76]. Blade elongation, ε, is taken as a measure of the blade creep strain and modeled as a stochastic process using the Norton Law. The blade is discarded when the accumulated elongation reaches a pre-determined value, namely the failure threshold, which assures that the risk of blade failure is below the desired safety limit. The detailed creep growth model implemented and a deeper discussion of this case study, can be found in paper III of Part II. The data used in this case study have been numerically simulated in order to allow testing the three approaches on a large number of different blade degradation trajectories and thus evaluate their capability of correctly estimating the uncertainty on the provided RUL prediction. Examples of some simulated creep growth trajectories are shown in Figure 8.

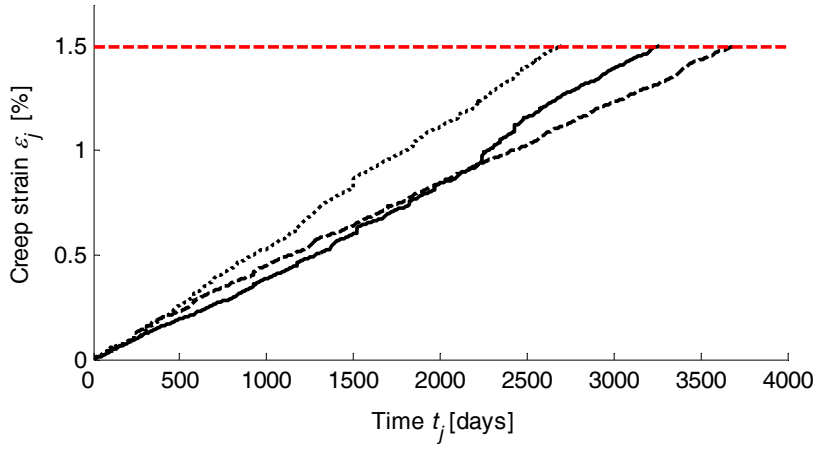

Figure 8: Examples of creep growth trajectories.

The main sources of degradation-related information for the creep growth process are listed in Table 5.

Table 5: main sources of information for prognostics of a creeping turbine blade

Source	Description	Mathematical representation
A	The creep growth model and the distributions of the model parameters	Paper II, Part II
B	Measurements of the creep strain of the currently creeping blade taken at J different time instants t_j	$\mathbf{z}_{1:J} = z_{\theta,1:J} = \varepsilon_{1:J} + noise$
C	Historical measurements of the creep strain of a set of R blades failed for creep, taken at n^r different time instants t_j	$\mathbf{z}^r_{1:n^r} = z^r_{\theta,1:n^r}$, $r = 1:R$
D	The value of the failure threshold	$\theta_{th} = \varepsilon_{th} = 1.5\%$
E	The observations equation (direct creep strain measurements with a white Gaussian noise υ_j)	$\mathbf{z}_{1:J} = \varepsilon_{1:J} + \upsilon_j$
F	The length of life of the set of *S* failed blades.	t^r_F, $r = 1:R$

At every time t_J during the life of a turbine blade, the set of observations $\mathbf{z}_{1:J}$ is assumed to be available; the objective of the analysis is to obtain a prognostic prediction for the test trajectory at different time instants t_J, $J = 1,2,\ldots$. Three situations, corresponding to the three prognostic cases of Section 4, have been artificially constructed for the turbine blade case study. In case 1, the sources of information *A*, *B*, *D* and *E* of Table 5 are assumed to be available and the model-based PF approach presented in Section 4.1.1 is adopted to predict the distribution of the equipment RUL. In cases 2, where the sources of information available are *B*, *C* and *F* of Table 5, and in case 3, where we have available sources of information *B*, *C*, *D* and *E*, the bootstrapped ensemble of models presented in Section 4.3.2 is used to provide the RUL prediction $\hat{rul}_J$ with an estimate of the prediction error variance $\sigma^2_{e_J}$. In case 2, $R = 13$ training trajectories and the duration of their lives are available for training and validating an ensemble of $H = 25$ linear least square models $f^h(\mathbf{z}_J \mid \mathbf{D}^h_{z/rul})$ for direct RUL prediction. In case 3, the dataset $\mathbf{D}^h_{z/\Delta t}$ is drawn from the available data

about the test trajectory only, and used to build and validate an ensemble of $H = 25$ linear least square degradation-based prognostic models. Notice that, in this case, since a minimum number of historical data must be available to build the prognostic model, the prediction is performed only after time instant t_{30}, i.e., when at least 30 observation are available.

Figure 9 shows the evolution of the true RUL value rul_J of the blade RUL (continuous thick line), its predicted value $\hat{rul}_J$ (dots) and the corresponding prediction interval with $\alpha = 0.68$ (continuous thin line) obtained during the life of a turbine blade by the three prognostic approaches. These results show that it is possible to provide accurate RUL predictions with a reliable quantification of their uncertainty in all three cases considered. The best results are obtained in case 1, which is the one with the maximum amount of information available. The prediction intervals provided in case 3 are characterized by large oscillations and low accuracy, especially at the beginning of the trajectory, i.e., when few training data are available. Furthermore, the RUL prediction itself is noisy. This effect could be reduced by properly filtering the predictions, e.g., by using the Kalman Filter ensemble approach presented in Section 4.3.1.

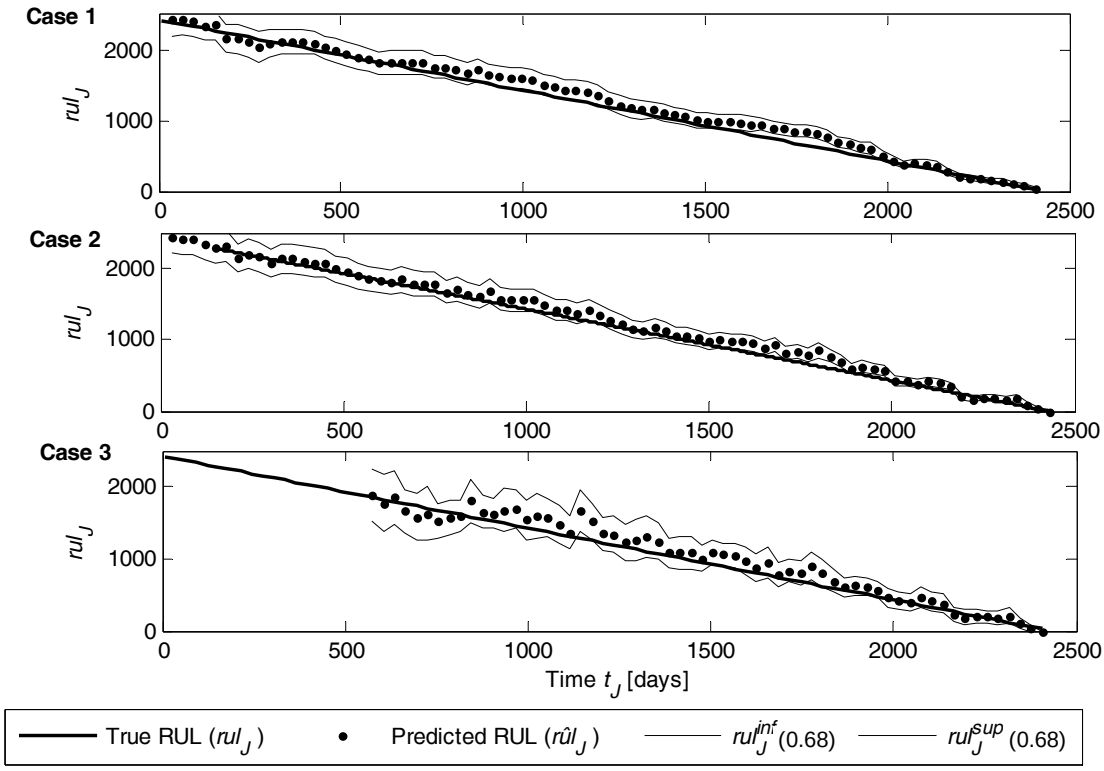

Figure 9: true rul_J (continuous thick line) of a turbine blade, predicted value $\hat{rul}_J$ (dots) and prediction interval $CI_J(0.68)$ (continuous thin line) for the three prognostic approaches.

Many further analyses, presented in Paper III of Part II, have been performed in the context of this case study to perform a robust evaluation of the performances of the three approaches and to determine whether the estimates of the prediction intervals provided by them properly describe the uncertainty in the RUL predictions.

Also, in order to provide some indications to the decision maker about which approach should be adopted in a case in which all the sources of information listed in Table 5 were available (so that all the three approaches presented can be applied), we have performed an analysis of the sensibility of the performance of the three prognostic approaches to the quality and quantity of the information available. To this purpose, we have considered the following indicators of the quality and quantity of the information:

A. the amplitude of the noise affecting the creep strain measurements;
B. the number of past measurements of the current trajectory available for making the RUL prediction;
C. the accuracy of the physical model of the degradation process;
D. the number of historical degradation trajectories available.

The performance of the three approaches is evaluated considering 250 test trajectories. Figure 10 shows the variation of the relative mean square error ($\overline{rMAE}$) when: (A) the amplitude of the noise affecting the creep strain measurements is varied from 0 to 0.03 (upper-left); (B) the current time instant at which prognostic is performed increases (upper-right); (C) a bias of variable amplitude is added to one of the parameters of the physical model used by PF to simulate its reduced accuracy (bottom, left); (D) the number of historical trajectories available for training the empirical model increases (bottom, right).

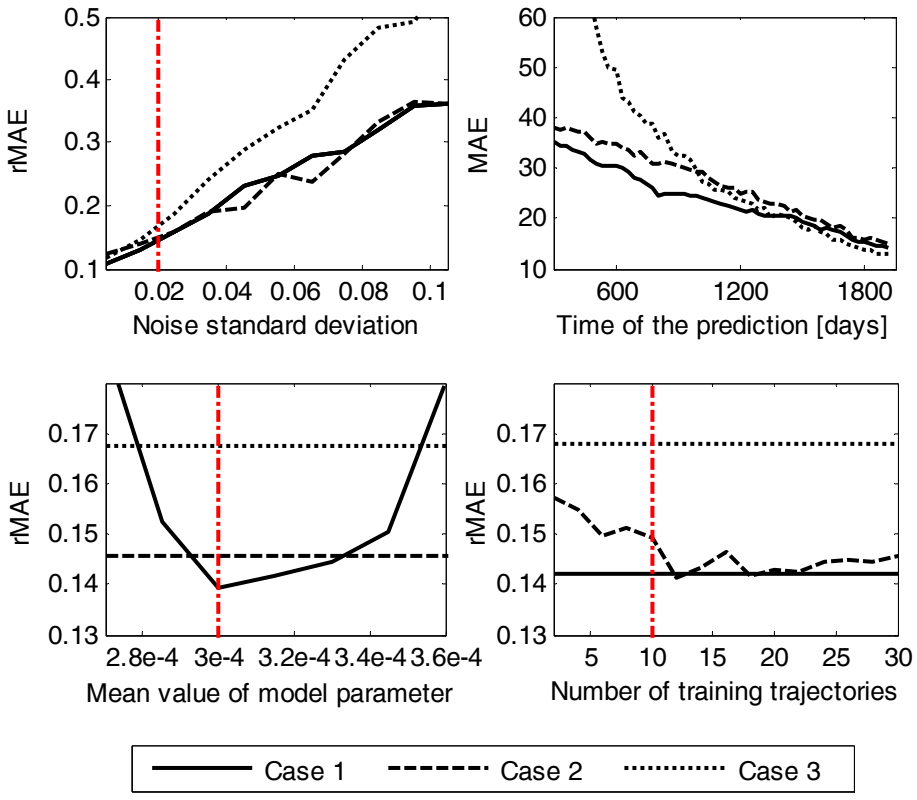

Figure 10: performance of the proposed approaches in different settings of information available. The vertical (red) line indicates the value assigned to the parameter in the numerical application of Figure 9.

Considering the results obtained performing this sensibility analysis (see also Paper II of Part II for an in-depth discussion of them), we have been able to provide the following guidelines for the choice of the prognostic approach: if one is very confident about the accuracy of the available physical degradation model,

approach 1 should be preferred; on the contrary, if one doubts about the model accuracy, approach 2 is, in general, the most accurate, especially if the number of historical trajectories available is large. However, if the measurement noise is small, the system is close to failure and many degradation measurements have been taken during the current degradation trajectory, approach 3 can provide better accuracy.

Consider however that, when multiple approaches with comparable degree of accuracy are available, an alternative strategy to the choice of the best performing approach consists in the combination of the different approaches outcomes. An application of the method presented in Section 4.2.3 for combining different complementary prognostic approaches, handling different sources of information available at the same time, is presented in the next Section in the context of a case study about prognostics of clogging filters in a BWR nuclear power plant.

5.2 Filter clogging

In this Section, we consider the problem of predicting the RUL of filters used to clean the sea water entering the condenser of the BWR reactor of a Swedish nuclear power plant [87-90]. During operations, filters undergo clogging and, once clogged, can cumulate particles, seaweed, and mussels from the cooling water in the heat exchanger. For this reason, prompt and effective cleaning of the filters is desirable; predictive maintenance can help achieving this result, keeping maintenance costs reasonably low.

From data collected on field, we have available $R = 8$ sequences of observations $\mathbf{z}^r_{1:n^r}$, $r = 1:R$ taken during the clogging process of as many historical filters. Each observation $\mathbf{z}^r_j = [\Delta P^r_j, \dot{M}^r_j, T^r_j]$ contains the measurements of the pressure drop ΔP^r_j, the flow across the filter $\dot{M}^r_j$, and the sea water temperature T^r_j collected at time t^r_j during the clogging process of the r-th filter. It has been well established in the literature concerning the study of filter clogging by solid aerosols [78] and liquid aerosols [79] that the clogging of a filter medium leads to an increase in pressure drop over the filter as long as the filtration velocity, and thus the flow, is kept constant. It is also known that the pressure drop is proportional to the square of the filtration velocity. Given these results, one can take as indicator of the state of clogging of filter r at time t_j the ratio [80]:

$$z^r_{\theta,j} = \frac{\Delta P^r_j}{(\dot{M}^r_j)^2} \tag{21}$$

Figure 11 shows the sequences of clogging indications $z^r_{\theta,1:n^r}$, $r = 1:R$ collected on field during the clogging process of the filters. We can see from this Figure that the clogging process is affected by large uncertainties, which can be ascribed to the very variable conditions of the sea water; in this context, the challenge is to provide sufficiently narrow confidence interval for the value of the filters RUL.

Due to the absence of physical knowledge about the failure threshold, it has been arbitrarily set to the value $\theta^{th} = 175$, based on the observation of the filters clogging state, indicated by $z^r_{\theta,j}$, at the time they undergoes maintenance. Having set the value of the failure threshold, we can derive the life durations t^r_F of the available reference filters.

In this case study, both the direct RUL prediction and the degradation-based approaches applies. The SBR approach described in Section 4.2.1, the GPR approach presented in Section 4.2.2, have thus been compared and their outcomes aggregated using the combination strategy proposed in Section 4.2.3.

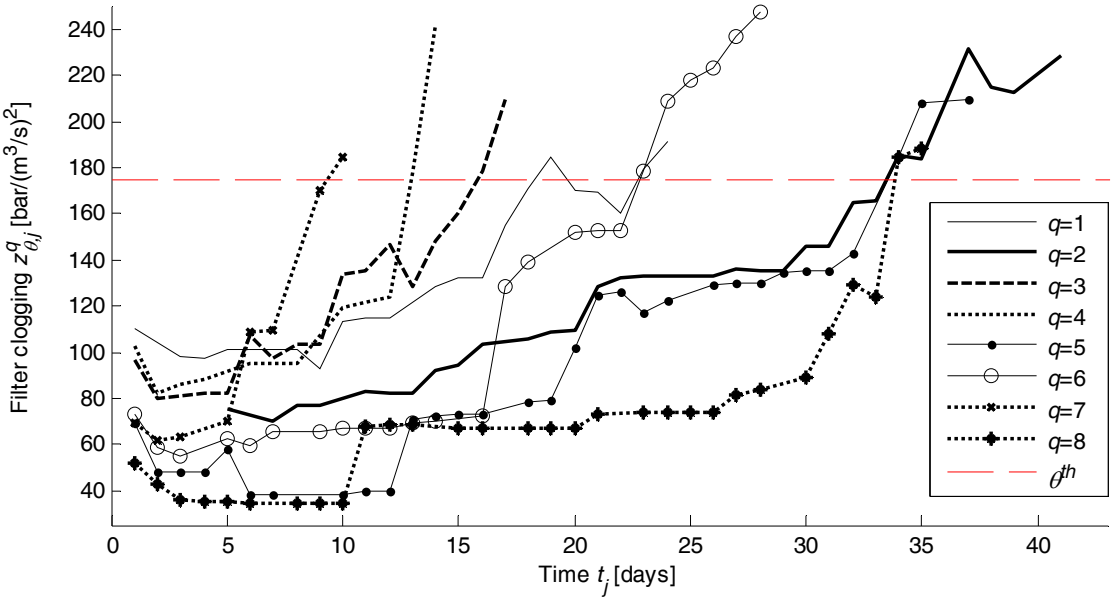

Figure 11: available clogging trajectories $z^r_{\theta,1:n^r}$, $r = 1:R$.

Figure 12 shows, for three representative clogging trajectories $r = 1$, 5, and 7, the true RUL value rul_J (continuous red line), the RUL prediction $\hat{rul}_J$ (dots) and the lower bound $rul_J^{\inf}(\alpha)$ of the left bounded prediction interval with $\alpha = 0.9$ (dotted line) supplied by the three approaches. Predictions are performed at each time instance during the evolution of a trajectory r, using the remaining $R-1=7$ trajectories $r' \neq r$ for training. Due to the large uncertainty of the process, for all three approaches the RUL prediction accuracy is rather low and the confidence intervals large, although the error and the amplitude of the prediction interval decrease with time. The GPR approach provides prediction intervals which are narrower that those of the similarity-based approach. This latter in fact provides a lower prediction bound $rul_J^{\inf}(\alpha)$ which is often very close to zero; as pointed out in [88] a vacuous SBR prediction (i,e., a prediction interval equal to the domain $\Omega_{RUL_J^{test}}$ of RUL_J^{test}), does not means that the evidence of very early failure is significant, but only that the evidence drawn from the reference trajectories is not sufficient to assert with the desired belief $\alpha = 0.9$ that the RUL value belongs to any subset of the RUL domain $\Omega_{RUL_J^{test}}$. In other words, the prediction $rul_J^{\inf,SB} = 0$ is a statement of *ignorance* about the value of RUL_J^{test}. Such intervals are too large to be effectively used by an operator to make a decision about the best time for undertaking a maintenance action. However, the large intervals predicted by the SBR approach can provide a correct indication that the information conveyed by the training trajectories is not relevant for a specific test trajectory, e.g., because they are too dissimilar. This can be seen for trajectory $r = 7$ where the GPR approach provides narrower confidence intervals, but such intervals do not include the true RUL value. Notice also that, for this trajectory, the prediction provided by SBR is more accurate than provided by GPR. Thus, the SBR approach provides always coverage very close to one, although at the price of too large prediction intervals. The

proper combination of the two methods allows obtaining the desired coverage with smaller prediction intervals than those provided by the SBR approach. In paper VI of Part II it is shown that this is made possible by the choice of the least committed isopignistic BDF transformation which relaxes the information conveyed by PDF $p_{RUL_J^{test}}(rul_J \left| \mathbf{D}_{\tau/z}^{+} \right.)$ predicted by the GPR method, thus making its relevance comparable to that of the SBR-predicted BBA.

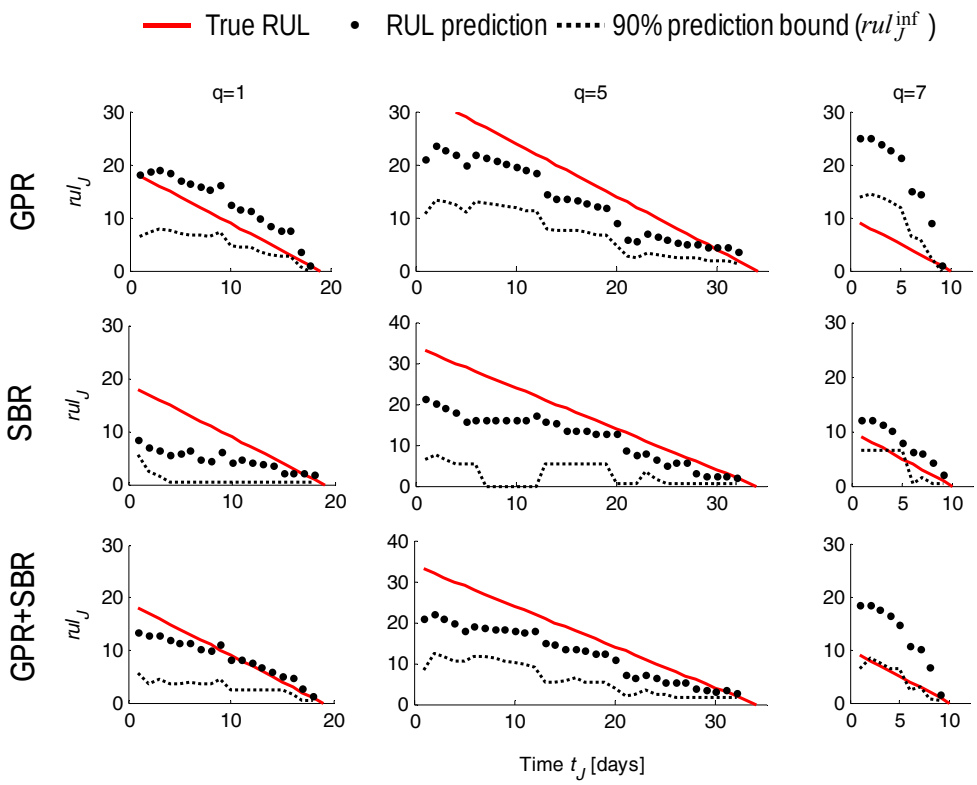

Figure 12: Predictions $\hat{rul}_J^{GPR}$ and $rul_J^{\inf,GPR}(0.9)$ supplied by the GPR approach.

Finally, it must be mentioned that, in the combination of the SBR and GPR model outcomes, the accuracy is reduced with respect to that of the individual SBR outcome. We expect that different strategies for the aggregation of the SBR and GPR RUL predictions, accounting, e.g., for the historical performance of the two methods, have the potential of improving the accuracy of the combined prediction.

6 Conclusions

Due to its importance in making PHM exploitable in industrial applications, a first important issue tackled in this PhD work has been the pre-treatment of the available prognostic information, which in real industrial application can be strongly affected by noise, inaccuracies and other sources of errors. In this context, the main contribution of this work has been to devise a procedure for validating the reliability of the available observations and improve the quality of those found unreliable, based on the other available observations. A fuzzy C Means clustering analysis has been applied to verify the consistency of the reliable and uncertain parameters values observed. When found unreliable, parameter observed values are improved by using an ensemble of Kernel regression models wisely aggregated. To this aim, an original aggregation strategy based on the AHP algorithm has been proposed for those cases where reliable values for the parameters under estimation are not available for evaluation model performance; on the other side, when they are available, the use of a local performance-based aggregation strategy is suggested. This procedure, applied to the real data related to choke valves erosion, has allowed recognizing some unreliable parameters (i.e., the oil, water and gas mass flow rates) and providing more accurate estimates of their values, which have the potential of improving the quality of the valve erosion assessment. The contribution given by this thesis work to prognostic data pre-treatment answers to some problem among the large variety of problems that may arise in drawing exploitable degradation-related information from sensor measurements, which remains one of the main open issues in the field of prognostics [3,8].

Besides data pre-treatment, the main contributions of this PhD thesis work concern the development of prognostic approaches for tackling different situations of information available, capable of supplying accurate predictions of the equipment RUL with, at the same time, a reliable quantification of the RUL prediction uncertainty. Quantitative considerations have been made with regards to simulated case studies concerning the linear and non linear growth of creep in steel equipment and to a real case study about filter clogging. Table 6 summarizes the different situation of information available considered and the prognostic methods developed to tackle each of them. The merits and open issue of the methods presented are also listed in the table. Moreover, since some method can provide a prediction of the future degradation trajectory and the RUL probability distribution, whereas others are limited to providing a prediction of the RUL expected value, with, eventually, a credible interval for its true value, Table 6 points out which outcomes are provided by the different method. In all situations, we have been able to supply accurate prediction of the RUL expected value and reliable prediction intervals for the true RUL value.

In case 1, the capability of Bayesian approaches, and particle filtering in particular, of providing accurate predictions when a physical model of the degradation process is available has been largely proven in the literature about model-based prognostics. The results obtained in this work show that the PF approach provides a good approximation of the exact distribution of the equipment RUL in the case in which an accurate model reproducing the equipment degradation process is available. When using model-based approaches, imprecision of the model in the reproduction of the degradation process due to simplifications, incorrect model structure or assumptions on the equipment specific geometries or material properties, etc., can be amplified over time, causing uncertain estimates of the RUL distribution. It has been shown in this work that, using particle filtering, it is possible to account for model parameters uncertainty in the prediction of the RUL distribution. Moreover, it is known that, by including model parameters in the state vector, one

can perform model adaptation in conjunction with state tracking. In any case, it is very difficult for a physics-based model to account for all aspects of a degradation process; for example, it is common to neglect some of the interactions between different degradation mechanisms or the possible existence of self-healing mechanisms which can reverse the degradation process and are likely to increase the uncertainty of the future degradation evolution. All these non-modeled phenomena can be accounted for by adding further noise to the process model which will result in a larger confidence interval associated to the RUL estimate. Further research is needed to quantify the impact of modeling errors on the final prediction of model-based approaches.

In case two, where a physics-based model of the degradation process is not available, two methods, similarity based regression and Gaussian process regression, have been proposed to provide, respectively, direct and a degradation-based RUL predictions, based on the available training trajectories. These methods, which provides satisfactorily results when several training trajectories with limited variability are available, can provide RUL prediction intervals with, respectively, too large amplitude or insufficient coverage when applied to degradation processes characterized by a large variability of the degradation trajectories followed by similar pieces of equipment. The coverage values obtained from the GPR methods, which are lower than expected, have been ascribed to the fact that the method do not correctly accounts for the model uncertainty. In this view, the use of an ensemble of different GPR models is envisaged as a possible improvement. Furthermore, future research should focus on the reduction of the prediction uncertainty of the SBR prognostic model. To reduce the prediction uncertainty when the SBR and GPR can both be applied, the aggregation of different prognostic methods using different pieces of available information has been considered. In the case study about filter clogging, the outcome of SBR prognostics, which makes use of life duration data, and those of GPR prognostics, which uses direct degradation indications, have been combined in the framework of the belief function theory. This approach has allowed increasing the reliability of the prediction intervals, while keeping them reasonable narrow.

Finally, in case 3, due to the little information available, the use of an ensemble approach has been proposed as an effective solution for increasing the robustness and accuracy of the prediction and quantifying the RUL prediction uncertainty. A Kalman filter-based ensemble and a bootstrapped ensemble of data-driven degradation models have been developed and applied to the linear creep growth case study. Results have proven the capability of the KF ensemble of increasing the prediction accuracy by filtering out the measurement and prediction noises, without imposing any restriction about the linearity of the degradation model. The main limitation of this method is that it does not supply a measure of the prediction uncertainty. On the other side, the bootstrapped ensemble approach provides reliable estimates of the prediction uncertainty. However, the bootstrap method requires building an empirical model for the estimate of the RUL prediction error variance which is then used outside the region covered by the training data: although good extrapolations have been obtained in the linear creep growth case study, the feasibility of the approach on more complex, non-linear models should be verified. Notice, however, that, when a single censored sequence of degradation indications is available, it can be very difficult, in practice, to develop a purely data-driven non-linear degradation model. Thus, in practice, an assumption of linearity of the degradation process has often to be adopted.

Table 6: schematic view of the proposed prognostic methods

Case 1	
Information available	*A.* Degradation model *B.* Current observations' sequence *D.* Failure threshold *E.* Measurement equation
Prognostic approach	Model-based approach
Computational method	Particle Filtering
Predicted outcome	Future degradation evolution RUL probability distribution RUL expected value Credible interval for the true RUL value
Merits	Accurate and precise predictions Treatment of model parameters uncertainty Particle degeneration problem addressed by resampling
Open issues	Implementation of an effective parameter updating strategy Guidelines for the setting of PF parameters (process and observation noises) The method does not account for errors in the model structure and assumptions
Case 2	
Information available	*B.* Current observations' sequence *C.* Historical observations' sequences *F.* Life duration data
Prognostic approach	Data-driven approach - Direct RUL prediction
Computational method	Similarity-based regression using belief mass assignments for uncertainty representation
Predicted outcome	RUL expected value Credible interval for the true RUL value
Results	Accurate predictions Guidelines for the setting of the model parameters (similarity and trust parameters γ and λ)
Open issues	The prediction intervals provided can be very large or even coincident with the RUL domain
Information available	*B.* Current observations' sequence *C.* Historical observations' sequences *D.* Failure threshold *E.* Measurement equation
Prognostic approach	Data-driven approach – Degradation-based prognostics
Computational method	Gaussian process regression
Predicted outcome	Future degradation evolution RUL probability distribution RUL expected value Credible interval for the true RUL value
Merits	Accurate and precise predictions Modeling of non-linear degradation processes with weak *prior* assumptions in the absence of physics-based knowledge about the degradation mechanism
Open issues	Extension needed to account for model uncertainty in the predicted RUL distribution and credible interval Guidelines for selecting the *prior* on the mean and covariance functions are needed

Information available	*B.* Current observations' sequence *C.* Historical observations' sequences *D.* Failure threshold *E.* Measurement equation *F.* Life duration data
Prognostic approach	Data-driven approach – combined direct RUL prediction and degradation-based prognostics
Computational method	Representation of belief density function on continuous real axis $\Re$
Predicted outcome	RUL expected value Credible interval for the true RUL value
Merits	Improved reliability of the predicted credible interval with respect to GPR model while keeping its amplitude relatively small with respect to SBR model.
Open issues	Further testing on a greater number of real and simulated data is needed Extension needed to include multi-modal RUL distributions
Case 3	
Information available	*B.* Current observations' sequence *D.* Failure threshold *E.* Measurement equation
Prognostic approach	Data-driven approach –Degradation-based prognostics
Computational method	Kalman Filter-based ensemble
Predicted outcome	RUL expected value
Merits	Filtering of the measurement and prediction noise Improved RUL prediction accuracy with respect to the individual best model Guidelines for the setting of KF parameters (process and observation noise) Capability of incremental-learning Applicability to non linear degradation processes
Open issues	Does not provide a measure of the RUL prediction uncertainty In case of non-linear degradation processes it can be hard to develop an accurate degradation model, due to the limited information available
Information available	*B.* Current observations' sequence *D.* Failure threshold *E.* Measurement equation
Prognostic approach	Data-driven approach –Degradation-based prognostics
Computational method	Bootstrapped ensemble
Predicted outcome	RUL expected value Credible interval for the true RUL value
Results	Provides a measure of the RUL prediction uncertainty when not automatically given by the regression method
Open issues	Extrapolation outside the region of training data Its applicability to non-linear degradation processes should be verified In case of non-linear degradation processes it can be hard to develop an accurate degradation model, due to the limited information available

References

1. US Nuclear Regulatory Commission. Oversight of Nuclear Power Plant Safety Has Improved, but Refinements Are Needed, Report to Congressional Requesters, GAO-06-1029, Sep 27, 2006.
2. N.F. Rieger, T.H. McClosky, and R.P. Dewey. The High Cost of Failure of Rotating Equipment, Proceedings of Machinery Failure Prevention Technology Conference 44, 1990, Virginia Beach (VA).
3. J.B. Coble. Merging Data Sources to Predict Remaining Useful Life – An Automated Method to Identify Prognostic Parameters. PhD dissertation, University of Tennessee, 2010.
4. G. Vachtsevanos, F. L. Lewis, M. Roemer, A. Hess and B. Wu. Intelligent Fault Diagnosis and Prognosis for Engineering Systems, 1st edition, John Wiley & Sons, Hoboken, 2006.
5. D.B. Jarrell, D.R Sisk and L.J. Bond, Prognostics and Condition-Based Maintenance: A New Approach to Precursive Metrics, Nuclear Technology 2004, Vol. 145, pp. 275-286.
6. E. Zio, and M. Compare. Evaluating maintenance policies by quantitative modeling and analysis. Reliability Engineering System Safety 2010, Vol. 109, pp. 53–65.
7. M. Schwabacher and K. Goebel. A Survey of Artificial Intelligence for Prognostics. Association for the Advancement of Artificial Intelligence Fall Symposium, Arlington VA.
8. J. Coble, and J.W. Hines. Identifying Optimal Prognostic Parameters from a Genetic Algorithms Approach. Annual Conference of the Prognostics and Health Management Society 2009, 27 sept – 1 oct, San Diego.
9. E. Zio, Prognostics and Health Management of Industrial Equipment. In: Kadry S, editor. Diagnostics and Prognostics of Engineering Systems: Methods and Techniques, IGI-Global, 2012
10. W. Hines, and A. Usynin. Current Computational Trends in Equipment Prognostics. International Journal of Computational Intelligence Systems 2008, Vol. 1(1), pp. 94–102.
11. T. Brotherton, G. Jahns, J. Jacobs and D. Wroblewski. Prognosis of faults in gas turbine engines, IEEE Aerospace Conference Proceedings, 2000, 18-25 Mar, Big Sky, MT.
12. T. Wang. Trajectory Similarity Based Prediction for Remaining Useful Life Estimation, PhD dissertation, University of Cincinnati, 2010.
13. J. Luo, K. Pattipati, L. Qiao, and S. Chigusa. Model-based Prognostic Techniques Applied to a Suspension System, IEEE Transactions on Systems, Man, and Cybernetics 2008, Vol. 38(5), pp. 1156-1168.
14. U. Pulkkinen. A stochastic model for wear prediction through condition monitoring, in: K. Holmberg, A. Folkeson (Eds.), Operational reliability and systematic maintenance, Elsevier, London (New York), 1991, pp. 223-43.
15. A. Ray and S. Tangirala. A nonlinear stochastic model of fatigue crack dynamics, Probabilistic Engineering Mechsnics 1997, Vol. 12(1), pp. 33-40.
16. F. Cadini, E. Zio and D. Avram. Monte Carlo-based filtering for fatigue crack growth estimation, Probabilistic Engineering Mechanics 2009, Vol. 24, pp. 367-373.
17. A. Doucet, J.F.G. de Freitas, and N.J. Gordon. Sequential Monte Carlo methods in practice. Springer-Verlag 2001, NY.
18. A. Doucet, On sequential simulation-based methods for Bayesian filtering, Technical report. Department of Engineering, University of Cambridge, CUED-F-ENGTR310, 1998.
19. B.D. Anderson, and J.B. Moore. Optimal filtering, Prentice Hall, Englewood Cliffs (NJ), 1979.
20. G. Kitagawa. Non-Gaussian state-space modeling of nonstationary time series, J. of the Am. Stat. Assoc.,82, 1987, 1032-63.
21. T. Khan. Particle filter based prognosis study for predicting remaining useful life of steam generator tubing, IEEE Conference on Prognostics and Health Management (PHM), 20-23 June 2011, Montreal, QC.
22. M.S. Arulampalam, S Maskell, N Gordon, and T Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Process, 50(2), 2002, 174-88.
23. D.R. Cox and D. Oakes. Analysis of Survival Data. 1984, Chapman and Hall.
24. D. Bendell, D. Wightman and E. Walker. Applying Proportional Hazard Modeling in Reliability: Reliability Engineering and System Safety 1991, Vol. 34, 35-53.
25. M.E. Orchard, and G.J. Vachtsevanos. A Particle Filtering Approach for On-Line Failure Prognosis in a Planetary Carrier Plate. International Journal of Fuzzy Logic and Intelligent Systems 2007, Vol. 7(4), pp. 221-227.
26. K. Goebel, B. Saha, and A. Saxena. A Comparison of Three Data-Driven Algorithms for Prognostics, Proceedings of Machinery Failure Prevention Technology Conference 62, 2008, pp.119-131.
27. V. Sotiris, and M. Pecht. Support Vector Prognostics Analysis of Electronic Products and Systems. Paper presented at the Association for the Advancement of Artificial Intelligence Conference, 2007, 22-26 July, Vancouver, Canada.
28. J.Q. Shi, R. Murray-Smith, and D.M. Titterington. Hierarchical Gaussian process mixtures for regression, Statistics and Computing 2005, Vol. 15, pp. 31–41.
29. E. Zio, and F. Di Maio. A Fuzzy Similarity-Based Method for Failure Detection and Recovery Time Estimation. International Journal of Performability Engineering 2010, Vol. 6(5), pp. 407-424.
30. L. Peel. Data Driven Prognostics using a Kalman Filter Ensemble of Neural Network Models. Proceedings of the International Conference on Prognostics and Health Management, 2008, 06-09 oct, Denver, CO.
31. O. Felix, and B.A.E. Heimes. Systems. Recurrent Neural Networks for Remaining Useful Life Estimation. Paper presented at the International Conference on Prognostics and Health Management, Denver, USA.

32. G.P. Zhang. Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing 50 (2003) 159 – 175.
33. F. Di Maio, J. Hu, P. Tse, K. Tsui, E. Zio, and M. Pecht. Ensemble-approaches for clustering health status of oil sand pumps, Expert Syst. with Appl., doi: 10.1016/j.eswa.2011.10.008.
34. R. Polikar. Ensemble Based Systems in Decision Making, IEEE Circuits and Syst. Mag., 6(3), 2006, 21-45.
35. T. Heskes. Practical confidence and prediction intervals, in: M.Mozer, M. Jordan, and T. Heskes (Eds.), Advances. Neural Information Processing Systems 9, MIT Press, Cambridge, 1997, pp. 466–472.
36. J.Z. Sikorska, M. Hodkiewicz, and L. Ma. Prognostic modeling options for remaining useful life estimation by industry. Mechanical Systems and Signal Processing 2011, Vol. 25, pp. 1803–1836.
37. N. Gorjian, L. Ma, M. Mittinty, P. Yarlagadda, and Y. Sun. Review on Degradation Models in Reliability Analysis, Proceedings of the 4th World Congress on Engineering Asset Management, 2009, 28-30 Sept, Athens, Greece.
38. W. Caesarendra. Machine degradation prognostic based on RVM and ARMA/GARCH model for bearing fault simulated data. Prognostics and Health Management Conference, 2010, 12-14 Jan, Macau, China.
39. P. Boskoski, M. Gasperin, and D. Petelin. Bearing fault prognostics based on signal complexity and Gaussian process models IEEE Conference on Prognostics and Health Management (PHM), 2012, 18-21 June, Denver (CO).
60. S. Petit-Renaud, and T. Denoeux. Nonparametric regression analysis of uncertain and imprecise data using belief functions. International Journal of Approximate Reasoing, Vol 35, 2004, pp.1-28.
41. C. Rasmussen, and C. Williams. Gaussian processes for machine learning. MIT Press 2006, Cambridge, MA.
42. D.J.C. MacKay. Introduction to Gaussian processes. In:Bishop CM (ed) Neural networks and machine learning 1998, Vol. 168. NATO ASI Series, Springer, Berlin, pp 133–165
43. B.H. Nystad, G. Gola, J.E. Hulsund, and D. Roverso. 2010. Technical Condition Assessment and Remaining Useful Life Estimation of Choke Valves subject to Erosion. Proc. PHM Society 2010 Ann. Conf., October 11-13, Portland, OR.
44. L. Ngkleberg, and T. Sontvedt. 1995. Erosion in choke valves-oil and gas industry applications. Wear, Vol. 186-187, Part 2, pp. 401-412.
45. P.M. Birchenough, D. Cornally, S.G.B. Dawson, P. McCarthy, and S. Susden. 1994. Assessment of Choke Valve Erosion in a High-Pressure, High-Temperature Gas Condensate Well Using TLA. Paper SPE 28887 presented at the European Petroleum Conference , 25-27 October , London, UK.
46. P. Vaidya. Prognosis - subsea oil and gas industry. Annual Conference of the Prognostics and Health Management Society 2010, 10-16 Oct, Portland, OR.
47. Metso Automation. Flow Control Manual. 4th edition (2005).
48. J.C. Dunn. 1974. A fuzzy relative of the isodata process and its use in detecting compact, well separated clusters. Cybernetics and Systems, Vol 3, Issue 3, pp. 32–57.
49. A.K. Jain, M.N. Murty, and P.J. Flyn. 1999. Data Clustering: A Review. ACM Computing Surveys (CSUR), Vol. 31, Issue 3, pp. 264–323.
50. E.A. Nadaraya. 1964. On Estimating Regression. Theory of Probability and Its Applications, Vol. 10, pp. 186–190.
51. C.G. Atkeson, A.W. Moore, and S. Schaal. 1997. Locally Weighted Learning. Artificial Intelligence Review, Vol. 11, pp. 11–73.
52. M.P. Perrone, and L.N. Cooper. 1992. When networks disagree: ensemble methods for hybrid neural networks, National Science Fundation, USA.
53. P.P. Bonissone. 2010. Soft Computing: A Continuously Evolving Concept, International Journal of Computational Intelligence Systems, Vol.3, No. 2, pp. 237-24.
54. T.L. Saaty. 1980. The analytic Hierarchy Process, Planning, Priority Setting, Resource Allocation. McGraw-Hill, New York.
55. P. Baraldi, A. Cammi, F. Mangili, and E. Zio. "Local Fusion of an Ensemble of Models for the Reconstruction of Faulty Signals," IEEE Transactions on Nuclear Science, Vol. 57 (2), 2010, pp. 793-806.
56. J.L. Myers, A.D. Well, and R.F. Lorch. 2010. Jr. Research Design and Statistical Analysis. Taylor & Francis, New York.
57. P.A. Devijver, and J. Kittler. 1982. Pattern Recognition: A Statistical Approach. Prentice/Hall, Englewood Cliffs, NJ.
58. E. Zio, and P. Baraldi. 2005. Identification of nuclear transients via optimized fuzzy clustering. Ann. of Nucl. Energy, Vol. 32, pp.1068–1080.
59. M.P. Wand, and W.R. Schucany. 1990. Gaussian-based kernels for curve estimation and window width selection. Canadian Journal of Statistics, Vol. 18, pp. 197–204.
60. A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel. Evaluating prognostics performance for algorithms incorporating uncertainty estimates. In Proc. IEEE Aerosp Conf, 2010.
61. A. Hess, G. Calvello, and P. Frith. Challenges, issues, and lessons learned chasing the 'Big P': real predictive prognostics Part 1. Proc IEEE Aerosp Conf, 2006 Mar 5-12; Big Sky, MT.
62. L. Tang, G.J. Kacprzynski, K. Goebel, and G. Vachtsevanos. Methodologies for uncertainty management in prognostics. Proc IEEE Aerosp Conf, 2009 Mar 7-14; Big Sky, MT.
63. R. Liu, L. Ma, R. Kang, and N. Wang. The modeling method on failure prognostics uncertainties in maintenance policy decision process. Proc 9th Int Conf on Reliab, Maint and Saf (ICRMS), 2011 Jun 12-15; Guiyang, China.
64. A. Urbina, S. Mahadevan, and T.L. Paez. Quantification of margins and uncertainties of complex systems in the presence of aleatoric and epistemic uncertainty, Reliab Eng Syst Saf 2011;96(9):1114-25.
65. S. Mohanty, A. Chattopadhyay, P. Peralta, and S. Das. Bayesian Statistic Based Multivariate Gaussian Process Approach for Offline/Online Fatigue Crack Growth Prediction. Experimental Mechanics 2011, Vol. 51, pp. 833–843.

66. A. Widodo, and B.S. Yang. Machine health prognostics using survival probability and support vector machine, Expert Systems with Applications 2011, Vol 38(7), pp. 8430–8437.
67. P. Smets. Belief functions on real numbers. International Journal of Approximate Reasoning 2005, Vol 40 (3), pp. 181-223.
68. G. Shafer. A mathematical theory of evidence, Princeton University Press. Princeton, NJ, 1976.
69. A.P. Dempster. Upper and lower probabilities induced by a multivariate mapping, Annals of Mathematical Statistics AMS-38 (1976) 325-339.
70. P. Smets, The transferable belief model for quantified belief representation, in: D.M. Gabbay, P. Smets (Eds.), Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, Kluwer Academic Publishers, Dordrecht, 1998, pp. 267–301.
71. R. Mann, R. Freeman, M. Osborne, R. Garnett, C. Armstrong, J. Meade, D. Biro, T. Guilford, and S. Roberts. Objectively identifying landmark use and predicting flight trajectories of the homing pigeon using Gaussian processes, Journal of the Royal Society Interface (2011) 8, 210–219
72. B. Ristic and P. Smets, Belief function theory on the continuous space with an application to model based classification. In: IPMU-2004 (Ed.), Information Processing and Management of Uncertainty, 2004, pp. 1119–1126
73. R. Polikar, L. Udpa, S.S. Udpa, and V. Honavar. Learn++: An incremental learning algorithm for supervised neural networks. IEEE transactions on systems, man, and cybernetics-part C: Applications and reviews, Vol. 31, No. 4. 2001
74. L. Breiman. "Bagging predictors," Machine Learning, Vol. 24 (2), pp. 123-140.
75. S.K. Yang, and T.S. Liu. "State estimation for predictive maintenance using Kalman filter," Reliability Engineering & System Safety, Vol. 66 (1), 1999, pp. 29–39
76. T.J. Carter. Common failures in gas turbine blades. Eng Failure Analysis, 2005;12:237-47.
77. M. Saez, N. Tauveron, T. Chataing, G. Geffraye, L. Briottet, and N. Alborghetti. Analysis of the turbine deblading in an HTGR with the CATHARE code. Nucl Eng and Des 2006;236:574–86.
78. C.B. Song, H.S. Park, and K.W. Lee. Experimental study of clogging with monodisperse PSL particles. Power Technology 163, pp 152-159, 2006.
79. P. Contal, J. Simao, D. Thomas, T. Frising, S. Callé, J.C. Appert-Collin, and D. Bémer. Clogging of fibre filters by submicron droplets. Phenomena and influence of operating conditions. Aerosol Science 35, pp 263-278, 2004.
80. B.H. Nystad. Condition-Based Maintenance (CBM) – filter clogging at OKG 1, a case study, HWR-961, OECD Halden Reactor Project, 2009.
81. P. Baraldi, F. Mangili, E. Zio, A Kalman Filter-based Ensemble Approach for Turbine Creep Prognostics. IEEE Trans. on Reliability, 2012.
82. P. Baraldi, F. Cadini, F. Mangili, and E. Zio. Model-Based and Data-Driven Prognostics under Different Available Information, accepted for publication in Probabilistic Engineering Mechanics, 2013
83. P. Baraldi, F. Mangili, and E. Zio. Investigation of uncertainty treatment capability of model-based and data-driven prognostic methods using simulated data. Reliability Engineering and System Safety, 2013, Vol 112C, pp. 94-108.
84. P. Baraldi, F. Cadini, F. Mangili, and E. Zio. Prognostics under different available information. Abstract accepted to the 2013 Prognostics and System Health Management Conference (PHM-2013), Milan, Italy, 8-11 September.
85. P. Baraldi, F. Mangili, and E. Zio. Ensemble of Bootstrapped Models for the prediction of the Remaining Useful Life of a Creeping Turbine Blade. 2012 IEEE Int. PHM Conf., Denver, Colorado, 18-21 June.
86. P. Baraldi, F. Mangili, and E. Zio. Creep growth prognostics using Gaussian Processes. Abstract accepted to the 2013 Prognostics and System Health Management Conference (PHM-2013), Milan, Italy, 8-11 September.
87. P. Baraldi, F. Mangili, E. Zio, G. Gola, and B.H. Nystad. A prognostics approach based on Gaussian Process Regression for degradation modeling, 2013.
88. P. Baraldi, F. Di Maio, F. Mangili, E. Zio, G. Gola, and B.H. Nystad. Similarity-based Prediction of Remaining Useful Life using belief function theory for uncertainty treatment, 2013
89. P. Baraldi, F. Di Maio, F. Mangili, and E. Zio. Data-driven methods for prognostics in clogging filters of BWR reactors. Abstract accepted to the 2013 Prognostics and System Health Management Conference (PHM-2013), Milan, Italy, 8-11 September.
90. P. Baraldi, F. Di Maio, F. Mangili, E. Zio, G. Gola, and B.H. Nystad. An approach to combining uncertainty measures provided by different data-driven prognostic methods with application to clogging filters of BWR reactors
91. P. Baraldi, F. Mangili, E. Zio, G. Gola, and B.H. Nystad. Ensemble of kernel regression models for assessing the health state of choke valves in offshore oil platforms, submitted to the Int J. of Computational Intelligence Systems, 2012.
92. P. Baraldi, E. Zio, F. Mangili,G. Gola, and B. H. Nystad. An ensemble based approach for process parameter estimation in offshore oil platforms. EHPG Meeting, Sandefjord, Norway, 2-7 October 2011.
93. G. Gola, B.H. Nystad, P. Baraldi, F. Mangili, and E. Zio, Health Assessment of Choke Valves undergoing Erosion with a Hybrid Ensemble Approach. PSAM 11 & ESREL 2012, 25-29 June, Helsinki, Finland.
94. Baraldi, E. Zio, F. Mangili, G. Gola, and B.H. Nystad. A hybrid ensemble approach for process parameter estimation in offshore oil platforms. FLINS Conf., 26-29 August, Istanbul, Turkey.

Report A[§]

An ensemble-based approach for process parameter estimation and health state assessment in offshore oil platforms

Piero Baraldi, Francesca Mangili, Enrico Zio, Giulio Gola, Bent H. Nystad

LASAR[**] and Halden Reactor Project Work Report, 2013

[§] This report is being submitted as official Halden Work Report (HWR)

[**] LASAR (Laboratorio di Analisi di Segnale ed Analisi di Rischio): Laboratory of Signal Analysis and Risk Analysis of the Department of Nuclear Engineering of the Polytechnic of Milan

An ensemble-based approach for process parameter estimation and health state assessment in offshore oil platforms

Piero Baraldi[a], Francesca Mangili[a], Enrico Zio[b,a]

[a]Dipartimento di Energia, Politecnico di Milano, Italy

[b]Chair on Systems Science and the Energetic challenge, European Foundation for New Energy-Electricite' de France, Ecole Centrale Paris and Supelec, France

Giulio Gola[c,d], Bent H. Nystad[c]

[c]Institutt for energiteknikk, Halden, Norway

[d]IO Center for Integrated Operations, Trondheim, Norway

Abstract

In offshore oil platforms, choke valve erosion is a major issue. To asses choke valve health state, the valve flow coefficient is commonly used. However, the traditional procedure for evaluating the erosion indicator introduces some inaccuracies which undermine the possibility of using it for prognostic purposes. The main reason for that was identified in the fact that the actual flow coefficient value is analytically calculated as a function of measured and allocated parameters. The latter are actually measured only during well test, whereas, on a daily basis, they are derived with large uncertainties from a physics-based model. In the present work, a clustering procedure is applied to the data collected on a single eroding choke to verify the coherence between the information conveyed by the measured parameters conjectured to be reliable (pressure drop and choke opening) and the three allocated ones (oil, water and gas mass flow rates); results proved the physics-based model to be often inaccurate. To overcome this hurdle, we resort to an ensemble of data-driven models built using a Kernel Regression (KR) technique. Results show that a hybrid approach which appends the data-driven ensemble to the physics-based model can improve the flow rates estimates. In order to exploit the KR models only when they actually outperform the physic-based model, a local performance-based approach is adopted for the aggregation of the different model outcomes.

A further problem addressed in this work is the non-monotonicity of the erosion indicator observed during well tests. Some hypotheses about the causes of this problem are discussed and possible improvements suggested.

Table of Contents

1 Introduction

Predicting the evolution of the equipment degradation allows efficient planning of maintenance operations (Garvey et al., 2009; Vachtsevanos et al., 2006; Jarrell et al., 2004). In general, a prognostic model can be developed based on information directly or indirectly related to equipment degradation (Hines & Usynin, 2008). In practice, however, field data are affected by noise, sensor faults and extrapolation errors and need to be verified and possibly corrected before they are used for developing the prognostic model.

In this paper, we consider the degradation of choke valves located topside at wells on the Norwegian Continental Shelf (Nystad et al., 2010; Haugen et al., 1995). The difference between the actual valve flow coefficient and its theoretical value is retained as the indicator of the choke valve health state and is used to assess the level of erosion affecting the choke. While the theoretical value of the valve flow coefficient depends only on the choke opening, the actual valve flow coefficient is analytically calculated on a daily basis as a function of the pressure drop through the choke which is directly measured and oil, gas and water flow rates which are allocated based on the measured total production from a number of wells and on physical parameters (pressures and temperatures) related to the single well. Such flow rates are actually measured only during a number of well tests carried out throughout the valve life. In practice, the resulting indicator of the choke valve state is very noisy and lacks the physical monotonicity of the erosion process; allocated values of oil, gas and water flow rates are conjectured to be the cause of the large inaccuracies and uncertainties in the calculation of the actual valve flow coefficient. To verify this, data are processed by the Fuzzy C Means (FCM) clustering algorithm (Dunn, 1974; Jain et al., 1999). FCM is applied to the projections of the five-dimensional dataset into the subspace of the two measured parameters (pressure drop and choke opening) and the subspace of the three allocated parameters (oil, water and gas flow rates). The two partitions are compared to investigate the coherence of the information conveyed by the parameters. A supervised clustering algorithm based on Mahalanobis metrics (Zio & Baraldi, 2005) is used to obtain a partition of the entire five-dimensional dataset as close as possible to that obtained based only on the two measured parameters. A measure of the importance of the parameters in the clustering is calculated and used to verify the coherence of the information conveyed by the less reliable allocated parameters with that conveyed by the two reliable ones.

Since the values of oil, gas and water flow rates are found unreliable, a method for correcting them based on the relations among all parameters has been developed. To this aim, an ensemble of Kernel Regression (KR) models is devised. KR is a distance-based regression algorithm (Nadaraya, 1964; Atkeson et al., 1997); an ensemble of four KR models is used to avoid the need of selecting the optimal model and to increase the robustness and reduce the uncertainty of the estimate (Perrone & Cooper, 1992, Bonissone, 2010). Diversity is injected in the ensemble by differentiating the training procedure for each KR model. The aggregation of the KR model outcomes is obtained through an original procedure based on the weighted average of the single model outcomes with weights calculated using the Analytic Hierarchy Process (AHP) (Saaty, 1980).

We have noticed that, depending on the well, either the physics-based model for flow rates allocation or the data-driven KR models ensemble can provide the most accurate estimates of the flow rates to be used for computing the flow coefficient and assessing the valve erosion. This is probably due to the specific conditions of the well. Physics-based approaches rely on the assumption of deep understanding of the system behaviours and detailed knowledge of geometry, material properties and other characteristics of the system.

On the other side, data-driven approaches are only accurate when applied to the same, or similar, operating condition for which data have been collected. Outside the training space of the empirical models, physics-based models should be preferred, because their results are derived from universal first principles. It is almost universally agreed in the forecasting literature that no single method is best in every situation (Zhang, 2003; Chatfield, 1988); moreover, real-world problems are often complex in nature and any single model may not be able to capture different patterns equally well. Thus, in practice, it is usually difficult to decide whether one particular model is more effective than the other in all the operational range. By combining different methods, the problem of model selection can be bypassed (Zhang, 2003). The combinations of physics-based and data-driven models are usually termed hybrid models. The result of combining the estimates of physics-based model and data-driven methods is to balance out their different errors and to augment the robustness and interpretability physics-based models with the sensitivity of data-driven methods (Zio, 2012; Coble & Hines, 2009; Polikar, 2006). The modeling framework underpinning hybrid methods is certainly more complicated, but offers clear advantages on the reliability and accuracy of the predictions.

In this work, the hybrid approach is applied to improve the accuracy of the mass flow rates estimates by appending to the ensemble of Kernel Regression (KR) models the physics based model providing the allocated values of the mass flow rates. In order to exploit the KR models only when actually needed, a local performance-based approach is adopted for the aggregation of the different model outcomes. The performance of the proposed approaches are evaluated on real data from 26 choke valves of 5 different wells, and compared to those of the physics based model and the KR ensemble.

In this work, it has been noticed that, even when only well tests are used to compute the erosion indicator and, consequently, its value is not affected by the inaccuracies in the allocated mass flow rates, still, it do not show high monotonicity. We conjectured this to be due to other sources of uncertainty and errors that affect the information available and undermine the possibility of using it for prognostic purposes. Some hypothesis about their nature are suggested and discussed in this work.

The remaining part of this work is framed as follows. The traditional procedure for the construction of a health indicator assessing the choke valve erosion state is presented in Section 2 where the clustering procedures is introduced to verify the reliability of the allocated parameters; in Section 3, an ensemble of KR models aggregated by a AHP-based procedure is proposed to improve the accuracy of the allocated flow rates and applied to the choke dataset; in Section 4, the hybrid ensemble approach is presented and results discussed; in Section 5 the quality of the erosion indicator and the possibility of using it for prognostic purposes are discussed; finally, conclusion and potential perspectives for future work are drawn in Section 6.

2 Choke valve erosion assessment

In oil and gas industries, choke valves are normally located on top of each well and are used to balance the pressure on several wells into a common manifold to control flow rates and protect the equipment from unusual pressure fluctuations.

In Figure 1, left, a choke valve is sketched. The throttle mechanism consists of two circular disks, each with a pair of circular openings to create variable flow areas. One of the disks is fixed in the valve body, whereas the other is rotated either by manual operation or by actuator, to vary or close the opening. For large pressure

drops, the well streams which contain gas, liquid and sand particles can reach 400-500 m/s and produce heavy metal loss mainly due to solids, liquid droplets, cavitation, and combined mechanisms of erosion-corrosion, resulting in choke lifetimes of less than a year. In Figure 1, right, the picture of an eroded choke valve is shown. The main parameters determining erosion potential in the chokes are the fluid velocity and the resulting angle of sand through the choke discs. Erosion management is vital to avoid failures that may result in loss of containment, production being held back, and increased maintenance costs. Moreover, several chokes are located subsea, where the replacement cost is high. Then, the need has increased for reliable models to estimate erosion and lifetime of choke valves, in order to allow implementing effective maintenance strategies (Wold et al., 2010; Ngkleberg & Sontvedt, 1995; Birchenough, 1994).

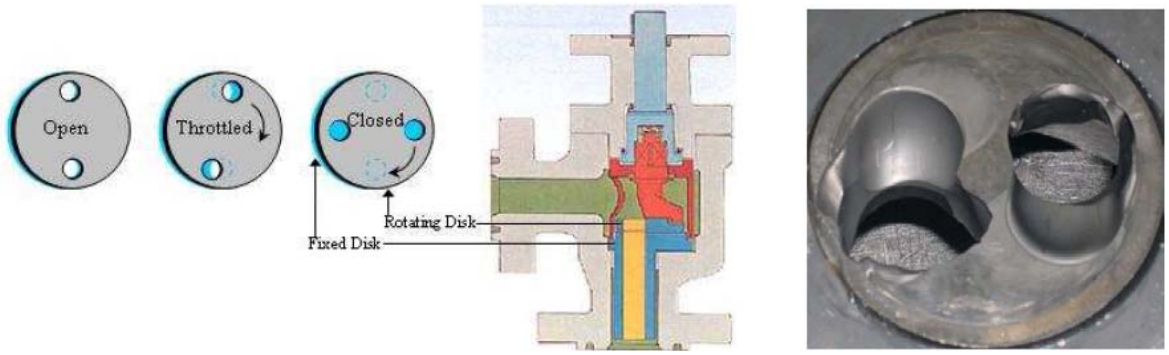

Figure 1: Typical choke valve of rotating disk type (http://www.vonkchokes.nl/) (left). Example of eroded choke disk(right).

A common indicator of the valve flow capacity is the flow coefficient C_V, which is related to the effective flow cross-section of the valve. For a specific valve opening θ, erosion produces a gradual increase of the valve area available for the flow transit, thus determining an increase of C_V (eq. 1). For this reason, knowing the value of the flow coefficient is fundamental for assessing the health state of the choke. During operation, C_V is not directly measured but computed, for a two-phase flow, as (Metso Automation, 2005):

$$C_V = \frac{\dot{m}_{tot}}{27.3 \cdot F_p \sqrt{\Delta P \cdot \rho_{mix}}}; \quad \rho_{mix} = \left(\frac{f_g}{\rho_g \cdot J^2} + \frac{f_w}{\rho_w} + \frac{f_o}{\rho_o} \right)^{-1} \tag{1}$$

where $\dot{m}_{o,w,g}$ are the oil, water ans gas flow rates, respectively, $f_{o,w,g} = \dot{m}_{o,w,g} / \dot{m}$ the corresponding fluid fractions and $\rho_{o,w,g}$ the corresponding densities, ρ_{mix} the mixture density, $\dot{m}_{tot} = \dot{m}_o + \dot{m}_w + \dot{m}_g$ is the total mass flow rate of the oil-water-gas mixture, J is the gas expansion factor, $F_p(\theta)$ is the piping geometry factor accounting for the geometry of the valve/pipe reducer assembly and ΔP is the pressure drop through the choke. Eq. (1) and the values of $\rho_{o,w,g}$, J, $F_p(\theta)$ and N_6 are derived from fluid dynamics; parameters ΔP, θ, $\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$ are measured or allocated during operation, i.e., calculated by a physics-based model of the piping process.

2.1 Choke valve dataset

For a correct assessment of the choke erosion state, it is fundamental to obtain frequent and reliable measurements or estimates of the parameters ΔP, θ, $\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$ used to compute the flow coefficient C_V. Nevertheless, only the pressure drop ΔP across the choke and the valve opening θ are known with good level of accuracy on a daily basis, since the first is directly measured during standard daily inspections (SI) whereas the second is a control parameter imposed by the actuator; hereafter, we will refer to these parameter as observed parameters. On the other side, measures of water, oil and gas flow rates are taken downstream of the choke only during well tests (WT) with a multiphase flow separator. On a daily basis, the values of $\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$ are allocated for a single well by a physics-based model accounting for the measured total production from a number of wells and on physical parameters (pressures and temperatures) related to the specific well. Figure 2 schematizes the procedure for the estimation of the flow coefficient.

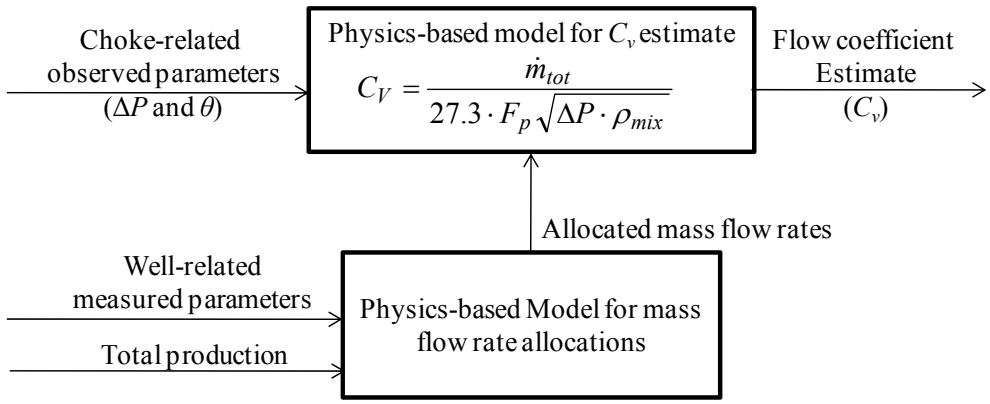

Figure 2: pictorial view of the procedure for the estimation of C_V.

The value of the parameters in input to the physics-based model has not been recorded. Only the values of the choke-related parameters ΔP, θ, $\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$ collected during WT and SI have been recorded during a protracted period of time for five different wells. Tables 1 and 2 outline the available information: the daily allocated values of $\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$, the daily observed values of ΔP and θ and the values of $\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$ measured during well tests. Since degraded valve disks are replaced, data collected for a single well refer to different disks.

Table 1: Available information

	Standard Inspections (SI)	**Well Test Inspections (WT)**
ΔP	Measured	Measured
θ	Control signal	Control signal
$\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$	Allocated	Measured

Table 2: Number of SI and WT patterns for each choke.

Well	Disk	N_{SI}	N_{WT}
1	1	259	7
2	1	597	36
	2	899	37
	3	358	14
3	1	191	9
	2	194	10
	3	211	14
	4	977	44
	5	570	19
4	1	318	20
	2	96	6
	3	131	6
	4	112	7
5	1	499	24
	2	373	17
	3	333	16
	4	134	7
	5	286	15
	6	153	11
	7	81	6
6	1	226	17
	2	344	14
	3	238	12
	4	870	28

Figure 3 shows the trend of the parameter values collected during SI (continuous line) and WT (stars) for the eroding choke disk of well 1. Figure 4 shows, for the same disk, the values of the health indicator δC_V computed using daily SI data (continuous line) and WT measurements (stars).

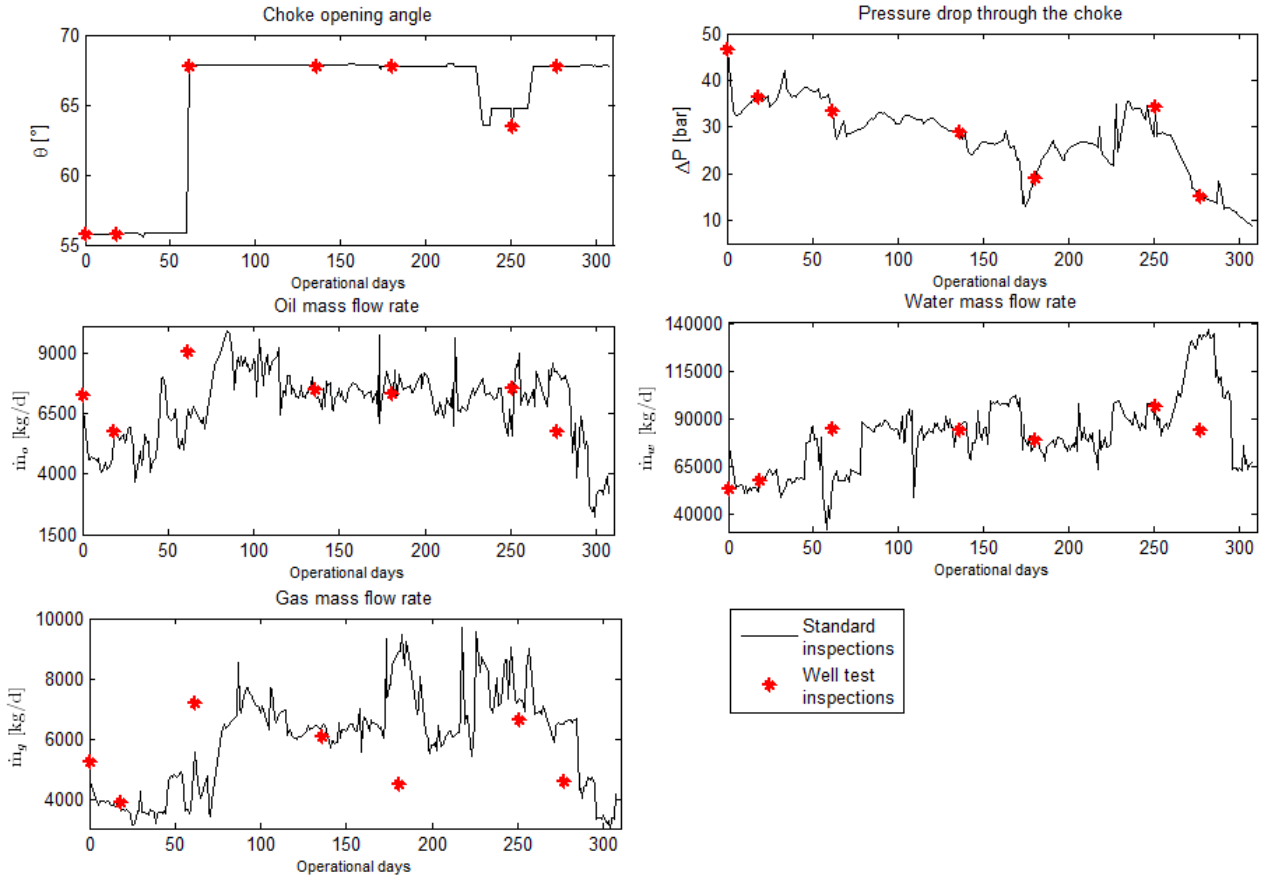

Figure 3: Parameters trends (continuous line represents SI, stars indicate WT)

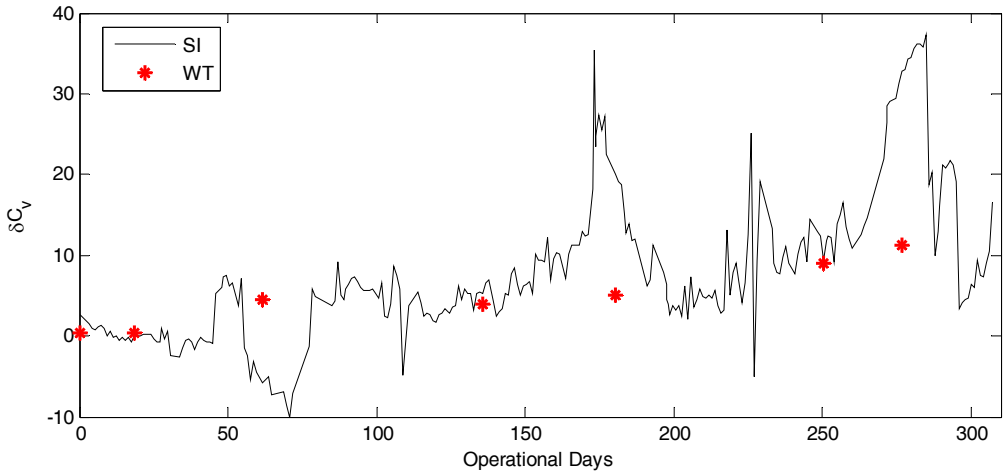

Figure 4: Health indicator δC_V using SI (continuous line) and WT (stars).

2.2 Data validation procedure

In this Section, a cluster analysis procedure is proposed to evaluate the reliability of the allocated parameters. Notice that all the results provided in this Section are obtained using only the data available for well 1.

In general, δC_V is expected to be monotonic since erosion cannot decrease in time unless maintenance actions are performed. A quantitative index of monotonicity is the Spearman's rank correlation used in statistics to assess how well the relationship between two variables can be described using a monotonic function (Mayers et al., 2010). The curve of δC_V computed using the SI data, is highly noisy and presents remarkable oscillations. The Spearman's rank correlation coefficient r_S between δC_V and time t_k at which the measurements are taken is computed as:

$$r_S = 1 - \frac{6\sum_{k=1}^{N}\left(R_{\delta C_V}(\mathbf{x}_k) - k\right)^2}{N(N^2-1)} \tag{2}$$

where $\mathbf{x}_k = [\Delta P, \theta, \dot{m}_o, \dot{m}_w, \dot{m}_g]$ is the five-dimensional vector containing the parameter values collected at time t_k, and $R_{\delta C_V}(\mathbf{x}_k)$ and k are the ranks (i.e., the relative positions) of pattern $\mathbf{x}_k$ when all patterns are ordered with respect to the values of δC_V and t_k, respectively. Values of r_S close to 1 are expected for a monotonic quantity. Results show that δC_V behaves monotonically (r_S=0.9643) only when WT measurements are used to compute it. On the contrary, the lower monotonicity (r_S=0.7401) obtained when δC_V is calculated using SI data suggests that some of the allocated mass flow rate values may be unreliable. A cluster analysis is performed in this Section in order to verify this hypothesis. In Section 2.2.1, the main steps of the procedure of cluster analysis proposed are presented; in Section2.2.2, the results of its application to the choke valve erosion case study are discussed.

2.2.1 Cluster analysis

Let **X** be a generic set of N patterns $\mathbf{x}_k = [\mathbf{x}_k^r \quad \mathbf{x}_k^u]$, k=1,…,N, of P=5 parameters which can be divided in a vector $\mathbf{x}_k^r = [\Delta P, \theta]$ of p^r=2 reliable parameters, and another vector $\mathbf{x}_k^u = [\dot{m}_o, \dot{m}_w, \dot{m}_g]$ of p^u=3 possibly unreliable parameters.

In general, the distinction between reliable and unreliable parameters can be achieved considering expert judgment, data analysis or by resorting to data validations techniques which allow detecting anomalous behaviors in datasets. In the choke valve case study, parameters ΔP and θ are classified as reliable according to expert judgment, whereas the allocated parameters, $\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$ are judged unreliable.

The aim is here to propose a procedure for verifying whether the information provided by the unreliable parameters in $\mathbf{x}_k^u$ is coherent with that of the reliable parameters in $\mathbf{x}_k^r$. This is done by considering the relative positions of the patterns in the p^r-dimensional subspace S^r of the reliable parameters, and in the p^u-dimensional subspace S^u of the unreliable parameters.

An effective technique to find a structure in a collection of unlabeled objects is unsupervised clustering, consisting in the organization (partition) of the patterns into non-overlapping, non-empty groups (clusters) so that patterns of the same cluster are *similar* between them and *dissimilar* to the patterns belonging to other clusters (Devijver & Kittler, 1982). For the validation of the unreliable parameters, two different partitions (Γ^r and Γ^u) of the dataset **X** into C clusters are considered: Γ^r is obtained using the unsupervised Fuzzy C-Means (FCM) clustering technique in the reliable parameters space S^r, whereas Γ^u obtained by applying the same technique in the unreliable parameters space S^u.

The FCM technique is an unsupervised clustering technique which makes no use of a priori known information on the true classes of the patterns (Yuan et al., 1995). The clustering is based on the minimization of a weighed sum Y of the distances $d(\mathbf{x}_k, \mathbf{v}_c)$ between the patterns $\mathbf{x}_k$ and the cluster centers $\mathbf{v}_c$,

$$Y = \sum_{c=1}^{C} \sum_{k=1}^{N} [\mu_c(\mathbf{x}_k)]^{\omega} d^2(\mathbf{x}_k, \mathbf{v}_c) \tag{3}$$

where the weight $\mu_c(\mathbf{x}_k)$ denote the membership of $\mathbf{x}_k$ to cluster c, and ω is a parameter which controls the degree of fuzziness of the clusters (often a value between 1 and 2 is found suitable in application) (Zio & Baraldi, 2005). In the traditional algorithm (Dunn, 1974), the distance is Euclidean. The membership values $\mu_c(\mathbf{x}_k)$ and the cluster centers $\mathbf{v}_c$ are computed via an iterative procedure reported, for completeness, in Appendix A.

The information used to build the partition Γ^r is incomplete, since only p^r out of P parameters are used; on the other hand, the cluster structure thereby identified is assumed as reference in the comparison with the partition Γ^u, since it is built using only the p^r reliable parameters in $\mathbf{x}_k^r$. Due to the incompleteness of the Γ^r information base, one could observe disagreement between Γ^r and Γ^u not only when the values of the unreliable parameters in $\mathbf{x}_k^u$ used to build Γ^u are incorrect, but also when they give information which, despite being correct, is uncorrelated with that given by the reliable parameters in $\mathbf{x}_k^r$. For example, two different clusters can coincide when projected on S^r and be well separated on S^u instead; in such a situation,

one can obtain significantly different partitions Γ^r and Γ^u, despite the correctness of the unreliable parameters. Since in the choke valve case study the unreliable parameters $\mathbf{x}_k^u = [\dot{m}_o \quad \dot{m}_w \quad \dot{m}_g]$ are somehow correlated to the reliable parameters $\mathbf{x}_k^r = [\Delta P \quad \theta]$ (see eq. (1)), the possibility of observing uncorrelated parameters has not been considered in this work.

Operatively, the cluster analysis is performed as follows:

1. Identify the optimal number of clusters C to be used for the partitions Γ^r and Γ^u. This is obtained by considering the minimum of the compactness and separation validity function $s(C)$:

$$s(C) = \frac{\frac{1}{N}\sum_{c=1}^{C}\sum_{k=1}^{N}\mu_c^{\omega}(\mathbf{x}_k)d(\mathbf{x}_k, \mathbf{v}_c)}{\min_{i,j} d(\mathbf{v}_i, \mathbf{v}_j)} \tag{4}$$

   which represents the ratio between the cluster compactness, measured by the average distance of the patterns from their cluster centers and the separation between the clusters, measured by the minimum distance between two cluster centers. Notice that the numerator tends to decrease when the compactness increases and the denominator tends to increase when the separation increases. For this reason, in order to obtain a partition characterized by highly compact and well separated clusters, one has to find the optimal number of clusters which minimizes the validity function $s(C)$.
2. The fuzzy partitions Γ^r and Γ^u of the N data into C clusters are obtained using the FCM clustering algorithm (see Appendix A).
3. The clusters of Γ^r and Γ^u are bi-univocally associated $c_r \leftrightarrow c_u$ by minimizing the partition distance $D(\Gamma^r, \Gamma^u)$ between the partitions Γ^r and Γ^u. In this respect, the distance $D(\Gamma^r, \Gamma^u)$ defined in Zio and Baraldi (2005) has been used:

$$D(\Gamma^r, \Gamma^u) = \sum_{c=1}^{C}\sum_{k=1}^{N_{SI}} \frac{\left|\mu_c^r(\mathbf{x}_k) - \mu_c^u(\mathbf{x}_k)\right|}{2N} \tag{5}$$

   where $0 \leq \mu_c^{r,u}(\mathbf{x}_k) \leq 1$ is the membership of the k-th pattern to the c-th cluster of the partition Γ^r and Γ^u.
4. Crisp partitions Ω^r and Ω^u are obtained from the fuzzy partitions Γ^r and Γ^u, respectively, by assigning a pattern $\mathbf{x}_k$ to a given cluster c if its degree of membership to the cluster, $\mu_c(\mathbf{x}_k)$, exceeds a predefined threshold $\gamma \in (0,1)$, which represents the required degree of confidence for the assignment. If the condition $\mu_c(\mathbf{x}_k) > \gamma$ is not fulfilled for any cluster or if it is verified for more than one cluster, the pattern is not associated to any cluster. The crisp partitions Ω^r and Ω^u are compared by considering the difference between the sets of patterns $\mathbf{X}_{c_r}$ and $\mathbf{X}_{c_u}$ assigned to the associated clusters c_r and c_u. A large difference in the assignment of the patterns to the clusters is taken as a symptom that the information conveyed by the unreliable parameters may be misleading.

2.2.2 Results

According to this procedure, the dataset $\mathbf{X}^{SI}$ of the N_{SI} =259 SI available patterns of well 1, $\mathbf{x}_k^{SI}$, k=1,…, N_{SI}, is projected into the subspaces $S^r=\Delta P\times\theta$ and $S^u=\dot{m}_o\times\dot{m}_w\times\dot{m}_g$ of the measured (reliable) and allocated (unreliable) parameters of the choke valve case study, respectively. Two partitions Γ^r and Γ^u of the dataset $\mathbf{X}^{SI}$ into C=5 clusters are obtained using the FCM algorithm with degree of fuzziness ω=2.

The clusters of Γ^r and Γ^u are then coupled by minimizing the partition distance $D(\Gamma^r, \Gamma^u)$ in eq. (5) and the same cluster index c=1,…,5 is assigned to each member of the pair of associated clusters. The minimal value found for the partition distance is 0.47 which is high considering that, by definition, the maximum partition distance is 1. With a degree of confidence γ=0.4, 255 patterns out of the total 259 patterns of $\mathbf{X}^{SI}$ are assigned without ambiguity to the clusters of Γ^r and 219 to the clusters of Γ^u. The remaining patterns are ambiguous. Ambiguous patterns in Γ^r, which differ from those in Γ^u, are located at the boundaries between clusters 1 and 3 and clusters 2 and 3, and for this reason their membership to both clusters is large.

Figure 5 shows the partitions Γ^r and Γ^u of the 259 SI patterns in the space S^r. It can be seen that in Γ^r, the clusters are clearly separated, contrarily to what happens in Γ^u. Moreover, one can observe large differences in clusters' composition, e.g. many patterns that belong to cluster 1 in Γ^r are assigned to cluster 5 in Γ^u; patterns of clusters 2, 3 and 4, which are well separated in Γ^r, are, instead, mixed in Γ^u.

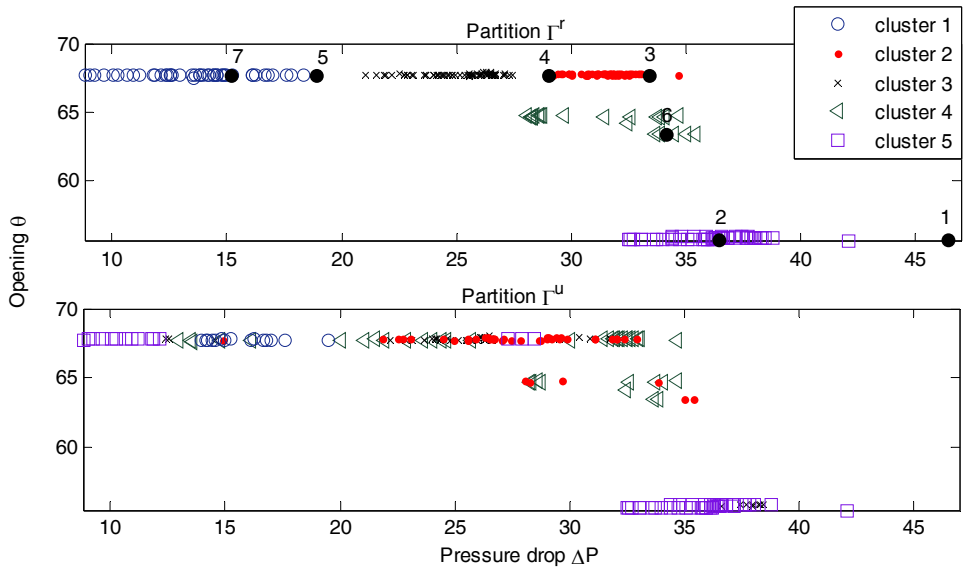

Figure 5: Visualization on the space $S^r=\Delta P\times\theta$ of the patterns assigned to the five clusters in Γ^r (top) and Γ^u (bottom). In the top graph, the WT patterns are also shown (black dots, numbered in chronological order).

Table 3 compares the number of patterns assigned to the same cluster in Γ^r and Γ^u (4th column) to the total number of patterns assigned separately to each cluster of Γ^r and Γ^u (2nd and 3rd column, respectively). Notice that, globally, less than half of the patterns (47%) assigned to a cluster of Γ^r are assigned to the associated cluster of Γ^u (last row in the Table).

Table 3: Number of patterns assigned to each cluster in Γ^r (2nd column), in Γ^u (3rd column), in both Γ^r and Γ^u (4th column) and percentage of patterns assigned to the same cluster in both partitions with respect to the number of patterns assigned to that cluster in Γ^r

Cluster c	Γ^r	Γ^u	Γ^r & Γ^u	$(\Gamma^r$ & $\Gamma^u)/(\Gamma^r)$
1	45	15	14	31.11%
2	56	49	15	26.79%
3	77	48	32	41.56%
4	25	47	15	60.00%
5	52	60	43	82.69%
Σ	255	219	119	46.67%

2.2.3 Supervised evolutionary clustering

To confirm the conclusions drawn in the previous Section, a further analysis based on a supervised clustering technique is here performed. Firstly, a partition Γ^s, as similar as possible to Γ^r, is obtained using a supervised evolutionary clustering technique based on Mahalanobis metrics in the space of all parameters.

A set $\mathbf{X}^{lab}$ of N_{lab} labeled training data is built by choosing, among the N_{SI} patterns of $\mathbf{X}^{SI}$, those belonging to one of the C clusters in Γ^r with a membership $\mu_{c_r}(\mathbf{x}_k) > 0.9$ and labeling them with the index c of the cluster they are assigned to. The evolutionary algorithm searches for the optimal metrics to be used by the FCM in order to achieve clusters as close as possible to the clusters of the labeled patterns.

In this view, each cluster c is defined by an individual geometric distance $d^2_{\mathbf{M}_c}$ through a dedicated Mahalanobis metric, defined by a definite positive matrix $\mathbf{M}_c$:

$$d^2_{\mathbf{M}_c}(\mathbf{x}_k, \mathbf{v}_c) = (\mathbf{x}_k - \mathbf{v}_c)^{\mathrm{T}} \mathbf{M}_c (\mathbf{x}_k - \mathbf{v}_c) \tag{6}$$

The classification task amounts to an optimization problem in which the metrics, i.e., the geometric distance functions, become additional parameters to be determined besides the fuzzy partition. The supervised target of the optimization is that of minimizing the partition distance D(Γ,Γ*) between the a priori known partition Γ and the obtained partition Γ^* as defined in eq. (5).

For the optimization, we integrate an evolutionary algorithm for determining the C optimal geometric distance functions (Yuan et al., 1995) with the FCM algorithm for determining the optimal fuzzy partition based on such distance. For more details on the algorithm one can refer to Zio & Baraldi (2005).

A measure of importance $I_{\mathbf{M}_c}(x_p)$ of a parameter x_p, p=1,…,P for the assignment of a pattern to a cluster c is:

$$I_{\mathbf{M}_c}(x_p) = \sum_{j=1}^{5} g^2_{c_{j,p}} \tag{7}$$

where $g_{c_{j,p}}$, $j,p=1,\ldots,P$ are the coefficients of the lower triangular matrix $\mathbf{G}_c=[g_{c_{j,p}}]$ for cluster c obtained from the decomposition of the Mahalanobis matrix $\mathbf{M}_c$ into its Cholesky factors $\mathbf{G}_c$, i.e., $\mathbf{M}_c=\mathbf{G}_c{}^{\mathrm{T}}\mathbf{G}_c$ (Zio and Baraldi, 2005).

The importance values $I_{\mathbf{M}_c}(\dot{m}_o)$, $I_{\mathbf{M}_c}(\dot{m}_w)$ and $I_{\mathbf{M}_c}(\dot{m}_g)$ associated to the allocated parameters are compared to those associated to the reliable ones, $I_{\mathbf{M}_c}(\Delta P)$ and $I_{\mathbf{M}_c}(\theta)$: if the importance of allocated and reliable parameters is similar, one can conclude that they both convey useful information for defining the partition Γ^s; vice versa, if the importance of the allocated parameters is lower than that of the measured parameters, one should doubt about their reliability, since the information they convey appears to be incoherent with that of the measured parameters.

Table 4 reports the measures of importance $I_{\mathbf{M}_c}$ obtained for the five parameters and for each cluster. The allocated parameters have low importance compared to the measured ones, meaning that they do not significantly contribute to the assignment of the patterns to any of the clusters.

Table 4: measures of importance $\mathbf{I}_{\mathbf{M}_c}$ of the different parameters

	Reliable parameters		**Allocated parameters**		
Cluster c	ΔP	θ	$\dot{m}_o$	$\dot{m}_w$	$\dot{m}_g$
1	2.221	1.770	0.095	0.048	0.105
2	2.410	5.933	0.000	0.001	0.002
3	2.175	4.443	0.050	0.009	0.011
4	0.362	7.847	0.013	0.696	0.008
5	0.288	3.802	0.044	0.097	0.199

The analysis performed in this Section has shown that the information conveyed by the allocated parameters, $\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$, i.e., the oil, water and gas mass flow rates, respectively, lowers the quality of the choke valve health indicator δC_V. For this reason, an empirical method for providing more accurate estimates of the mass flow rates has been developed and is presented in the following Section.

3 Improving the quality of the allocated parameters

After verifying that the allocated values of $\dot{m}_o$, $\dot{m}_w$, and $\dot{m}_g$ in the choke valve case study are noisy and unreliable, a procedure for improving the accuracy of these estimates is here proposed. This is done by means of empirical models which learn from a training set the relationships between the parameters, and provide as output an estimate $\hat{\mathbf{x}}_k^u$ of the allocated parameters in the input pattern $\mathbf{x}_k$. Different regression techniques such as those based on the use of principal component analysis (Baraldi et al., 2009), artificial neural networks (Fantoni & Mazzola, 1996; Marseguerra et al., 2006), support vector machines (Sun et al., 2005), evolving clustering methods (Chevalier et al., 2009) have been applied to this purpose. In this work, Kernel Regression models (Nadaraya, 1964; Atkeson et al., 1997) have been chosen.

3.1 Kernel regression

Kernel Regression (KR) models provide estimates by developing local models in the neighborhoods of the test patterns they are fed with. Estimates are obtained as weighted averages of the training patterns, with weights decreasing as the distance between the test and the training pattern increases. In this view, training patterns closer to the test pattern are conjectured to be more similar to it, thus giving the most relevant contribution to its estimate.

Let $\mathbf{X}^{trn}=\{\mathbf{x}_k\}$, $k=1,\ldots, N_{trn}$ be the training set used for the estimate of the test pattern $\mathbf{x}_{tst}$. To develop the KR model, parameters are divided into a predictor group (PG) and a response group (RG) (with the two groups eventually overlapping). For the estimate of $\mathbf{x}_{tst}$, the KR algorithm assigns to each training pattern $\mathbf{x}_k$ a weight $w_k=\mathrm{K}[d_{PG}(\mathbf{x}_{tst},\mathbf{x}_k)]$, where K is the kernel function which produces the weight for a given distance $d_{PG}(\mathbf{x}_{tst},\mathbf{x}_k)$, between the training and the test patterns, computed considering only the parameters of the predictor group. The estimate $\hat{\mathbf{x}}_{tst}^{RG}$ of the RG parameters of the test patterns is obtained as a weighted average of the RG parameters of the training patterns:

$$\hat{\mathbf{x}}_{tst}^{RG} = \frac{\sum_{k=1}^{N_{trn}} w_k \cdot \mathbf{x}_k^{RG}}{\sum_{k=1}^{N_{trn}} w_k} \tag{8}$$

The kernel function K must be such that training patterns with small distances from the test pattern are assigned large weights and vice versa. Among the several functions which satisfy this criterion, the Gaussian kernel is commonly used (Wand & Schucany, 1990):

$$K(d_{PG}) = \frac{1}{\sqrt{2\pi h^2}} \exp\left(-\frac{{d_{PG}}^2}{2h^2}\right) \tag{9}$$

where the parameter h defines the kernel bandwidth and is used to control how close training patterns must be to the test pattern to be assigned a large weight. In order to compute d_{PG}, the PG parameters are rescaled in the range [0, 1].

In the present case study, the choice of training dataset and predictor parameters is critical. In principle, both the WT and the SI patterns can form the training dataset; due to their greater reliability, only the WT should be used for training the KR models. However, when few WT data are available, such as for well 1, the allocated value can help completing the information available for the data-driven model. Then, we devised the four different models listed in Table 5, by differentiating the training set.

Finally, the response group is formed by the unreliable parameters that need to be estimated, i.e., $\mathbf{x}_k^{RG} = \mathbf{x}_k^u$.

Table 5: Model training procedures in case 1.

Model	Training set	Predictor parameters
1	Well test data X^{WT}	Measured $\mathbf{x}_k^{PG} = [\Delta P, \theta]$
2	Standard inspections data X^{SI}	Measured $\mathbf{x}_k^{PG} = [\Delta P, \theta]$
3	Well test data X^{WT}	Measured & allocated $\mathbf{x}_k^{PG} = [\Delta P, \theta, \dot{m}_o, \dot{m}_w, \dot{m}_g]$
4	Standard inspections data X^{SI}	Measured & allocated $\mathbf{x}_k^{PG} = [\Delta P, \theta, \dot{m}_o, \dot{m}_w, \dot{m}_g]$

3.2 Ensemble approach

Since the performance of the models depends on the characteristics of the parameter to be estimated and the intensity of the noise (Baraldi et al., 2011), it is difficult to identify a single best model.

Using an ensemble of models allows overcoming this dilemma. Indeed, the general idea underlying ensembles is to create many models and combine their outputs in order to achieve a performance better than that provided by each individual model in the ensemble (Perrone & Cooper,1992). Models' prediction diversity plays a fundamental role when ensemble approaches are devised. In fact, individual models committing diverse errors can be opportunely combined in such a way that the error of the aggregated prediction is smaller than the error of any of the individual models.

Different techniques for the aggregation of the outcomes of individual models have been proposed in the literature, the most common being statistics methods like the simple mean, the median and the trimmed mean (Baraldi et al., 2009; Polikar, 2007). Other aggregation techniques which allow improving the ensemble performance consider weighted averages of the model outcomes with weights proportional to the performance of the individual models. In this respect, both global approaches (in which the performance is computed on all the available patterns) and local approaches (which measure the performance only on the patterns closed to the test one) have been proposed (Baraldi et al., 2010). Since in the choke valve case study a complete input-output set of patterns is represented by the WT only, where only few WT are available, model weighting cannot be based on a measure of the performance of the individual model. For this reason, a new strategy is proposed in Section 3.2.1 based on the use of the Analytic Hierarchy Process (AHP) (Saaty, 1980).

3.3 Outcome aggregation with Analytic Hierarchy Process

AHP is a multi-criteria decision method that uses hierarchic structures to represent a decision problem and provides ranking of different choices (Saaty, 1980). It consists of two main steps: 1) structuring a hierarchy; 2) assigning priorities to the elements of each hierarchy level by comparative judgments of the elements based on a pre-defined scale.

In this application, AHP is used to assign performance weights to the models of the ensemble using relative performance measurements, without resorting to an absolute measurement of the model performance. The hierarchy structure sketched in Figure 6 is used. The four models on level 3 are compared with respect to the two criteria Z_1 and Z_2 of the level 2 towards the goal (level 1) of obtaining high model accuracy.

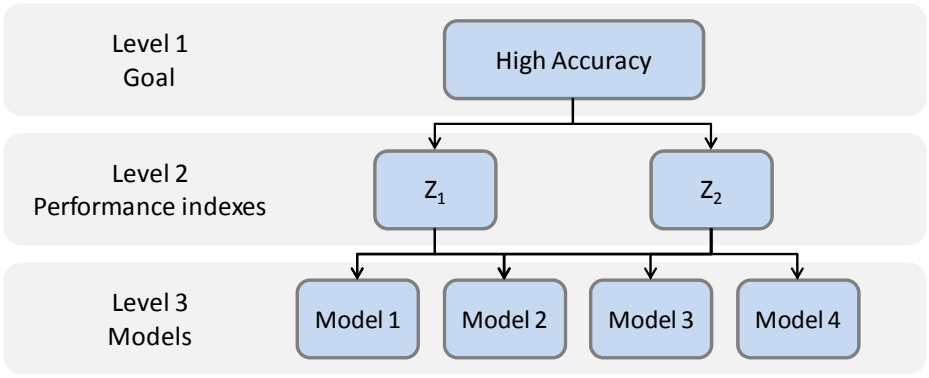

Figure 6: Model weighting hierarchy structure

The basic tool for assigning priorities to the elements of a level of the hierarchy are matrices of pairwise comparisons based on the criteria defined at the previous level. For the hierarchy of Figure 6, two matrices of comparisons $\mathbf{A}_{Z_1}$ and $\mathbf{A}_{Z_2}$ have to be defined, each one containing elements a_{ij} representing the relative importance of model i when compared to model j based, respectively, on criteria Z_1 and Z_2.

Once a matrix of comparisons $\mathbf{A}_{Z_l}$ is defined, the vector of priorities $\boldsymbol{\pi}_{Z_l}$ of the models in level 3 of the hierarchy with respect to criterion Z_l is given by the eigenvector associated to the maximum eigenvalue of matrix $\mathbf{A}_{Z_l}$. The priority vectors obtained for each criterion are weighted with the priority assigned to the corresponding criterion and averaged to obtain the overall priority vector $\boldsymbol{\pi}=[\pi_1, \pi_2, \pi_3, \pi_4]$ assigning the priority π_m to model m.

In the proposed aggregation method, the priorities assigned to each model are used as weights to aggregate the models' outcomes through a weighted average:

$$\hat{\mathbf{x}}_{tst}^{u} = \frac{\sum_{m=1}^{4} \pi_m \hat{\mathbf{x}}_{tst}^{u,m}}{\sum_{m=1}^{4} \pi_m} \tag{10}$$

where $\hat{\mathbf{x}}_{tst}^{u,m}$ is the estimate provided by model m of the unreliable parameters in $\mathbf{x}_{tst}^{u}$.

In this application, the first criterion Z_1, chosen to evaluate the relative importance $a_{ij}(\mathbf{x}_{tst})$ of model i with respect to model j in the reconstruction of a test pattern $\mathbf{x}_{tst}$, is the relative similarity of the two models outcomes $\hat{\mathbf{x}}_{tst}^{u,i}$ and $\hat{\mathbf{x}}_{tst}^{u,j}$ to the remaining models outcome $\hat{\mathbf{x}}_{tst}^{u,m}$, $m \neq i, j$. Assuming that the outcomes of the models left out of the pairwise comparison are distributed around the correct value, this criterion assigns larger weights to the model (i or j) whose outcome is more similar to those of the models left out.

The similarity of two patterns $\hat{\mathbf{x}}_{tst}^{u,i}$ and $\hat{\mathbf{x}}_{tst}^{u,m}$ has been estimated by the inverse of their Euclidean distance $d(\hat{\mathbf{x}}_{tst}^{u,i}, \hat{\mathbf{x}}_{tst}^{u,m})$; the relative importance $a_{ij}^{m}(\mathbf{x}_{tst})$ of a model i with respect to model j when model m is taken as reference is defined by:

$$a_{ij}^{m}(\mathbf{x}_{tst}) = d(\hat{\mathbf{x}}_{tst}^{u,j}, \hat{\mathbf{x}}_{tst}^{u,m}) / d(\hat{\mathbf{x}}_{tst}^{u,i}, \hat{\mathbf{x}}_{tst}^{u,m}) \tag{11}$$

and the entry a_{ij} of the comparison matrix **A** is given by the product of the relative importance values $a_{ij}^m(\mathbf{x}_{tst})$ m=1,…,4, $m \neq i, j$:

$$a_{ij} = \prod_{m \neq i,j} a_{ij}^m(\mathbf{x}_{tst}) \tag{12}$$

According to the AHP method, the quality of a matrix of comparison can be evaluated considering its consistency. Matrix $\mathbf{A}_{Z_1}$ is consistent if the following equation is satisfied for any i, j and k (Saaty, 1980):

$$a_{ij} a_{jk} = \frac{\pi_i}{\pi_j} \frac{\pi_j}{\pi_k} = \frac{\pi_i}{\pi_k} = a_{ik} \tag{13}$$

In our case, substituting eqs. (11) and (12) in eq. (13) gives:

$$a_{ij} a_{jk} = \prod_{m \neq i,j} \frac{d^{jm}}{d^{im}} \prod_{m \neq j,k} \frac{d^{km}}{d^{jm}} = \left[\frac{d^{jk}}{d^{ik}} \prod_{m \neq i,j,k} \frac{d^{jm}}{d^{im}} \right] \left[\frac{d^{ki}}{d^{ji}} \prod_{m \neq j,k,i} \frac{d^{km}}{d^{jm}} \right] = \frac{d^{jk}}{d^{ji}} \prod_{m \neq i,j,k} \frac{d^{km}}{d^{im}} = \prod_{h \neq i,k} \frac{d^{km}}{d^{im}} = a_{ik}$$

where $d^{ij} = d(\hat{\mathbf{x}}_{tst}^{u,i}, \hat{\mathbf{x}}_{tst}^{u,j})$ and, by definition, $d^{ij} = d^{ji}$. This shows that, in the proposed approach, matrix $\mathbf{A}_{Z_1}$ is consistent.

A second criterion Z_2 for evaluating the performance of a model takes into account the root mean square error (RMSE) in reconstructing the reliable parameters in $\mathbf{x}_{tst}^r$, i.e. the root mean square difference between the reconstructed and measured values. This second criterion is based on the fact that robust and reliable models should be able to correctly reconstruct the reliable parameters of $\mathbf{x}_{tst}^r$ despite the noise on the unreliable parameters of $\mathbf{x}_{tst}^u$. Since all model performances are evaluated with respect to the same reference, i.e. the reliable measurements in $\mathbf{x}_{tst}^r$, the pairwise comparison is not needed in this case, and the vector of priorities $\boldsymbol{\pi}_{Z2}$ is computed by taking for each model h=1,…,4, the inverse of its RMSE, i.e. $\pi_{Z_2}^m = 1/RMSE^m$.

Finally, the two criterions Z_1 and Z_2 of level 2 of the hierarchy are given the same importance and thus the priority vector $\boldsymbol{\pi}$ is given by:

$$\boldsymbol{\pi} = [0.5 \quad 0.5] \cdot \begin{bmatrix} \boldsymbol{\pi}_{Z_1} \\ \boldsymbol{\pi}_{Z_2} \end{bmatrix} \tag{14}$$

3.4 Results

The ensemble approach proposed in this Section is applied to the single choke disk of well 1 to improve the quality of the erosion indicator δC_V, in this case where very few WT are available. First, the mass flow rates $\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$ allocations are corrected, then the estimates are used to calculate the δC_V from eq. (1). Notice that, since the coefficients of the physical model in eq. (1) are not available to the authors, their value

has been derived by fitting this model to the data available for well 1. The resulting evolution of δC_V is shown in Figure 7 (continuous line).

In Baraldi et al., (2010), it is shown that in the case of very noisy parameters, the estimation error can be reduced by iterating the estimation procedure: the estimate of the noisy parameters obtained at the previous iteration is repeatedly given in input to the estimation model. Then, to improve the δC_V estimate, at one iteration, the values of the allocated parameters in $\mathbf{x}_k^u$ estimated by the ensemble at the previous iteration are given in input to the ensemble together with the original values of the reliable parameters in $\mathbf{x}_k^r$. The δC_V computed using the mass flow rates estimates obtained when this procedure is iterated 10 times is shown in Figure 7 (dots).

The values of δC_V obtained using the estimated values, both in case of 1 or 10 iterations, are more monotonic and more similar to those obtained in correspondence of the WT inspections (dots). Nevertheless, neither the AHP ensemble nor any of the single models considered can produce a totally monotonic indicator and some anomalous behaviors remains (e.g., some peaks such as the one occurring between 150 and 200 operational days which corresponds to a decrease in the pressure drop not followed by a decrease of the allocated values of the mass flow rates).

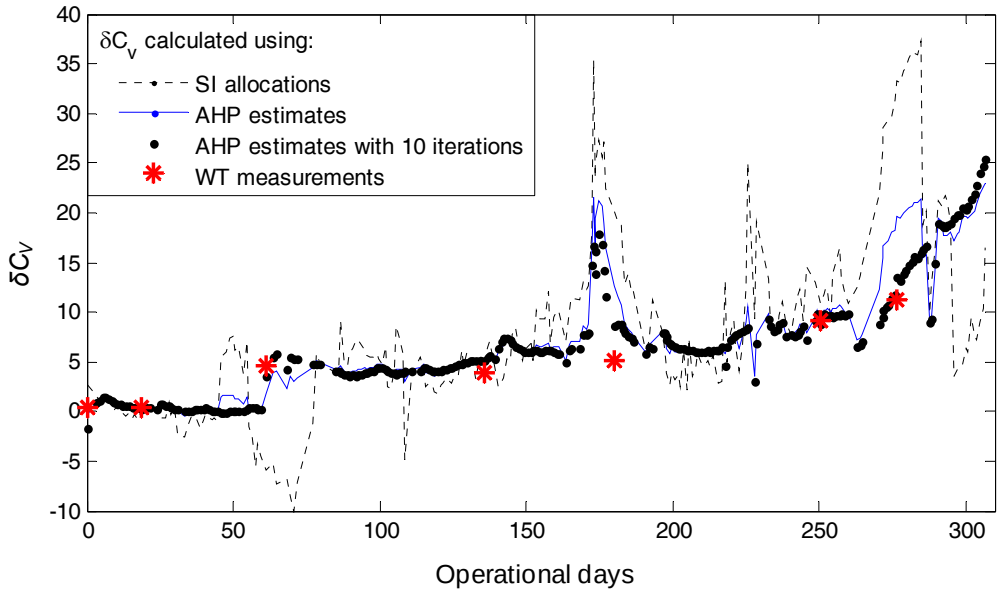

Figure 7: Comparison of the erosion indicator obtained using the allocated values of the mass flow rates (dshed line), those estimated by the AHP ensemble with 1 (continuous line) and 10 iterations (dots), and those measured during WT (asterisks)

Table 6 compares the value of the Spearman's rank correlation coefficient r_S measuring the monotonicity of the health indicator obtained using the SI dataset, the estimates of the four individual models and those of the AHP ensemble iterated 1 and 10 times. Results show that estimating $\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$ allows increasing the monotonicity of the health indicator δC_V with respect to that obtained by directly using the values computed during standard inspections. Furthermore, notice that in this case model 3 generates a health indicator slightly more monotone than that obtained by using the AHP ensemble. Nevertheless, since ensemble

approaches have proved to increase robustness compared to single best models, the estimates obtained by the latter should be preferred.

Table 6: Monotonicity r_s of the health indicator calculated using the SI dataset, the individual models estimates and those of the AHP ensembles with 1 (1-AHP) and 10 (10-AHP) iterations.

SI data	Model 1	Model 2	Model 3	Model 4	1-AHP ensemble	10-AHP ensemble
0.740	0.847	0.903	0.920	0.843	0.915	0.919

The KR models in Table 5 are then applied to the remaining 5 wells to obtain estimates of the mass flow rates $\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$. In analogy with well 1, for the test pattern $\mathbf{x}_{tst}$ collected for a specific choke, only patterns concerning the same choke are considered. This is done to better evaluate the method performance when few WT data are available. In practice, when estimating the test pattern $\mathbf{x}_k$ for a specific choke c, only the patterns $\mathbf{x}_j$, j=1,…,k-1 previously collected during the life of the c-th choke are used as training patterns.

Performance are given by the root mean square error in the estimation of the flow rates, normalized in the range [0,1]. In order to verify the performance of the models, we have considered a test set formed by the N_{WT} patterns of $\mathbf{X}^{\mathrm{SI}}$ collected the same day of a well test. Indeed, since an accurate measure of the process parameters under estimation is available for these patterns, they can be used to assess the performance of the estimation models. The subset retained for testing is formed by the N_{tst} patterns for which the value of the opening and the pressure drop considered for the allocation, i.e., which are given in input to the models for estimating the mass flow rates, are very similar to the values actually recorded at the time of the well test. Also, only data collected after the 5$^{\mathrm{th}}$ WT are retained for testing, so that at least 5 WT data are always available for training the KR ensemble models. Table 7 gives the number N_{tst} of well tests of each well used for validating the methods and compares the performance of the allocations, the four KR models, and the AHP ensemble.

Table 7: Number of validation WT patterns for each well.

Well	N_{tst}	RMSE (10^{-2})					
		SI	Mod 1	Mod 2	Mod 3	Mod 4	AHP
1	58	7.23	4.60	7.70	4.47	7.24	5.11
2	40	2.28	3.37	3.27	2.94	2.66	2.73
3	8	1.53	9.22	5.22	6.75	1.71	6.24
4	34	3.34	5.10	4.65	3.89	3.41	3.79
5	25	5.44	5.85	5.82	5.75	5.55	4.55
Average	165	4.68	4.82	5.59	4.28	4.82	4.23

Results show that in average the AHP ensemble outperforms both the allocations and all single models. However one can notice that the ensemble estimates do not always outperform the allocation. This can be due to two main reasons:

1) for some choke valve the allocations are highly accurate, and thus they cannot be improved through the KR estimates;
2) in some cases, e.g., those characterized by abrupt changes in the operating conditions of the choke such as large variations of the choke opening, the patterns used for the model training may not cover the range of parameter values of interest for the test pattern.

With respect to the situation at point 1) one can have a *prior* indication of the accuracy achieved by the allocations and the KR ensemble by considering the corresponding RMSE obtained in the estimates of the previous well test of the same choke; in this view, a different strategy for process parameters estimation is implemented in the next Section by resorting to a hybrid ensemble aggregated by a local aggregation strategy. On the other side, with respect to the situation at point 2) an estimate of the confidence of the ensemble outcome could be provided by analyzing the position of the test pattern with respect to the distribution of the training patterns.

4 Hybrid modeling approach

As shown in the previous Section the allocated values of $\dot{m}_o$, $\dot{m}_w$, and $\dot{m}_g$ can be noisy and unreliable (Baraldi et al., 2011); however the data-driven on-line procedure proposed for improving the accuracy of their estimates does not always outperform the allocations. For this reason, in this Section we propose a procedure based on a hybrid approach (Penha et al., 2002) combining the physics-based model used for the allocations and the ensemble of data-driven models.

There are two major approaches to hybrid modeling: the series and the parallel approach (Penha et al.2002). In the serial hybrid approach (Figure 8, left), data-driven models are used to model parameters in the physics-based model which is used to model the system (Psichogios & Ungar, 1992). In the parallel hybrid approach (Figure 8, right) data-driven models are trained to predict the residuals not explained by the first principle model (Thompson & Kramer, 1992).

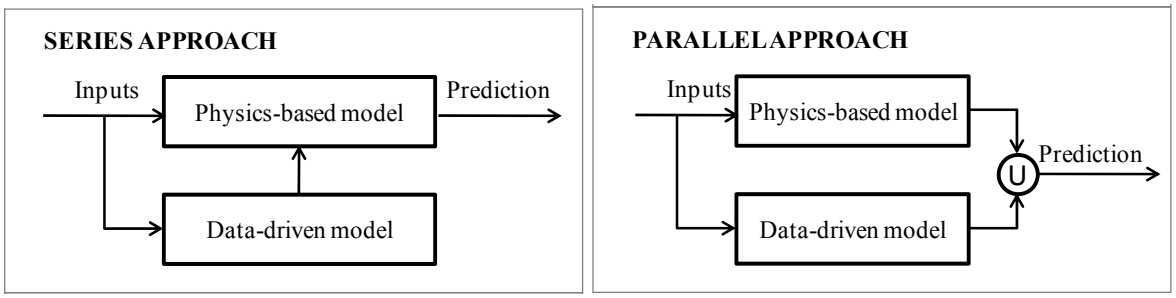

Figure 8: Schematic of series (left) and paraller (right) hybrid approaches

In this work, we wish to apply a hybrid modeling approach for the estimation of the mass flow rates. Neither the series and the parallel approaches can be applied in their basic formulation, since, as explained in Section 2, the inputs fed to the physics-based model for estimating the mass flow rates have not been recorded and thus no dataset is available for training a data-driven model (DD) in the framework described by Figure 8. For this reason, the different approach presented in Figure 9 has been devised: the values of the parameters $\dot{m}_o$, $\dot{m}_w$, and $\dot{m}_g$ in $\mathbf{x}^{u,PB}$ allocated by the physics-based model and the two measured parameter ΔP and θ are fed in input to a data-driven model, whose output is a new data-driven estimate $\mathbf{x}_k^{u,DD}$ of the allocated parameters. The data-driven estimate $\mathbf{x}_k^{u,DD}$ is then aggregated to the physics-based one, $\mathbf{x}_k^{u,PB}$, within a hybrid ensemble (HE) approach similar to the parallel configuration of Figure 8, to obtain improved estimates $\mathbf{x}_k^{u,HE}$ of the flow rates.

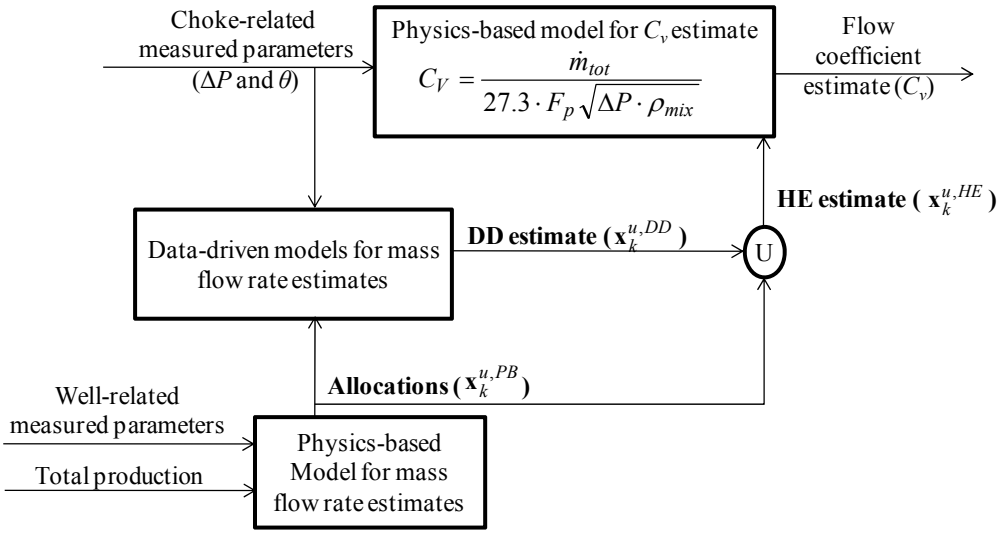

Figure 9: Schematic of the hybrid modelling approach proposed.

Once again, the ensemble of KR models has been adopted for the construction of the DD model. However, since for most of the wells considered an acceptable number of WT data is available, it is preferred to use only WT data for training the data-driven models. Many different models can still be devised by differentiating the predictor parameters based on which distances are computed in the KR algorithm. The four models listed in Table 8 have been retained for the construction of the ensemble, based on some considerations about the quality of the information conveyed by the different parameters (e.g., parameters θ and ΔP are reliable and thus used in all four models) and on the performance of the model in estimating the mass flow rates of a validation dataset (e.g., it has been verified that a model using as predictor parameters only the pressure drop ΔP and the opening θ performs very poorly).

These four models are used, within an ensemble approach, in a parallel configuration with the physics-based model (Figure 10); this allows overcoming the dilemma of selecting the optimal model, and increases the robustness of the final estimate by using the DD model only when needed. To achieve this, an effective aggregation strategy has been implemented, different than the AHP aggregation strategy which was developed for a situation characterized by the lack of a sufficient number of historical data upon which basing the evaluation of the different models performance. Indeed, when such data are available,

performance-based aggregation methods, e.g., the local aggregation, are in general preferred, since they are likely to produce more accurate results.

Table 8: Model training procedures in case 2.

Model	Training set	Predictor parameters
1	Standard inspections data X^{SI}	$\mathbf{x}_k^{PG} = [\Delta P, \theta, \dot{m}_o, \dot{m}_w, \dot{m}_g]$
2	Standard inspections data X^{SI}	$\mathbf{x}_k^{PG} = [\Delta P, \theta, \dot{m}_o]$
3	Standard inspections data X^{SI}	$\mathbf{x}_k^{PG} = [\Delta P, \theta, \dot{m}_w]$
4	Standard inspections data X^{SI}	$\mathbf{x}_k^{PG} = [\Delta P, \theta, \dot{m}_g]$

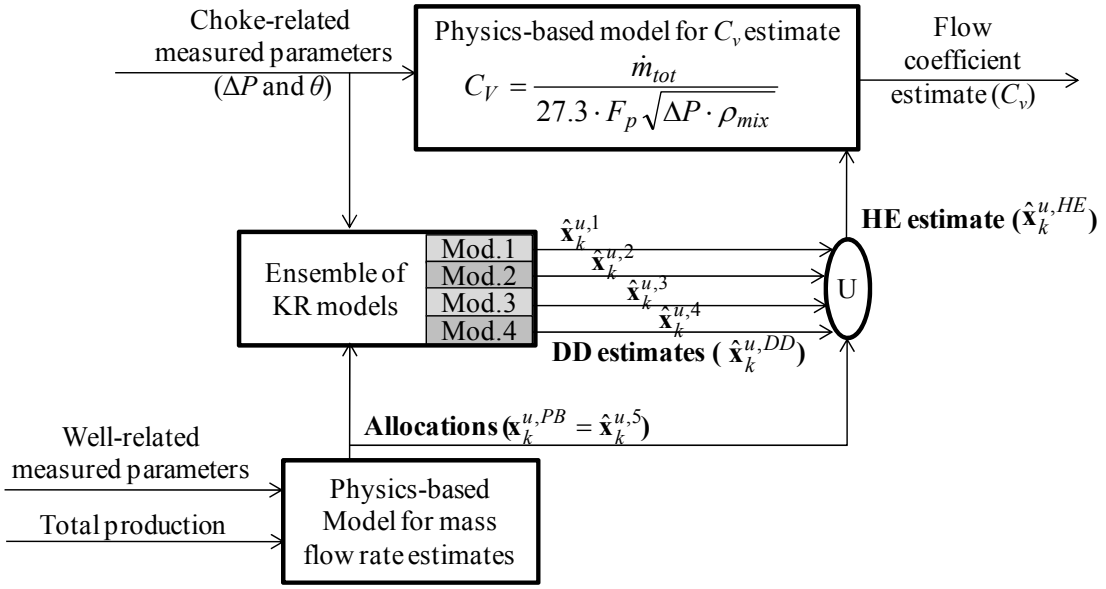

Figure 10: Schematic of the hybrid ensemble of model.

4.1 Local Aggregation

In this work, a local performance-based technique is applied for aggregating the outcomes of a hybrid ensemble of models. This allows exploiting the four data-driven models of the ensemble only when they actually outperform the physics-based model, thus avoiding corrupting its allocations when they are accurate.

In the local aggregation approach, the aggregation of the different model outcomes is guided by the local performance of each model, i.e., its reconstruction accuracy on patterns of training similar (and for this reason also called neighbors) to that under test. These methods rely on the idea that each model can perform well in some regions of the parameter space and poorly in others.

For each parameter p to estimate, the local performance aggregation approach here adopted assigns to the generic model m in the ensemble a weight $w_{p,m}$ proportional to the model performance evaluated on the Nn=3 training samples closest to the test pattern $\mathbf{x}_{tst}$. The estimation error made by model m in providing

the estimate $\hat{x}_i^{p,m}$ of the *p*-th parameter is then obtained by comparing $\hat{x}_i^{p,m}$ to the corresponding WT measurement $x_i^{p,WT}$; the local weight $w_{p,m}$ assigned to model *m* is taken as the inverse of its mean square estimation error over the *Nn* patterns closest to $\mathbf{x}_{tst}$:

$$w_{p,m} = \left[\frac{1}{N} \sum_{i=1}^{Nn} \left(\hat{x}_i^{p,m} - x_i^{p,WT} \right)^2 \right]^{-1} \tag{15}$$

The final estimate $\hat{x}_{tst}^{p}$ of parameter *p* is obtained as the weighted average of the multiple model estimates $\hat{x}_{tst}^{p,m}$.

$$\hat{x}_{tst}^{p} = \frac{\sum_{m=1}^{5} w_{p,m} \hat{x}_{tst}^{p,m}}{\sum_{m=1}^{5} w_{p,m}} \tag{16}$$

4.2 Results

The hybrid ensemble is here applied to estimate the mass flow rates $\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$ from patterns collected at wells 2 to 6. Notice that the model used to estimate the mass flow rates at well *W* is trained by considering only WT data from the same well, since it has been observed that well behaviors vary from one well to another. In practice, when estimating the test pattern $\mathbf{x}_{tst}=\mathbf{x}_k$ for a specific well *W*, only the WT patterns $\mathbf{x}_j$, j=1,…,k-1 previously collected during the life of the *w*-th well are used by the KR model to generate the estimates $\dot{m}_{o,w,g}$. Due to the fact that degraded choke valves are replaced, patterns collected for a single well may refer to different chokes.

The total number of available well tests used for training and the number of well tests used for testing is given in Table 9. The ensemble performance is evaluated by considering the RMSE between the estimated values $\dot{m}_{o,w,g}$ and the corresponding WT measurements, normalized in the range [0,1]. Similarly, the RMSE is computed also for the physics-based allocations and for an ensemble made only by the KR models (data-driven ensemble). Table 10 compares the RMSE values.

Table 9: Number of training and validation WT patterns for each well.

Well	N_{WT}	N*tst*
2	87	68
3	96	59
4	39	20
5	96	54
6	71	36

Table 10: Comparison of the performance of the hybrid ensemble with that of the physics-based model and of the data-driven ensemble

	RMSE (10^{-2})					
	Well 2	Well 3	Well 4	Well 5	Well 6	average
Physics-based	7.118	2.742	1.056	3.248	5.796	4.435
Data-driven ensemble	4.438	3.892	3.395	4.208	6.277	4.441 (+0.1%)
Hybrid ensemble	3.877	2.984	1.194	3.665	5.480	3.623 (-18%)

Results show that the data-driven ensemble outperforms the physics-based model only for well 2, producing a 38% reduction of the RMSE, whereas in the other four wells the physics-based model generates more accurate estimates than the data-driven ensemble. The hybrid ensemble mediates between the physics-based and data-driven approaches: it slightly reduces the good performance of the physics-based model on wells 3, 4 and 5, but, at the same time, enhances significantly the performance of both the data-driven and the physics-based model on well 2. This confirms that the local-performance aggregation technique is able to automatically distinguish those cases in which the physics-based model works properly from those in which it is convenient to integrate the data-driven models.

In Figure 11 the measured valued of oil, water and gas flow rates for the test patterns are compared with the estimates obtained by the physics-based model and by the hybrid ensemble. Results shows that the hybrid ensemble and the physics-based model estimates are in general very similar except for the oil flow in well 1 where the ensemble estimates significantly outperform the physics-based model estimates.

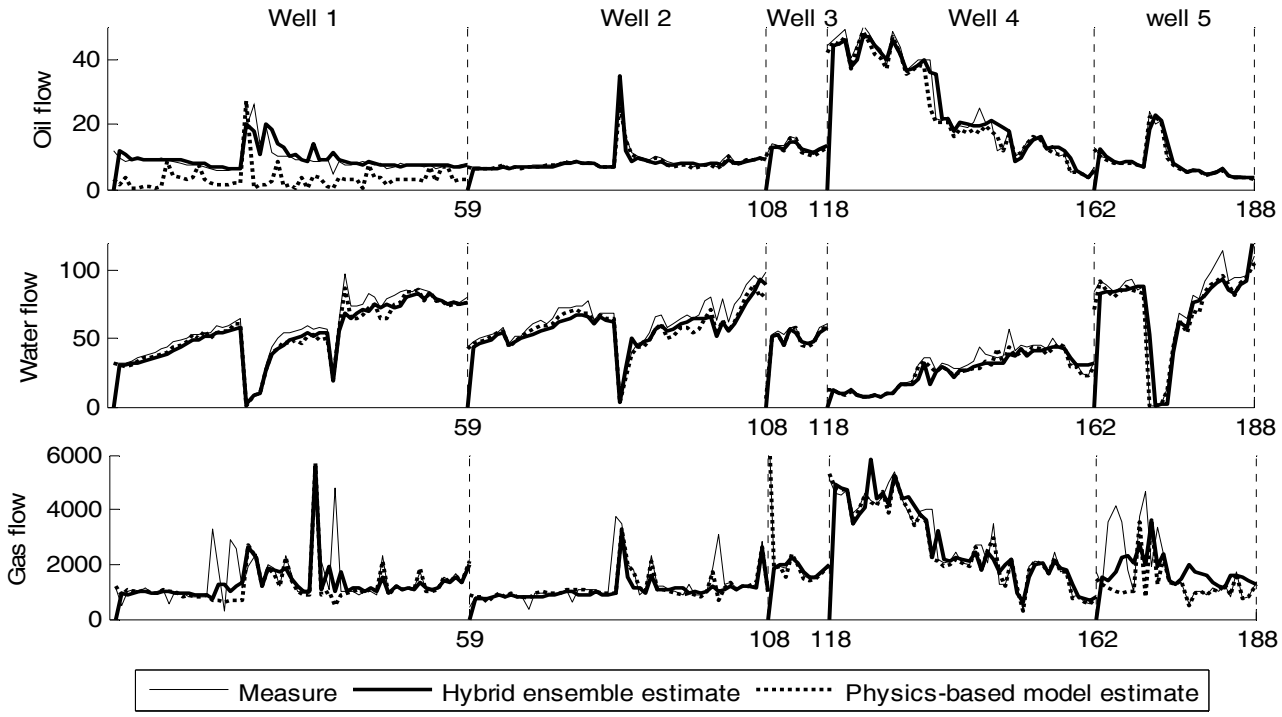

Figure 11: comparison of the measured valued of oil, water and gas flow rates with the estimates obtained by the physics-based model and the hybrid ensemble.

5 Erosion indicator

In this Section we discuss the possibility of using δC_V as an erosion indicator.

Table 11 reports the Spearman's rank correlation coefficient r_s, indicating the monotonicity of δC_V computed in correspondence of the well test for each eroding disk with more than 3 WT available. These results show that the δC_V can lack of monotonicity even when the mass flow rates are directly measured (see for example disk 1 of well 1 or disk 7 of well 4), and thus the erosion indicator is not affected by the large inaccuracies of their allocated values. In Table 11, the cells corresponding to those disks which were assessed as being *worn-out* or *with leakage* at the time of their changing are highlighted by a grey background; notice that in this case the coefficient r_s is always larger than 0.8, showing a good degree of monotonicity δC_V. The last line of the table shows the average value of r_s computed considering all disks (white background) or the disks with worn-out state (grey background). Table 12 shows the same results in the case all SI data are considered. The monotonicity in this case is slightly reduced.

Table 11: Monotonicity r_s of δC_V calculated using the WT data

	Disk 1	Disk 2	Disk 3	Disk 4	Disk 5	Disk 6	Disk 7
Well 1	0,46	0,76	0,89				
Well 2	1,00	0,89	0,99	0,59	0,74		
Well 3	1,00	1,00	0,86	0,86			
Well 4	0,56	0,82	0,95	1,00	0,96	0,90	-0,49
Well 5	0,93	0,73	0,96	0,86			
Average		0,79			0,92		

Table 12: Monotonicity r_s of δC_V calculated using the the SI data

	Disk 1	Disk 2	Disk 3	Disk 4	Disk 5	Disk 6	Disk 7
Well 1	0,46	0,70	0,59				
Well 2	0,95	0,81	0,93	0,42	0,83		
Well 3	0,07	0,83	0,90	0,86			
Well 4	0,65	0,70	0,93	0,94	0,92	0,93	0,51
Well 5	0,82	0,85	0,77	0,89			
Average		0,75			0,86		

It would be interesting to evaluate the monotonicity r_s also for the δC_V obtained using the estimated mass flow rates. This is not possible since we do not have the model used by the Company that supplied us the data, to derive the value of the flow coefficient from the five parameters ΔP, θ, $\dot{m}_o$, $\dot{m}_w$ and $\dot{m}_g$. Then, a model for the flow coefficient model is empirically derived from the available WT data by mean of an ensemble of artificial neural networks (ANNs) (Gola et al., 2012). Table 13 shows the value of r_s for the sequences of δC_V values obtained from the estimated mass flow rates using the ANNs ensemble. Results show that in this case the monotonicity has a net drop; this is probably due to the fact that the ANN model is not enough accurate. Further development of this model is required if one whishes to use the outcomes of the hybrid ensemble for improving the accuracy of the erosion indicator.

Table 13: Monotonicity r_s of δC_V calculated using the hybrid ensemble estimates

	Disk 1	Disk 2	Disk 3	Disk 4	Disk 5	Disk 6	Disk 7
Well 1	0,41	0,69	0,41				
Well 2	0,90	-0,08	0,79	0,14	0,58		
Well 3	0,45	0,89	0,54	0,55			
Well 4	0,35	0,23	0,64	0,76	0,76	0,88	-0,89
Well 5	0,72	0,65	0,93	0,77			
Average		0,53			0,65		

Figure 12 shows the evolution in time of the δC_V indicator for the 12 disks changed when worn-out. One can see that the failure time has large variation from about 100 to 800 working days. This depend on the condition of the wells whose sand production, the main cause of disks erosion, can have large oscillations. Also, the δC_V value at the beginning and at the end of the disk life varies significantly from one disk to another, thus making difficult to associate to the δC_V value a physical level of erosion of the disk.

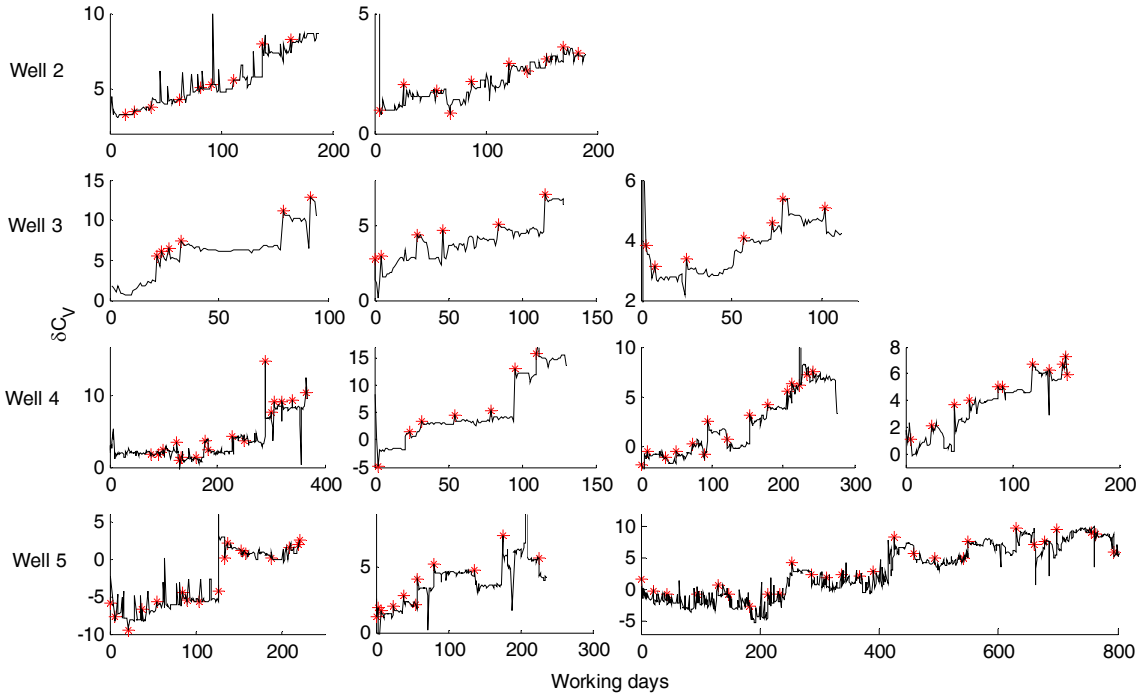

Figure 12: δC_V indicator for the 12 valves changed when worn-out.

Table 14 shows the value of the δC_V recorded at the first and the last days of the disk lives. Notice that these values have large variations, from a minimum of -13.89 and 2.54, respectively, to a maximum of 7.8 and 13.69. The variance in the health indicator value at failure gives a measure of its prognosability: a wide spread in the δC_V values at failure makes difficult to accurately fix the value of a failure threshold stating the boundary between a disk in good condition and an eroded one. Then, it is not possible to extrapolate from the

sequence of δC_V values the failure time of the system as the time at which the value of the δC_V indicator exceeds the failure threshold.

Table 14: δC_V values recorded the first (last) day of the choke disk lives

	Disk 1		Disk 2		Disk 3		Disk 4		Disk 5		Disk 6		Disk 7	
Well 1	-1,23	(-)	-0,11	(-)	7,80	(-)								
Well 2	3,23	(8.69)	2,03	(3.31)	2,32	(-)	5,30	(-)	-1,93	(-)				
Well 3	0,12	(-)	0,75	(10.66	1,78	(6.32)	3,00	(4.22)						
Well 4	-13,89	(-)	3,69	(9.58)	0,43	(-)	-1,85	(13.69)	-1,58	(3.34)	0,50	(5.89)	0,52	(-)
Well 5	-6,87	(2.54)	0,52	(-)	0,79	(4.30)	-0,48	(6.84)						

Min =	-13.89 (2.54)	**Max =**	7.8 (13.69)	**Mean =**	0.21 (6.61)	**Std =**	4.20 (3.44)

Since they can be observed even when only well test data are considered, the low monotonicity and prognosability of the δC_V as an erosion indicator is clearly not due to the poor accuracy of the allocated values of the mass flow rates, and has to be related to further sources of uncertainty and errors affecting the available information. We conjecture them to be:

1. the δC_V indicator depends also on other parameters (e.g., the valve opening) than the valve erosion;
2. the actual erosion state of the valve can be assessed only when the choke valve is changed; moreover this assessment is optional and completely based on the operator opinion.
3. the procedure for collecting data during well tests is error-prone, e.g., data are hand written, the possible failure of the separator used for flow rates measurements can remain undetected, etc;

With reference to the first itemized source of uncertainty, in Figure 13 the δC_V indicator calculated in correspondence of the well tests is shown as a function of the opening. From this figure, a weak correlation between the two parameters, with the δC_V decreasing as the opening increases, can be perceived and suggests that the δC_V indicator is a function not only of the erosion but also of the valve opening. The fact that the correlation coefficient between the δC_V indicator and the valve opening, shown in Table 15, is largely negative for most of the choke disks supports our hypothesis. If true, this could affect its performance as a prognostic indicator by contributing to the decrease of its monotonicity and prognosabiliy, since different values of the δC_V indicator would correspond to the same level of disk erosion.

The empirical identification of the relationship between the δC_V indicator and the valve opening, is a hard task, since the δC_V value depends, at the same time, from the opening and the disk erosion state, which is unknown for all the historical data available. In facts, the actual erosion state of the valve is assessed only qualitatively and only when the valve is changed; moreover this assessment is optional and it is not based on a rigorous classification of the erosion levels but rely completely on the operator opinion (second itemized source of uncertainty).

One can also notice that, even for fixed values of the opening, the δC_V calculated in correspondence of well test measurements does not always increase monotonically. This introduces some doubts about the reliability of the data collection procedure also during the well test (third itemized source of uncertainty). To solve this problem and thus improve the quality of erosion assessment, a validation of such procedure is deemed necessary.

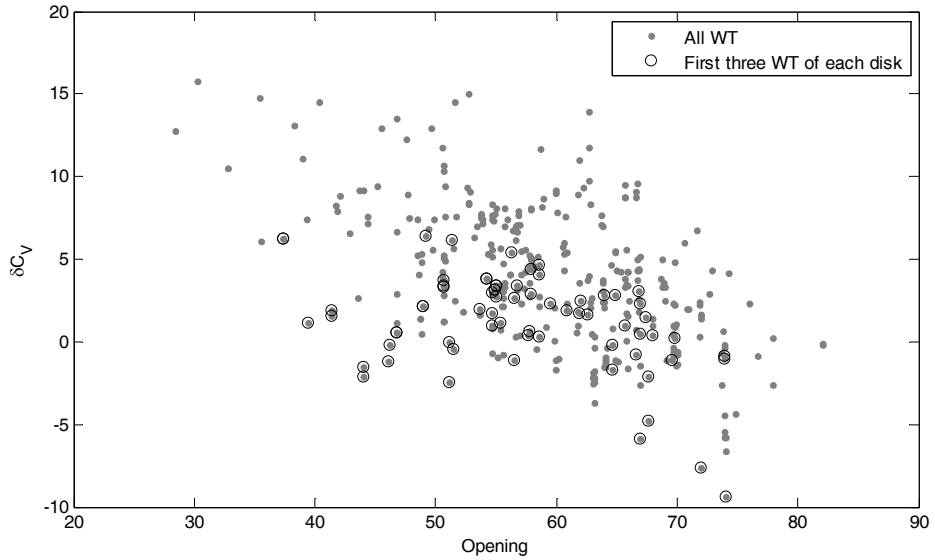

Figure 13: delta δC_V indicator versus opening

Table 15: Correlation with the opening of δC_V calculated using the WT (SI) data

	Disk 1		Disk 2		Disk 3		Disk 4		Disk 5		Disk 6		Disk 7	
Well 1	.38	(.12)	-.51	(-.59)	-.56	(-.60)								
Well 2	-.92	(.90)	.60	(-.35)	-.71	(-.24)	-.10	(-.26)	-.83	(-.80)				
Well 3	.58	.27	-.91	(-.86)	-.42	(-.19)	.38	(-.14)						
Well 4	-.93	(-.41)	-.78	(-.74)	-.82	(-.67)	-.94	(-.92)	-.65	(-.59)	-.77	(-.66)	-.83	(-.35)
Well 5	-.81	(-.73)	-.57	(-.10)	.46	(.59)	-.56	(-.42)						
Average	-0.44 (-0.42)						-0.44(-0.49)							

6 Conclusions and open issues

In this report, we have tackled the problem of providing a reliable health indicator of a choke valve used in offshore oil platforms which undergoes erosion. The health indicator is derived from the valve flow coefficient which is a valve parameter that regulates the analytical relationship between the pressure drop across the choke and the flow of oil, water and gas through the choke. However, the traditional procedure for calculating this indicator introduces some uncertainties and errors which undermine the possibility of using it for prognostic purposes. Some of the main problems which have been identified in this work are:

1. The actual valve flow coefficient is analytically calculated on a daily basis as a function of the pressure drop through the choke which is directly measured and oil, gas and water flow rates which are allocated based on the measured total production from a number of wells and on physical parameters (e.g., pressures and temperatures) related to the single well. Allocated values of oil, gas and water flow rates are conjectured to be the cause of the large inaccuracies and uncertainties in the calculation of the actual valve flow coefficient.
2. The difference between the theoretical and actual value of the valve coefficient (δC_V) is usually retained as an indication of the choke valve erosion. However, since the flow coefficient depends on the opening, the δC_V itself might depend on it, and not, as pursued, on the erosion level only.

3. The procedure for collecting data during well tests is error-prone, e.g., data are hand written, the separator used for flow rates measurements or the actuator imposing an opening value to the choke can fail, etc.
4. The actual erosion state of the valve can be assessed only when the choke valve is changed; moreover this assessment is optional and relies completely on the operator opinion.

The main contribution of this work has been given with respect to the problem at item 1., by devising a procedure to evaluate the accuracy of the allocated parameters and improve it based on the other available measurements (pressure drop and choke opening), which are conjectured to be reliable, and on a number of well tests performed throughout the valve life which provide few reliable measurements also for the oil, water and gas flow rates.

Fuzzy C Means clustering has been applied to verify the consistency of the measured and allocated parameters. A comparison of the FCM partitions obtained in the space of the measured and allocated parameters has been made and the importance of each parameter in the data partitioning by a supervised evolutionary clustering has been evaluated. The results of the analyses performed on the choke valve data have indicated the low quality of the allocated values of the mass flow rates. This has led to the development of a method for improving their accuracy. To this aim, Kernel Regression models have been devised. Different training procedures have been adopted to generate diverse models within an ensemble approach. To aggregate the outcomes of the individual models, an original technique based on the Analytic Hierarchy Process (AHP) method has been used. The application of the proposed method to the choke valve case study has allowed significant improvement of the oil, water and gas mass flow rates calculation and, as a consequence, it has improved the quality of the health indicator. However it has been noted that when allocations are accurate, the KR ensemble can worsen the accuracy of the mass flow rates estimates. To overcome this problem, a hybrid ensemble approach has been proposed. In the application considered, the hybrid ensemble is constituted by one physics-based model and four data-driven models based on kernel regression. The aggregation of the outcomes of the models in the ensemble is done by a weighting technique based on the local performance of the ensemble models. The results obtained from the application of the approach on a number of choke valves have confirmed that in those circumstances where the physics-based model is inaccurate, an ensemble approach appending to it a selection of diverse data-driven models can increase the accuracy of the estimates. The hybrid ensemble correctly favors the most accurate between data-driven and physics-based models, thus allowing the improvement of the parameter estimate.

Concerning the problem at item 2, the available data are not suited for deriving in an accurate and reliable way the relation between flow coefficient, opening and erosion. This, instead, could be achieved by using a set of flow coefficient values collected in correspondence of a fixed erosion level for different values of the choke openings. Also, a set of data assessing the level of erosion of a disk at different time steps along its operating life, would supply useful information for deriving the correlation between the δC_v indicator and valve erosion.

Concerning the problem at item 3, the simplest way to validate the well test measurement could is to repeat each measurement multiple times and verify the agreement between the different measurements. A more robust and reliable validation strategy requires the development of a physics-based or data-driven model estimating the value of the mass flow rates from other well parameters (e.g., pressures, temperatures, etc.).

Discrepancies between the values estimated by the model and those obtained from the well test, could then be retained as symptoms of well test measurement errors (Du et al., 1997; Heyen & Kalitventzeff, 2006).

Finally, concerning the problem at item 4, a standardized rating scale for evaluating the erosion level should be set up and used to assess the disk erosion, at least every time the disk is changed. Also, tests for assessing the valve performance at different erosion levels would help defining a threshold on the erosion level which can be accepted.

Appendix A: The unsupervised Fuzzy C Means technique

The Fuzzy C Means (FCM) technique is an unsupervised clustering technique, since it makes no use of a priori known information on the true classes of the data. The clustering is based on the minimization of a weighed sum Y of the distances $d(\mathbf{x}_k, \mathbf{v}_c)$ between the patterns $\mathbf{x}_k$ and the cluster centers $\mathbf{v}_c$,

$$Y = \sum_{c=1}^{C} \sum_{k=1}^{N} [\mu_c(\mathbf{x}_k)]^{\omega} d^2(\mathbf{x}_k, \mathbf{v}_c) \tag{A1}$$

where the weight $\mu_c(\mathbf{x}_k)$ denotes the membership of $\mathbf{x}_k$ to clusters c and ω is a parameter which controls the degree of fuzziness of the clusters (often a value of 2 has been found suitable as in Zio & Baraldi, 2005). In the traditional algorithm (Dunn, 1974) the distance is Euclidean:

$$d^2(\mathbf{x}_k, \mathbf{v}_c) = d_I^2(\mathbf{x}_k, \mathbf{v}_c) = (\mathbf{x}_k - \mathbf{v}_c)^{\mathrm{T}} \mathbf{I} (\mathbf{x}_k - \mathbf{v}_c) \tag{A2}$$

where $\mathbf{I}$ is the identity matrix.

The membership values $\mu_c(\mathbf{x}_k)$ which minimize Y (eq. (A1)) for a given a set of centers $\mathbf{v}_c$, c=1,…,C, are computed as in eq. (A3) and used in eq. (A4) to compute a new optimal set of clusters centers, which are in return used in eq. (A3) to update the membership values. The iterative procedure provides the optimal fuzzy partition of the dataset.

$$\mu_c(\mathbf{x}_k) = \left[\frac{1}{d_I^2(\mathbf{x}_k, \mathbf{v}_c)} \right]^{1/(\omega-1)} \Bigg/ \sum_{i=1}^{C} \left[\frac{1}{d_I^2(\mathbf{x}_k, \mathbf{v}_i)} \right]^{1/(\omega-1)} \tag{A3}$$

$$\mathbf{v}_c = \frac{\sum_{k=1}^{N} [\mu_c(\mathbf{x}_k)]^{\omega} \mathbf{x}_k}{\sum_{k=1}^{N} [\mu_c(\mathbf{x}_k)]^{\omega}} \tag{A4}$$

Based on the set of optimal centers $\mathbf{v}_c$, $c=1,\ldots,C$, a generic pattern $\mathbf{x}_k$ is assigned to cluster c provided that its membership $\mu_c(\mathbf{x}_k)$ exceeds a threshold $\gamma \in (0,1)$ representing the degree of confidence that $\mathbf{x}_k$ belongs to c. If the condition $\mu_c(\mathbf{x}_k) > \gamma$ is never fulfilled or if it is verified for more than one value of c, the pattern is not associated to any cluster.

References

Atkeson, C.G., Moore, A.W., and Schaal, S. 1997. Locally Weighted Learning. Artificial Intelligence Review, Vol. 11, pp. 11–73.

Baraldi, P., Zio, E., Gola, G., Roverso, D., and Hoffmann, M. 2009. Aggregation of randomized model ensemble outcomes for reconstructing nuclear signals from faulty sensors, Proc. ESREL Conf., Sept 7-10, Prague, CZ.

Baraldi, P., Cammi, A., Mangili, F., and Zio, E. 2010. Local Fusion of an Ensemble of Models for the Reconstruction of Faulty Signals, IEEE Trans. on Nucl. Sci., Vol.57, Issue 2, pp. 793 - 806.

Baraldi, P., Zio, E., Mangili, F., Gola, G., Nystad, B.H. 2011. Ensemble of kernel regression models for assessing the health state of choke valves in offshore oil platforms. Submitted for pubblication to IJCIS.

Birchenough, P.M., Cornally, D., Dawson, S.G.B., McCarthy, P., and Susden, S. 1994. Assessment of Choke Valve Erosion in a High-Pressure, High-Temperature Gas Condensate Well Using TLA. Paper SPE 28887 presented at the European Petroleum Conference , 25-27 October , London, UK. doi: 10.2118/28887-MS.

Bonissone, P.P., 2010. Soft Computing: A Continuously Evolving Concept, International Journal of Computational Intelligence Systems, Vol.3, No. 2, pp. 237-24.

Chatfield, C. 1988. What is the 'best' method of forecasting?, J. Appl. Statist. Vol. 15, pp. 19–39.

Chevalier, R., Provost, D., and Seraoui, R. 2009. Assessment of Statistical and Classification Models For Monitoring EDF's Assets, Sixth American Nuclear Society Int. Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies, Knoxville, USA, 2009.

Coble, J., and Hines, J. W. 2009. Analysis of Prognostic Opportunities in Power Industry with Demonstration. Paper presented at the 6th American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Controls and Human Machine Interface Technology, April, 5-9, Knoxville, Tennessee, US.

Devijver, P.A., and Kittler, J. 1982. Pattern Recognition: A Statistical Approach. Prentice/Hall, Englewood Cliffs, NJ.

Du, Y.G., Hodouin, D., Thibault, J. 1997. Use of a Novel Autoassociative ,Neural Network for Nonlinear Steady-State Data Reconciliation, American Institute of Chemical Engineers Journal, Vol. 43(7), pp. 1785-1796.

Dunn, J.C. 1974. A fuzzy relative of the isodata process and its use in detecting compact, well separated clusters. Cybernetics and Systems, Vol 3, Issue 3, pp. 32–57.

Fantoni, P.F., and Mazzola, A. 1996. Multiple-Failure Signal Validation in Nuclear Power Plants using Artificial Neural Networks, Nuclear technology, Vol. 113, Issue 3, pp. 368-374.

Garvey, D.R., Baumann, J., Lehr, J., Hughes, B., and Hines, J.W. 2009. Pattern Recognition Based Remaining Useful Life Estimation of Bottom Hole Assembly Tools. Paper SPE 118769 presented at the SPE/IADC Drilling Conference and Exhibition, 17-19 March, Amsterdam, NL. doi: 10.2118/118769-MS.

Gola, G., Nystad, B.H., Baraldi, P., Mangili, F., Zio, E. 2012. Health Assessment of Choke Valves undergoing Erosion with a Hybrid Ensemble Approach, 11th International PSAM 2011 & ESREL 2012 Conference, 25-29 June, Helsinki, Finland.

Haugen, K., Kvernvold, O., Ronold, A., Sandberg, R. 1995. Sand Erosion of Wear Resistant Materials: Erosion in Choke Valves. Wear, Vol. 186-187, Part 1, pp. 179-188.

Jain, A.K., Murty, M.N., and Flyn, P.J. 1999. Data Clustering: A Review. ACM Computing Surveys (CSUR), Vol. 31, Issue 3, pp. 264–323.

Jarrell, D.B., Sisk, D.R., and Bond, L.J. 2004. Prognostics and Condition-Based Maintenance: A New Approach to Precursive Metrics, Nuclear Technology, Vol. 145, pp. 275-286.

Heyen, G., and Kalitventzeff, B., 2006. Process Monitoring and Data Reconciliation, In L. Puigjaner, G. Heyen Eds., Part 3: Computer Aided Process and Product Engineering, Wiley-VCH Verlag, pp. 517–540.

Hines, J.W., and Usynin, A. 2008. Current Computational Trends in Equipment Prognostics. International Journal of Computational Intelligence Systems, Vol. 1, Issue 1, pp. 94–102.

Marseguerra, M., Zio, E., and Marcucci, F. 2006. Continuous Monitoring and Calibration of UTSG Process Sensors by Autoassociative Artificial Neural Network. Nuclear Technology, Vol. 154, Issue 2, pp. 224-236.

Metso Automation. 2005. Flow Control Manual. 4th edition.

Myers, J.L., Well, A.D., and Lorch, R.F. 2010. Jr. Research Design and Statistical Analysis. Taylor & Francis, New York.

Nadaraya, E.A. 1964. On Estimating Regression. Theory of Probability and Its Applications, Vol. 10, pp. 186–190.

Ngkleberg, L., and Sontvedt, T. 1995. Erosion in choke valves-oil and gas industry applications. Wear, Vol. 186-187, Part 2, pp. 401-412.

Nystad, B.H., Gola, G., Hulsund, J.E., and Roverso, D. 2010. Technical Condition Assessment and Remaining Useful Life Estimation of Choke Valves subject to Erosion. Proc. PHM Society 2010 Ann. Conf., October 11-13, Portland, OR.

Penha, R.L., J.W. Hines and B.R. Upadhyaya, "Application of Hybrid Modeling for Monitoring Heat Exchangers", 3rd Meeting of the Americas - America's Nuclear Energy Symposium, Miami, FL, October 16-18, 2002.

Perrone, M.P., and Cooper, L.N. 1992. When networks disagree: ensemble methods for hybrid neural networks, National Science Fundation, USA.

Polikar, R. 2006. Ensemble Based Systems in Decision Making: IEEE Circuits and Systems Magazine, Vol. 6 (3), pp. 21-45.

Polikar, R. 2007. Bootstrap-inspired techniques in computational intelligence, IEEE Signal Processing Magazine, Vol. 59, pp. 59–72.

Psichogios, D.C. and L.H. Ungar, A Hybrid Neural Network - First Principles Approach to Process Modeling, AIChE Journal, 1499--1512, October (1992

Saaty, T.L. 1980. The analytic Hierarchy Process, Planning, Priority Setting, Resource Allocation. McGraw-Hill, New York.

Sun, B.Y., Huang, D.S., and Fang, H.T. 2005. Lidar Signal Denoising Using Least-Squares Support Vector Machine. IEEE Signal Processing Letters, Vol. 12, Issue 2, pp.101-104.

Thompson, M.L., Kramer, M.A. 1994. Modeling chemical processes using prior knowledge and neural networks, AIChE J., Vol. 40 (8), pp. 1328-1340. doi:10.1002/aic.690400806.

Vachtsevanos, G., Lewis, F.L., Roemer, M., Hess, A., and Wu, B. 2006. Intelligent Fault Diagnosis and Prognosis for Engineering Systems, 1st edition, Hoboken, New Jersey, John Wiley & Sons.

Wand, M.P., and Schucany, W.R. 1990. Gaussian-based kernels for curve estimation and window width selection. Canadian Journal of Statistics, Vol. 18, pp. 197–204.

Wold, K., Hopkins, S., Jakobsen, T., Lilleland, S.E., Roxar, R.S., and Brandal, Ø. 2010. New Generation Software Integrates Intrusive and Non-intrusive Systems for Corrosion and Sand/erosion Monitoring. Paper SPE 130569 presented at the SPE International Conference on Oilfield Corrosion, 24-25 May, Aberdeen, UK. doi: 10.2118/130569-MS.

Yuan, B., Klir, G., Swan-Stone, J. 1995. Evolutionary fuzzy c-means clustering algorithm. Proc. Fourth IEEE Int. Conf. Fuzzy Syst., Vol 4, pp. 2221–2226.

Zhang, G.P. 2003. Time Series Forecasting using a hybrid ARIMA and neural network model, Neurocomputing. Vol. 50, pp. 159-175.

Zio, E., and Baraldi, P. 2005. Identification of nuclear transients via optimized fuzzy clustering. Ann. of Nucl. Energy, Vol. 32, pp.1068–1080.

Zio, E. 2012. Prognostics and health management of industrial equipment, in “Diagnostics and Prognostics of Engineering Systems: Methods and Techniques”, S. Kadry, Eds. IGI-Global.

Paper I

A Kalman Filter-based Ensemble Approach for Turbine Creep Prognostics

Piero Baraldi, Francesca Mangili, Enrico Zio

IEEE Transactions on Reliability 2012, Vol. 61(4), pp. 966-977

A Kalman Filter-based Ensemble Approach - with Application to Turbine Creep Prognostics

Piero Baraldi[a], Francesca Mangili[a], Enrico Zio[b,a]

[a]*Politecnico di Milano, Milano, Italy*

[b]*Chair on Systems Science and the Energetic challenge, European Foundation for New Energy-Electricite' de France, Ecole Centrale Paris and Supelec, Paris, France*

Abstract

The safety of nuclear power plants can be enhanced, and the costs of operation and maintenance reduced, by means of prognostic and health management systems which enable detecting, diagnosing, predicting and proactively managing the equipment degradation toward failure. In this paper we propose a prognostic method which predicts the Remaining Useful Life (RUL) of a degrading system by means of an ensemble of empirical models. The RUL predictions of the individual models are aggregated through a Kalman Filter (KF)-based algorithm. The method is applied to the prediction of the RUL of turbine blades affected by a developing creep.

Keywords: Creep, Ensemble, Kalman Filter, Prognostics and health management.

Acronym

RUL	Remaining Useful Life
KF	Kalman Filter
BAGGING	Bootstrapping AGGregatING
MTTF	Mean Time To Failure
MD	MeDian
GWA	Globally Weighted Average
SM	Single Model
MSE	Mean Square Error

Notation

τ	generic time instant
τ_i	i-th measurement time instant
t	time instant at which the prediction of the RUL is performed
τ_f	failure time

$\hat{\tau}_f$	predicted failure time
$\hat{\tau}_f{}^m$	failure time predicted by the m-th model of the ensemble
$d(\tau_i)$	direct measure of the system degradation level reached at time τ_i
d_{th}	system failure threshold
$\mathbf{D}_N$	dataset of N measured values $d(\tau_i)$
$\mathbf{D}^g$	g-th subset of $\mathbf{D}_N$ containing K temporally consecutive measurements; $g = 0,1,...,G-1$
G	number of subsets in which $\mathbf{D}_N$ is partitioned
$\mathbf{D}^{g,h}$	h-th bootstrapped replicate of dataset $\mathbf{D}^g$; $h = 1,2,...,H$
H	number of bootstrapped replicates generated from $\mathbf{D}^g$
$\theta_B \in (0,1]$	BAGGING fraction
$f(\tau;\mathbf{D})$	empirical model of the degradation evolution trained on dataset $\mathbf{D}$
$f^m(\tau;\mathbf{D}^m)$	m-th empirical model of the ensemble, trained on dataset $\mathbf{D}^m$, $m = gK + h$
$\hat{d}(\tau) = f(\tau;\mathbf{D})$	prediction of the degradation level at times τ produced by model f
$M = H \cdot G$	total number of ensemble models
$RUL(t)$	actual value of the RUL at time t
$\overline{RUL}(t)$	mean value of the RUL distribution at time t
$\hat{RUL}^m(t)$	prediction of $\overline{RUL}(t)$ produced by the m-th model of the ensemble
$\hat{RUL}(t)$	prediction of $\overline{RUL}(t)$ produced by the ensemble
$w^m(t)$	weight associated to the m-th model in the prediction of $\hat{RUL}(t)$
$\mathbf{z}(t)$	KF observations
$\mathbf{x}(t)$	KF state of the system at time t
$\mathbf{A}$	KF transition matrix
$\mathbf{H}$	KF observation matrix
$\mathbf{v}(t)$	process noise
$\mathbf{u}(t)$	observation noise
$\mathbf{Q}$	process noise covariance matrix
q	element (1,1) of matrix $\mathbf{Q}$
$\mathbf{R}$	observation noise covariance matrix
$\hat{\mathbf{x}}(t)$	KF *a posteriori* system state estimate
$\mathbf{P}(t)$	KF *a posteriori* error covariance matrix estimate
$\sigma^2_{RUL}(t)$	element (1,1) of matrix $\mathbf{P}(t)$
$\hat{\mathbf{x}}^-(t)$	KF *a priori* system state estimate
$\mathbf{P}^-(t)$	KF *a priori* error covariance matrix estimate
$\mathbf{K}(t)$	Kalman gain
$r(t)$	RUL rate of modification in time
$v(t)$	white Gaussian noise associated to the evolution in time of $\overline{RUL}(t)$
$\sigma^2_v(t)$	variance of $v(t)$
$u^m(t)$	observation noise due to the prediction error of the m-th model
$\sigma^2_{u^m}(t)$	variance of $u^m(t)$
$mse^m_{1\to N_p}$	performance of model m up to the last available measurement $d(\tau_{N_p})$
$\varepsilon(\tau_i)$	creep strain measured at time τ_i

ε_{th}	failure threshold for creep strain
$\delta\varepsilon_i$	creep strain increments between times τ_i and τ_{i+1}
Φ	turbine blade creep activation energy
α	Arrhenius law pre-exponential factor
n	Norton law creep stress exponent
$\Re$	ideal gas constant
T	turbine blade operating temperature
s	turbine blade applied stress
ω	rotational speed of the turbine
ρ	turbine blade density
r_{hub}	hub radius of the turbine
r_{tip}	tip radius of the turbine
δs	fluctuations in the stress applied to a specific blade
$MTTF$	estimated mean value of the blades failure time (τ_f)
$\sigma^2_{\tau_f}$	estimated variance of the blades failure time (τ_f)
b^m and d_o^m	parameters of the linear model defining $f^m(\tau;\mathbf{D}^m)$ for the creep growth process
$e(t)$	RUL prediction error
MSE	mean square RUL prediction error
$SCORE$	asymmetric indicator of the performance in the RUL prediction

1 Introduction

Prognostics and health management (PHM) aims at solutions for effective maintenance strategies, to reduce the risk of failure of equipments with low maintenance costs. These solutions are founded on knowledge and information about the equipment degradation evolution to *predict the equipment* Remaining Useful Life (RUL), i.e., the amount of time the equipment will continue to perform its function [1] [2].

In general, different forms of information and data may be available for the prognostic assessment of the trajectory to failure of an equipment undergoing degradation, e.g., time-to-failure data of similar equipments, direct or indirect measures of the degradation path of the monitored equipment or of a set of similar equipments under similar operating conditions, information on exogenous operational and environmental parameters, *deterministic, empirical or semi-empirical models of the degradation process,* etc. Depending on the situation, different methods, or their combination, may be applied [3] [4].

In this work, the following information is assumed to be available for the development of the prognostic model:

(I) the maximum degradation level beyond which the equipment fails (i.e., the failure threshold);
(II) a collection of values of parameters related to the equipment degradation level evolution.

The historical values of the equipment degradation level (information in II) are used to build a prognostic model which predicts the future degradation evolution; then, the equipment RUL can be estimated as the time remaining until the predicted degradation level reaches the failure threshold (information in I).

The primary objective of this work is to develop the prognostic model in the case in which the available values of the parameters indicating the equipment degradation level are affected by uncertainty. This situation is encountered in many applications where the direct measurement of the equipment degradation level is either not precise or not possible. In the latter case, the equipment degradation is typically inferred from other measured parameters. For example, in rotating equipments the crack depth is usually not directly measurable during operation, but vibrations can provide an indirect indication. Techniques for estimating the equipment degradation level from indirect parameters can be found in [5]; they are not considered in the present work, where we assume that a directly measurable parameter indicates the equipment degradation level, and investigate the effects on the RUL prognosis of the uncertainty affecting its value.

The approach here undertaken to satisfy such requirement exploits an ensemble of prognostic models. One motivation behind the use of an ensemble of models is that it increases the robustness of the prediction. By measuring the performance of the individual models and giving more importance to the predictions of the best performing ones, a robust and accurate RUL estimate can be obtained [6]. Furthermore, the ensemble is open to accommodate new information by simply adding to the old models new models built on the new information [7]. This allows reducing the computational efforts associated to the model updating.

For using an ensemble of prognostic models one must (a) build the diverse models and (b) combine their RUL predictions [6]. Concerning (a), in this work the individual models are all of the same type but trained on different datasets constructed by Bootstrapping AGGregatING (BAGGING) the degradation measurements available [8]. Concerning (b), the aggregation of the RUL model outcomes is performed by a Kalman Filter (KF)-based method which allows weighting the models according to their performance and, at the same time, filters out the noise in the predictions [9]. In this work, the Kalman Filter method is originally used, beyond its traditional purpose, as a device for the effective performance-based aggregation of the outcomes of the ensemble models.

The proposed approach is compared to i) a statistical approach based on the use of the mean time to failure (MTTF) ii) a single predictive model iii) an ensemble of predictive models based on different aggregation methods, such as the Median (MD) and the Globally Weighted Average (GWA) [10].

The case study considered for the comparison regards the prediction of the RUL of a turbine blade undergoing a creep degradation process [11]. The creep strain level values have been numerically simulated using a traditional model of the creep growth. Artificial data have been used in order to allow testing the method on a large number of different blade degradation trajectories and comparing the performance of the different approaches considering different levels of noise on the creep strain measurements.

The remaining part of the paper is organized as follows: Section 2 defines the prognostic problem and the general approach proposed to tackle it; in Section 3, the ensemble approach is described, focusing on the BAGGING technique used for the generation of diverse models and the Kalman Filter (KF)-based aggregation technique; in Section 4, the method is applied to a simulated case study concerning the prediction of the creep evolution in a turbine blade; finally, Section 5 states some conclusions and draws on potential future steps of the work.

2 Degradation modeling for prognostics

Let us consider an equipment undergoing a degradation process and let $d(\tau_i)$ be the measure of the degradation level reached at time τ_i. When the degradation reaches the threshold value d_{th}, the equipment fails; let τ_f indicate the time at which the failure occurs.

At the generic time *t* a dataset $\mathbf{D}_N$ of N values $d(\tau_i)$, measured at previous times $\tau_i \le t$, $t = 1,2,...,N$, is available. The objective is to estimate the RUL of the equipment, i.e., $\tau_f - t$.

To address the problem, one can develop an empirical model $f(\tau;\mathbf{D}_N)$ which, based on the N measurements in $\mathbf{D}_N$ available at the current time t, predicts the future component degradation level $\hat{d}(\tau)$ at times $\tau > t$, i.e, $\hat{d}(\tau) = f(\tau;\mathbf{D}_N)$. The prediction can be extended up to the time $\hat{\tau}_f$ at which the predicted degradation level $\hat{d}(t_f)$ exceeds the failure threshold d_{th}, i.e., $\hat{\tau}_f = f^{-1}(d_{th};\mathbf{D}_N)$. Since degradation is a stochastic process, both the future degradation level and the failure time are random variables. In the following, the model predictions $\hat{d}(\tau)$ and $\hat{\tau}_f$ should be considered estimates of the mean value of the distributions of the degradation level and failure time, respectively.

Notice that as time *t* proceeds, new measurements $d(\tau_i)$, $i = N+1, N+2,...$ become available and can be incorporated in the degradation model to update the estimate $\hat{\tau}_f$ of the failure time.

The choice of the modelling technique to be used for building the model $f(\tau;\mathbf{D}_N)$ depends on the specificity of the problem under study and on the information available. A plethora of approaches have been proposed to develop prognostic models: a typical distinction is made between model-based and historical process data-driven approaches. Within the former category, experimental data can be used to calibrate the parameters of the model within the state-observer formulation typical of a filtering problem with given state model. On the other hand, there are methods that do not use any explicit form of model and rely exclusively on process history data. Empirical techniques like artificial neural networks [9] [12], Support or Relevance Vector Machines [13] [14], local Gaussian regression [15], fuzzy similarity-based methods [16] are typical examples.

3 Ensemble approach to prognostics

It is shared experience that in forecasting, no single method is best for all situations [17]: real-world problems are often complex in nature and any single model may not be able to capture different patterns equally well. Ensembles of models build their outcome from a combination of the outcomes of the individual models. They have been proven an effective solution, since the errors of the individual models are compensated by the other models [6] [7]. Such compensation of the errors strongly depends on the diversity among the individual models of the ensembles [6] [8] [11] [20]: if these perform differently in different regions of the input space, the errors tend to balance out in the combination. Hence, in practice, an ensemble of models outperforms a single model only if the ensemble individual models are diverse [6], [7], [18]. to this purpose, different techniques to obtain diversity have been proposed [8] [9] [22] [23] and will be discussed in Section 3.1. On the other hand, when computational costs are an issue, a single model is preferred, given the computational burden of developing ensemble approaches. Notice also that an ensemble can naturally accommodate new information by adding new models tailored to the new information, without

discarding the old models in the ensemble. Updating the overall ensemble model in this incremental way requires reduced computational efforts in comparison to re-building a model from scratch based on all available (old and new) data [21] [7].

Section 3.1 describes how the diversity among the base ensemble models is obtained in the prognostic modeling approach here proposed; Section 3.2 illustrates the KF-based technique adopted to aggregate the single model outcomes.

3.1 Generation of multiple models by BAGGING

Diversity among the individual models of the ensemble can be obtained by different approaches: using different modeling techniques [22], differentiating the training parameters or the models structures [23], projecting the training data into different training spaces [9] [24] [25], or using different datasets to train individual models [8] [26]. A detailed discussion is beyond the purpose of this work; useful considerations can be found in [6]. In this work, diversity between the models has been obtained by resampling of the training data using the bootstrapping AGGregatING (BAGGING) method [6] [8] [27] [28]. It consists in generating multiple versions of a predictor by making bootstrap replicates of the dataset and using these as training sets of different empirical models. The vital element of the BAGGING technique is the instability of the prediction method: if perturbing the training set can cause significant changes in the predictor constructed, then bagging can improve accuracy [8].

In this work, in order to further increase the diversity between the models of the ensemble, the original training dataset $\mathbf{D}_N = \left[d(\tau_1),...,d(\tau_i),...,d(\tau_N)\right]$ is partitioned into G subsets each of which contains K temporally consecutive measurements: $\mathbf{D}^g = \left[d(gK+1), d(gK+2),...,d(gK+K)\right]$, $g = 0,1,...,G-1$. Next, the BAGGING technique is applied to each subset $\mathbf{D}^g$ to generate a number H of bootstrapped replicates $\mathbf{D}^{g,h}$, $h = 1,2,...,H$, of each subset $\mathbf{D}^g$ by randomly sampling (with replacement) for each replicate a fraction $\theta_B \in (0,1]$ of the K data contained in $\mathbf{D}^g$. Fig. 1 reports the procedure followed to generate the different training datasets.

For each of the $M = H \cdot G$ bootstrapped training datasets $\mathbf{D}^{g,h}$, $h = 1,2,...,H$, $g = 0,1,...,G-1$, a model $f_m(\tau;\mathbf{D}^m)$ is built (trained). Note that for simplicity of notation, the two indexes h and g have been grouped together in a single index $m = gK + h$. The differently trained models predict different failure times $\hat{\tau}_f{}^m = \hat{d}_m{}^{-1}(d_{th};\mathbf{D}^m)$ and, thus, different values of the mean value of the remaining useful life distribution $\hat{RUL}^m = \hat{\tau}_f{}^m - t$.

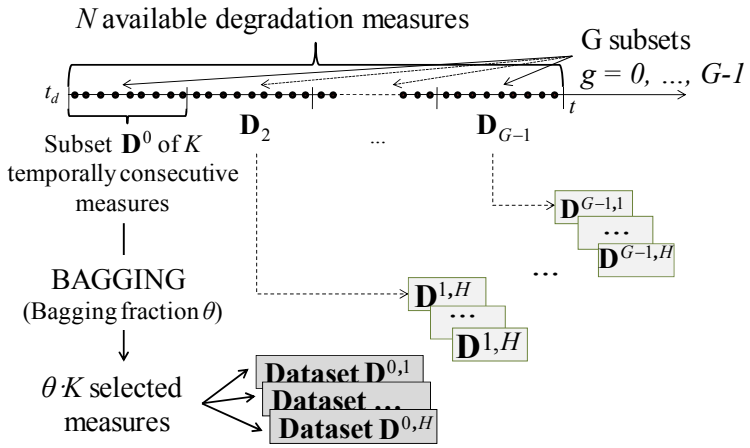

Figure 1: Procedure for the generation of diverse training datasets.

3.2 Aggregation of the models outcomes by the Kalman Filter

Common techniques for the aggregation of multiple models outcomes are based on statistics, e.g., the simple mean (SM) and the median (MD), or on models performance, e.g., the globally weighted average (GWA) [10] [29] [30].

In general, the aggregation requires associating a weight w^m to the prediction $R\hat{U}L^m$ made by each ensemble model f_m and combining the predictions by means of a weighted average:

$$R\hat{U}L = \frac{\sum_m w^m \cdot R\hat{U}L^m}{\sum_m w^m} \tag{1}$$

According to the SM approach, all weights are equal, i.e., $w^m = 1/(H \cdot G)$, $m = 1,.2,..., H \cdot G$; the GWA requires the computation of weights inversely proportional to the RUL prediction performance of each ensemble model; finally, for the computation of the MD, all weights are set equal to zero except for the weight w^{m_C} corresponding to the reconstruction $R\hat{U}L^{m_C}$ which lies in the centre of the distribution of the m models predictions $R\hat{U}L^m$, $m = 1,.2,..., M$.

In case of large noise affecting the degradation measurements $d(\tau_i)$, these aggregation techniques may lead to noisy predictions of the equipment RUL since all the ensemble models are built considering bootstrapped replicates of the same noisy data.

A possibility to improve the accuracy and robustness of the $R\hat{U}L(t)$ predictions in such cases is to properly filter them. In principle, the degradation measurements can be filtered instead of the RUL predictions, but in case of non-linear degradation processes this would be a more challenging problem since it cannot be solved with the basic (linear) KF. In this respect, this work proposes an effective strategy, inspired by the discrete KF [9], for both aggregating the multiple model outcomes and filtering the ensemble predictions.

In general, a KF [31] is a recursive method which uses a system dynamics model, a measurement equation and a set of measurements $\mathbf{z}(t)$, called observations, to estimate the state $\mathbf{x}(t)$ of the system at time t. In the discrete KF, the system's dynamics model and the measurement equation take the form in (2)(2):

$$\begin{cases} \mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) + \mathbf{v}(t) \\ \mathbf{z}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{u}(t) \end{cases} \tag{2}$$

Where t is a discrete time step index, $\mathbf{A}$ and $\mathbf{H}$ are the transition and observation matrices, and $\mathbf{v}(t)$ and $\mathbf{u}(t)$ are the process and observation noises, assumed to have a zero mean normal distribution with covariance matrices $\mathbf{Q}$ and $\mathbf{R}$, respectively.

The algorithm for the solution of (2) goes through a predict and an update phase. The predict phase (equations (3) and (4) below) uses the state estimate $\hat{\mathbf{x}}(t-\Delta t)$ and the estimate of the error covariance matrix $\mathbf{P}(t-\Delta t)$ from the previous time step to produce their *a priori* estimate $\hat{\mathbf{x}}^{-}(t)$ and $\mathbf{P}^{-}(t)$ at the current time step t:

$$\hat{\mathbf{x}}^{-}(t) = \mathbf{A}\hat{\mathbf{x}}(t-\Delta t) \tag{3}$$

$$\mathbf{P}^{-}(t) = \mathbf{A}\mathbf{P}(t-\Delta t)\mathbf{A}^{\mathbf{T}} + \mathbf{Q} \tag{4}$$

In the update phase (equations (5)-(7) below), the *a priori* prediction is combined with the current observations $\mathbf{z}(t)$ to obtain the *a posteriori* estimates $\hat{\mathbf{x}}(t)$ and $\mathbf{P}(t)$ of the state and error covariance matrix:

$$\mathbf{K}(t) = \mathbf{P}^{-}(t)\mathbf{H}^{\mathbf{T}}\left[\mathbf{H}\mathbf{P}^{-}(t)\mathbf{H}^{\mathbf{T}} + \mathbf{R}\right]^{-1} \tag{5}$$

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}^{-}(t) + \mathbf{K}(t)\left[\mathbf{z}(t) - \mathbf{H}\hat{\mathbf{x}}^{-}(t)\right] \tag{6}$$

$$\mathbf{P}(t) = \left[I - \mathbf{K}(t)\mathbf{H}\right]\mathbf{P}^{-}(t) \tag{7}$$

where $\mathbf{K}(t)$ is the Kalman gain at time t.

In the application of the Kalman Filter to the aggregation of the RUL predictions provided by the multiple ensemble models, the state vector $\mathbf{x}$ is constituted by the mean value $\overline{RUL}(t)$ of the equipment RUL at the t-th time step and by the term $r(t) = \left[RUL(t) - RUL(t+\Delta t)\right]/\Delta t$ representing the RUL rate of modification in time, i.e., $\mathbf{x}(t) = \left[\overline{RUL}(t) \quad r(t)\right]^{\mathbf{T}}$. Thus, assuming $\Delta t = 1$, the system can be described through a linear dynamic model as in (8) below, where the RUL rate of change $r(t)$ is set constantly equal to -1, since the RUL of a component decreases of one time unit for every time unit passed. Nevertheless, the mean value of the RUL distribution, which is affected by the stochasticity of the evolution of the degradation level between

time $t-1$ and t, does not deterministically decrease of 1 time unit at every time step: for this reason, an element of uncertainty is introduced in (8) through the additive white Gaussian noise $v(t)$.

$$\begin{aligned}\begin{bmatrix}\overline{RUL}(t)\\ r(t)\end{bmatrix} &= \begin{bmatrix}\overline{RUL}(t-1)+r(t-1)+v(t)\\ r(t-1)\end{bmatrix}\\ &= \begin{bmatrix}\overline{RUL}(t-1)-1\\ -1\end{bmatrix}+\begin{bmatrix}v(t)\\ 0\end{bmatrix}\end{aligned} \tag{8}$$

By comparison of (8) with the first equation of the system dynamics model in (2)(2), the observation matrix is $\mathbf{A}=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$ and the process noise covariance matrix is $\mathbf{Q}=\begin{bmatrix}\sigma_v^2(t) & 0\\ 0 & 0\end{bmatrix}$, where $\sigma_v^2(t)$ is the variance of $v(t)$.

The measurement vector $\mathbf{z}(t)$ of the Kalman Filter is assumed to be formed by the $RUL(t)$ predictions of the *M* models in the ensemble, i.e., $\mathbf{z}(t)=[\hat{RUL}^1(t) \quad ... \quad \hat{RUL}^m(t) \quad ... \quad \hat{RUL}^M(t)]^{\mathbf{T}}$.

The observation matrix H has been taken equal to:

$$\mathbf{H}=\begin{bmatrix}1 & 0\\ 1 & 0\\ \vdots & \vdots\\ 1 & 0\end{bmatrix} \tag{9}$$

in order to obtain that each observation $\hat{RUL}^m(t)$ is equal to the actual *RUL* at time t plus an observation noise $u^m(t)$.

Notice that although $\mathbf{z}(t)$ is called measurement vector in the KF terminology, in this case it is formed by the model outcomes which are not measurable quantities. As a consequence, in this application the observation noise covariance matrix $\mathbf{R}$ should contain the prediction error covariance of the models. Since, however, the prediction error covariances of the ensemble models $\sigma_{u^m}^2(t)$ are unknown and cannot be correctly computed from data, given that at the current time *t* the true value of the equipment RUL is not available, they have been roughly estimated. To this purpose, $\mathbf{R}$ has been taken equal to:

$$\mathbf{R}=diag(mse_{1\to N_p}^1,...,mse_{1\to N_p}^m,...,mse_{1\to N_p}^M) \tag{10}$$

with the model f_m performance $mse_{1\to N_p}^m$ computed by considering the relative difference between the time $\hat{\tau}_i^m$ at which the degradation level $d(\tau_i)$ is reached according to model f_m, i.e., $\hat{\tau}_i^m = f_m^{-1}(d(\tau_i))$ and the time τ_i, for all the N_p measurements $d(\tau_i)$ collected at time $\tau_i \le t$:

$$mse_{1\to N_p}^{m} = \frac{1}{N_p}\sum_{i=1}^{Np}(\hat{\tau}_i^m - \tau_i)^2 \tag{11}$$

Thus, the weight assigned to each model is proportional to its performance up to the last available measurement time N_p and matrix $\mathbf{R}$ can be updated each time a new measurement becomes available. Furthermore, assuming independence between the errors of the different ensemble models f_m, matrix $\mathbf{R}$ is taken diagonal.

With respect to the matrix $\mathbf{Q}$, representing the process covariance matrix, it has been assumed equal to:

$$\mathbf{Q} = \begin{bmatrix} q & 0 \\ 0 & 0 \end{bmatrix} \tag{12}$$

where q is an arbitrary constant which affects the smoothness of the resultant state estimate time series (lower values of q imply higher amount of smoothing).

Using (10) and (12) for the estimation of $\mathbf{R}$ and $\mathbf{Q}$, the $RUL(t)$ prediction is obtained according to (3)-(7). The drawback of this approach is that the first element of the resulting matrix $\mathbf{P}(t)$:

$$\mathbf{P}(t) = \begin{bmatrix} \sigma_{RUL}^2(t) & 0 \\ 0 & 0 \end{bmatrix} \tag{13}$$

can no more be interpreted as an estimate of the prediction error variance. Thus, the proposed KF approach which constitutes an effective technique for the aggregation of the outcomes of the ensemble models, does not provide an estimation of the uncertainty affecting the RUL.

In Appendix, it is shown how the aggregation is performed in the simplified case of an ensemble made of $M = 2$ models. The RUL estimate at time t turns out to be a weighted combination of the estimates generated by the m models of the ensemble plus a sort of $(m+1)$-th estimate given by the previous time estimate $\hat{RUL}(t-1)$ diminished of the one time unit which has passed: $\hat{RUL}^{m+1}(t) = \hat{RUL}(t-1) - 1$. The weights assigned to the different models depend on the observation noise covariance matrix $\mathbf{R}$, and thus are proportional to the model performance in (11), whereas that assigned to the $(m+1)$-th estimate depends on the process noise covariance matrix $\mathbf{Q}$.

Finally, notice that the approach does not imply any constraint on the linearity of the degradation model $f(\tau;\mathbf{D})$, since only the outputs of the predictive models are used in the KF scheme; thus, the non-linearity of the degradation process can be described by non-linear degradation models, with no additional complication for the KF-based aggregation procedure. Moreover, the approach removes the necessity of filtering the degradation measurements which is a challenging problem for non-linear degradation processes.

4 Case study

In this Section, the proposed prognostic method is applied to the RUL prediction of a turbine blade in which creep damage is developing. Creep is an irreversible deformation process affecting materials exposed to a load below the elastic limit for a protracted length of time and at high temperature. Notice that a turbine undergoing this degradation process can experience the loss of its blades, one of the most feared failure modes of turbomachinery since it is accompanied by abrupt changes in the power conversion system and in the reactor flow conditions [32]. Fig. 2 shows an example of high-pressure turbine deblading occurred in a German power plant [32].

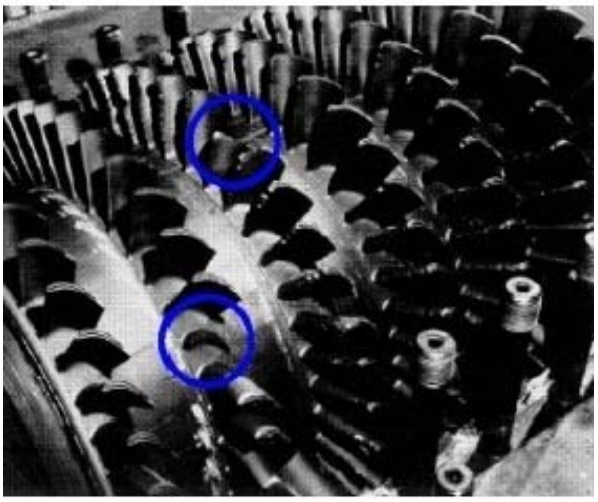

Figure 2: Deblading in a high pressure turbine [32].

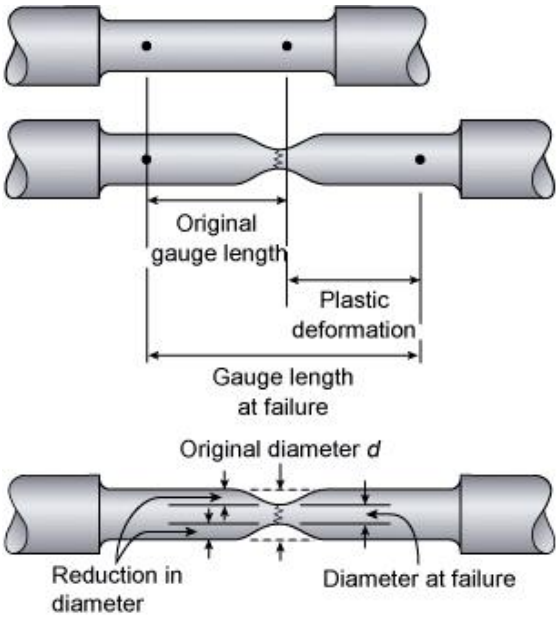

Figure 3: Schematics of a specimen before and after a creep test [33].

As shown in Fig. 3, the uniaxial creep deformation consists in an augmentation of the original length and a reduction of the diameter. In this work, the adimensional quantity ε, defined as the percentage of elongation of the turbine blade in the longitudinal direction with respect to its original length, is considered as measure of the creep strain. Methods for measuring blades deformation can be found in [34] and [35].

4.1 Creep growth model

The creep evolution has been simulated using the Norton Law, assuming that the dependence from the temperature follows the Arrhenius law [36]:

$$\frac{d\varepsilon}{dt} = \alpha \cdot \exp\left(-\frac{\Phi}{\Re T}\right) \cdot s(t)^n \tag{14}$$

where $d\varepsilon/dt$ is the creep strain rate, Φ is the activation energy, α and n are material inherent characteristics varying from one blade to another, $\Re$ is the ideal gas constant, T is the blade operating temperature and s

is the applied stress. For simplicity, the blade temperature is supposed equal to the gas temperature and the stress s is derived from the rotational speed ω of the turbine:

$$s(t) = \rho \frac{r_{tip}^{\ 2} - r_{hub}^{\ 2}}{2} \omega(t)^2 \tag{15}$$

where ρ is the blade density and r_{hub} and r_{tip} are respectively the hub and tip radiuses. The rotational speed ω and the gas temperature T are external parameters depending on the power setting of the gas turbine.

To numerically simulate the creep evolution, (14) is discretized in the time domain:

$$\varepsilon(t + \Delta t) = \varepsilon(t) + \alpha \cdot \exp\left(- \frac{\Phi}{\Re T(t)} \right) \cdot s(t)^n \cdot \Delta t \tag{16}$$

assuming $\Delta t = 5$ days.

The variations of parameters α and n from one blade to another have been simulated by sampling their values from normal distributions at the beginning of each new simulated degradation path, whereas the variation in time of the rotational speed ω and the gas temperature T is simulated by sampling their values $\omega(t)$ and $T(t)$ from normal distributions at each discrete time instant t. Finally the fluctuations in the stress applied to a specific blade, which are due to fabrication defects, aging and corrosion of the blade, to vibrations of the system or turbulences of the gas flow are modeled through a random variable δs added to the stress s in (16).

A turbine blade is considered within its useful life if the creep elongation strain in the longitudinal direction of the turbine blade is less than 1 or 2% of its initial length. Thus, the failure threshold for creep strain ε_{th} is set equal to the value of 1.5.

The values of the parameters $T(t)$, $\omega(t)$, r_{hub} and r_{tip} have been set with reference to the helium gas turbine of a Gas Turbine Modular Helium Reactor (GT-MHR) developed by an international consortium, with a targeted 286MWe generation per module [37]; the material inherent characteristics α, n and ρ are taken assuming that the blade is made of Ni-base cast Superalloy 713LC [37]. The distributions used for the parameters are reported in Table I.

4.2 The prognostic problem

The objective of the present case study is to predict the RUL of a degrading turbine blade using the following information:

- an estimate of the mean ($MTTF = 822.6$ days) and variance ($\sigma_{\tau_f}^2 = 5.83 \cdot 10^3$ days^2) of the blades failure time (τ_f) distribution. These values have been obtained by considering a set of 50 simulated degradation paths.
- the failure threshold value $d_{th} = \varepsilon_{th} = 1.5$;

- the first $N = 50$ direct measurements of the blade creep strain $\mathbf{D} = [d(1),...,d(50)] = [\varepsilon(1),...,\varepsilon(50)]$ (recorded every 20 days). Two different cases will be considered in the following Sections: a) $[\varepsilon(1),...,\varepsilon(50)]$ are directly obtained from eq. (16) without considering any measurement error (hereafter, they will be named undisturbed data) and b) an artificial white noise with standard deviation of 0.02 has been added to $[\varepsilon(1),...,\varepsilon(50)]$ in order to simulate the measurement error (noisy data). Fig. 4 shows an example of simulated creep growth path (left) corresponding to the undisturbed case and the associated noisy creep strain measurements (right).

Table I: Type of Distribution, Mean Value and Standard Deviation used for the Creep Growth Model Parameters

Variable	Symbol	Distribution	Units	Parameters of the distribution
Activation energy	Φ	Deterministic	kJ/mol	$\Phi = 290$
Norton Law parameters	α	Normal	$(N/m^2)^{-n}/h$	$\mu_\alpha = 3 \cdot 10^{-4}$, $\sigma_\alpha = 5\%$
	n	Normal	-	$\mu_n = 6$, $\sigma_n = 0.2\%$
Failure threshold for creep strain	ε_{th}	Deterministic	%	$\varepsilon_{th} = 1.5$
Operating temperature	$T(t)$	Normal	K	$\mu_T = 1100$, $\sigma_T = 1\%$
Rotational speed	$\omega(t)$	Normal	rpm	$\mu_\omega = 3000$, $\sigma_\omega = 1\%$
Density	ρ	Deterministic	Kg/m^3	$\rho = 8000$
Hub radius	r_{hub}	Deterministic	m	$r_{hub} = 0.7$
Tip radius	r_{tip}	Deterministic	m	$r_{tip} = 0.87$
Stress fluctuations	δs	Gamma (η, ξ)	MPa	$\eta = 2; \xi = 10$

The procedure proposed in Section 3.1 has been applied to the 50 available measurements $[\varepsilon(1),...,\varepsilon(50)]$ using the ensemble parameter values reported in Table II.

The initial state for the KF is set considering the estimated mean and variance of the blades failure time, i.e., $RUL(t=0) = MTTF$, $\sigma^2_{RUL}(t=0) = \sigma^2_{\tau_f}$.

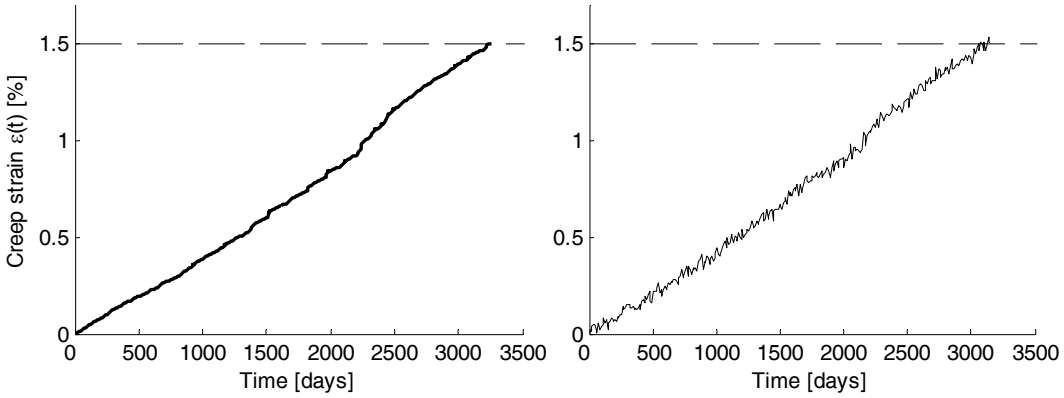

Figure 4: An example of undisturbed creep growth path (left) and corresponding noisy measurement sequence (right).

Table II: Ensemble Model Parameters

Parameter	Undisturbed data	Noisy data
G	2	2
K	25	25
H	5	5
ϑ_B	1	5
q	0.1	0.1

The creep strain measures available, $\varepsilon(\tau_i)$, are not uncorrelated, since they refer to a single degradation path, where the degradation level reached at any time steps depends on the past evolution of the degradation process. Thus, bootstrapping from the original dataset $\mathbf{D} = [\varepsilon(1),...,\varepsilon(50)]$ does not produce dataset replicates which are effective for constructing significantly diverse models. For the linear process considered in this work, the creep strain increments $\delta\varepsilon_i = \varepsilon(\tau_{i+1}) - \varepsilon(\tau_i)$ at different time steps are uncorrelated and time-independent; thus one can consider the dataset $\mathbf{D}' = [\delta\varepsilon_1,...,\delta\varepsilon_{50}]$ instead of $\mathbf{D} = [\varepsilon(1),...,\varepsilon(50)]$, without loss of useful information, and use it for training the models. For the RUL estimation, a linear model $f^m(\tau;\mathbf{D}^m) = \varepsilon_o^m + b^m\tau$ of the creep growth is built: the parameters b^m and ε_o^m are estimated by applying the least squares method to a fictitious degradation trajectory obtained as the cumulative sum of the increments in $\mathbf{D}^m$. Then, at the current time t (which is here assumed to coincide with the last measurement time τ_{N_p}), the RUL, i.e., the time needed to reach the failure threshold ε_{th}, is estimated by each model m as:

$$\hat{RUL}_m(t) = \frac{\varepsilon_{th} - \varepsilon(\tau_{N_p})}{b_m} \tag{17}$$

Notice that this is equivalent to assume that the degradation evolves linearly with slope b^m starting from the last available measurement $\varepsilon(\tau_{N_p})$ and that the parameter ε_o^m has no influence on the RUL prediction of model m.

4.3 Analysis of the Kalman Filter ensemble performance

In order to verify the proposed KF approach, the method has been applied to 500 simulated degradation paths different from those used in 1) for estimating the mean and variance of the blade failure time τ_f. The performance achieved by the proposed KF approach has been compared with those obtained using:

- the traditional statistical method based on the estimate of the mean time to failure (MTTF), according to which the RUL prediction is given by $\hat{RUL}(t) = MTTF - t = 822.6 - t$;
- a single linear predictive model (SM) built using all the $N = 50$ available creep strain measurements;
- the same ensemble of models aggregated using two classical aggregation methods: the median (MD) and the globally weighted average (GWA), which have been demonstrated to outperform the simple mean [10].

For each of the 500 simulated degradation paths, the different methods have been applied in correspondence of N_{tst} time instants, $\tau_{N_p} \geq \tau_{50}$. Notice that the new creep strain measurements acquired as time passes, $\varepsilon(\tau_{51}), \varepsilon(\tau_{52}), ..., \varepsilon(\tau_{N_p})$, are not used to update the values of the model coefficients b^m which are kept fixed to their values found by using the first $N = 50$ measurements. However, the last available information on the creep strain, $\varepsilon(\tau_{N_p})$ is used in eq. (11) for updating the computation of the model performances used in matrix **R** (10) and in eq. (17) for computing $\hat{RUL}_m(t)$. The performance of each method has been evaluated through two indicators: the mean square prediction error MSE and the SCORE:

$$RMSE = \frac{1}{500}\sum_{p=1}^{500}\frac{1}{N_{tst}}\sum_{N_p=50}^{N_{tst}+50} e(\tau_{N_p})^2 \tag{18}$$

$$SCORE = \begin{cases} \dfrac{1}{500}\displaystyle\sum_{p=1}^{500}\frac{1}{N_{tst}}\sum_{N_p=50}^{N_{tst}+50} \exp\left(-\frac{e(\tau_{N_p})}{13}\right) - 1 & e < 0 \\ \dfrac{1}{500}\displaystyle\sum_{p=1}^{500}\frac{1}{N_{tst}}\sum_{N_p=50}^{N_{tst}+50} \exp\left(-\frac{e(\tau_{N_p})}{10}\right) - 1 & e \geq 0 \end{cases} \tag{19}$$

where

$$e(\tau_n) = \hat{RUL}(\tau_n) - RUL(\tau_n) \tag{20}$$

The SCORE indicator is based on an asymmetric and exponential scoring function [38], such that late predictions are penalized more heavily than early predictions.

Table III shows the performance obtained when the methods are applied in correspondence of measurements collected from τ_{50} to the failure of the blade. Fig. 5 compares the performance of the KF ensemble with that achieved by the other methods in terms of the ratio, $R_{success}$, between the number of degradation paths in which the KF ensemble outperforms each of the other strategies and the total number of degradation paths considered (500). Finally, Fig. 6 compares the sequence of RUL predictions obtained with a single model, with a KF ensemble of models and with the MTTF-based method for a single degradation path with noisy measurements.

The obtained results show that the methods based on the modeling of the degradation process outperform on average the MTTF method. This is due to the fact that they take into account the information regarding the measured degradation level, which becomes progressively more important for the prognosis, as time passes. Notice that the KF ensemble is the only ensemble aggregation method which outperforms the single model and the MTTF methods both in the case of undisturded and noisy data, with sensible improvements in the latter case. The superior performance of the KF ensemble has two main reasons: the ability of the ensemble of balancing out the errors of its individual models and the noise filtering operated by the Kalman filter. To this purpose, Fig. 6 shows that the noise affecting the time series of the single model predictions is effectively reduced using the KF ensemble.

Table III: Comparison of the Prediction Performance of Different Prognostics Methods, with Models Built Using the First N Measurements only

Method	Undisturbed data $[\varepsilon(1),...,\varepsilon(50)]$		Noisy data $[\varepsilon(1),...,\varepsilon(50)]$	
	√MSE	SCORE	√MSE	SCORE
MTTF	11.703	3.263	11.703	3.263
Single Model	6.241	0.896	6.834	1.014
Median	6.076	0.897	9.167	5.208
GWA	5.785	0.797	7.397	1.037
KF	6.128	0.919	6.400	0.862

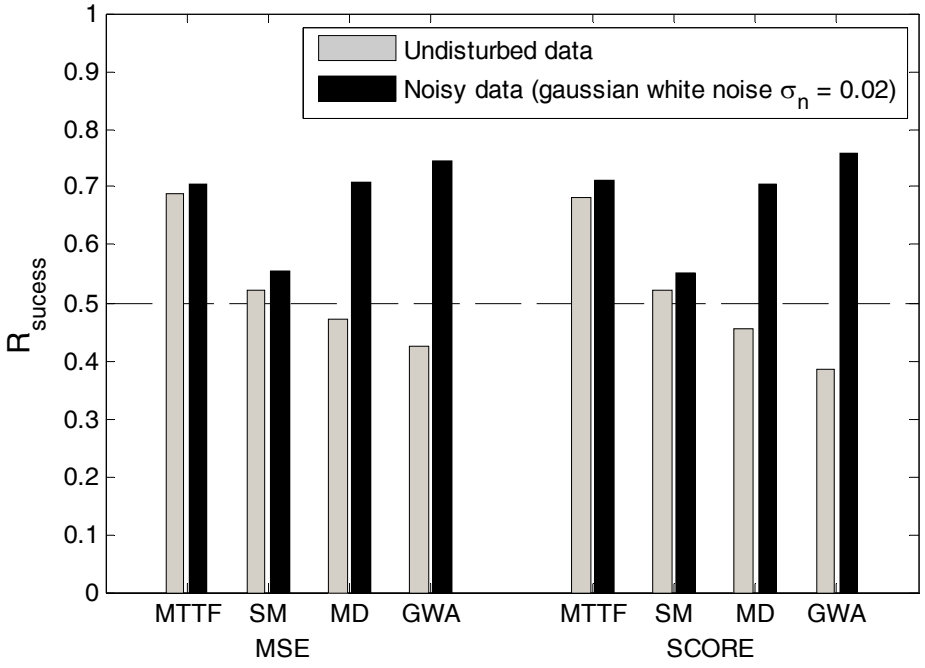

Figure 5: Fraction of degradation paths in which the KF ensemble outperforms the

Results also show that in case of noise the KF ensemble outperforms the other methods, whereas in the case of undisturbed data, the GWA and the median give better results since the filtering function of the KF ensemble seems to slightly damage the information contained in the correct measurements available instead of correcting the noise. On the other side, in case of noise, the traditionally aggregated ensembles perform poorly. With respect to a statistical-based aggregation method, such as the median, this is due to the fact that in case of noise the distribution of the model prediction is asymmetric and characterized by a large variance (Fig. 7). On the other hand, a performance based aggregation method, such as the globally weighted average, is characterized by large oscillations of the weights associated to the different individual models of the ensemble. This effect can be observed in Fig. 8 where the variations of the weights assigned to four different models of the ensemble during a creep degradation path in the case of undisturbed (top) and noisy (bottom) measurements are shown. Notice that the time evolution of the weights assigned to the generic *m*-th model of the ensemble at two successive time steps τ_{N_p} and $\tau_{N_p}+1$ is characterized by large oscillations in the case of noise. This is due to the fact that the weights are proportional to the model performance which may vary

remarkably since the predicted failure time, $\hat{\tau}_i^m$, depends from the noisy creep strain measurement $\varepsilon(\tau_{N_p})$ according to:

$$\hat{\tau}_i^m = \tau_{N_p} + \frac{\varepsilon(\tau_i) - \varepsilon(\tau_{N_p})}{b^m} \tag{21}$$

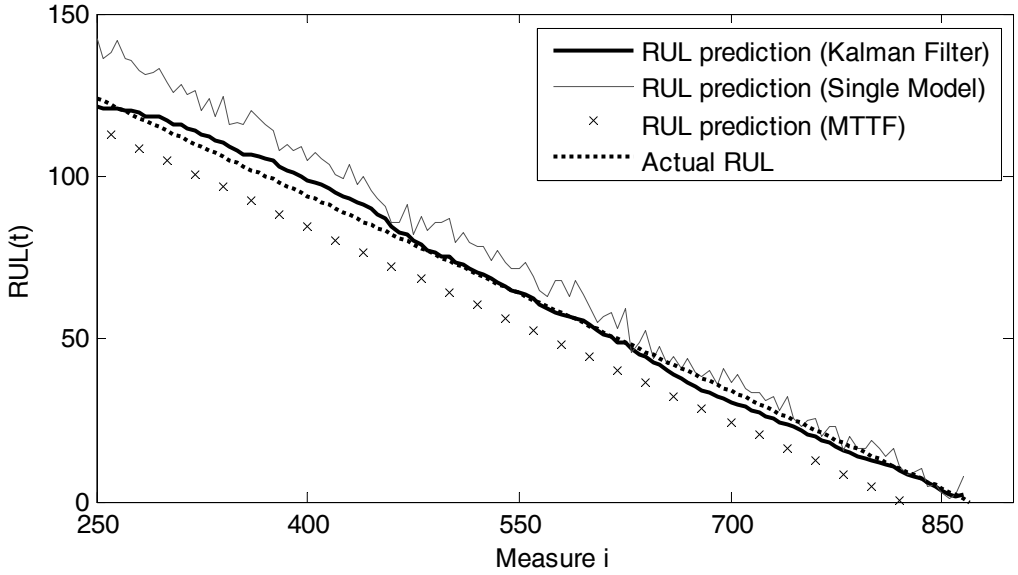

Figure 6: Comparison of KF Ensemble, Single Model and MTTF predictions of the RUL for a test degradation path.

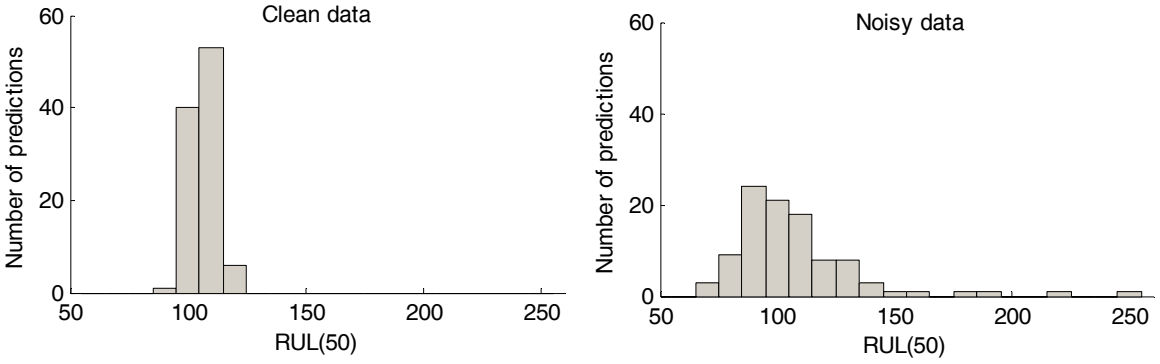

Figure 7: Histogram of the RUL predictions of 100 different models at t=50 in case of clean (left) and noisy (right) data.

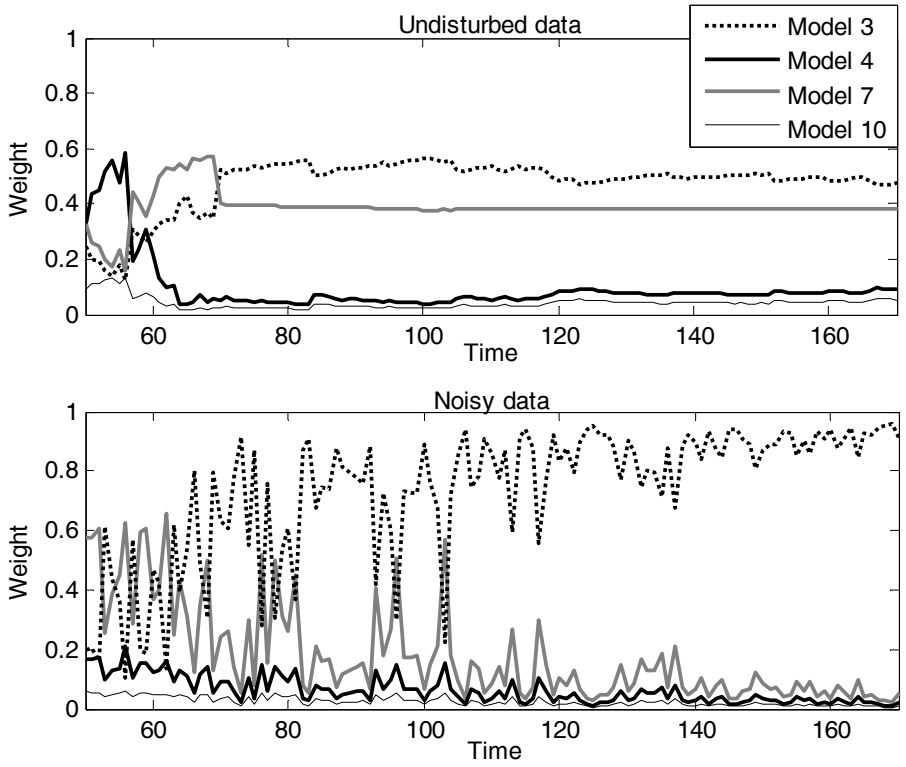

Figure 8: Weights assigned to 4 different models of the ensemble during a degradation path in case of undisturbed (upper) and noisy (bottom) data.

4.4 The choice of the parameters

In this Section we discuss the influence of the parameter setting on the KF ensemble performance. To this aim, the performance of the KF ensemble has been evaluated by varying, one at a time, the value of parameters ϑ_B, G or H. The results reported in Fig. 9 refer to the MSE committed in the RUL prediction performed at successive time steps from τ_{50} to the time of blade failure in 200 simulated blade degradation paths different from those used in the previous Section. Figure 9 compares the square root of the MSE (eq. (18)) made by the Single Model (continuous line) with that made by the KF ensemble for different values of ϑ_B (left), G (middle), H (right) in case of noisy data.

Figure 9 (left) shows that low values of ϑ_B should be preferred. This is due to the fact that low values of ϑ_B generate more diversity in the bootstrap replicates used to train the ensemble models [39]. With respect to the number of data subsets G, Figure 9 (middle) shows that the ensemble MSE tends to decrease until G reaches the value of 6. Since higher values of G cause more diversity between the data subsets used for training the classifiers, also in this case the best choice of the parameter is influenced by the diversity of the models. Notice, however, that, as the number of subsets becomes larger, the ensemble performance tends to decrease given that the number of patterns in the training subsets becomes very small. For example, the number of patterns with G=10 is less than 5, which is not sufficient for training accurate models. Fig. 9 (right) shows that the KF ensemble performance remarkably increases until H reaches the value of 10. Since the computational burden of developing a KF ensemble increases linearly with H, an optimal trade-off between performance and computational time should be found and one may accept to use values of H lower than 10, if it is necessary to reduce the computational burden.

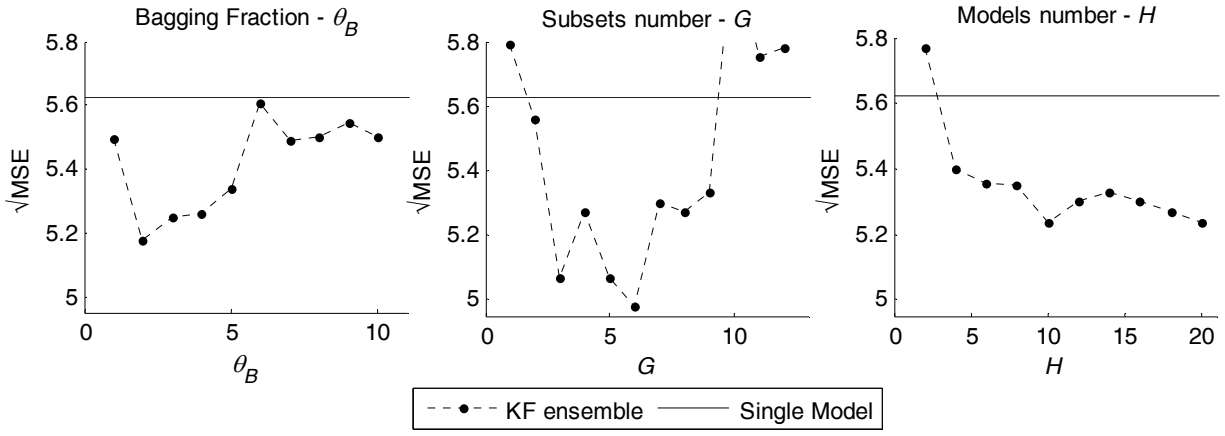

Figure 9: RMSE of the KF Ensemble predictions for different values of ϑ_B (left), K (middle), H (right) in case of noisy data.

With respect to the choice of the initial conditions ($RUL(t=\tau_N)$, $\sigma^2_{RUL}(t=\tau_N)$), in this work we have used the available estimates of the mean and variance of the blades failure time. In other cases, the choice of the initial condition $RUL(t=\tau_N)$ is expected to be derived from expert knowledge whereas the value of parameter $\sigma^2_{RUL}(t=\tau_N)$ can be established considering that the higher it is, the lower the confidence in the assumption about the initial state $RUL(t=\tau_N)$.

Finally, also the parameter q affects the RUL prediction since it determines the relative importance of the ensemble model outcomes with respect to that of the prediction $\hat{RUL}^{m+1}(t) = \hat{RUL}(t-1) - 1$ (Section 3.2). Fig. 10 shows the variation of the ensemble performance with q: we can see that values between 0.1 and 1 provide good performance whereas values lower than 0.1 imply too much smoothing and, on the contrary, values larger than 1 prevent from filtering out the noise.

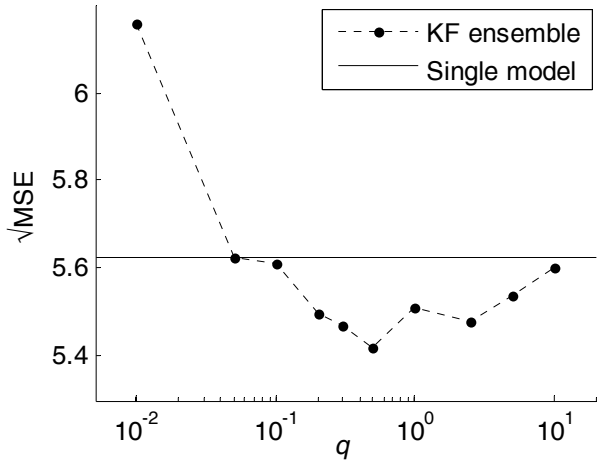

Figure 10: RMSE of the KF Ensemble predictions for different values of q in case of noisy data.

4.5 Incremental updating of the overall ensemble

In the prognostic scheme previously presented, the information contained in the new measurements $\varepsilon(\tau_i)$, $i = N+1, N+2, ..., N_p$, which become available as time passes, is used only for updating the weights in the KF and GWA aggregation methods. This allows saving computational time but can be not sufficiently accurate in the case in which the degradation trend varies remarkably. On the other side, retraining all the models of the ensemble at each time step could be prohibitively costly for complex degradation processes.

As stated in Section 3, an ensemble can naturally accommodate new information, without discarding the old models in the ensemble. Thus, a trade-off between computational costs and accuracy can be found by adding to the ensemble some few models trained by considering the new available information. To show this, $H = 5$ new models are added to the ensemble every time $N_w = 25$ new creep strain measurements become available. Each model is built by applying the BAGGING technique to the dataset made of the last $N_w = 25$ measurements.

Table IV: Comparison of the Prediction Performance of Different Prognostics Methods, with Models Updated Using Measurements Taken at Time t >50

Method	No Noise		Noise	
	√MSE	SCORE	√MSE	SCORE
Single Model	6.078	0.835	6.642	0.946
KF	5.969	0.859	6.126	0.788

In Table IV, the performance obtained with the updated KF ensemble are compared to those obtained by a single model whose parameter are recomputed every $N_w = 25$ time steps using all the available degradation measurements. Also in this case the KF ensemble outperforms the single model, although the linear degradation process considered in this case study can obviously be well approximated by a single model if enough data are available and an accurate estimate of the model parameters can be achieved.

5 Conclusions

Different forms of information and data may be available for the prognosis of the RUL of an equipment undergoing degradation. In this work, we have considered a practical situation in which some degradation measurements taken during the first part of the equipment degradation path are available and we know the failure threshold, i.e., the maximum degradation level which still allows the correct functioning of the equipment. The prognostics has been developed resorting to an ensemble of linear models, individually created by BAGGING and then aggregated using a Kalman Filter-inspired strategy.

The effectiveness of the proposed method has been tested in a simulated case study regarding the RUL prediction of turbine blades with a developing creep. The results obtained have confirmed that the use of an ensemble of models improves the accuracy of the RUL prediction with respect to a single model built using all the available measurements.

Furthermore, the Kalman Filter-inspired aggregation strategy allows reducing the noise affecting the prediction. In this respect, the criticality of properly assigning the Kalman Filter parameters has been thoughtfully addressed and some guidelines for setting the parameters according to the quality and quantity of the information available have been provided. In particular, the analyst is required to provide: a) an a priori estimate of the component RUL at time $t = \tau_N$, b) a parameter $\sigma^2_{RUL}(t = \tau_N)$ which is related to the analyst confidence on this estimate (higher the parameter value, lower the confidence) and c) a parameter q related to the smoothness of the time evolution of the RUL prediction (higher is q, smoother is the RUL prediction time series).

Finally, the proposed approach has been shown to provide an effective way for updating the prognostic model when new measurements of the degradation level of the equipment are collected.

In this work, the comparison between the single model and the proposed ensemble has been made by considering in both cases a linear model for the RUL prediction. Notice that the use of advanced data-driven modeling techniques for the RUL prediction is expected to increase the accuracy of the prognostic model and, thus, of both approaches. It is, however, expected that the performance of the ensemble will still be more satisfactory than that of the single model due to the ensemble ability of balancing out the errors of single models and of filtering the noise. In this respect, future work will be devoted to the verification of the ensemble performance employing different types of prognostic models and considering real data in the case study.

Appendix: two-models KF ensemble

For further understanding of the mechanisms underpinning the KF aggregation, let us consider a two-models ensemble which produces at (discrete) time instant *t* the two predictions $\hat{RUL}^1(t)$ and $\hat{RUL}^2(t)$, one for each model:

$$\mathbf{x}(t) = \begin{bmatrix} \overline{RUL}(t) \\ r = -1 \end{bmatrix}; \ \mathbf{z}(t) = \begin{bmatrix} \hat{RUL}^1(t) \\ \hat{RUL}^2(t) \end{bmatrix} \tag{A1}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}; \ \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}; \tag{A2}$$

$$\mathbf{Q} = \begin{bmatrix} q & 0 \\ 0 & 0 \end{bmatrix}; \ \mathbf{R}(t) = \begin{bmatrix} \sigma_1^2(t) & 0 \\ 0 & \sigma_2^2(t) \end{bmatrix}; \ \mathbf{P}(t) = \begin{bmatrix} p(t) & 0 \\ 0 & 0 \end{bmatrix} \tag{A3}$$

Equations (A1) (A2) and (A3) give the state and observation vectors $\mathbf{x}(t)$ and $\mathbf{z}(t)$, the transition and observation matrices $\mathbf{A}$ and $\mathbf{H}$, and the covariance matrixes $\mathbf{R}$, $\mathbf{Q}$ and $\mathbf{P}$, where $\sigma^2_{1,2}(t) = mse^{1,2}_{1 \to N_p}$

Developing the computations in (3) and (4) with $\Delta t = 1$, one obtains for the predict phase:

$$\hat{\mathbf{x}}^{-}(t) = \mathbf{A}\hat{\mathbf{x}}(t-1) = \begin{bmatrix} \overline{RUL}(t-1) - 1 \\ -1 \end{bmatrix} \tag{A4}$$

$$\mathbf{P}^{-}(t) = \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^{\mathbf{T}} + \mathbf{Q} = \begin{bmatrix} p(t-1)+q & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} p^{-}(t) & 0 \\ 0 & 0 \end{bmatrix} \tag{A5}$$

where $p^{-}(t) = p(t-1) + q$.

Similarly, expanding (5) (6) and (7) one obtains for the update phase:

$$\begin{aligned} \mathbf{K}(t) &= \mathbf{P}^{-}(t)\mathbf{H}^{\mathbf{T}}\left[\mathbf{H}\mathbf{P}^{-}(t)\mathbf{H}^{\mathbf{T}} + \mathbf{R}\right]^{-1} \\ &= \begin{bmatrix} p^{-} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} p^{-}+\sigma_1^2 & p^{-} \\ p^{-} & p^{-}+\sigma_2^2 \end{bmatrix}^{-1} = \frac{1}{p^{-}\sigma_1^2 + p^{-}\sigma_2^2 + \sigma_1^2\sigma_2^2}\begin{bmatrix} p^{-}\sigma_2^2 & p^{-}\sigma_1^2 \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{A6}$$

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}^{-}(t) + \mathbf{K}(t)\cdot\left[\mathbf{z}(t) - \mathbf{H}\cdot\hat{\mathbf{x}}^{-}(t)\right] = \begin{bmatrix} x_1(t) \\ -1 \end{bmatrix} \tag{A7}$$

$$\begin{aligned} x_1(t) &= \overline{RUL}(t-1) - 1 + \frac{1}{p^{-}\sigma_1^2 + p^{-}\sigma_2^2 + \sigma_1^2\sigma_2^2}\cdot\begin{bmatrix} p^{-}\sigma_2^2 & p^{-}\sigma_1^2 \end{bmatrix}\begin{bmatrix} \hat{RUL}^1(t) - \overline{RUL}(t-1) + 1 \\ \hat{RUL}^2(t) - \overline{RUL}(t-1) + 1 \end{bmatrix} \\ &= \frac{1}{w^1 + w^2 + w^K}\{w^1\hat{RUL}^1(t) + w^2\hat{RUL}^2(t) + w^K[\overline{RUL}(t-1) - 1]\} \end{aligned} \tag{A8}$$

$$\mathbf{P}(t) = \left[I - \mathbf{K}(t)\mathbf{H}\right]\cdot\mathbf{P}^{-}(t) = \begin{bmatrix} 1 - \dfrac{(\sigma_1^2+\sigma_2^2)}{\sigma_1^2+\sigma_2^2+\sigma_1^2\sigma_2^2/p^{-}} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} p^{-} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \dfrac{1}{w^1+w^2+w^K} & 0 \\ 0 & 0 \end{bmatrix} \tag{A9}$$

where $w^1 = 1/\sigma_1^2$, $w^2 = 1/\sigma_2^2$ and $w^K = 1/p^{-}$ can be seen as weights assigned to models 1, 2 and to the KF system dynamics model, respectively. Thus, the aggregated prediction of the RUL resulting from (A8) is a weighted average of the ensemble models prediction plus the additional prediction $\hat{RUL}^3(t) = \overline{RUL}(t-1) - 1$ generated during the predict phase of the KF method.

References

[1] D.B Jarrell, D. R Sisk, and L.J. Bond, "Prognostics and Condition-Based Maintenance: A New Approach to Precursive Metrics", Nuclear Technology, Vol. 145, 2004, pp. 275-286.

[2] A.K.S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition based maintenance", Mechanical Systems and Signal Processing, Vol. 20, 2006, pp. 1483-1510.

[3] E. Zio. "Prognostics and health management of industrial equipment", in Diagnostics and Prognostics of Engineering Systems: Methods and Techniques, 2012, S. Kadry, Eds. IGI-Global.

[4] Y. Z. Rosunally, S. Stoyanov, C. Bailey, P. Mason,S. Campbell, G. Monger, and I. Bell, "Fusion Approach for Prognostics

Framework of Heritage Structure", IEEE Transactions on reliability, Vol. 60 (1), 2011, pp. 3-13.

[5] J. Coble, and J. Hines, "Fusing Data Sources for Optimal Prognostic Parameter Selection", 6th topical meeting on Nuclear plant instrumentation, control, and human-machine interface technologies, Knoxville, Tennessee, April 5-9, 2009.

[6] R. Polikar, "Ensemble Based Systems in Decision Making", IEEE Circuits and Systems Magazine, Vol. 6 (3), 2006, pp. 21-45,.

[7] P. Baraldi , R. Razavi-Far , and E. Zio, "Classifier-Ensemble Incremental-Learning Procedure for Nuclear Transient Identification at Different Operational Conditions", Reliability Engineering & System Safety, vol. 96 (4), 2010, pp. 480-488.

[8] L. Breiman, "Bagging predictors", Machine Learning, Vol. 24, 1996, pp. 123-140.

[9] L. Peel, "Data Driven Prognostics using a Kalman Filter Ensemble of Neural Network Models", International Conference on Prognostics and Health Management, Denver, CO, October 6-9, 2008

[10] P. Baraldi, E. Zio, G. Gola, D. Roverso, and M. Hoffmann, "Robust nuclear signal reconstruction by a novel ensemble model aggregation procedure", International Journal of Nuclear Knowledge Management, Vol. 4 (1), 2010, pp. 32-41.

[11] T. Tinga, "Advanced maintenance modeling based on physical failure mechanism", 38th European Safety, Reliability & Data Association Seminar, Pecs, Hungry, May 4-5, 2010.

[12] D. Jiang and C. Liu, "Machine Condition Classification Using Deterioration Feature Extraction and Anomaly Determination", IEEE Transactions on reliability, Vol 60 (1), 2011, pp. 41-48.

[13] F. Di Maio, J. Hu, P. Tse, K. Tsui, E. Zio, and M. Pecht, "Ensemble-approaches for clustering health status of oil sand pumps", Expert Systems with Applications, Vol. 39 (5), 2012, pp. 4847–4859.

[14] V.A. Sotiris, P.W. Tse, and M.G. Pecht, "Anomaly Detection Through a Bayesian Support Vector Machine, IEEE Transactions on reliability", Vol 59 (2), 2010, pp. 277-286.

[15] S. Saha, B. Saha, A. Saxena, K. Goebel, "Distributed prognostic health management with gaussian process regression", IEEE Aerospace Conference, Big Sky, MT, March 6-13, 2010.

[16] E. Zio, F. Di Maio. "A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system", Reliability Engineering & System Safety, Vol. 95 (1), 2010, pp 49–57.

[17] C. Chatfield, "What is the 'best' method of forecasting?", Journal of Applied Statistics, Vol. 15, 1988, pp. 19–39.

[18] P. Baraldi, A. Cammi, F. Mangili, E. Zio, "An ensemble approach to sensor fault detection and signal reconstruction for nuclear system control", Annals of Nuclear Energy, Vol. 37 (6), 2010, pp. 778-790.

[19] E. Zio, P. Baraldi, G. Gola, "Feature-based classifier ensembles for diagnosing multiple faults in rotating machinery", Applied Soft Computing, Vol. 8 (4), 2008, pp. 1365–1380.

[20] P. Baraldi, G. Gola, E. Zio, D. Roverso, M. Hoffmann, "A randomized model ensemble approach for reconstructing signals from faulty sensors". Expert Systems With Application, Vol. 38 (8), 2011, pp. 9211-9224.

[21] R. Polikar, L. Udpa, S. S. Udpa, and V. Honavar, "Learn++: An incremental learning algorithm for supervised neural networks", IEEE transactions on systems, man, and cybernetics-part C: Applications and reviews, Vol. 31 (4), 2001, pp. 497-508.

[22] G.P. Zhang, "Time Series Forecasling using a hybrid ARIMA and neural network model", Neurocomputing, Vol. 50, 2003, pp. 159-175.

[23] L. K. Hansen, and P. Salamon, "Neural network ensembles," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12 (10), 1990, pp. 993–1001.

[24] T. K. Ho, "Random subspace method for constructing decision forests," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20 (8), 1998, pp. 832–844.

[25] R. Bryll, R. Gutierrez-Osuna, and F. Quek, "Attribute bagging: improving accuracy of classifiers ensembles by using random feature subsets", Pattern Recognition, Vol. 36, 2003, pp. 1291–1302.

[26] R. Polikar, "Bootstrap-Inspired Techniques in Computational Intelligence", IEEE Signal Processing magazine, Vol. 24 (4), 2007, pp. 59-72.

[27] P. Baraldi, R. Razavi-Far, E. Zio, "Bagged Ensemble of FCM Classifiers for Nuclear Transient Identification", Annals of Nuclear Energy; Vol. 38 (5), 2011, pp 1161-1171.

[28] E. Zio, "A study of the bootstrap method for estimating the accuracy of artificial neural networks in predicting nuclear transient processes", IEEE Transactions on Nuclear Science, Vol. 53 (3), 2006, pp. 1460-1478.

[29] P. Baraldi, E. Zio, G. Gola, D. Roverso, M. Hoffmann, "Two novel procedures for aggregating randomized model ensemble outcomes for robust signal reconstruction in nuclear power plants monitoring systems", Annals of Nuclear Energy, Vol. 38 (2-3), 2011, pp. 212-220.

[30] P. Baraldi, A. Cammi, F. Mangili, E. Zio, "Local Fusion of an Ensemble of Models for the Reconstruction of Faulty Signals", IEEE Transactions on Nuclear Science, Vol. 57 (2), 2010, pp.793-806.

[31] R. E. Kalman, "A new approach to linear filtering and prediction problems," Journal of Basic Engineering, Vol. 82, 1960, pp. 35–45.

[32] M. Saez, N. Tauveron, T. Chataing, G. Geffraye, L. Briottet, and N. Alborghetti, "Analysis of the turbine deblading in an HTGR with the CATHARE code", Nuclear Engineering and Design, Vol. 236, 2006, pp. 574–586.

[33] http://www.twi.co.uk/content/jk69.html

[34] J. Ferleger. Method and apparatus for determining turbine blade deformation, U.S. Patent 5238366 issued August 24, 1993.

[35] S. Hoyte, E. Gebhardt, E. Bauknight, S. Draper, J. Grant Method and systems for measuring blade deformation in turbines, U.S. Patent 7762153 issued July 27, 2010.

[36] R.W. Swindeman, and M.J. Swindeman, "A comparison of creep models for nickel base alloy for advanced energy systems", International Journal of Pressure Vessels and Piping, Vol. 85, 2008, pp. 72–79.

[37] R. Couturier, and C. Escaravage, "High temperature alloys for the HTGR Gas Turbine: Required properties and development needs", IAEA Technical Committee Meeting on Gas Turbine Power Conversion Systems for Modular HTGRs, Palo Alto, CA, November 14-16, 2000.

[38] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation", International Conference on Prognostics and Health Management, Denver, CO, October 6-9, 2008.

[39] G. Martínez-Muñoz, and A. Suárez, "Out-of-bag estimation of the optimal sample size in bagging", Pattern Recognition, Vol. 43, 2009, pp. 143-152.

Paper II

Model-Based and Data-Driven Prognostics under Different Available Information

Piero Baraldi, Francesco Cadini, Francesca Mangili, Enrico Zio

Accepted for publication on Probabilistic Engineering Mechanics, 2013

Model-Based and Data-Driven Prognostics under Different Available Information

Piero Baraldi[a], Francesco Cadini[a], Francesca Mangili[a], Enrico Zio[b,a]

[a]*Politecnico di Milano, Milano, Italy*

[b]*Chair on Systems Science and the Energetic challenge, European Foundation for New Energy-Electricite' de France, Ecole Centrale Paris and Supelec, Paris, France*

Abstract

In practical industrial applications, different prognostic approaches can be used depending on the information available for the model development. In this paper, we consider three different cases: 1) a physics-based model of the degradation process is available; 2) a set of degradation observations measured on components similar to the one of interest is available; 3) degradation observations are available only for the component of interest.

The objective of the present work is to develop for these three cases properly tailored prognostic approaches and to evaluate them in terms of the assumptions they require, the accuracy of the Remaining Useful Life (RUL) predictions they provide and their ability of providing measures of confidence in the model outcomes. The first case is effectively handled within a particle filtering (PF) scheme, whereas the second and third cases are addressed by bootstrapped ensembles of empirical models. From the methodological point of view, the main novelty of the work consists in the development of a bootstrap method able to assess the confidence in the RUL prediction in the third case characterized by the unavailability of any degradation observations until failure.

A case study is analyzed, concerning the prediction of the RUL of turbine blades affected by a developing creep.

Keywords: Prognostics, particle filtering, bootstrapped ensemble, turbine blade, creep.

1 Introduction

Prognostics aims at supplying reliable predictions about the Remaining Useful Life (RUL) of a component or system undergoing degradation. This is expected to improve planning of maintenance actions, increase safety and lower costs [1] [2].

Different forms of information and data may be available for the assessment of the evolution to failure of a degrading system, e.g., time-to-failure data of similar systems, direct or indirect measures of the degradation states reached during its evolution or during the evolution of a set of similar systems under similar operating

conditions, information on exogenous operational and environmental parameters, deterministic, empirical or semi-empirical models of the degradation process, etc. Depending on the situation, different prognostic methods may be applied [3] [4].

In this work, we consider three practical situations with decreasing information available for the prognostic task, and propose accurate and robust prognostic methods for each of them.

In general, prognostic methods can be classified in model-based and data-driven methods [5]. Model-based methods use an explicit mathematical model of the degradation process to predict the future evolution of the degradation state and, thus, the RUL of the system [6]. Examples of degradation models are the non-linear stochastic model of fatigue crack dynamics [7] [8] or the creep growth model based on the Norton law [9]. In practice, even when the model of the degradation process is known, the RUL estimate may be difficult to obtain, since the degradation state of the system may not be directly observable and/or the measurements may be affected by noise and disturbances. In these cases, model-based estimation methods aim at inferring the dynamic degradation state and provide a reliable quantification of the estimation uncertainty on the basis of the sequence of available noisy measurements. Many approaches rely on Bayesian methods [10] [11]: the exact Kalman filter has been largely used in case of linear state space models and independent, additive Gaussian noises, whereas analytical or numerical approximations of the Kalman filter (such as the Extended Kalman filter, the Gaussian-sum filters or the grid-based filters) have been applied in most realistic cases where the dynamics of degradation is non-linear and/or the associated noises are non-Gaussian [12]. Numerical approximations based on the Monte Carlo sampling technique have gained popularity for their flexibility and ease of design [13].

In the first case considered in this work, hereafter referred to as case 1, we have available a stochastic model of the degradation process and we know the value of the failure threshold, i.e., the maximum degradation beyond which the system loses its function. Also, a sequence of observations of the system degradation state are available and an observation equation describes the relation between the observations and the system degradation state. On this basis, a Monte Carlo-based filtering technique, called particle filtering (PF), is set-up to predict the distribution of the system RUL and online-update it when new observations are collected. The proposed approach improves the one previously proposed in [14] and [15] by taking into account the uncertainty on the parameters of the model of the degradation process and addressing the particle degeneration problem by means of the resampling algorithm [16].

On the other side, data-driven methods are used when an explicit model of the degradation process is not available, but sufficient historical data have been collected. These methods are based on statistical models that 'learn' trends from the data [17]. In this respect, artificial neural networks are often used [5] [18] [19]; other examples are Autoregressive Moving Average techniques [20], Relevance Vector Machines [18] [20] [21], fuzzy similarity-based methods [22]. Recently, ensemble approaches, based on the aggregation of multiple model outcomes, have been introduced due to the superior robustness and accuracy with respect to single models [23] and the possibility of estimating the uncertainty of the predictions [24].

In this work, data-driven methods have been developed to tackle two different situations of information available (hereafter referred to as cases 2 and 3). In case 2, a number of observations of degradation evolution and the failure times of a set of similar systems operating under similar conditions are available; in case 3 only observations of the degradation of the system for which we want to predict the RUL and the

value of the failure threshold are available. In both cases, the proposed prognostic approaches are based on the regression of the system degradation state by using an ensemble of bootstrapped models [24] which allows providing the uncertainty of the estimated RUL, caused by the uncertainty in the data, the variability of the system behavior and the empirical model error. From the methodological point of view, the main contribution of the present work consists in the approach developed to deal with case 3 which, differently from case 2, is characterized by the unavailability of degradation data until the component failure and, thus, of the input (degradation value) output (RUL) pairs used in case 2 for estimating the uncertainty in the RUL prediction.

The three cases are studied with reference to the creep growth process in the blades of a helium gas turbine of a Gas Turbine Modular Helium Reactor (GT-MHR) [25] [26].

The problem of selecting the most appropriate prognostic approach in the case in which a mix of the information considered in cases 1,2 and 3 is available, has also been addressed by comparing the performance of the three proposed approaches and by investigating their sensitivity to the accuracy of the model of the degradation process and to the amount and accuracy of the empirical data available.

The remainder of the paper is organized as follows: in Section 2 the objectives of the prognostic activity are presented; in Section 3, the sources of information for prognostics are discussed; in Section 4, the three cases considered are described; in Section 5, the prognostic methods developed to tackle the three cases are presented; in Section 6, the problem of blade creeping in high temperature turbines is illustrated and the prognostic results obtained in the three different cases considered are discussed; in Section 7 the problem of selecting the correct approach for specific situations of information available is discussed; finally, in Section 8 some conclusions are drawn and potential for future work suggested.

2 Information and data for prognostics

Let us discretize, for ease of exposition, the continuum time variable t into a sequence of time instants t_i, i=1,2,… assumed to be equally spaced.

The aim of prognostics is to estimate the Remaining Useful Life RUL_i of a degrading system, i.e., the time left from the current time t_i before the system degradation crosses the failure threshold. Since degradation evolution is intrinsically random, the system RUL_i is a random variable and, thus, the objective of applying a prognostic method to a system whose current degradation state is d_i is to estimate the probability distribution $p_{RUL_i}(rul_i \mid d_i)$ of RUL_i at time t_i.

Table I summarizes the main sources of information upon which prognostics can be based [4]:

- A physical model of the degradation mechanism (source A, Table I), e.g., described by a first-order Markov process:

$$d_j = g(d_{j-1}, \boldsymbol{\gamma}_{j-1}) \; ; \; d_0 \sim p_{D_0}(d_0) \tag{1}$$

where d_j is the degradation state at time t_j, $p_{D_0}(d_0)$ is the initial distribution of the degradation at time t_0, g is a possibly non-linear function describing the value of a one-time-step degradation increment and $\boldsymbol{\gamma}_j$, $j = 1,2,...$ is a sequence of mutually independent vectors of state noises. The model g can contain parameters referring to system inherent characteristics (material, physical, chemical, geometrical, etc.), which may vary from one individual system to another of the same type: this variability is described by probability distribution functions. The model can also describe the dependence of the degradation process from external parameters (environmental, operational, etc.), which may vary during the system life. Although these parameters are not directly related to the system degradation state, they may influence its evolution. Some of these parameters may be directly observable and measured by sensors, others may not; for some, there may be a priori knowledge of their behavior in time or statistical knowledge of their distribution.

Table I: main sources of information for prognostics

Source	Description	Mathematical representation
A	Dynamic model of the degradation process	eq.(1)
B	Sequence of observations related to the degradation of the system collected at $t_j = 1,2,...,i$	$\mathbf{z}_{1:i}$
C	Historical sequences of observations related to the degradation of a set of S failed systems collected at N_s time instants t_j ; s=1,…,S; j=1,…,N_s	$\mathbf{z}^s_{1:N_s}$, s=1,…,S
D	Value of the failure threshold	d_{th}
E	Measurement equation	eq.(2)
F	Durations of lives of the set of S failed systems s.	L^s, s=1,…,S

- A set of observations $\mathbf{z}_{1:i}$, collected at different time instants $t_{1:i}$, during the life of the system whose RUL we want to predict (source of information B, Table I) or of a population of identical or similar systems (source of information C, Table I). Among the observable process parameters in $\mathbf{z}$ there can be a direct measure of the degradation state of the system (e.g., depth of a crack fracture, elongation of a creeping component, etc.) or they can be only indirectly related to it (e.g., the time of travel or the intensity of ultrasonic waves for non-destructive inspections).
- The value of the failure threshold d_{th} (source D, Table I).
- The observation equation (source E, Table I), i.e., the physical model describing the relation between the observation $\mathbf{z}_j$ containing the values of the observable process parameters measured by sensors at some time instant t_j and the actual degradation state d_j of the system:

$$\mathbf{z}_j = h(d_j, \mathbf{n}_j) \tag{2}$$

where h is a known function, in general non-linear, and $\mathbf{n}_j$ is a vector of measurement noises.

- The life durations $\{L_s\}_{s=1}^{S}$ of a number S of similar systems which have failed (source F, Table I); notice that, the actual value of the RUL of the *s*-th failed system can be computed at any time $t_i < L_s$ as

$$rul_i^s = L_s - t_i \tag{3}$$

3 Three prognostic cases with different sources of information

Three cases are considered in this work, in which a set of measurements $\mathbf{z}_{1:i}$ collected during the life of the system whose RUL we want to predict (source of information B) is available in combination with other different sources of information (Table II).

Table II: information available in each of the three prognostic cases considered

	Source of information	Case 1	Case 2	Case 3
A	Dynamic model	X		
B	Current observations' sequence	X	X	X
C	Historical observations' sequences		X	
D	Failure threshold	X		X
E	Measurement equation	X		X
F	Life duration data		X	

In case 1, the physical model of the evolution of the degradation state is known, as well as the distribution and evolution in time of all its characteristic and external parameters (source A). Other sources of information available are the value of the failure threshold d_{th} (source D) and the observation equation (source E) linking the observations with the degradation state. This situation is typical for well known degradation mechanisms, such as the crack or creep growth processes, which have been widely studied in laboratory.

In case 2, a set of observations $\{\mathbf{z}_{1:N_s}^s\}_{s=1}^{S}$, of S similar systems (source C) and the duration of their lives (source F) are available. This situation is typical for short-life systems, for which many trajectories to failure can be observed.

Finally, in case 3, the information available is the observation equation (source E) and the value of the failure threshold d_{th} (source D). This situation can occur in case of very reliable systems, e.g., those used in the nuclear industry, which have a very long life duration and are usually renewed before failure happens.

4 Modeling approaches

This Section illustrates the three modeling approaches undertaken to cope with the three prognostic cases outlined in Section 3 (Table II).

4.1 Case 1: Particle Filtering

In case 1, at time t_i, the current degradation state d_i is not directly known, but the stochastic system dynamic model of eq. (1), the observation equation of eq. (2), the sequence of i observations $\mathbf{z}_{1:i}$ related to the system degradation state and the value of the failure threshold d_{th} are available. Thus, instead of estimating $p_{RUL_i}(rul_i \mid d_i)$ we are forced to restrict our objective to estimating the probability density function (pdf) $p_{RUL_i}(rul_i \mid \mathbf{z}_{1:i}, D(t_i) < d_{th})$, conditioned on the observations $\mathbf{z}_{1:i}$ and on the fact that at time t_i the equipment has not yet failed, i.e., $D(t_i) \le d_{th}$,

In this setting, defining $D(t_j)$ the random variable which describes the degradation state at time t_j, it is desired to infer the unknown pdf $p_{D(t_j)}(d_j \mid \mathbf{z}_{1:i}, D(t_i) < d_{th})$ of the degradation d_j at the future times $t_j > t_i$ on the basis of all the previously estimated distribution of the state values $p_{D(t_{0:j-1})}(d_{0:j-1} \mid \mathbf{z}_{1:i})$ and of all the observations $\mathbf{z}_{1:i}$. The RUL cumulative probability distribution $F_{RUL_i}(rul_i \mid \mathbf{z}_{1:i}, D(t_i) \le d_{th})$ is then computed from $p_{D(t_j)}(d_j \mid \mathbf{z}_{1:i}, D(t_i) < d_{th})$ as the probability that the failure threshold d_{th} is exceeded before time $t_i + rul_i$:

$$\begin{aligned} F_{RUL_i}(rul_i \mid \mathbf{z}_{1:i}, D(t_i) < d_{th}) &= \text{Prob}(RUL_i < rul_i \mid \mathbf{z}_{1:i}, D(t_i) \le d_{th}) \\ &= \text{Prob}(L - t_i < rul_i \mid \mathbf{z}_{1:i}, D(t_i) \le d_{th}) \\ &= \text{Prob}(L < t_i + rul_i \mid \mathbf{z}_{1:i}, D(t_i) \le d_{th}) \\ &= \int_{d_{th}}^{+\infty} p_{D(t_j = t_i + rul_i)}(d_j \mid \mathbf{z}_{1:i}, D(t_i) \le d_{th}) \, \mathrm{d}\, d_j \end{aligned} \tag{4}$$

In the prognostic problem, we resort to PF for estimating $p_{D(t_j)}(d_j \mid \mathbf{z}_{1:i}, D(t_i) \le d_{th})$ and solving the integral in eq.(4). In particular, the Sampling Importance Resampling (SIR) version of PF is here adopted, whose analytical details are provided in Appendix A.

The SIR PF method is based on sampling a large number *K* of trajectories $\{d_{0:i}^k\}_{k=1}^K$ (called particles), by recursively sampling the state d_j^k from the transition pdf $p_{D(t_{j+1})}(d_{j+1}^k \mid d_j^k)$ which can derived from the physical model in eq. (1). Then, the posterior pdf $p_{D(t_i)}(d_i \mid \mathbf{z}_{1:i})$ can be approximated as [10]:

$$p_{D(t_i)}(d_i \mid \mathbf{z}_{0:i}) \approx \sum_{k=1}^{K} w_i^k \delta(d_i - d_i^k) \tag{5}$$

where w_i^k is the importance weight associated to the sampled state sequence $d_{0:i}^k$, $k = 1,2,...,K$. The weight w_i^k can, then, be recursively computed as:

$$w_i^k = \frac{p_{Z(t_i)}(\mathbf{z}_i \mid d_i^k) w_{i-1}^k}{\sum_{k=1}^{K} p_{Z(t_i)}(\mathbf{z}_i \mid d_i^k) w_{i-1}^k} \tag{6}$$

where $p_{Z(t_i)}(\mathbf{z}_i \mid d_i^k)$ is the *likelihood* of the observation $\mathbf{z}_i$.

To predict the pdf of the degradation states at future times t_j, j=i+1,i+2,…, the prediction stage is iterated for each particle, by recursively appending the sampled trajectory $d_{0:i}^k$ with a new degradation states d_{i+1}^k, d_{i+2}^k,…, d_j^k, while keeping the weights fixed to their values w_i^k calculated at the time t_i of the last observation. Indeed, the pdf $p_{D(t_j)}(d_j \mid \mathbf{z}_{1:i})$ can be approximated as:

$$p_{D(t_j)}(d_j \mid \mathbf{z}_{0:i}) \approx \sum_{k=1}^{K} w_i^k \delta(d_j - d_j^k) \tag{7}$$

Finally, the pdf $p_{D(t_j)}(d_j \mid \mathbf{z}_{0:i}, d_i^k \le d_{th})$ conditioned on the fact that $D(t_i) \le d_{th}$ can still be approximated resorting to eq. (7) but taking into account only those particles whose degradation at time t_i is below the threshold, i.e., $d_i^k \le d_{th}$. Operatively, this entails setting to zero the weights of these particles and normalizing the remaining ones, thus getting a new set of weights $\tilde{w}_i^k$.

Notice that the approximated pdf thus obtained is a discrete probability mass function where only the degradation values $d_{t_i+rul_i}^k$ assumed by the particles at time $t_i + rul_i$ have a finite probability equal to their weights $\tilde{w}_i^k$; then, the integral in eq. (4) corresponds to the summation of the weights of the particles whose degradation at time $t_i + rul_i$ exceeds the threshold d_{th}:

$$\begin{aligned} F_{RUL_i}(rul_i \mid \mathbf{z}_{1:i}, d_i < d_{th}) &= \int_{d_{th}}^{+\infty} \sum_{k=1}^{K} \tilde{w}_i^k \delta(d_j - d_{t_i+rul_i}^k)\, \mathrm{d}\, d_j \\ &= \sum_{k=1}^{K} \tilde{w}_i^k H(d_{t_i+rul_i}^k - d_{th}) \end{aligned} \tag{8}$$

where $H(d_{t_i+rul_i}^k - d_{th})$ is the Heaviside step function.

The application of the particle filtering procedure to the estimation of $p_{RUL_t}(rul_i \mid \mathbf{z}_{1:i})$ is detailed in the pseudo-code of Figure 1.

Unfortunately, the procedure illustrated suffers from the so called degeneracy phenomenon: after few samplings, the weight variance increases and most of the K weights in eq. (8) become negligible so that the corresponding trajectories do not contribute to the estimate of the pdf of interest [10][16]. As a result, the approximation of the target distribution $p_{RUL_t}(rul_i \mid \mathbf{z}_{1:i}, D(t_i) < d_{th})$ becomes very poor and significant computational resources are spent trying to update particles with minimum relevance.

A possible solution to this problem is offered by the bootstrap resampling algorithm, which is detailed in the pseudo-code of Figure 2 [11]. When degeneracy occurs, e.g. after few iterations of the weight updating procedure, K samples are drawn with replacement from the swarm of K particles; the k-th particle is sampled with a probability proportional to its weight value $\tilde{w}_i^k$ and the sequence of degradation state $d_{1:i}^k$ until time

t_i is retained for the resampled particle *k'* and recursively augmented with new degradation states $d_{j+1}^{k'}$. The *K* resampled particles are then assigned the same weight $1/K$ is assigned to all of them. Then, the filtering procedure continues with the original trajectories $d_{1:j}^{k}$ and the associated weights $\tilde{w}_i^k$ replaced by new trajectories $d_{1:j}^{k'}$ with weights $\tilde{w}_i^{k'} = 1/K$.

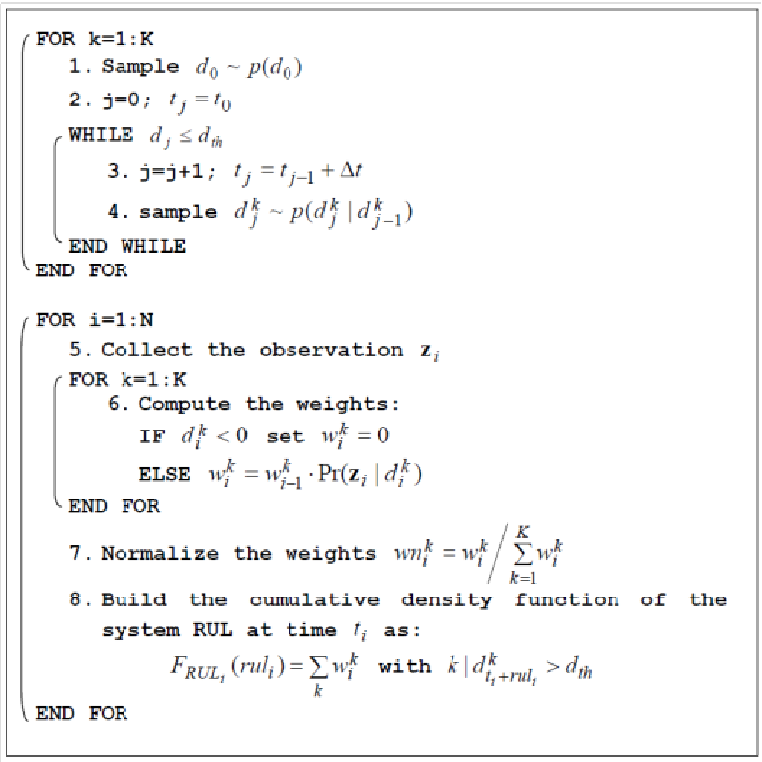

Figure 1: Particle filtering operative procedure for estimation of the RUL cumulative distribution.

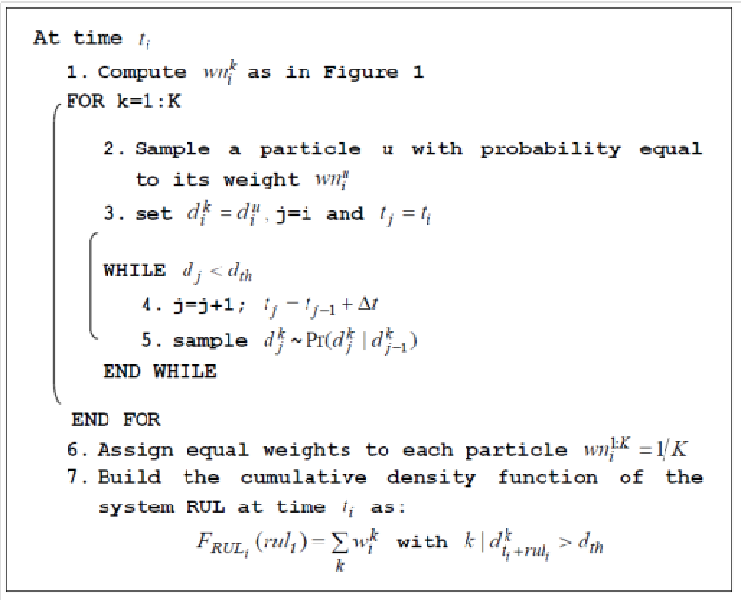

Figure 2: Procedure for performing resampling at time t_i.

4.2 Case 2: Data-driven prognostics based on an ensemble of bootstrapped models trained on degradation and life duration data

In case 2, the information available at time t_i is a set of degradation observations $\{\mathbf{z}^s_{1:N_s}\}^S_{s=1}$, taken during the trajectory to failure of S similar systems, the duration of their life L^s, and the observations $\mathbf{z}_{1:i}$ related to the degradation state of the system of interest.

In this context of information available, we are not able to estimate the probability distribution, $p_{RUL_i}(rul_i \mid \mathbf{z}_{1:i})$, of RUL_i for a system that at time t_i is in the degradation state d_i. In practice, our objective is limited to obtain:

1. an estimate $\hat{rul}_i$ of the expected value μ_{RUL_i} of RUL_i;
2. an estimate $\hat{\sigma}^2_{\hat{rul}_i}$ of the variance of the prediction error $\sigma^2_{\hat{rul}_i} = E[(\hat{rul}_i - rul_i)^2]$; this quantity can be interpreted as a measure of the accuracy with which the predicted value $\hat{rul}_i$ is expected to describe the actual rul_i.

The idea is to develop an empirical model:

$$f(\mathbf{z}_i) = \hat{rul}_i \tag{9}$$

of the relationship between the degradation observation available at time t_i, $\mathbf{z}_i$, and RUL_i. This empirical model receives in input the current observation $\mathbf{z}_i$ and produces as output the RUL prediction, $\hat{rul}_i$ and an estimate $\sigma^2_{\hat{rul}_i}$ of the variance of the predicted error.

In order to develop the model, a dataset:

$$\mathbf{D}_{i/o} = \{\{\mathbf{z}^s_i; rul^s_i\}^{N_s}_{i=1}\}^S_{s=1} \tag{10}$$

is extracted from the set of observations $\{\mathbf{z}^s_{1:N_s}\}^S_{s=1}$, by associating to the observation $\mathbf{z}_i$ taken at time t_i during the s-th trajectory to failure, the corresponding RUL:

$$rul^s_i = L^s - t_i \tag{11}$$

The dataset $\mathbf{D}_{i/o}$ can be used to train an empirical model built using one among the many data-driven modeling methods existing today (e.g., polynomial regression, non-parametric regression, neural networks, etc.).

In general, the regression problem can be framed as follows: given a set of data pairs $\{\mathbf{z}_i, y_i\}^N_{i=1}$, generated from:

$$y_i = h(\mathbf{z}_i) + \upsilon_i \tag{12}$$

where y_i is the target value, $h(\mathbf{z}_i)$ the true input/output relation and υ_i a process noise with zero mean and standard deviation $\sigma^2_{\upsilon_i}(\mathbf{z}_i)$. Our aim is to train a model $f(\mathbf{z}_i)$ using the data $\{\mathbf{z}_i, y_i\}_{i=1}^N$, which approximate $h(\mathbf{z}_i)$. According to [24] the developed empirical model can be interpreted as an estimate of the mean distribution of the target values given an input vector $\mathbf{z}_i$. In this context, Heskes [24] proposed a method, for providing a measure of confidence in the prediction $f(\mathbf{z}_i)$ (see appendix B for a detailed description of the method). In our application, assuming RUL_i a random variable with mean $\mu_{RUL_i}(\mathbf{z}_i)$ and variance $\sigma^2_{RUL_i}(\mathbf{z}_i)$, we can write the relationship between pairs $\{\mathbf{z}_i^s; rul_i^s\}$ as:

$$rul_i = \mu_{RUL_i}(\mathbf{z}_i) + \upsilon_i \tag{13}$$

where $\mu_{RUL_i}(\mathbf{z}_i)$ is the expected value of the RUL value given the observation $\mathbf{z}_i$ and υ_i is a random variable with zero mean and standard deviation $\sigma^2_{\upsilon_i}(\mathbf{z}_i) = \sigma^2_{RUL_i}(\mathbf{z}_i)$. Comparing eq. (13) with eq. (12) we have $y_i = rul_i$ and $h(\mathbf{z}_i) = \mu_{RUL_i}(\mathbf{z}_i)$. Accordingly, $f(\mathbf{z}_i)$ is interpreted as an estimator of $\mu_{RUL_i}(\mathbf{z}_i)$ and, as described in [24], the uncertainty in the prediction $f(\mathbf{z}_i)$ is quantified by the prediction error variance $\sigma^2_{\hat{rul}_i}(\mathbf{z}_i)$ which can be decomposed into two terms:

$$\begin{aligned}\sigma^2_{\hat{rul}_i}(\mathbf{z}_i) &= E[(\hat{rul}_i - rul_i)^2] = \\ &= E\left\{[f(\mathbf{z}_i) - \mu_{RUL_i}(\mathbf{z}_i)]^2\right\} + E\left\{[\mu_{RUL_i}(\mathbf{z}_i) - rul_i]^2\right\} \\ &= \sigma^2_m(\mathbf{z}_i) + \sigma^2_{RUL_i}(\mathbf{z}_i)\end{aligned} \tag{14}$$

where the term $\sigma^2_m(\mathbf{z}_i)$ is the model error variance describing the regression error made by the model $f(\mathbf{z}_i)$ in estimating the true RUL mean value $\mu_{RUL_i}(\mathbf{z}_i)$, and the term $\sigma^2_{RUL_i}(\mathbf{z}_i)$ is the RUL variance caused by the uncertainty on the future degradation of the system and describing the accuracy of $\mu_{RUL_i}(\mathbf{z}_i)$ in predicting the target rul_i.

Notice that the application of the method described in Appendix B to the estimate of the model error variance $\sigma^2_m(\mathbf{z}_i)$, and the RUL variance $\sigma^2_{RUL_i}(\mathbf{z}_i)$ requires the partition of the input/output dataset $\mathbf{D}_{i/o}$ into a training and a validation datasets, $\mathbf{D}^{trn}_{i/o}$ and $\mathbf{D}^{val}_{i/o}$: the training dataset is used to train the regression model $f(\mathbf{z}_i)$, whereas the validation dataset is used to test $f(\mathbf{z}_i)$ and collect examples of its prediction error. Since the two datasets $\mathbf{D}^{trn}_{i/o}$ and $\mathbf{D}^{val}_{i/o}$ have to be independent in order to avoid underestimating the variance $\sigma^2_{RUL_i}(\mathbf{z}_i)$, we have considered a validation dataset $\mathbf{D}^{val}_{i/o} = \{\{\mathbf{z}_i^s; rul_i^s\}_{i=1}^{N_s}\}_{s=1}^{S_{val}}$ made by input/output pairs taken from trajectories different from those used to build the training dataset $\mathbf{D}^{trn}_{i/o}$.

In practice, the overall approach to estimate $\hat{rul}_i$ and $\hat{\sigma}^2_{\hat{rul}_i}$ requires to:

- train an ensemble of models $\{f^b(\mathbf{z}_i \mid \mathbf{D}^b_{i/o})\}_{b=1}^B$ using bootstrapped replicates $\mathbf{D}^b_{i/o}$ of the training dataset $\mathbf{D}^{trn}_{i/o}$;

- test the bootstrapped ensemble on the validation dataset $\mathbf{D}_{i/o}^{val}$ to compute the prediction residuals $r^2(\mathbf{z}_i^s)$ as in eq. (B4) of Appendix B (with $y_i = rul_i^s$);
- use the set of residuals input/output pairs $\{\{\mathbf{z}_i^s; r^2(\mathbf{z}_i^s)\}_{i=1}^{N_s}\}_{s=1}^{S_{val}}$ to train the model $\chi(\mathbf{z}_i) = \hat{\sigma}_{RUL_i}^2(\mathbf{z}_i)$ describing the dependence of $\hat{\sigma}_{RUL_i}^2(\mathbf{z}_i)$ from $\mathbf{z}_i$ [24][27];
- when a new observation $\mathbf{z}_i$ about the degradation state of a functioning system is collected compute the output $\hat{rul}_i^b$ of each models $f^b(\mathbf{z}_i \mid \mathbf{D}_{i/o}^b)$ of the ensemble;
- compute the prediction $\hat{rul}_i$, i.e., the estimate of the RUL expected value μ_{RUL_i}:

$$\hat{rul}_i = \frac{1}{B}\sum_{b=1}^{B} f^b(\mathbf{z}_i \mid \mathbf{D}_{I/O}^b) \tag{15}$$

- compute the estimate $\hat{\sigma}_m^2(\mathbf{z}_i)$ of the model error variance:

$$\hat{\sigma}_m^2(\mathbf{z}_i) = \frac{1}{B}\sum_{b=1}^{B}\left[f^b(\mathbf{z}_i \mid \mathbf{D}_{I/O}^b) - \hat{rul}_i\right]^2 \tag{16}$$

- apply the model $\chi(\mathbf{z}_i)$ to the input $\mathbf{z}_i$ to obtain the RUL variance estimate $\hat{\sigma}_{RUL_i}^2(\mathbf{z}_i)$;
- sum up the two variance components to obtain the prediction error variance $\sigma_{\hat{rul}_i}^2$:

$$\sigma_{\hat{rul}_i}^2(\mathbf{z}_i) = \hat{\sigma}_m^2(\mathbf{z}_i) + \hat{\sigma}_{RUL_i}^2(\mathbf{z}_i) \tag{17}$$

4.3 Case 3: Data-driven prognostics based on an ensemble of bootstrapped models trained on degradation data only

This case is characterized by the availability of the observations $\mathbf{z}_{1:i}$ related to the degradation state of the system of interest at j different measurement time instants up to the current time t_i, the relative observation equation and the value of the failure threshold d_{th}. Given the observation equation, an estimate of the degradation state d_j can always be derived from the observation $\mathbf{z}_j$. For simplicity of illustration, we consider here only a situation where the observation $\mathbf{z}_j$ is a direct measure of the degradation state d_j, eventually affected by a zero-mean noise, and thus no further estimate of d_j is needed.

The modeling approach proposed in the previous Section, based on the availability of input/output pairs formed by the observations $\mathbf{z}_i$ and the corresponding RUL value rul_i, cannot be directly applied to this case. For this reason, an approach which uses the time series $\mathbf{z}_{1:i}$ of the past observations to build a model of the time evolution of the degradation process is proposed. Notice that the approach differs from that used in case 1 since the physical stochastic model describing the true dynamics of the degradation process (eq. (1)) is unknown and should be replaced by an empirical deterministic model derived from the few available data. Coherently, the estimate of the prediction error variance $\sigma_{\hat{rul}_i}^2(\mathbf{z}_i)$ should account also for the error made when approximating the true degradation process with the empirical model.

A generic model of the evolution of the degradation state of the system, achieved by fitting the most suited degradation model, e.g., linear and non-linear regression models, general degradation path models, etc. [28] to the sequence of data $\mathbf{z}_{1:i}$, can be written as:

$$\hat{d}_j = \eta(t_j) \tag{18}$$

where $\hat{d}_j$ is the degradation value at time t_j predicted by the model.

The prediction $\hat{rul}_i$ of the system RUL at time t_i is then obtained from the relation

$$\eta(t_i + \hat{rul}_i) = d_{th} \tag{19}$$

An estimate of the prediction error variance $\sigma^2_{\hat{rul}_i}(\mathbf{z}_i)$ cannot be obtained by means of the method proposed in case 2, since there are no available pairs $\{\mathbf{z}_j, rul_j\}_{j=1}^{N'}$ for which rul_j is known, and thus the prediction residuals $\{r^2(\mathbf{z}_j)\}_{j=1}^{N'}$ cannot be computed in correspondence of any of the observation $\mathbf{z}_{1:i}$. To overcome this problem, we consider a model

$$\Delta\hat{t}_{j,j'} = \tilde{\eta}(\mathbf{d}_{j,j'}) \tag{20}$$

which receives in input a vector of two degradation states $\mathbf{d}_{j,j'} = [d_j \;\; d_{j'}]$ and returns in output the estimate $\Delta\hat{t}_{j,j'}$ of the time interval needed to reach the degradation state d_j starting from $d_{j'}$. In general, model $\tilde{\eta}(\mathbf{d}_{j,j'})$ can be derived from model $\eta(t_j)$ according to:

$$\Delta\hat{t}_{j,j'} = \tilde{\eta}(\mathbf{d}_{j,j'}) = \eta^{-1}(d_{j'}) - \eta^{-1}(d_j) \tag{21}$$

The prediction $\hat{rul}_i$ is obtained from this model by setting $d_j = \mathbf{z}_i$ and $d_{j'} = d_{th}$; in this view, the RUL prediction at time t_i corresponds to the estimate of the time interval $\Delta t_{i,th}$ needed to increase the degradation state from d_i to the failure threshold d_{th}. Model $\tilde{\eta}(\mathbf{d}_{j,j'})$ is assumed to be an unbiased estimator of the mean value $\mu_{\Delta t_{j,j'}}(\mathbf{d}_{j,j'})$ of the random variable $\Delta t_{j,j'}$; the variance $\sigma^2_m(\mathbf{d}_{j,j'})$ of the difference between the estimate $\tilde{\eta}(\mathbf{d}_{j,j'})$ and $\mu_{\Delta t_{j,j'}}(\mathbf{d}_{j,j'})$ represents the uncertainty associated to the model $\tilde{\eta}$; the variance $\sigma^2_{\Delta t_{j,j'}}$ of the difference between $\mu_{\Delta t_{j,j'}}(\mathbf{d}_{j,j'})$ and the actual target value $\Delta t_{j,j'}$ represents the uncertainty in the evolution of the degradation process from d_j to $d_{j'}$. Being $\sigma^2_{\Delta t}$ and $\mu_{\Delta t}$ functions of the input $\mathbf{d}_{j,j'}$, eq. (12) becomes:

$$\Delta t_{j,j'} = \mu_{\Delta t_{j,j'}}(\mathbf{d}_{j,j'}) + \upsilon_{j,j'} \tag{22}$$

where $\upsilon_{j,j'}$ represent a process noise with zero mean and standard deviation $\sigma^2_{\Delta t_{j,j'}}(\mathbf{d}_{j,j'})$.

The bootstrap method used for case 2 and described in Appendix B, can now be applied considering, instead of the quantities y_i, $h(\mathbf{z}_i)$ and $f(\mathbf{z}_i)$, the quantities $\Delta t_{j,j'}$, $\mu_{\Delta t_{j,j'}}(\mathbf{d}_{j,j'})$ and $\tilde{\eta}(\mathbf{d}_{j,j'})$, respectively.

As underlined in Section 4.2, to avoid underestimating the prediction error, the validation datasets should not contain measurements belonging to degradation trajectories used for training. Since only a single trajectory is now available, the solution proposed is to partition the dataset $\mathbf{z}_{1:i}$ into two sequences of consecutive measurements, $\mathbf{D}^{trn} = \{\mathbf{z}_{1:N_{trn}}\}$ and $\mathbf{D}^{val} = \{\mathbf{z}_{N_{trn}+1:i}\}$. An ensemble of B models $\{\tilde{\eta}^b(\mathbf{d}_{j,j'})\}_{b=1}^{B}$ is then generated by training each model on a bootstrapped replicate $\mathbf{D}^b$ of $\mathbf{D}^{trn}$ and validated on validation dataset $\mathbf{D}^{val}_{i/o}$ derived from $\mathbf{D}^{val}$:

$$\mathbf{D}^{val}_{i/o} = \{\{\mathbf{d}_{j,j'} = [\mathbf{z}_j \quad \mathbf{z}_{j'}], \Delta t_{j,j'} = [t_j \quad t_{j'}]\}_{j=N_{trn}+1}^{i-1}\}_{j'=j+1}^{i} \tag{23}$$

The prediction residuals $r^2(\mathbf{d}_{j,j'})$ are then computed as in eq. (B4) of Appendix B , where y_i is replaced by $\Delta t_{j,j'}$ and $f^b(\mathbf{z}_i)$ is replaced by $\tilde{\eta}^b(\mathbf{d}_{j,j'})$ and used to build the empirical model $\hat{\sigma}^2_{\Delta t_{j,j'}} = \chi(\mathbf{d}_{j,j'})$ estimating the variance $\sigma^2_{\Delta t_{j,j'}}(\mathbf{d}_{j,j'})$ of $\Delta t_{j,j'}$. The RUL variance $\hat{\sigma}^2_{RUL_i}(\mathbf{z}_i)$ is then obtained from model χ fed with the input $\mathbf{d}_{i,th} = [\mathbf{z}_i \quad d_{th}]$.

Thus, when a new observation $\mathbf{z}_i$ is collected at time t_i, the outcomes of the ensemble models are use to generate the RUL prediction:

$$\hat{rul}_i = \frac{1}{B}\sum_{b=1}^{B}\tilde{\eta}^b(\mathbf{d}_{i,th}) \tag{24}$$

and the prediction error variance estimate

$$\sigma^2_{\hat{rul}_i} = \hat{\sigma}^2_m(\mathbf{z}_i) + \sigma^2_{RUL_i}(\mathbf{z}_i) = \frac{1}{B}\sum_{b=1}^{B}[\tilde{\eta}^b(\mathbf{d}_{i,th}) - \hat{rul}_i]^2 + \chi(\mathbf{d}_{i,th}) \tag{25}$$

Notice that the training data used to build model χ cover a range of values for the input $\mathbf{d}_{j,j'}$ in general different from that of the input $\mathbf{d}_{i,th}$ to which the model is applied to obtain the estimate $\hat{\sigma}^2_{RUL_i}$. This represents a limit to the quality of the estimate $\hat{\sigma}^2_{RUL_i}$, since the performances of empirical models tend to degrade when they are applied to input patterns belonging to regions far away from those containing the patterns used to train the model.

5 Numerical application

In this Section, the three different cases presented in Section 3 are considered with reference to the prognostics of a turbine blade in which creep damage is developing [26]. Creep is an irreversible deformation process affecting materials exposed to a load below the elastic limit for a protracted length of time and at high temperature. Notice that a turbine undergoing this degradation process can experience the loss of its blades, one of the most feared failure modes of turbomachinery since it is accompanied by abrupt changes in the power conversion system and in the reactor flow conditions [29]. Figure 7 shows an example of high-pressure turbine deblading occurred in a German power plant [29].

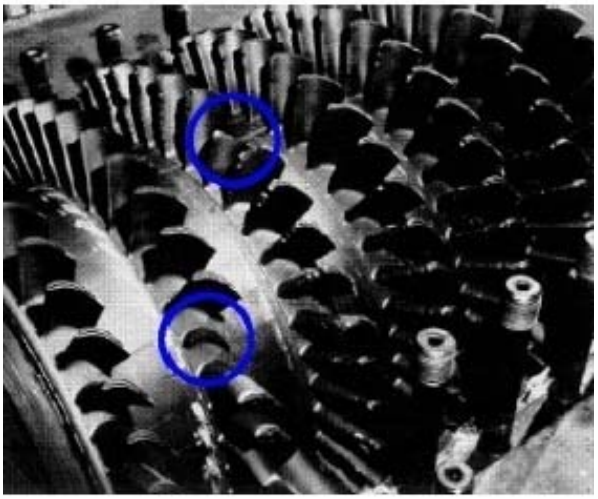

Figure 3: Deblading in a high pressure turbine [30].

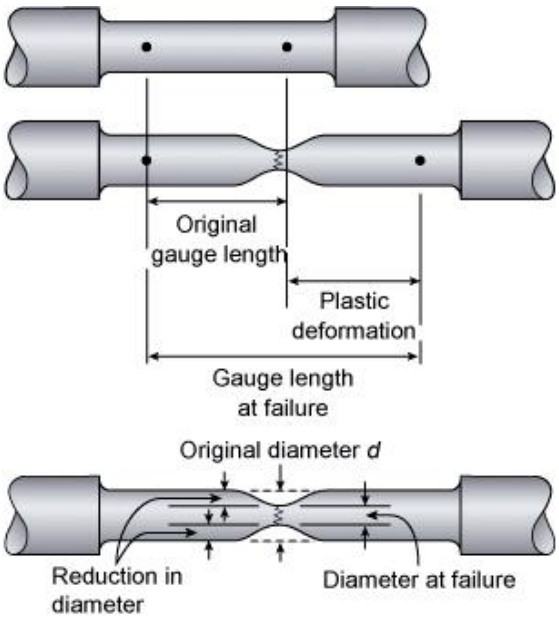

Figure 4: Schematics of a specimen before and after a creep test (http://www.twi.co.uk/content/jk69.html).

As shown in Figure 4, the uniaxial creep deformation consists in an augmentation of the original length and a reduction of the diameter. In this work, the adimensional quantity ε, defined as the percentage of elongation of the turbine blade in the longitudinal direction with respect to its original length, is considered as measure of the creep strain.

5.1 Information available for prognostics

The main sources of degradation-related information for the creep growth process, listed in Table IV, are further detailed in this Section.

Information A: creep growth model

Creeping in turbine blades is a stochastic degradation process which can be modeled through the Norton Law, assuming that the dependence from the temperature follows the Arrhenius law [9]:

$$\frac{d\varepsilon}{dt} = A \cdot \exp\left(-\frac{Q}{RT}\right) \cdot \varphi^n \tag{25}$$

where $d\varepsilon/dt$ is the creep strain rate, Q is the activation energy, A and n are material characteristics varying from one blade to another, R is the ideal gas constant, T is the blade operating temperature and φ is the applied stress. For simplicity, the blade temperature is supposed equal to the gas temperature and the stress φ is derived from the rotational speed ω of the turbine:

$$\varphi = \rho \frac{r_{tip}^{\ 2} - r_{hub}^{\ 2}}{2} \omega^2 \tag{26}$$

where ρ is the blade density and r_{hub} and r_{tip} are the hub and tip radiuses, respectively. The rotational speed ω and the gas temperature T are external parameters depending on the power setting of the gas turbine.

Table III: type of distribution, mean value and standard deviation used for the creep growth model parameters

Variable	Symbol	Distribution	Units	Parameters of the distribution
Activation energy	Q	Deterministic	kJ/mol	Q=290
Norton Law parameters	A	Normal	$(N/m^2)^{-n}/h$	$\mu_A=3\cdot10^{-4}$; σ_A=5%
	N	Normal	-	μ_n=6; σ_n=0.2%
Operating temperature	T_i	Normal	K	μ_T=1100; σ_T=1%
Rotational speed	ω_i	Normal	rpm	μ_ω=3000; σ_ω=1%
Density	P	Deterministic	Kg/m^3	ρ =8000
Hub radius	r_{hub}	Deterministic	m	r_{hub} =0.7
Tip radius	r_{tip}	Deterministic	m	r_{tip} =0.87
Stress fluctuations	$\delta\varphi$	Gamma	MPa	θ=2; k=10

For Δt sufficiently small compared to the time horizon of the analysis (here $\Delta t = 5$ days, with respect to the time horizon of several thousands), the state space model in eq. (26) can be discretized to give:

$$\varepsilon_{j+1} = \varepsilon_j + A \cdot \exp\left(-\frac{Q}{RT_j}\right) \cdot \varphi_j^{\ n} \cdot \Delta t_{j,j+1}, \qquad \varepsilon_0 = 0 \tag{27}$$

The characteristic parameters A and n vary from one blade to another, whereas the external parameters, i.e., the rotational speed ω and the gas temperature T vary continuously in time; all these parameters are assumed to have normal distributions. Finally, the fluctuations in the stress applied to a specific blade, which are due

to fabrication defects, aging and corrosion of the blade, vibrations of the system or turbulences of the gas flow, are modeled through a random variable $\delta\varphi$ added to the stress φ in eq. (27).

The values and distributions of the parameters T, ω, r_{hub} and r_{tip} have been set with reference to the helium gas turbine of a Gas Turbine Modular Helium Reactor (GT-MHR) developed by an international consortium, with a targeted 286MWe generation per module[25]; the material inherent characteristics A, n and ρ are taken assuming that the blade is made of Ni-base cast Superalloy 713LC [25]. The distributions used for the parameters are reported in Table III.

Information B: creep strain measurements

This source of information consists in a sequence of observations $\mathbf{z}_{1:i}$ of creep strain performed on the blade on which we want to apply the three prognostic approaches, hereafter called 'test trajectory'. Given the unavailability of real experimental data, in this work the creep growth trajectory is simulated using eq. (27). The variation in time of the rotational speed ω, the gas temperature T and the stress fluctuations $\delta\varphi$ are simulated by sampling their values ω_j, T_j and $\delta\varphi_j$ from the relative distributions (Table III) at each time instant t_j. Every 30 days a creep strain measurement $\mathbf{z}_j$, corresponding to the creep strain ε_j, is simulated by using eq. (28). A total number of 87 creep strain measurements have been simulated for a turbine blade with parameters $A=3\cdot10^{-4}$ and $n=6$.

In order to verify the performance of the prognostic approaches, the simulation of the test trajectory has been conducted until the time L at which the creep strain reaches the failure threshold. The difference between L and the time t_i at which the prognosis is performed is the actual remaining useful life of the turbine blade and will be referred to as "true RUL", and represented by the notation rul_i (Column 1, Table V).

Information C: historical creep strain measurements

This source of information consists in a number $S=13$ of historical sequences of creep growth observations from similar blades. In analogy to what is done for information B, the degradation trajectories have been simulated using eq. (27). The variations of the characteristic parameters A and n from one blade to another have been simulated by sampling their values from normal distributions at the beginning of each new simulated degradation path. Some examples of simulated creep growth paths are shown in Figure 8.

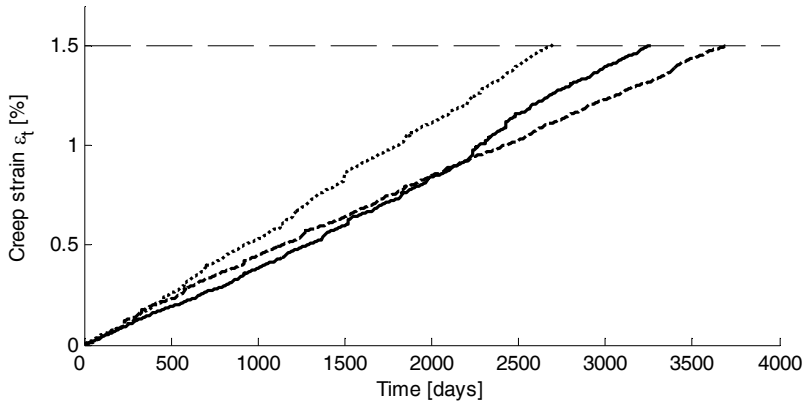

Figure 5: Examples of creep growth paths.

For each trajectory, a number N_s of direct creep strain measurement $\mathbf{z}_{1:N_s}$, one every 30 days, are simulated according to eq. (28) (Information E).

Information D: failure threshold

A turbine blade is considered within its useful life if the creep elongation strain in the longitudinal direction of the turbine blade is less than 1% or 2% of its initial length. Thus, the failure threshold for creep strain ε_{th} is set equal to the value of 1.5%.

Information E: measurement equation

For simplicity, we assume to be able to directly measure the value of the creep strain once every 30 days. Consequently, the observation equation is:

$$\mathbf{z}_j = \varepsilon_j + \upsilon_j \tag{28}$$

where υ_j is a white Gaussian measurement noise with standard deviation $\sigma_\upsilon = 0.02$.

Information F: life duration data

The time step at which the creep strain ε_j exceeds the failure threshold ε_{th} is collected for each of the S=13 simulated degradation trajectories and represents the life duration L_s of the simulated turbine blade.

Table IV: main sources of information for prognostics of a creeping turbine blade

Source	Description	Mathematical representation
A	The creep growth model and the distributions of the model parameters	eq. (27) and Table III
B	Measurements of the creep strain of the currently creeping blade taken at i different time instants t_j	$\mathbf{z}_{1:i} = \varepsilon_{1:i} + \upsilon_{1:i}$
C	Historical measurements of the creep strain of a set of S blades failed for creeping, taken at N_s different time instants t_j	$\mathbf{z}^s_{1:NS}$, s=1,…,S
D	The value of the failure threshold	$d_{th} = \varepsilon_{th}$
E	The measurement equation and the noise distribution	eq. (28) and $\Pr(\sigma_\upsilon)$
F	The length of life L^s of the set of S failed blades.	L^s, s=1,…,S

5.2 The three prognostic problems and corresponding modeling approaches

According to the three cases presented in Section 2, three prognostic problems have been tackled with respect to the turbine blade case study described in the previous Section 5.1. In all cases, the objective of the analysis is to predict at time t_i, i=1,…,87, the RUL distribution for the test trajectory. At every time t_i during the life of the turbine blade, the set of observations $\mathbf{z}_{1:i}$ is assumed to be available (source of

information B) and the predictions of the RUL is updated according to the new available information, i.e., the last observation $\mathbf{z}_i$.

In case 1, the sources of information A, B, D and E in Table IV are available. The particle filter has been applied and a number $U = 1000$ of particles are simulated starting from $\varepsilon_0 = 0$. Particle resampling is performed once every 5 measurements. The particle filter has been preferred to the Kalman filter since the distribution of the process noise is not Gaussian as a consequence of the combination of speed, temperature and stress fluctuations in the creep growth process described by eq. (27).

In case 2, sources of information B, C and F in Table IV are available. $S_{trn} = 10$ trajectories among the $S = 13$ totally available are used for building an ensemble of $B = 25$ linear least square models

$$\hat{rul}^b(\mathbf{z}_i \mid D^b_{i/o}) = \alpha^b_0 + \alpha^b_1 \cdot \mathbf{z}_i, \tag{29}$$

whereas the remaining $S_{val} = 3$ trajectories are used to validate the ensemble and build the training dataset for the model

$$\chi(\mathbf{z}_i) = \hat{\sigma}^2_{RUL_i} = \gamma_0 + \gamma_1 \cdot \mathbf{z}_i + \gamma_2 \cdot \mathbf{z}_i{}^2 \tag{30}$$

estimating the RUL variance $\sigma^2_{RUL_i}(\mathbf{z}_i)$.

In case 3, prognostic results are achieved based on the sources of information D and E of Table IV together with the information on the test trajectory (source of information B). In this case, the prognostic model has been developed only after time t_{30} in order to have available a dataset $\mathbf{D} = \{\mathbf{z}_{1:i}\}$ of at least $i = 30$ direct creep strain measurements. This dataset has been partitioned into a training dataset $\mathbf{D}^{trn}$ containing the first 75% of the available measurements and a validation dataset $\mathbf{D}^{val}$ containing the remaining 25%. An ensemble of $B = 25$ linear least square models

$$\eta^b(t_j \mid \mathbf{D}^b) = \beta^b_0 + \beta^b_1 \cdot t_j \tag{31}$$

is built and the models

$$\tilde{\eta}^b[\Delta\mathbf{z}_{j,j'} \mid \mathbf{D}^b] = \Delta\mathbf{z}_{j,j'} / \beta^b_1 \tag{32}$$

are derived from it. Notice that, in a linear process, the time needed to increase the degradation state from ε_j to $\varepsilon_{j'}$ is proportional to the degradation increment $\Delta\varepsilon_{j,j'} = \varepsilon_{j'} - \varepsilon_j$ and does not depend on the initial and

final degradation states. The ensemble of models is tested on the validation dataset made of input/output pairs

$$\mathbf{D}_{i/o}^{val} = \left\{\left\{\Delta\mathbf{z}_{j,j'} = \mathbf{z}_{j'} - \mathbf{z}_j ; \Delta t_{j,j'} = t_{j'} - t_j \right\}_{j=N_{trn}+1}^{i-1}\right\}_{j'=j+1}^{i}, \tag{33}$$

and the prediction residuals $r_{j,j'}$ obtained are used to train the linear model

$$\chi(\Delta\mathbf{z}_{j,j'}) = \lambda\Delta\mathbf{z}_{j,j'} \tag{34}$$

for the variance of Δt. The predictions $\hat{rul}_i^b$ and the estimate $\hat{\sigma}^2_{RUL_i}$ are obtained respectively from the model ensemble $\tilde{\eta}^b$ and from model χ in correspondence of the input $\Delta\mathbf{z}_{i,th} = \varepsilon_{th} - \mathbf{z}_i$. This way, the data used for training model χ concern creep strain increments which for the first two thirds of the trajectory are smaller than the increment $\Delta\mathbf{z}_{i,th}$ considered for obtaining the prognostic results, so that the empirical model $\chi(\Delta\mathbf{z}_{j,j'})$ is used in an input region not described by the training data.

Finally, the prediction $\hat{rul}_i$ and the relative prediction error variance are obtained from eqs. (23) and (24). Each time t_i, i=31,…,87, a new measurement becomes available, a new ensemble of models is built and a new RUL prediction is obtained.

5.3 Results

Table V reports the RUL predictions obtained by applying the three prognostic approaches of cases 1, 2 and 3 to a degrading blade. The first row refers to the RUL prediction performed at time $t_{50} = 1475$ days on the basis of the measurements $\mathbf{z}_{1:50}$ of the test trajectory, the second to the prediction performed at time $t_{80} = 2375$ days on the basis of the measurements $\mathbf{z}_{1:80}$. Column 1 reports the true RUL value, rul_i, observed for the turbine blade under test, whereas columns 2 and 3 report the expected value μ_{RUL_i} and the variance $\sigma^2_{rul_{ii}}$ of the distribution $\Pr(RUL_i \mid \varepsilon_i)$. This latter distribution represents the irreducible uncertainty of the RUL prediction which is caused by the stochastic future evolution of the creep strain. $\Pr(RUL_i \mid \varepsilon_i)$ has been obtained by simulating P=1000 degradation trajectories all characterized by the values A and n of the blade under test and by a creep strain ε_i at time t_i. Notice that the predictions $\hat{rul}_i$ of the three approaches provide satisfactory estimates of μ_{RUL_i}, whereas in all the cases the prediction error variances $\hat{\sigma}^2_{\hat{rul}_i}$ tend to overestimate $\sigma^2_{rul_{ii}}$. This is due to the fact that according to eq. (14), $\hat{\sigma}^2_{\hat{rul}_i}$ takes into consideration both the uncertainty due to the future stochastic evolution of the test trajectory represented by $\sigma^2_{RUL_i}$ and the uncertainty due to the prognostic model regression error σ^2_m. It is interesting to observe that an analyst which has to decide the maintenance policy to be applied to the turbine blade would like to have the least uncertain prediction of the RUL. Thus, in the case in which the analyst were in the position to choose one of the three prognostic approaches, he/she would prefer the one which guarantees the lowest uncertainty, i.e., the one whose prediction error variance is smaller.

Table V: estimates of μ_{RUL_i} and $\sigma^2_{rul_i}$ at time $t_{50} = 1475$ (first row) and $t_{80} = 2375$ (second row) in the three prognostic cases considered.

				Case 1		Case 2		Case 3	
t_i	rul_i	μ_{RUL_i}	$\sigma^2_{RUL_i}$	$\hat{rul}_i$	$\hat{\sigma}^2_{\hat{rul}_i}$	$\hat{rul}_i$	$\hat{\sigma}^2_{\hat{rul}_i}$	$\hat{rul}_i$	$\hat{\sigma}^2_{\hat{rul}_i}$
1475	1110	1092	90	1085	107	1079	109	1075	238
2375	210	264	42	247	45	167	63	248	57

In correspondence of each prediction $\hat{rul}_i$, it is also possible to estimate the prediction interval $PI(\alpha)$

$$C_i^{\inf}(\alpha) < rul_i < C_i^{\sup}(\alpha) , \tag{35}$$

i.e., the interval expected to contain the true RUL value rul_i with a probability of $1-\alpha$. According to the three approaches, this interval can be obtained as follows:

- In case 1, $C_i^{\inf}(\alpha)$ and $C_i^{\sup}(\alpha)$ are the $\alpha/2$ and $1-\alpha/2$ percentiles, respectively, of the RUL distribution estimated with the particle filtering method.
- In cases 2 and 3, assuming that the prediction error has a Gaussian distribution, the interval can be computed according to the theory of the bootstrap method [24] as:

$$\hat{rul}_i - c^{\alpha}_{conf}\,\hat{\sigma}_{\hat{rul}_i} < rul_i < \hat{rul}_i + c^{\alpha}_{conf}\,\hat{\sigma}_{\hat{rul}_i} \tag{36}$$

  where c^{α}_{conf} is the $1-\alpha/2$ percentile of a Student's *t*-distribution with number of degrees of freedom equal to the number B of bootstrapped models.

Figure 6 shows the evolution of the true value of the blade RUL (continuous thick line), its estimated value $\hat{rul}_i$ (dots) and the corresponding prediction interval for $\alpha = 0.32$ (continuous thin line) obtained during the turbine blade life at times t_i, i=1,…,80. In case three, since a minimum number of historical data must be available to build the predictive model, the prediction is performed only after t_{30}. Notice that in this case the prediction intervals are characterized by large oscillations and low accuracy, especially at the beginning of the trajectory, i.e., when few training data are available. Furthermore, the RUL prediction itself is very noisy. This effect can be reduced by properly filtering the predictions. To this purpose, since the time evolution of the RUL is a linear process ($rul(t) = rul(t-1) - 1$), and assuming a Gaussian noise affecting the prediction, Kalman filtering can be applied [26].

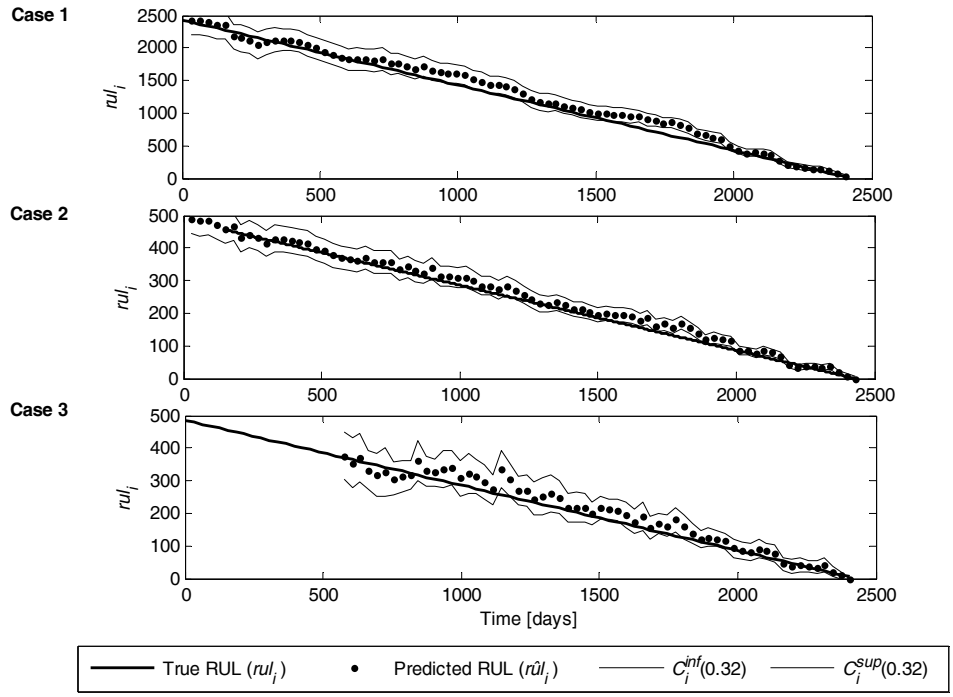

Figure 6: true RUL (continuous thick line) of a turbine blade with its predicted value $\hat{rul}_i$ *(dots) and prediction interval (continuous thin line) for the three prognostic cases.*

In order to perform a robust analysis of the performances of the three approaches, the model in eq. (27) has been used to generate 250 different creep growth trajectories. For each trajectory, the prognostic indicators $\hat{rul}_i(\mathbf{z}_i)$ and the confidence interval $C_i^{\inf}(0.32)$ have been computed at N_{tst} different time steps, once every 150 days, based on the past measurements collected once every 30 days.

For each degradation test trajectory, two performance indicators are computed:

1. the mean relative absolute error *rMAE*:

$$rMAE = \frac{1}{N_{tst}} \sum_{i=1}^{N_{tst}} \left| \frac{rul_i - \hat{rul}_i}{rul_i} \right| \tag{37}$$

   which evaluates the accuracy of the estimate $\hat{rul}_i$ with respect to the true rul_i of the system. Notice that since $\hat{rul}_i$ estimates the expected value of μ_{RUL_i} and not the true value of the rul_i, this value is not expected to be zero even for the best possible prognostic model.

2. the coverage:

$$Cov = \frac{1}{N_{tst}} \sum_{i=1}^{N_{tst}} c_i \ , \quad c_i = \begin{cases} 1 & C_i^{\inf}(0.32) < rul_i < C_i^{\sup}(0.32) \\ 0 & otherwise \end{cases} \tag{38}$$

This indicator is used to verify whether the estimation of the prediction interval $PI(0.32)$ actually contains with probability 1-0.32=0.68 the true RUL of the system. Coverage values around 0.68 indicate satisfactory estimation of the prediction interval.

The average values $\overline{rMAE}$ and $\overline{Cov}$ of the performance indicators obtained in the three cases over the 250 test trajectories are reported in Table VI.

Table VI: prognostic performance in the three prognostic cases considered

	$\overline{rMAE}$	$\overline{Cov}$
Case 1	0.150±0.009	0.663±0.018
Case 2	0.172±0.009	0.613±0.019
Case 3	0.170±0.009	0.682±0.014

The best results are obtained in case 1, which is the one with the maximum amount of information available. In this case the prediction is accurate (low rMAE) and the uncertainty of the prediction well estimated (coverage close to 0.68). The accuracy of the prediction in case 3 slightly outperforms the one in case 2, although a smaller amount of information is available to build the model. This can be explained by considering that in case 2 the prediction is based on knowledge about the creeping behaviors of a population of similar, but not identical blades, i.e., characterized by different values of parameters A and n; on the contrary, in case 3 the empirical model is trained using degradation data concerning only the turbine blade of interest and thus all training data refer to the same values of parameters A and n. To confirm this hypothesis, in Table VII two cases 1b and 2b analogous to cases 1 and 2 are considered: in case 1b it is assumed that the exact values of parameters A and n are known for each blade, whereas in case 2b the degradation trajectories used to build the training dataset are simulated using the same value of A and n considered in the test trajectory. We observe that the accuracy of the prediction is increased and in both cases 1b and 2b the results are better than in case 3 given the larger amount of information available.

Table VII: prognostic performance in case 1 when parameters A and n are assumed known and in case 2 when parameters A and n are kept constant for all historical training trajectories

	$\overline{rMAE}$	$\overline{Cov}$
Case 1b	0.135±0.009	0.669±0.019
Case 2b	0.145±0.007	0.623±0.016

6 Different information settings

In the previous Sections, we have considered three well defined situations of information available and we have developed three, properly tailored, prognostic approaches. However, in real applications, it is common to face hybrid situations characterized by the availability, at the same time, of multiple sources of information. Furthermore, since some sources of information can be partially inaccurate or affected by large

uncertainty, the identification of the correct prognostic approach to be applied can become a non trivial problem. In order to provide some indications to the decision maker, we consider a case in which all the sources of information listed in Table II are available (so that all the three proposed approaches can be applied) and we perform an analysis of the sensibility of the performance of the three prognostic approaches to the quality and quantity of the information available. To this purpose, we have considered the following indicators of the quality and quantity of the information:

E. the amplitude of the noise affecting the creep strain measurements;
F. the number of past measurements of the current trajectory available for making the RUL prediction;
G. the accuracy of the physical model of the degradation process;
H. the number of historical degradation trajectories available.

The performance of the three approaches is evaluated considering 250 test trajectories. Figure 7 (upper-left) shows the variation of the relative mean square error (*rMAE*) when the amplitude of the noise affecting the creep strain measurements is varied from 0 to 0.03. Notice that the performances of the three approaches decrease as the amplitude of the noise increases, and that the third approach is the most sensible to this parameter.

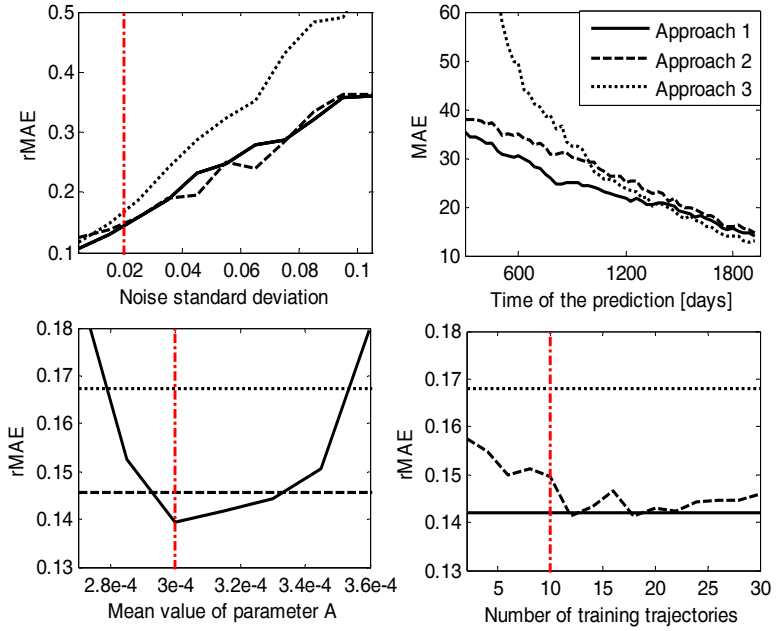

Figure 7: performance of the proposed approaches in different settings of information available. The vertical (red) line indicates the value assigned to the parameter in the numerical application of Section 5.

With respect to the second indicator, Figure 7 (upper-right) shows the mean absolute error of the approaches when they are applied at different time instants during the evolution of the degradation trajectories. Since every 30 days a new measurement is collected, the number of measurements available for making the RUL prediction increases as time passes. In this case, the performance is evaluated using the mean absolute error (*MAE*) instead of the *rMAE* which tends to be very noisy at the blade end-of-life, when the denominator rul_i of eq. (37) gets close to zero. The performance of approach 3 is the most affected by the time at which the prognosis is made: the performance is very poor at the beginning of the degradation trajectory, when very

few measurements are available for the construction of the empirical degradation model, but it increases significantly as time passes and, finally, when the creep strain becomes close to the failure threshold, approach 3 outperforms the other approaches.

The sensibility of approach 1 to the accuracy of the physical model (indicator 3.) has been estimated by using in the PF process model of eq. (27) a biased value of parameter A: its mean value has been taken in the range [2.7 10^{-4}; 3.6 10^{-4}] instead of equal to its true value $\mu_A=3\cdot10^{-4}$ (Table III). Figure 7 (bottom, left) shows that the particle filtering approach outperforms the others only if the available degradation model is very accurate: an error of 10% in the estimate of A is sufficient to decrease approach 1 performance under those of the other two methods. Similar results have been obtained varying the values of other parameters of the process model.

Finally, with respect to indicator 4, we have verified the performance of approach 2 varying the number of historical trajectories available for training the empirical model from 2 to 30. Figure 7 (bottom, right) shows that the *rMSE* made by approach 2 decreases when the number of historical trajectories increases from 2 to 12 and then tends to stabilize around a value very close to the performance of approach 1.

Considering the results obtained performing this sensibility analysis, we can provide the following guidelines for the choice of the prognostic approach: if one is very confident about the accuracy of the available physical degradation model, approach 1 should be preferred; on the contrary, if one doubts about the model accuracy, approach 2 is, in general, the most accurate, especially if the number of historical trajectories available is large. However, if the measurement noise is small, the system is close to failure and many degradation measurements have been taken during the current degradation trajectory, approach 3 can provide better accuracy.

When multiple approaches with comparable degree of accuracy are available, an alternative strategy to the choice of the best performing approach consists in the combination of the different approaches outcomes. This requires the development of a weighting strategy for the aggregation of the predictions made by the different approaches, based on their performances in the different situations of information available. To this aim, the aggregation techniques proposed in literature within the study of the ensemble system [23][26] will be considered in a future work.

7 Conclusions

Different forms of information and data may be available for the prognosis of the RUL of a system undergoing degradation. In this work, we have considered three practical situations with decreasing amount of information available: in the first case the model of the degradation process is available, in the second case the model is not available but can be empirically derived from a number of observations collected during the degradation trajectories to failure of similar systems, in the third case only direct measurements of the degradation state reached during the life of the system of interest are available.

In this work, we have discussed the choice of the prognostic method in different information settings, considering the accuracy and the ability of providing measures of confidence in the RUL prediction of different prognostic approaches. In the first considered case, where a physical model of the degradation

process is available, a particle filtering approach has been properly tailored to the prognostic problem, whereas a bootstrapped ensemble-based technique has been proposed and further developed to estimate the uncertainty of the RUL prediction in those situations where a priori knowledge of the mechanisms and models of the degradation process are missing (cases 2 and 3). For this, the prognostic problem has been reformulated, so that it was possible to define training and validation datasets of input/output pairs necessary for the construction of the prognostic model and the assessment of its accuracy; furthermore, solutions to ensure the independency between these two datasets have been developed. The merit of the proposed approach is that it allows producing a confidence interval for the RUL prediction, even when a measure of the prediction accuracy is not automatically provided by the regression method adopted.

The approaches proposed have been tested on a case study concerning the creep growth process in a high temperature turbine blade. The results show that both the particle filter and the bootstrapped ensemble methods provide a reliable prediction of the system RUL with a quantification of its uncertainty, the particle filter being the best performing method.

With respect to the ensemble of bootstrapped models trained with historical measurements of the degradation process in similar systems, the main limitation of the method is that it is not able to learn the peculiar characteristics of the system of interest but it tends to reproduce an 'average' degradation trajectory. To overtake this problem, a different modeling approach could be used, such as that based on the idea of fuzzy similarity [22], or a procedure for updating the ensemble with the information conveyed by new observations could be implemented [26].

For the application of the ensemble to the last case, in which only direct measurements of the degradation state reached during the life of the system of interest are available, the bootstrap method requires building an empirical model for the RUL variance estimate which is then used outside the region covered by the training data. Although good extrapolations have been obtained in the linear creep growth case study, the feasibility of the approach on more complex models should be verified.

Finally, it has been shown that in the general case where all sources of information are available, the performance of the three proposed approaches varies depending on the quantity and quality of the available information. Thus, aggregating their outcomes within an ensemble approach can improve the prediction accuracy; to this purpose, the choice of an effective weighting strategy is crucial and will be the objective of future research.

Appendix A: Particle Filtering

Given the first-order Markov process in eq. (1) and a set of observations $\mathbf{z}_{1:i}$ related to the equipment degradation state d_j by eq. (2), we aim at predicting the filtered posterior pdf $p_{D(t_i)}(d_i \mid \mathbf{z}_{1:i})$ at time t_i. Within a Bayesian framework, the filtered posterior pdf $p_{D(t_i)}(d_i \mid \mathbf{z}_{1:i})$ is recursively computed in two stages: prediction and update [16][31]. Given the pdf $p_{D(t_{i-1})}(d_{i-1} \mid \mathbf{z}_{1:i-1})$ at time t_{i-1}, the prediction stage involves using the transition probability $p_{D(t_i)}(d_i \mid d_{i-1})$ defined by the system equation (1) and the known distribution of the noise vector $\boldsymbol{\gamma}_{i-1}$ to obtain the prior probability distribution of the system state d_i at time t_i via the Chapman-Kolmogorov equation:

$$p_{D(t_i)}(d_i \mid \mathbf{z}_{1:i-1}) = \int p_{D(t_i)}(d_i \mid d_{i-1}, \mathbf{z}_{1:i-1}) p_{D(t_{i-1})}(d_{i-1} \mid \mathbf{z}_{1:i-1}) \,\mathrm{d}\, d_{i-1}$$
$$= \int p_{D(t_i)}(d_i \mid d_{i-1}) p_{D(t_{i-1})}(d_{i-1} \mid \mathbf{z}_{1:i-1}) \,\mathrm{d}\, d_{i-1} \tag{A1}$$

where the Markovian assumption underpinning the system model (1) has been used.

At time t_i, a new observation $\mathbf{z}_i$ is collected and used to update the prior distribution via Bayes rule, so as to obtain the required posterior pdf of the current state d_i [33]:

$$p_{D(t_i)}(d_i \mid \mathbf{z}_{1:i}) = \frac{p_{D(t_i)}(d_i \mid \mathbf{z}_{1:i-1}) p_{Z(t_i)}(\mathbf{z}_i \mid d_i)}{p_{Z(t_i)}(\mathbf{z}_i \mid \mathbf{z}_{1:i-1})} \tag{A2}$$

where the normalizing constant is

$$p_{Z(t_i)}(\mathbf{z}_i \mid \mathbf{z}_{1:i-1}) = \int p_{D(t_i)}(d_i \mid \mathbf{z}_{1:i-1}) p_{Z(t_i)}(\mathbf{z}_i \mid d_i) \,\mathrm{d}\, d_i \tag{A3}$$

The recurrence relations (A1) and (A2) form the basis for the exact Bayesian calculation of the pdf $p_{D(t_i)}(d_i \mid \mathbf{z}_{1:i})$ at time t_i.

Unfortunately, except for a few cases, including linear Gaussian state space models (Kalman filter) and hidden finite-state space Markov chains (Wohnam filter), it is not possible to evaluate analytically these distributions, since they require the evaluation of complex high-dimensional integrals.

An alternative and effective approach is that of resorting to Monte Carlo sampling methods for integration. This solution is based on sampling a large number K of trajectories $\{d_{0:i}^k\}_{k=1}^K$ (called particles), from a suitably introduced importance function $q(d_{0:i} \mid \mathbf{z}_{1:i})$. In the following we briefly describe how these simulated trajectories can be utilized for filtering out the unobserved trajectory of the real degradation process. For more details, one can refer to the specialized literature, e.g., [11] and [16].

The posterior probability $p_{D(t_i)}(d_i \mid \mathbf{z}_{1:i})$ we wish to calculate is the marginal of the probability $p_{D(t_{0:i})}(d_{0:i} \mid \mathbf{z}_{1:i})$, i.e., the multiple integral of this latter with respect to $d_o, d_1, ..., d_{i-1}$ in $[-\infty, \infty]$, which may be formally extended to include also the variable d_i by means of a δ-function, viz.,

$$p_{D(t_i)}(d_i \mid \mathbf{z}_{0:i}) = \int p_{D(t_{0:i-1})}(d_{0:i-1}, u \mid \mathbf{z}_{1:i}) \delta(d_i - u) \,\mathrm{d}\, d_{0:i-1} \,\mathrm{d}\, u \tag{A4}$$

Then, by using the large number of trajectories $\{d_{0:i}^k\}_{k=1}^K$ sampled from the importance function $q(d_{0:i} \mid \mathbf{z}_{1:i})$, the integral can be approximated as [10]:

$$p_{D(t_i)}(d_i|\mathbf{z}_{0:i}) = \int\left[\frac{p_{D(t_{0:i-1})}(d_{0:i-1},u\,|\,\mathbf{z}_{1:i})}{q(d_{0:i-1},u\,|\,\mathbf{z}_{1:i})}\delta(k_i-u)\right]q(d_{0:i-1},u\,|\,\mathbf{z}_{1:i})\,\mathrm{d}\,d_{0:i-1}\,\mathrm{d}\,u$$
$$\approx \frac{1}{K}\sum_{k=1}^{K} w_i^{*k}\delta(d_i - d_i^k) \tag{A5}$$

where w_i^k is the importance weight associated to the state sequence $d_{0:i}^k$, $k=1,2,...,K$, sampled from $q(d_{0:i}\,|\,\mathbf{z}_{1:i})$ and is given by

$$w_i^k = \frac{p_{D(t_i)}(d_{0:i}^k\,|\,\mathbf{z}_{1:i})}{q(d_{0:i}^k\,|\,\mathbf{z}_{1:i})} = \frac{p_{Z(t_i)}(\mathbf{z}_{1:i}\,|\,d_{0:i}^k)p_{D(t_i)}(d_{0:i}^k)}{p_{Z(t_i)}(\mathbf{z}_{1:i})q(d_{0:i}^k\,|\,\mathbf{z}_{1:i})} \tag{A6}$$

Typically $p_{Z(t_i)}(\mathbf{z}_{1:i})$ cannot be expressed in closed form. However, in [10] it is shown that the approximation in (A5) is equivalent to:

$$p_{D(t_i)}(d_i|\mathbf{z}_{0:i}) \approx \sum_{k=1}^{K}\frac{w_i^{*k}}{\sum_{k=1}^{K} w_i^{*k}}\delta(d_i - d_i^k) = \sum_{k=1}^{K} w_i^k\delta(d_i - d_i^k) \tag{A7}$$

where $\tilde{w}_i^k$ and w_i^k are, respectively, the *unnormalised* and *normalized* importance weights:

$$\tilde{w}_i^k = \frac{p_{Z(t_i)}(\mathbf{z}_{1:i}\,|\,d_{0:i}^k)p_{D(t_{0:i})}(d_{0:i}^k)}{q(d_{0:i}^k\,|\,\mathbf{z}_{1:i})},\quad w_i^k = \frac{w_i^{*k}}{\sum_{k=1}^{K} w_i^{*k}} \tag{A8}$$

It is often convenient to choose the importance density to be the transition probability:

$$q(d_i^k\,|\,\mathbf{z}_{1:i}) = p_{D(t_i)}(d_i^k\,|\,d_{i-1}^k) \tag{A9}$$

so that the importance function factorizes as follows

$$q(d_{0:i}^k\,|\,\mathbf{z}_{1:i}) = p_{D(t_i)}(d_i^k\,|\,d_{i-1}^k)p_{D(t_{0:i-1})}(d_{0:i-1}^k\,|\,\mathbf{z}_{1:i-1}) \tag{A10}$$

and one can obtain samples by augmenting each of the existing ones $d_{0:i-1}^k$ with the new state d_i^k sampled from $p_{D(t_i)}(d_i^k \mid d_{i-1}^k)$.

Using the Bayes rule, the hypothesis of Markovianity of the process and the fact that the observation $\mathbf{z}_i$ depends on the state d_i only, i.e., $p_{Z(t_i)}(\mathbf{z}_i \mid d_{0:i}^k) = p_{Z(t_i)}(\mathbf{z}_i \mid d_i^k)$, the wights $\tilde{w}_i^k$ defined in eq. (A8) can be rewritten as (details of the calculations can be found in [16]):

$$\begin{aligned} \tilde{w}_i^k &= \frac{p_{Z(t_i)}(\mathbf{z}_i \mid d_i^k) p_{D(t_i)}(d_i^k \mid d_{i-1}^k)}{p_{D(t_i)}(d_i^k \mid d_{i-1}^k)} \frac{p_{Z(t_{0:i-1})}(\mathbf{z}_{1:i-1} \mid d_{0:i-1}^k) p_{D(t_{0:i-1})}(d_{0:i-1}^k)}{p_{D(t_{0:i-1})}(d_{0:i-1}^k \mid \mathbf{z}_{1:i-1})} \\ &= p_{Z(t_i)}(\mathbf{z}_i \mid d_i^k) w_{i-1}^k \end{aligned} \tag{A11}$$

where $p_{Z(t_i)}(\mathbf{z}_i \mid d_i^k)$ is the *likelihood* of the observation $\mathbf{z}_i$, which can be derived from the observation equation (2).

The resulting normalized weights w_i^k are then:

$$w_i^k = \frac{p_{Z(t_i)}(\mathbf{z}_i \mid d_i^k) w_{i-1}^k}{\sum_{k=1}^{K} p_{Z(t_i)}(\mathbf{z}_i \mid d_i^k) w_{i-1}^k} \tag{A12}$$

Appendix B: Bootrapped ensemble-based estimate of the prediction uncertainty

Assume we are given a set of data pairs $\{\mathbf{z}_i, y_i\}_{i=1}^N$, generated according to

$$y_i = h(\mathbf{z}_i) + \upsilon_i \tag{B1}$$

where y_i is the target value, $h(\mathbf{z}_i)$ true input/output relation and υ_i a process noise with zero mean and standard deviation $\sigma_{\upsilon_i}^2(\mathbf{z}_i)$. When we train a model $f(\mathbf{z}_i)$ on such data, our aim is to approximate $h(\mathbf{z}_i)$; such a model can be interpreted as an estimate of the mean distribution of the target values given an input vector $\mathbf{z}_i$. In many practical application, all the more in prognostics, it is highly desirable to have a measure of confidence in the prediction $f(\mathbf{z}_i)$. As described in [24] the uncertainty in the prediction $f(\mathbf{z}_i)$ is quantified by the prediction error variance $\sigma_f^2(\mathbf{z}_i)$ which can be decomposed in two terms:

$$\begin{aligned}\sigma_f^2(\mathbf{z}_i) &= E[(f(\mathbf{z}_i) - y_i)^2] = \\ &= E\left\{[f(\mathbf{z}_i) - h(\mathbf{z}_i)]^2\right\} + E\left\{[h(\mathbf{z}_i) - y_i]^2\right\} \\ &= \sigma_m^2(\mathbf{z}_i) + \sigma_{\upsilon_i}^2(\mathbf{z}_i)\end{aligned} \tag{B2}$$

where the term $\sigma_m^2(\mathbf{z}_i)$ is the variance of the distribution of $f(\mathbf{z}_i) - h(\mathbf{z}_i)$ and is concerned with the accuracy of the model $f(\mathbf{z}_i)$ in estimating the true function $h(\mathbf{z}_i)$, whereas the term $\sigma_{\upsilon_i}^2(\mathbf{z}_i)$ is the variance of the distribution of $\upsilon_i = y_i - h(\mathbf{z}_i)$, and is concerned with the accuracy of $h(\mathbf{z}_i)$ in predicting the target y_i itself.

To generate an estimate of $\sigma_m^2(\mathbf{z}_i)$ one can resort to a bootstrapped ensemble of models $\{f^b(\mathbf{z}_i)\}_{b=1}^B$ built on bootrapped replicates of the original set of data pairs $\{\mathbf{z}_i, y_i\}_{n=1}^N$. These replicates are obtained by randomly sampling with replacement *N* data pairs from the original dataset. As derived in [32], the bootstrapped outputs $\{f^b(\mathbf{z}_i)\}_{b=1}^B$ provide us with the empirical estimate of the distribution of $f(\mathbf{z}_i) - h(\mathbf{z}_i)$. This estimate is given by the distribution of $f(\mathbf{z}_i) - \overline{f^b(\mathbf{z}_i)}$, where $\overline{f^b(\mathbf{z}_i)}$ is the average value of the predictions $\{f^b(\mathbf{z}_i)\}_{b=1}^B$ and is retained as the estimate of the true function $h(\mathbf{z}_i)$ to which we have no access. Thus, under the hypothesis that $\overline{f^b(\mathbf{z}_i)}$ is an unbiased estimator of $h(\mathbf{z}_i)$ and that the distribution $f(\mathbf{z}_i) - h(\mathbf{z}_i)$ is Gaussian (see [32] for more details), the term $\sigma_m^2(\mathbf{z}_i)$ can be estimated as

$$\hat{\sigma}_m^2(\mathbf{z}_i) = \frac{1}{B}\sum_{b=1}^{B}[f^b(\mathbf{z}_i) - \overline{f^b(\mathbf{z}_i)}]^2 \tag{B3}$$

To estimate total variance $\sigma_f^2(\mathbf{z}_i)$, we need to build a model that provide an estimate for the noise term $\sigma_{\upsilon_i}^2(\mathbf{z}_i)$ in correspondence of an input $\mathbf{z}_i$. Such a model is found by fitting the residuals

$$\{r^2(\mathbf{z}_i) = [y_i - \overline{f^b(\mathbf{z}_i)}]^2 - \sigma_m^2\}_{j=1}^{N'} \tag{B4}$$

which should be calculated on a validation dataset different than the training set to avoid overfitting and underestimating of the variance $\sigma_{\upsilon_i}^2(\mathbf{z}_i)$. In other words, we train a model $\chi(\mathbf{z}_i)$ on the set of input output pairs $\{\mathbf{z}_i, r^2(\mathbf{z}_i)\}_{n=1}^{N'}$ by maximizing the loglikelihood function

$$LL = \sum_{i=1}^{N'} \log\left(\frac{1}{\sqrt{2\pi}\chi(\mathbf{z}_i)}\exp\left(-\frac{r^2(\mathbf{z}_i)}{2\chi^2(\mathbf{z}_i)}\right)\right) \tag{B5}$$

written under the hypothesis that the residuals $r^2(\mathbf{z}_i)$ have a Gaussian distribution with zero mean and variance which can be shown to be equal to $\sigma_{\upsilon_i}^2(\mathbf{z}_i)$:

$$\begin{aligned} E[r^2(\mathbf{z}_j)] &= E\{[y_i - f(\mathbf{z}_i)]^2 - \sigma_m^2(\mathbf{z}_i)\} \\ &= E\{[y_i - f(\mathbf{z}_i)]^2 - \sigma_m^2(\mathbf{z}_i) \\ &= \sigma_f^2(\mathbf{z}_i) - \sigma_m^2(\mathbf{z}_i) \\ &= \sigma_{\upsilon_i}^2(\mathbf{z}_i) \end{aligned} \tag{B6}$$

where eq. (B2) has been used in the last equivalence of eq. (B6).

Finally, the prediction error variance in correspondence of the input $\mathbf{z}_i$ can be approximated by:

$$\sigma_f^2(\mathbf{z}_i) = \hat{\sigma}_m^2(\mathbf{z}_i) + \hat{\sigma}_{\upsilon_i}^2(\mathbf{z}_i) = \left\{ \frac{1}{B} \sum_{b=1}^{B} [f^b(\mathbf{z}_i) - \overline{f^b(\mathbf{z}_i)}]^2 \right\} + \chi^2(\mathbf{z}_i) \tag{B7}$$

References

[1] G. Vachtsevanos, F. L. Lewis, M. Roemer, A. Hess and B. Wu, Intelligent Fault Diagnosis and Prognosis for Engineering Systems, 1st edition, John Wiley & Sons, Hoboken, 2006.

[2] D. B. Jarrell, D. R Sisk and L. J. Bond, Prognostics and Condition-Based Maintenance: A New Approach to Precursive Metrics, Nucl. Technol., 145, 2004, 275-286.

[3] J. W. Hines and A. Usynin, Current Computational Trends in Equipment Prognostics, Int. J. Comput. Intell. Syst., 1(1), 2008, 94–102.

[4] E. Zio, Prognostics and Health Management of Industrial Equipment, submitted for publication.

[5] T. Brotherton, G. Jahns, J. Jacobs and D. Wroblewski, Prognosis of faults in gas turbine engines, IEEE Aerosp. Conf. Proc., 6, 2000, 163-171, 18-25 Mar, Big Sky, MT, USA.

[6] J. Luo, K. Pattipati, L. Qiao, and S. Chigusa, Model-based Prognostic Techniques Applied to a Suspension System, IEEE Trans. on Syst., Man, and Cybern, 38(5), 2008, 1156-1168.

[7] U. Pulkkinen, A stochastic model for wear prediction through condition monitoring, in: K. Holmberg, A. Folkeson (Eds.), Operational reliability and systematic maintenance, Elsevier, London (New York), 1991, pp. 223-43.

[8] A. Ray and S. Tangirala. A nonlinear stochastic model of fatigue crack dynamics, Prob. Engng Mech., 12(1), 1997, 33-40.

[9] R.W. Swindeman, and M.J. Swindeman, A comparison of creep models for nickel base alloy for advanced energy systems, Int. J. of Press. Vessel. and Pip.,85, 2008, 72–79.

[10] A. Doucet, On sequential simulation-based methods for Bayesian filtering, Technical report. Dept. of Engng, University of Cambridge, CUED-F-ENGTR310, 1998.

[11] A. Doucet, J.F.G. de Freitas and N.J. Gordon. Sequential Monte Carlo methods in practice. Springer-Verlag, New York, 2001.

[12] B.D. Anderson, J.B. Moore, Optimal filtering, Prentice Hall, Englewood Cliffs (NJ), 1979.

[13] G. Kitagawa, Non-Gaussian state-space modeling of nonstationary time series, J. of the Am. Stat. Assoc.,82, 1987, 1032-63.

[14] F. Cadini, E. Zio and D. Avram, Monte Carlo-based filtering for fatigue crack growth estimation, Prob. Engng Mech., 24, 2009, 367-373.

[15] T. Khan, Particle filter based prognosis study for predicting remaining useful life of steam generator tubing, IEEE Conference on Prognostics and Health Management (PHM), 20-23 June 2011, Montreal, QC.

[16] M.S. Arulampalam, S Maskell, N Gordon, and T Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Process, 50(2), 2002, 174-88.

[17] M.A. Schwabacher, A Survey of Data-Driven Prognostic, Infotech@Aerospace, 26 - 29 September 2005, Arlington, Virginia.

[18] K. Goebel, B. Saha, and A. Saxena, A Comparison of Three Data-Driven Algorithms for Prognostics, Proc. of the 62nd Meet. of the Soc. for Mach. Fail. Prev. Technol., 2008, pp.119-131.

[19] L. Peel, Data driven prognostics using a Kalman filter ensemble of neural network models, Proc. Int. Conf. on prong. and health manag., 2008.

[20] B. Saha, K. Goebel, and J. Christophersen, Comparison of Prognostic Algorithms for Estimating Remaining Useful Life of Batteries, Trans. of the Inst. of Meas. & Control, 31(3-4), 2009, 293-308.

[21] F. Di Maio, J. Hu, P. Tse, K. Tsui, E. Zio, and M. Pecht, Ensemble-approaches for clustering health status of oil sand pumps, Expert Syst. with Appl., doi: 10.1016/j.eswa.2011.10.008.

[22] E. Zio, and F. Di Maio, A Fuzzy Similarity-Based Method for Failure Detection and Recovery Time Estimation. Int. J. of Perform. Engng, 6(5), 2010, 407-424.

[23] R. Polikar, Ensemble Based Systems in Decision Making, IEEE Circuits and Syst. Mag., 6(3), 2006, 21-45.

[24] T. Heskes, Practical confidence and prediction intervals, in: M.Mozer, M. Jordan, and T. Heskes (Eds.), Advances. Neural Information Processing Systems 9, MIT Press, Cambridge, 1997, pp. 466–472.

[25] R. Couturier, and C. Escaravage, High temperature alloys for the HTGR Gas Turbine: Required properties and development needs, IAEA-TECDOC-1238, 2000.

[26] P. Baraldi, F. Mangili, and E. Zio, A Kalman Filter - based Ensemble Approach for Turbine Creep Prognostics. Submitted for publication, 2011.

[27] E. Zio, A Study of the Bootstrap Method for Estimating the Accuracy of Artificial Neural Networks in Predicting Nuclear Transient Processes. IEEE Trans. Nucl. Sci., 53(3), 2006, 1460-1478.

[28] N. Gorjian, L. Ma, M. Mittinty, P. Yarlagadda, and Y. Sun, Review on Degradation Models in Reliability Analysis, Proc. of the 4th World Congr. on Engng Asset Manag., Athens, 28-30 Sept. 2009.

[29] M. Saez, N. Tauveron, T. Chataing, G. Geffraye, L. Briottet, and N. Alborghetti, Analysis of the turbine deblading in an HTGR with the CATHARE code, Nucl. Engng and Design, 236, 2006, 574–586.

[30] M. Orchard, A Particle Filtering-based Framework for On-line Fault Diagnosis and Failure Prognosis, Ph.D. Thesis, Dept. of Electrical and Computer Engng, Georgia Institute of Technology, 2007.

[31] P. Li and V. Kadirkamanathan, “Particle Filtering based Likelihood Ratio Approach to Fault Diagnosis in Nonlinear Stochastic Systems”, in IEEE Trans. On Systems, Man, And Cybernetics, Part C: Applications and Reviews, vol. 31, no 3, pp. 337-343, 2001.

[32] Carney, J.G. Cunningham, P., Bhagwan, U., Confidence and prediction intervals for neural network ensembles, International Joint Conference on Neural Networks (IJCNN), Jul 1999, Vol. 2, pp. 1215-1218

Paper III

Investigation of uncertainty treatment capability of model-based and data-driven prognostic methods using simulated data

Piero Baraldi, Francesca Mangili, Enrico Zio

Reliability Engineering and System Safety 2013, Vol 112C, pp. 94-108

Investigation of uncertainty treatment capability of model-based and data-driven prognostic methods using simulated data

Piero Baraldi[a], Francesca Mangili[a], Enrico Zio[b,a]

[a]*Dipartimento di Energia, Politecnico di Milano, Italy*

[b]*Chair on Systems Science and the Energetic challenge, European Foundation for New Energy-Electricite' de France, Ecole Centrale Paris and Supelec, France*

Abstract

We look at different prognostic approaches and the way of quantifying confidence in equipment Remaining Useful Life (RUL) prediction. More specifically, we consider: 1) a particle filtering scheme, based on a physics-based model of the degradation process; 2) a bootstrapped ensemble of empirical models trained on a set of degradation observations measured on equipments similar to the one of interest; 3) a bootstrapped ensemble of empirical models trained on a sequence of past degradation observations from the equipment of interest only.

The ability of these three approaches in providing measures of confidence for the RUL predictions is evaluated in the context of a simulated case study of interest in the nuclear power generation industry and concerning turbine blades affected by developing creeps.

The main contribution of the work is the critical investigation of the capabilities of different prognostic approaches to deal with various sources of uncertainty in the RUL prediction.

Keywords: Prognostics, uncertainty, particle filtering, bootstrap ensemble, turbine blade, creep.

1 Introduction

In prognostics the current system condition is projected in time by a predictive model [1-2]. Since the prediction of the Remaining Useful Life (RUL) of degrading equipment is performed in the absence of future measurements concerning equipment degradation and operational conditions, the prognostic task is necessarily affected by large uncertainty. In this work, the sources of uncertainty affecting the RUL prediction are classified in three categories:

a. Randomness in the future degradation of the equipment. This intrinsic uncertainty in the degradation process has several causes such as the unknown future load profile, and operation and environmental conditions.

b. Modeling error, i.e., inaccuracy of the prognostic model used to perform the prediction. In model-based prognostic approaches, this source of uncertainty takes into account the assumptions and simplifications

made on the form and structure of the model, and the uncertainty on the model parameters. In data-driven approaches, it relates to the incomplete coverage of the data set used to train the empirical model.

Table I: nomenclature

Symbol	Description	Symbol	Description
t_i	i-th time instant	h	reference index of the ensemble models
d_i	equipment degradation at time t_i	H	number of empirical models
$\mathbf{z}_i$	physical observations related to d_i	r_i^s	prediction residual obtained in correspondence of the input $\mathbf{z}_i^s$
RUL_i	random variable (rv) representing system RUL at time t_i	χ	empirical model for the prediction of the residuals variance
δ_i	difference between RUL_i and its expected value E[RUL_i]	$\hat{d}_i$	prediction of d_i at time t_i
rul_i	realization of RUL_i	$\eta(t_i)$	empirical model for the prediction of d_i
$f(\mathbf{z}_i)$	prognostic model for the prediction of rul_i	$\mathbf{d}_{j,j'}$	vector of the degradation states at t_j and $t_{j'}$
$\hat{rul}_i$	prediction of rul_i	$\Delta t_{j,j'}$	rv representing the time interval to evolve from d_j to $d_{j'}$
$\sigma^2_{\hat{rul}}$	prediction error variance		
σ^2_A	prediction error variance due to randomness in the future degradation of the equipment	$\tilde{\eta}(\mathbf{d}_{j,j'})$	empirical model for the prediction of $\Delta t_{j,j'}$
		$\Delta\hat{t}_{j,j'}$	estimate of $\Delta t_{j,j'}$
σ^2_B	prediction error variance due to modeling error	$\delta_{\Delta t_{j,j'}}$	difference between $\Delta t_{j,j'}$ and its mean value
σ^2_C	prediction error variance due to uncertainty in equipment degradation data	$r_{j,j'}$	prediction residual obtained in correspondence of the input $\mathbf{d}_{i,j'}$
s	index of the equipment	ε_j	creep strain at time t_j
S	number of equipments	Q	creep activation energy
L_s	failure time of the s-th equipment	ω_j	turbine rotational speed
d_{th}	failure thresholds	K	constant relating the load to ω_j
Pr(x)	probability distribution function of the rv x	R	ideal gas constant
Pr(x|y)	conditional probability distribution function of the rv x given y	T_j	blade operating temperature at time t_j
		$\delta\varphi_j$	stress fluctuations
μ_x	mean value of the rv x	v_j	creep strain measurement noise
σ_x	standard deviation of the rv x	ε_{th}	creep strain failure threshold
g	transition function of the degradation state	$\alpha^h_{0,1}$ / $\beta^h_{0,1}$	coefficients of the h-th empirical model of approach 2/3
$\boldsymbol{\gamma}_j$	process noise vector		
$\hat{\sigma}^2_x$	estimate of σ^2_x	$\gamma_{0,1,2}$ / λ	coefficients of the empirical model of the residual variance χ of approach 2/3
p	index of the Monte Carlo sampled particle		
P	number of particles	$\Delta\varepsilon_{j,j'}$	degradation increment from t_j to $t_{j'}$
w_i^p	weight of the p-th particle at time t_i	$\Delta\mathbf{z}_{j,j'}$	observed degradation increment from t_j to $t_{j'}$
$\mathbf{D}$	dataset made by the observations $\mathbf{z}_i$	α	significance level of the confidence interval
$\mathbf{D}_{i/o}$	dataset of the input/output pairs ($\mathbf{z}_i^s$, rul_i^s)	$C_i^{\inf/\sup}(\alpha)$	inferior/superior bound of the (1-α)-confidence interval for the RUL prediction
trn/val	apex/subscript indicating training/ validation sets of data	c^{α}_{conf}	(1-$\alpha/2$) percentile of a Student's t-distribution with H degrees of freedom
$N^{trn/val}$	number of training/validation patterns		

c. Uncertainty in current and past equipment degradation data, which are used by the prognostic model to elaborate the RUL prediction. These data are usually acquired by sensors with some measurement noise or derived from diagnostic systems assessing the equipment health state with some degree of uncertainty.

Other possible sources of uncertainty, which are not considered in this work, are the imperfect knowledge of the value of degradation beyond which the equipment can no longer perform its functions (failure threshold), or the time at which the degradation process starts (degradation initiation). Furthermore, in this work we assume that the equipment degradation is caused by a single degradation mechanism, not considering the uncertainty on the degradation caused by the onset of other, possibly competing, mechanisms [3].

The challenge of managing uncertainties associated with prognostics has been recently addressed in [1-2,4-5]. Uncertainty management in prognostics entails to identify, classify and analyze uncertainty sources with the aim of associating to the RUL predictions provided by a prognostic model an estimate of its uncertainty [4-7], i.e., a measure of the expected degree of mismatch between the real and predicted equipment failure time. This information, provided in the form of a probability distribution of the equipment RUL, can be used by the decision maker to confidently plan maintenance actions, according to the desired risk tolerance [2].

In this context, the objective of the present paper is to contribute to the way of investigating the capabilities of different prognostic approaches to deal with the uncertainty in the RUL prediction. To this aim, the analysis is performed with respect to three previously developed approaches [8].

A first approach, hereafter named “approach 1”, is based on a mathematical model of the degradation process for the RUL prediction [9], embedded in a filtering method capable of accounting for the stochasticity of the process (source of uncertainty A) and the noise affecting the measurements (source of uncertainty C). Most filtering approaches rely on Bayesian methods and provide the probability distribution of the RUL [10-11]. The exact Kalman filter has been largely used in case of linear state space models and independent, additive Gaussian measurements and modeling noises, whereas analytical or numerical approximations of the exact solution (such as the Extended Kalman filter, the Gaussian-sum filters or the approximate grid-based filters [12]) have been applied in cases where the dynamics of degradation is non-linear and/or the associated noises are non-Gaussian [13]. Numerical approximations based on the Monte Carlo sampling technique have gained popularity for their flexibility and ease of design [14-17]. Among them, Particle Filtering (PF) is often considered a state-of-the-art technology in the prognostic field and used as a term of comparison for newer approaches. The model-based particle filter approach here considered was firstly applied to state estimation for diagnostics [18-19] and then applied to prognostics [20-21]. According to the particle filtering scheme proposed in [20], the RUL distribution prediction is performed by considering the stochastic model of the degradation process and the on-line observations of the equipment degradation. In [22], this particle filtering-based prognostic approach is discussed with respect to the design of a predictive maintenance strategy, whose advantages are then compared with those of other maintenance strategies.

We consider also two data-driven approaches [23], based on statistical models that ‘learn’ trends from historical data. In particular, we consider bootstrap ensemble approaches [24-25], which are based on the aggregation of multiple model outcomes and have gained interest due to their ability of estimating the uncertainty in the predictions. These approaches allow estimating the model uncertainty (source of uncertainty B) by considering the variability in the predictions of the diverse models of the ensemble [24]. On the other hand, the estimate of the uncertainty due to the stochasticity of the degradation process (source A) and the input noise (source C), requires to investigate the relation between the input and the error of the prognostic model based on its performance on a validation dataset. In what we will refer to as “approach 2”,

a bootstrap ensemble model is built to estimate the equipment RUL based on sequences of observations of evolution to failure of a set of similar equipments operating under similar conditions; in what we will refer to as "approach 3", a bootstrap ensemble model is built based on a sequence of degradation observations only of the equipment whose RUL we want to predict. Although approaches 2 and 3 are both based on the development of an ensemble of bootstrap models, they differ for the type of model used. Whereas in approach 2 we can directly model the RUL as a function of the observed parameters, in approach 3 we have to model the degradation evolution as a function of time, since direct RUL observations are not available.

The three approaches are investigated with reference to the creep growth process in the turbine blades of a Gas Turbine Modular Helium nuclear Reactor (GT-MHR) [26-27]. The data used in this case study have been numerically simulated using a traditional model of the creep growth. Artificial data have been used in order to allow testing the three approaches on a large number of different blade degradation trajectories and thus evaluate their capability of correctly estimating the uncertainty on the provided RUL prediction.

The remainder of the paper is organized as follows: Section 2 presents the terminology used in the paper and the problem setting; in Section 3, the decomposition of the prediction error variance into three terms corresponding to the randomness in the future degradation of the equipment, the modeling error, and the uncertainty in current and past equipment degradation measures is reported; in Section 4, the three considered prognostic approaches are described; in Section 5, the problem of blade creeping in high temperature turbines is illustrated and the capability of uncertainty management of the three prognostic approaches are discussed; finally, in Section 6 some conclusions are drawn and potential future work suggested.

2 Terminology and problem setting

In this work, we assume that the equipment is subject to a single degradation mechanism described as a random process; we do not consider the effects that other competing degradation mechanisms can have on the equipment degradation. Also, we assume that degradation cannot exceed a maximum acceptable level, hereafter referred to as "failure threshold", d_{th}, which is fixed and identical among similar equipment. Notice that in prognostics the failure threshold does not necessarily indicate complete failure of the system, but, for safety margins, it is often set at a conservative value of the degradation limit beyond which the risk of complete failure exceeds tolerance limits or the performance of the system does not fulfill the requirements [28]. Since the failure threshold is usually derived from expert knowledge or from experimental measurements of the equipment degradation at failure, its estimate is typically affected by uncertainty which contributes to increase the RUL prediction error variance. Methods to deal with the uncertainty on the failure threshold have been proposed in [29-30], and, for this reason, this aspect is not considered in the present work.

We indicate by d_i the equipment degradation level at time t_i and we assume that its direct measure is not available, but some physical observations $\mathbf{z}_i$ related to it are obtained. We indicate by $\mathbf{z}_{1:i}=(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_i)$ the past and present observations taken at times $t_1, t_2 \ldots, t_i$ for the equipment whose RUL we want to predict. In some cases, also the sequences of observations of evolution to failure of a set of S similar equipments

operating under similar conditions are available and we will indicate by $\mathbf{z}_{1:N_s}^s$, s=1,…,S, the observations taken at times $t_1,...,t_{N_s}$ for the s-th equipment whose failure time is L_s, where $L_s \geq t_{N_s}$.

The objective of prognostics is the estimation of the equipment RUL, i.e., the time left from the current time t_i before the equipment degradation, currently of value d_i, crosses the failure threshold d_{th}. As degradation evolves randomly in time, the equipment RUL at time t_i is a random variable which will be referred to as RUL_i. Thus, the objective of applying prognostics to an equipment of current degradation level d_i is to estimate the probability density function (pdf) $\Pr(RUL_i \mid d_i)$. The uncertainty described by such distribution regards the future stochastic evolution of the equipment degradation and, thus, it is irreducible.

A realization rul_i of the random variable RUL_i can be written as:

$$rul_i = \mu_{RUL_i|d_i} + \delta_i \tag{1}$$

where $\mu_{RUL_i|d_i}$ is the RUL expected value of the equipment with degradation d_i at time t_i, and δ_i is a random variable with zero mean and variance σ_A^2 which represents the uncertainty on the future evolution of degradation (source of uncertainty A).

Furthermore, in practice, the 'exact' model, g, of the equipment degradation process is not available (source of uncertainty B) and the degradation d_i at time t_i is not exactly known (source of uncertainty C). In this setting, the complete distribution of RUL_i cannot be derived and prognostics is limited to estimating:

- the expected value of RUL_i
- the variance of the prediction error as a measure of the accuracy with which the estimated expected value predicts the actual RUL value.

With respect to the estimate of the expected value of RUL_i, it will be indicated by $\hat{rul}_i$ and considered as our RUL prediction. The prognostic model which generates at time t_i the estimate $\hat{rul}_i$ of RUL_i on the basis of the observations $\mathbf{z}_i$ will be referred to as f, i.e. $\hat{rul}_i = f(\mathbf{z}_i)$. Finally, we indicate by $\sigma^2_{\hat{rul}_i}$ the estimate of the prediction error variance, defined by $\sigma^2_{\hat{rul}_i} = E[(\hat{rul}_i - RUL_i)^2]$.

3 Prediction error variance

According to [24], the prediction error variance $\sigma^2_{\hat{rul}_i}$ can be decomposed into two terms: σ_A^2, the variance related to the uncertainty on the future degradation of the equipment whose degradation at time t_i is d_i (source of uncertainty A), and σ^2_{B+C}, the variance related to the imprecision of the model $f(\mathbf{z}_i)$ (source of uncertainty B) and the noise on the data $\mathbf{z}_i$ (source of uncertainty C):

$$\begin{aligned} \sigma^2_{\hat{rul}_i} &= E[(RUL_i - \hat{rul}_i)^2] = \sigma_A^2 + \sigma^2_{B+C} = \\ &= E\left[\left(RUL_i - \mu_{RUL_i|d_i}\right)^2\right] + E\left[\left(\mu_{RUL_i|d_i} - f(\mathbf{z}_i)\right)^2\right] \end{aligned} \tag{2}$$

In some applications, it can be useful to distinguish the uncertainty due to the modeling error (source B) from that due to the noise on the input data (source C). To this aim, we introduce the quantity $\mu_{RUL_i|\mathbf{z}_i}$ which represents the RUL_i expected value of a degrading equipment for which at time t_i we have the observations $\mathbf{z}_i$, and we assume that $f(\mathbf{z}_i)$ is an unbiased estimator of $\mu_{RUL_i|\mathbf{z}_i}$.Thus we obtain:

$$\begin{aligned}\sigma^2_{B+C} &= E\left[\left(f(\mathbf{z}_i) - \mu_{RUL_i|d_i}\right)^2\right] = E\left[\left(f(\mathbf{z}_i) - \mu_{RUL_i|\mathbf{z}_i} + \mu_{RUL_i|\mathbf{z}_i} - \mu_{RUL_i|d_i}\right)\right)^2\right] \\ &= E\left[\left(f(\mathbf{z}_i) - \mu_{RUL_i|\mathbf{z}_i}\right)^2\right] + E\left[\left(\mu_{RUL_i|\mathbf{z}_i} - \mu_{RUL_i|d_i}\right)^2\right] \\ &= \sigma^2_B + \sigma^2_C\end{aligned} \tag{3}$$

Combining eqs. (2) and (3), one obtains:

$$\begin{aligned}\sigma^2_{\hat{rul}_i} &= E[(\hat{rul}_i - RUL_i)^2] = \\ &= E\left[\left(\mu_{RUL_i|d_i} - RUL_i\right)^2\right] + E\left[\left(f(\mathbf{z}_i) - \mu_{RUL_i|\mathbf{z}_i}\right)^2\right] + E\left[\left(\mu_{RUL_i|\mathbf{z}_i} - \mu_{RUL_i|d_i}\right)^2\right] \\ &= \sigma^2_A + \sigma^2_B + \sigma^2_C\end{aligned} \tag{4}$$

Notice that these results have been obtained by assuming that the different components of the prediction error are independent and thus the expected values $E[(RUL_i - \mu_{RUL_i|d_i})(\mu_{RUL_i|d_i} - f(\mathbf{z}_i))]$ and $E[(f(\mathbf{z}_i) - \mu_{RUL_i|\mathbf{z}_i})(\mu_{RUL_i|\mathbf{z}_i} - \mu_{RUL_i|d_i})]$ in eqs. (2) and (3), respectively, are zero.

4 Modeling approaches for RUL prediction

This Section illustrates briefly the three modeling approaches considered for RUL prediction.

4.1 Approach 1: Particle Filtering

In approach 1, a Monte Carlo-based filtering technique, called particle filtering [10,12], is used to predict the pdf $\Pr(RUL_i \mid \mathbf{z}_{1:i})$ of the equipment RUL at time t_i. The prediction is based on the following information: a sequence of observations $\mathbf{z}_{1:i}$ related to the equipment degradation at times t_1, t_2, …, t_i, the (observation) equation describing the relation between $\mathbf{z}_i$ and the degradation level d_i at time t_i, the failure threshold d_{th}, and the (stochastic) model of the equipment degradation dynamics, e.g., described by a first-order Markov process:

$$d_j = g(d_{j-1}, \boldsymbol{\gamma}_{j-1});\ d_0 \sim \Pr(d_0),\ j = 1,2,... \tag{5}$$

where $\Pr(d_0)$ is the initial distribution of the degradation at time t_0, g is the possibly non-linear state transition function and $\boldsymbol{\gamma}_j$ is the noise vector.

The estimation of the probability distribution $\Pr(d_i \mid \mathbf{z}_{1:i})$ of the degradation d_i at time t_i given the set of observations $\mathbf{z}_{1:i}$ is obtained by a recursive computational procedure divided into successive prediction and update stages [12]. In the prediction stage, supposing that the probability distribution function (pdf) $\Pr(d_{i-1} \mid \mathbf{z}_{1:i-1})$ at time t_{i-1} is available, the transition probability distribution $\Pr(d_i \mid d_{i-1})$ derived from the model in eq. (5) is used to obtain the prior pdf of the degradation state $\Pr(d_i \mid \mathbf{z}_{1:i-1})$ at time step t_i via the Chapman–Kolmogorov equation [12]. In the update stage, the posterior distribution $\Pr(d_i \mid \mathbf{z}_{1:i})$ is obtained using the incoming measurement $\mathbf{z}_i$ to update the prior distribution via the Bayes rule, based on the likelihood function $\Pr(\mathbf{z}_i \mid d_i)$ defined by the observation equation [12]. The updated posterior probability distribution $\Pr(RUL_i \mid \mathbf{z}_{1:i})$ can then be computed as the probability $\Pr(RUL_i \mid \mathbf{z}_{1:i}) = \Pr[d(t_i + RUL_i) > d_{th} \mid \mathbf{z}_{1:i}]$ that the degradation level at time $t_i + RUL_i$ exceeds the failure threshold d_{th} [20-21,31].

The recursive computation of the posterior $\Pr(d_i \mid \mathbf{z}_{1:i})$ involves an integral which in practical cases does not have a closed-form solution. For this reason, approximated solutions have been proposed, like the Extended Kalman Filter, the Gaussian sum filter, and grid-based methods [32-33]. Also, Monte Carlo sampling techniques have become of increasing interest. Among these, particle filtering provides a solution by approximating the integrals in the Bayesian recursive procedure with weighted summations over a high number of samples called particles [10,12].

The application of the particle filtering procedure to the estimation of $\Pr(RUL_i \mid \mathbf{z}_{1:i})$ is detailed in the pseudo-code given in Figure 1. The P particles $p = 1,...,P$ are future degradation trajectories built by recursively sampling the particle degradation state d_j^p at time t_j from the transition probability distribution $\Pr(d_j^p \mid d_{j-1}^p)$ derived from the degradation model, until the failure threshold d_{th} is exceeded and the length of life L^p of the particle is recorded. The value rul_i^p of the particle RUL at time step t_i can then be computed from $rul_i^p = L^p - t_i$. When an observation $\mathbf{z}_i$ is collected, each particle is assigned a weight w_i^p proportional to the likelihood $\Pr(\mathbf{z}_i \mid d_i^p)$ of observing $\mathbf{z}_i$ given the degradation level d_i^p reached by the particle at the time t_i [20]. The distribution $\Pr(RUL_i \mid \mathbf{z}_{1:i})$ is then approximated by an histogram of the P weighted values rul_i^p of the particle RULs at time t_i; the weighted average and the weighted standard deviation of the values rul_i^p, $p = 1,...,P$ represent the prediction $\hat{rul}_i$ of the expected value μ_{RUL_i} of RUL_i and the estimate $\hat{\sigma}^2_{\hat{rul}_i}$ of the prediction error variance $\sigma^2_{rul_i}$, respectively.

The sampling importance resampling (SIR) algorithm is used to avoid the degeneracy problem of the particle filtering algorithm, which consists in having all but one of the importance weights close to zero after several weight updates [12]. This algorithm requires sampling, after one or more updates of the particle weights, a new set of particles from the old one with probability for a particle to be sampled proportional to its weight (see pseudo-code in Figure 2). New degradation trajectories have to be sampled starting from the degradation state d_i^p of each particle resampled at the observation time t_i and new values of the particles duration of life L^p are recorded. For more details the interested reader may refer to the specialized literature (e.g., [10,12]).

FOR p=1:P

1. Sample $d_0 \sim \Pr(d_0)$

2. j=0; $t_j = t_0$

WHILE $d_j < d_{th}$

3. j=j+1; $t_j = t_{j-1} + \Delta t$

4. sample $d_j^p \sim \Pr(d_j^p \mid d_{j-1}^p)$

END WHILE

5. Register the particle failure time $t_f^p = t_j$

END FOR

FOR i=1:N

6. Collect the observation $\mathbf{z}_i$

FOR p=1:P

7. Compute the RUL of the particle $rul_i^p = t_f^p - t_i$

8. Compute the weights:

IF $rul_i^p < 0$ **set** $w_i^p = 0$

ELSE $w_i^p = w_{i-1}^p \cdot \Pr(\mathbf{z}_i \mid d_i^p)$

END FOR

9. Normalize the weights $wn_i^p = w_i^p \Big/ \sum_{k=1}^{P} w_i^k$

10. Build the probability density function of the equipment RUL at time t_i **as the histogram of the *P* weighted particle RULs** rul_i^p **.**

11. Compute $\hat{rul}_i = \text{Mean}(rul_i^p)$

12. Compute $\hat{\sigma}^2_{\hat{rul}_i} = \text{Var}(rul_i^p)$

END FOR

Figure 1: Particle filtering operative procedure for RUL estimation [8].

At time t_i

1. Compute rul_i^p **and** wn_i^p **as in Figure 1**

FOR p=1:P

2. Sample a particle *k* with probability equal to its weight wn_i^k

3. set $d_i^p = d_i^k$ **, j=i and** $t_j = t_i$

WHILE $d_j < d_{th}$

4. j=j+1; $t_j = t_{j-1} + \Delta t$

5. sample $d_j^p \sim \Pr(d_j^p \mid d_{j-1}^p)$

END WHILE

6. Register the new particle failure time $t_f^p = t_j$

END FOR

7. Assign equal weights to each particle $wn_i^{1:P} = 1/P$

8. Build the RUL probability density function and compute its moments $\hat{rul}_i$ **and** $\hat{\sigma}^2_{\hat{rul}}$ **as in Figure 1.**

Figure 2: Procedure for performing resampling at time t_i *[8].*

Notice that by this approach the distribution $\Pr(RUL_i \mid \mathbf{z}_{1:i})$ is estimated, which is different from the distribution $\Pr(RUL_i \mid d_i)$ of the equipment RUL at time t_i, given that the equipment has degradation d_i at that time. However, in the Bayesian framework, this is the maximum information we can have on RUL_i.

As for the uncertainty in the RUL prediction, in this approach the randomness of the degradation process (source of uncertainty A) is described by the model, whereas the observation equation accounts for the observation noise (source of uncertainty C). Thus, these two causes of uncertainty are accounted for in the RUL prediction through the procedure of particle sampling and weights updating, respectively. On the contrary, the contribution of model uncertainty to the RUL prediction uncertainty is not directly considered (source of uncertainty B), since it is assumed that the degradation dynamics model and the observation equation are exactly known. The effects of this uncertainty on the RUL prediction will be further discussed in Section 4.3.1. Notice, however, that if the uncertainty on the model parameters can be quantified and a probability distribution assigned to the value of the uncertain model parameter, the PF approach can be adjusted to handle also this source of uncertainty [34].

4.2 Approach 2: bootstrapped ensemble of empirical models trained on sequences of degradation observations and life time data

Approach 2 is based on the development of an empirical model f representing the relationship between the degradation observations $\mathbf{z}_i$ available at time t_i and the corresponding equipment RUL. The empirical model is built considering the observations of a set of S trajectories $\mathbf{z}^s_{1:N_s}$, s=1,…,S, of similar equipments which have each reached failure in a time L^s discretized in N_s steps. The empirical model receives in input the observations $\mathbf{z}_i$ and produces as output the RUL prediction, $\hat{rul}_i$. In order to develop the model, a dataset of input/output pairs $\mathbf{D}_{i/o} = \{(\mathbf{z}_i^s; rul_i^s);\ s=1,...,S; i = 1:N_s\}$ is extracted from the set of observations $\mathbf{z}^s_{1:N_s}$ by associating to the observations $\mathbf{z}_i^s$ at time t_i along the s-th trajectory to failure the corresponding realization of RUL_i, i.e., $rul_i^s = L^s - t_i$. The dataset $\mathbf{D}_{i/o}$ can then be used to train an empirical model based on one among the many data-driven modeling methods existing today (e.g., polynomial regression, non-parametric regression, neural networks, etc.). In their basic form, these methods provide in output a point prediction $\hat{rul}_i$ of the RUL without any information on the uncertainty of the estimate [35]. To overcome this limitation, the bootstrap method for estimating the accuracy in the prediction of a stochastic output whose mean value and variance are unknown functions of the input is used in this work. Under the hypothesis that the model $f(\mathbf{z}_i) = \hat{rul}_i$ is as an unbiased estimator of $\mu_{RUL_i|\mathbf{z}_i}$, i.e., $E[f(\mathbf{z}_i)] = \mu_{RUL_i|\mathbf{z}_i}$, the model error variance σ_B^2 can be rewritten as follows [24,36]:

$$\begin{aligned}\sigma_B^2(\mathbf{z}_i) &= E\left[\left(f(\mathbf{z}_i) - \mu_{RUL_i|\mathbf{z}_i}\right)^2\right] \\ &= E\left[\left(f(\mathbf{z}_i) - E[f(\mathbf{z}_i)]\right)^2\right] + E\left[\left(E[f(\mathbf{z}_i)] - \mu_{RUL_i|\mathbf{z}_i}\right)^2\right] \\ &= E\left[\left(f(\mathbf{z}_i) - E[f(\mathbf{z}_i)]\right)^2\right]\end{aligned} \tag{6}$$

An estimate of the model error variance, σ_B^2, is then obtained from an ensemble of models $f^h(\mathbf{z}_i \mid \mathbf{D}^h_{i/o})$, h=1,…,H trained using bootstrapped replicates $\mathbf{D}^h_{i/o}$ of a training dataset $\mathbf{D}^{trn}_{i/o}$, drawn from $\mathbf{D}_{i/o}$. Given a

generic input $\mathbf{z}_i$, the models of the ensemble generate H different predictions $\hat{rul}_i^h = f^h(\mathbf{z}_i \mid \mathbf{D}_{i/o}^h)$; their variance is assumed as the estimate $\hat{\sigma}_B^2(\mathbf{z}_i)$ of the model error variance $\sigma_B^2(\mathbf{z}_i)$ [24,37], whereas their average is taken as the best estimate $\hat{rul}_i$ of the equipment RUL.

With respect to the estimate of the remaining part of the RUL prediction variance, which is caused by the randomness of the degradation process and the observation noise (sources of uncertainty A and C), i.e. $\sigma_{A+C}^2 = \sigma_A^2 + \sigma_C^2$, an independent validation dataset $\mathbf{D}_{i/o}^{val}$ is used. In particular, the ensemble of empirical models $f^h(\mathbf{z}_i \mid \mathbf{D}_{i/o}^h)$ is applied to the observations in the validation dataset $\mathbf{D}_{i/o}^{val}$. The obtained RUL predictions $\hat{rul}_{1:N_s}^s$, s=S_{trn}+1,…,S are used to calculate, for each validation observation $\mathbf{z}_i^{val,s}$, the prediction residuals r_i^s:

$$r_i^s = (\hat{rul}_i^s - rul_i^s)^2 - \hat{\sigma}_B^2(\mathbf{z}_i^{val,s}) \tag{7}$$

The set of input/output pairs obtained by associating to the observations $\mathbf{z}_i^{val,s}$, $s = 1,...,S_{val}$, i=1,…,N_s, in $\mathbf{D}_{i/o}^{val}$ the corresponding residuals r_i^s is used for training an empirical model $\chi(\mathbf{z}_i) = \hat{\sigma}_{A+C}^2(\mathbf{z}_i)$ of the residual variance approximating the unknown relation between the input $\mathbf{z}_i$ and the variance of the residuals [24,38].

When a new observation $\mathbf{z}_i$ is collected, the following procedure is applied in order to obtain the estimate $\hat{rul}_i$ of the equipment RUL and of the corresponding variance $\hat{\sigma}_{\hat{rul}_i}^2$:

- Compute the output $\hat{rul}_i^h$ of each models $f^h(\mathbf{z}_i \mid \mathbf{D}_{i/o}^h)$ of the ensemble;
- Compute the point estimate of the RUL:

$$\hat{rul}_i = \frac{1}{H}\sum_{b=1}^{H} f^h(\mathbf{z}_i \mid \mathbf{D}_{i/o}^h) \tag{8}$$

- Compute the RUL prediction uncertainty as follows:

$$\hat{\sigma}_{\hat{rul}_i}^2 = \hat{\sigma}_B^2(\mathbf{z}_i) + \hat{\sigma}_{A+C}^2(\mathbf{z}_i) = \mathrm{var}[f^h(\mathbf{z}_i \mid \mathbf{D}_{i/o}^h)] + \chi(\mathbf{z}_i) \tag{9}$$

Then, $\hat{\sigma}_{\hat{rul}_i}^2$ accounts for all three sources of uncertainty listed in Section 1.

However, the degradation measurements depend on the entire past trajectory of degradation, which means that training and validation data taken from the same trajectory are not independent, causing an underestimation of the variance. For this reason, the validation dataset $\mathbf{D}_{i/o}^{val}$ is made by input/output pairs $(\mathbf{z}_i^s; rul_i^s)$, s=1,…,S_{val}, i=1,…,N_s, taken from trajectories different from those used in training. Furthermore, to ensure enough diversity of the models in the ensemble, the bootstrapped training datasets $\mathbf{D}_{i/o}^h$, h=1,…,H, are sampled from $\mathbf{D}_{i/o}^{trn}$ as follows: first, S_{trn} training trajectories are randomly sampled with replacement from the S_{trn} different trajectories of $\mathbf{D}_{i/o}^{trn}$; then, N^{trn} input/output pairs $(\mathbf{z}_i^s; rul_i^s)$ are sampled with replacement from the total amount of input/output pairs in $\mathbf{D}_{i/o}^{trn}$.

4.3 Approach 3: Bootstrapped ensemble of empirical models trained on a sequence of past degradation observations from the equipment of interest only

Approach 3 is based on the development of an empirical model of the degradation process based on the time series of its past observations $\mathbf{z}_{1:i}$ and used for identifying the time at which the degradation will exceed the failure threshold. For simplicity of illustration the observations $\mathbf{z}_{1:i}$ are assumed to be direct measures of the degradation $d_{1:i}$, eventually affected by noise.

The approach differs from approach 1 in that the stochastic model describing the dynamics of the degradation process is not available and actually the point is to develop it empirically. Coherently, the estimate of the prediction error variance $\sigma^2_{\hat{rul}_i}(\mathbf{z}_i)$ should account also for the error of approximation of the empirical model.

The approach differs from approach 2 in that there are no available pairs $(\mathbf{z}_i; rul_i)$ for which rul_i is known for training and validating the prognostic model $f(\mathbf{z}_i \mid \mathbf{D}_{i/o})$.

Empirical modeling of the degradation process can be achieved by fitting the most suited degradation model, e.g., linear and non-linear regression models, general degradation path models, etc. [39] to the available data. Let us call $\hat{d}_j = \eta(t_j)$ a generic model of the equipment degradation, derived from the sequence of data $\mathbf{z}_{1:i}$. The prediction $\hat{rul}_i$ of the equipment RUL at time t_i can be simply obtained from the relation $\eta(t_i + \hat{rul}_i) = d_{th}$. Once again an estimate of the prediction error variance $\sigma^2_{\hat{rul}_i}(\mathbf{z}_i)$ is needed, but cannot be obtained by means of the method proposed for approach 2 since there are no available pairs $(\mathbf{z}_i; rul_i)$ for which rul_i is known, and thus eq. (7) cannot be used to calculate the value of the prediction residual r_i in correspondence of the observation $\mathbf{z}_i$.

Let us, instead, consider a model $\Delta\hat{t}_{j,j'} = \tilde{\eta}(\mathbf{d}_{j,j'})$ receiving in input a vector of two degradation values $\mathbf{d}_{j,j'} = [d_j, d_{j'}]$ and returning in output the estimate $\Delta\hat{t}_{j,j'}$ of the time interval needed to reach degradation d_j starting from $d_{j'}$. Notice that model $\tilde{\eta}(\mathbf{d}_{j,j'})$ can be derived, in general, from model $\eta(t_j)$:

$$\Delta t_{j,j'} = \tilde{\eta}(\mathbf{d}_{j,j'}) = \eta^{-1}(d_{j'}) - \eta^{-1}(d_j) \tag{10}$$

The prediction $\hat{rul}_i$ is then obtained from this model by setting $d_j = \mathbf{z}_i$ and $d_{j'} = d_{th}$, which means that the RUL prediction at time t_i corresponds to the estimate of the time interval $\Delta t_{i,th}$ needed to increase the degradation from d_i to the failure threshold d_{th}.

The relation between the input $\mathbf{d}_{j,j'} = [d_j, d_{j'}]$ and the output $\Delta t_{j,j'}$ of the model in eq. (10) is:

$$\Delta t_{j,j'} = \mu_{\Delta t_{j,j'}} + \delta_{\Delta t_{j,j'}} \tag{11}$$

where $\delta_{\Delta t_{j,j'}} = \Delta t_{j,j'} - \mu_{\Delta t_{j,j'}}$ is a zero mean random variable with variance $\sigma^2_{\Delta t_{j,j'}}$ representing the uncertainty in the evolution of the degradation process from d_j to $d_{j'}$ and $\mu_{\Delta t_{j,j'}}$ is the mean value of the random variable $\Delta t_{j,j'}$. Both $\sigma^2_{\Delta t_{j,j'}}$ and $\mu_{\Delta t_{j,j'}}$ are, in general, functions of the input $\mathbf{d}_{j,j'}$.

The observations $\mathbf{z}_{1:i}$ are used to build input/output pairs $(\mathbf{d}_{j,j'} = [\mathbf{z}_j, \mathbf{z}_{j'}]; \Delta t_{j,j'} = t_{j'} - t_j)$, $j = 1,...,i-1$; $j' = j+1,...,i$ and the bootstrapping of approach 2 can be applied to estimate the variance of the prediction error of the model in eq. (10) by building training and validation datasets of input/output pairs. As underlined in Section 3.2, to avoid underestimating the prediction error, the validation datasets should not contain measurements belonging to degradation trajectories used for training. Since only a single trajectory is now available, the solution proposed is to partition the dataset $\mathbf{D}$ into two sequences of consecutive measurements, $\mathbf{D}^{trn} = \{\mathbf{z}_{1:N_{trn}}\}$ and $\mathbf{D}^{val} = \{\mathbf{z}_{N_{trn}+1:i}\}$ and to use $\mathbf{D}^{trn}$ for building the model $\hat{d} = \eta(t)$ and $\mathbf{D}^{val}$ for building the dataset of input/output pairs $\mathbf{D}^{val}_{i/o} = \{(\mathbf{d}_{j,j'}; \Delta t_{j,j'} = t_j - t_{j'})\}$, $j = N_{trn}+1,...,i-1$; $j' = j+1,...,i$ to be used for estimating the prediction error.

An ensemble of models $\eta^h(t \mid \mathbf{D}^h)$, h=1,…,H, is trained using bootstrap replicates $\mathbf{D}^h$ of the training dataset $\mathbf{D}^{trn}$, and the ensemble of models $\tilde{\eta}^h(\mathbf{d}_{j,j'} \mid \mathbf{D}^h)$ is derived from eq. (10). The average and variance of the ensemble model prediction are retained as the estimates $\Delta\hat{t}_{j,j'}$ and $\hat{\sigma}^2_B(\mathbf{d}_{j,j'})$ of, respectively, the time interval $\Delta t_{j,j'}$ and the error variance of model $\tilde{\eta}(\mathbf{d}_{j,j'} \mid \mathbf{D})$. The ensemble of models is applied to the validation dataset $\mathbf{D}^{val}_{i/o}$ in order to obtain a set of prediction residuals $r_{j,j'}$:

$$r_{j,j'} = (\Delta\hat{t}_{j,j'} - \Delta t_{j,j'})^2 - \hat{\sigma}^2_B(\mathbf{d}_{j,j'}) \tag{12}$$

Finally, an empirical model $\chi(\mathbf{d}_{j,j'}) = \hat{\sigma}^2_{A+C}(\mathbf{d}_{j,j'})$, estimating the part of the error variance $\sigma^2_{\Delta t_{j,j'}}$ due to the stochasticity of the degradation process and the observation noise (sources of uncertainty A and C), is trained using the input/output pairs $(\mathbf{d}_{j,j'}; r_{j,j'})$. The sum of the RUL and noise variance equipments, $\hat{\sigma}^2_{A+C}(\mathbf{z}_i)$, is then obtained from this model by setting $\mathbf{d}_{j,j'} = \mathbf{d}_{i,th} = [\mathbf{z}_i, d_{th}]$.

Notice that the training data $(\mathbf{d}_{j,j'}; r_{j,j'})$ used to build the models cover a range of values for the input $\mathbf{d}_{j,j'}$ in general different from that of the input $\mathbf{d}_{i,th} = [\mathbf{z}_i, d_{th}]$ to which the model is applied to obtain the estimate $\hat{\sigma}^2_{A+C}(\mathbf{z}_i)$. This can represents a limit to the quality of the estimate $\hat{\sigma}^2_{A+C}(\mathbf{z}_i)$, since in general the performance of empirical models are good when applied to input regions well described by training data, and decrease moving away from these regions.

When a new observation $\mathbf{z}_i$ of the degradation d_i is collected at time t_i, the multiple RUL predictions $\hat{rul}^h_i = \tilde{\eta}^h[\mathbf{d}_{i,th} \mid \mathbf{D}^h]$ and the RUL variance estimate $\hat{\sigma}^2_{A+C}(\mathbf{z}_i) = \chi(\mathbf{d}_{i,th})$ are used to obtain the prediction $\hat{rul}_i$ and the relative estimate $\sigma^2_{\hat{rul}}$ of the prediction error variance:

$$\hat{rul}_i = \frac{1}{H}\sum_{h=1}^{H} \hat{rul}^h_i \tag{13}$$

$$\hat{\sigma}^2_{\hat{rul}_i} = \hat{\sigma}^2_B(\mathbf{z}_i) + \hat{\sigma}^2_{A+C}(\mathbf{z}_i) = \mathrm{var}[\tilde{\eta}^h(\mathbf{d}_{i,th} \mid \mathbf{D}^h)] + \chi(\mathbf{d}_{i,th}) \tag{14}$$

As for approach 2, all three sources of uncertainty listed in Section 1 are taken into account in the estimate $\hat{\sigma}^2_{\hat{rul}_i}$.

5 Numerical application

The three different approaches presented in Section 4 are verified with respect to the RUL prediction of a simulated turbine blade undergoing degradation. The application focuses on the turbine of a generation IV high temperature gas reactor, which is characterized by rather extreme turbine operational conditions such as working temperatures exceeding 900°C. The predominant damage mechanisms affecting turbines operating at such elevated temperatures include creep deformation, corrosion and fatigue [40]. The interaction of these and other mechanisms generates a degradation process that leads to crack initiation which rapidly leads to failure due to the quick accumulation of stress cycles caused by the high rotational speed. Notice that a turbine undergoing this degradation process can experience the loss of its blades, one of the most feared failure modes of turbomachinery since it is accompanied by abrupt changes in the power conversion equipment and in the reactor flow conditions [41]. Figure 3 shows an example of high-pressure turbine deblading occurred in a German power plant [42]. Also, fracture in rotary machines can result in turbine missiles, i.e., irregularly shaped projectiles travelling at high velocities which can impact on barriers in nuclear power plants causing severe damages to the facilities, and threatening public safety [43]. This and the high cost of turbine blade replacement are strong reasons for performing prognostics on creeping turbine blades.

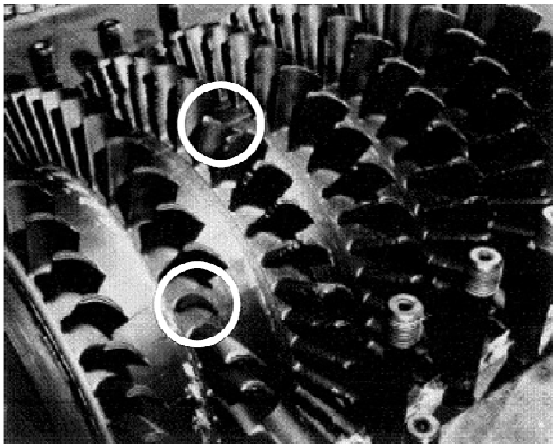

Figure 3: Deblading in a high pressure turbine [42].

Indeed, health monitoring and prognostics may allow scheduling maintenance actions before the blade degradation evolves into cracks. In this context, the dimensionless quantity ε, defined as the percentage of elongation of the turbine blade in the longitudinal direction with respect to its original length, can be used as an indicator of the blade degradation state [44]. The blade is discarded when the accumulated elongation reaches a pre-determined value, namely the failure threshold, which assures that the risk of blade failure is below the desired safety limit. In [45] various sensing technologies for measuring blade deformations are outlined and their possible application to turbine blade and disk health monitoring are discussed. In particular, methods measuring blade tip clearance (BTP), which defines the gap between the tip of blades and the casing [45], appear to be promising for measuring blade plastic elongation. In this respect, the main technologies available today are based on capacitive measurements, eddy current or microwave sensing. Capacitive-based technologies [46-47] are already on the market [48-49] and provide clearance and vibration

monitoring sensors to gas turbines used in the power generation and aerospace industries. In [50], a magnetic sensor relying on high-frequency eddy currents actively induced in passing blades, is tested on field trials with jet engines and it is demonstrated the capability of generating online clearance measurement for each blade. Similar results are obtained in [51] using a microwave sensing system, which is claimed to be capable of performing blade monitoring in the harsh environment of the first turbine stages.

In [45] it is observed that long-term trends in BTP can be measured and its future application to monitoring creep-related blade deformation is anticipated. In [44], the possibility of using BTP in blade failure risk analysis and diagnosis is analyzed and a blade prognostic approach based on BTP linear regression is proposed. In [50] a general framework for PHM of turbine blades is proposed, considering blade tip clearance and other damage indicators such as vibrations, blade angular position, etc.

Notice that, given the advancements in blade tip clearance sensing technology, a sufficient amount of data for training and validation of prognostic models can be expected to be available in the future. Furthermore, notice that the amount of real data necessary to validate the prognostic approaches developed can be largely reduced by resorting to the leave-one-out cross-validation procedure [52].

5.1 Degradation model

Modeling the degradation of a turbine blade is a hard task, especially if one needs to take into account all mechanisms involved and their interactions. For the purpose of this work, we limit ourselves to considering the accumulation of creep damage. Creep is an irreversible deformation process affecting materials exposed to a load below their elastic limit for a protracted length of time and at high temperatures. In the high pressure turbine first stage, blades creep is a major problem due to the high operational temperatures, and is often the life-limiting process [53]. Blade elongation, ε, is taken as a measure of the blade creep strain.

In this work, the creep evolution is modeled using the Norton Law discretized with a step $\Delta t = 5$ days, assuming that the dependence from the temperature follows the Arrhenius law [27,54]:

$$\varepsilon_{j+1} = \varepsilon_j + A \cdot \exp\left(-\frac{Q}{RT_j}\right) \cdot \left[K\omega_j{}^2 + \delta\varphi_j\right]^n \cdot \Delta t\,, \qquad \varepsilon_0 = 0 \tag{15}$$

where ε_j is the creep strain at time t_j, Q is the activation energy, A and n are material inherent characteristics varying from one blade to another, K is a constant relating the load to the rotational speed ω_j, R is the ideal gas constant, T_j is the blade operating temperature and $\delta\varphi_j$ is a random variable modeling the fluctuations in the stress applied to a specific blade, which are due to fabrication defects, aging and corrosion of the blade, vibrations of the equipment or turbulences of the gas flow. Oscillations of the rotational speed ω_j, and of the blade operating temperature T_j, are represented by considering their deviations from the mean values μ_ω and μ_T as noises. In practice, $\delta\omega_j = \omega_j - \mu_\omega$ and $\delta T_j = T_j - \mu_T$ are Gaussian random variables with mean values zero and assigned standard deviations. Thus, the noise vector $\boldsymbol{\gamma}_j$ in eq. (5) can be set equal to $\boldsymbol{\gamma}_j = [\delta\omega_j; \delta T_j; \delta\varphi_j]$.

The values of the parameters T_j, ω_j and K have been set with reference to the helium gas turbine of a Gas Turbine Modular Helium Reactor (GT-MHR) developed by an international consortium, with a targeted 286MWe generation per module [26]; the material inherent characteristics A and n are taken assuming that the blade is made of Ni-base cast Superalloy 713LC [26]. The distributions used for the parameters are reported in Table II.

Table II: type of distribution, mean value and standard deviation used for the creep growth model parameters

Variable	Symbol	Distribution	Units	Parameters of the distribution
Activation energy	Q	Deterministic	kJ/mol	Q=290
Norton Law parameters	A	Normal	$(N/m^2)^{-n}/day$	$\mu_A=7.2\cdot10^{-3}$; σ_A=5%
	n	Normal	-	μ_n=6; σ_n=0.2%
Operating temperature	T_j	Normal	K	μ_T=1100; σ_T=1%
Rotational speed	ω_j	Normal	rpm	μ_ω=3000; σ_ω=1%
Load parameter	K	Deterministic	Kg/m	ρ=1068
Stress fluctuations	$\delta\varphi$	Gamma	MPa	θ=2; k=10

Eq. (15) represents a stochastic process whose unknowable future evolution (cause A, Section 1) produces an irreducible uncertainty in the RUL prediction. Parameters A and n instead represent an uncertainty in the model (source B, Section 1). In fact, to a specific blade correspond fixed parameters A and n but their exact values are not known in practice; to include this source of uncertainty in the model, we assume to know with a certain precision the range of values of these parameters and associate to them a probability distribution (Table II).

For simplicity, it is assumed that it is possible to directly measure at inspection time t_j the value of the creep strain ε_j. Thus, the observation equation is:

$$\mathbf{z}_j = \varepsilon_j + \upsilon_j \tag{16}$$

where υ_j is a white Gaussian measurement noise with standard deviation σ_υ=0.02. Then, the likelihood $\Pr(\mathbf{z}_j \mid d_j^p)$ used in the particle filtering approach is Gaussian with mean d_j^p and variance σ_υ^2. This noise represents a source of uncertainty (source C, Section 1) in the final RUL prediction.

The failure threshold for creep strain ε_{th} is set equal to the value of 1.5%.

Given the unavailability of real experimental data, a sequence of creep strain measurements $\mathbf{z}_{1:i}$ on the blade of interest, hereafter called 'test trajectory', is simulated using eq. (15). The variation in time of the rotational speed ω, the gas temperature T and the stress fluctuations $\delta\varphi$ are simulated by sampling their values ω_j, T_j and $\delta\varphi_j$ from the relative distributions (Table II) at each time instant t_j. Every 30 days a measurement $\mathbf{z}_j$, corresponding to the creep strain ε_j, is simulated by using eq. (16). A total number of 87 creep strain measurements have been simulated for a turbine blade with parameters $A=3\cdot10^{-4}$ and n=6.

In order to verify the performance of the prognostic approaches, the simulation of the test trajectory has been conducted until the time L at which the creep strain reaches the failure threshold. The difference between L and the time t_i at which the prognosis is performed is the RUL of the turbine blade; it will be referred to as "true RUL", and represented by the notation rul_i .

Also, a number S=13 of historical creep growth trajectories of similar blades have been simulated using eq. (15). To induce variability in the behavior of the similar blades the values of the characteristic parameters A and n from one blade to another have been sampled from normal distributions (Table II) at the beginning of each new simulated degradation trajectory. Some examples of simulated creep growth trajectory are shown in Figure 4.

For each trajectory, a sequence of N_s direct creep strain measurement $\mathbf{z}_{1:N_s}$, one every 30 days, are simulated according to eq. (16).

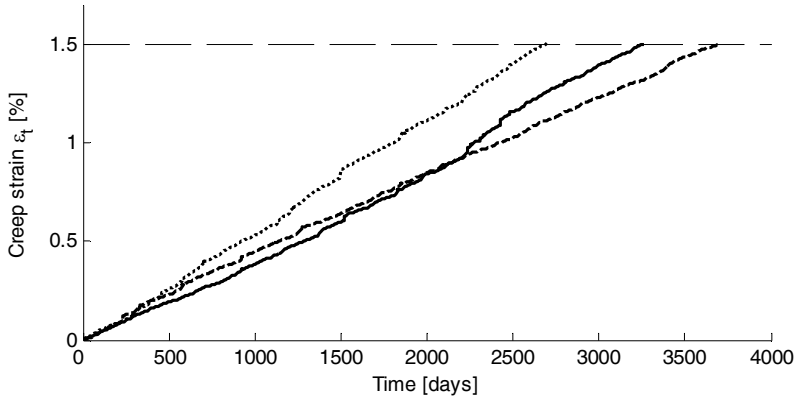

Figure 4: Examples of creep growth trajectories.

5.2 Results

During the life of the turbine blade, at every time t_i , the set of observations $\mathbf{z}_{1:i}$ is assumed to be available; the objective of the analysis is to predict at time t_i , i=1,…,87, the RUL distribution for the test trajectory. Three situations have been artificially constructed for the turbine blade case study described in the previous Section 4.1, corresponding to the three prognostic approaches of Section 3.

In the PF approach 1, the model of eq. (15) is used to simulate a number $P = 1000$ of particles starting from $\varepsilon_0 = 0$. Particle resampling is performed once every 5 measurements. Note that the particle filter has been preferred to the Kalman filter since the distribution of the process noise is not Gaussian as a consequence of the combination of speed, temperature and stress fluctuations in the creep growth process described by eq. (15).

In the bootstrapped ensemble approach 2, which uses multiple sequences of degradation observations, S_{trn}=10 trajectories among the S = 13 totally available are used for building an ensemble of H=25 linear least square models $\hat{rul}^h(\mathbf{z}_i \mid D^h_{i/o}) = \alpha^h_0 + \alpha^h_1 \cdot \mathbf{z}_i$, whereas the remaining S_{val}=3 trajectories are used to validate the ensemble and build the training dataset for the least squares model $\chi(\mathbf{z}_i) = \gamma_0 + \gamma_1 \cdot \mathbf{z}_i + \gamma_2 \cdot \mathbf{z}_i^2$ estimating the RUL variance σ^2_{A+C} .

In the bootstrapped ensemble approach 3, which uses the time series of degradation observations, the prognostic model has been developed only after time t_{30} in order to have available a dataset $\mathbf{D} = \{\mathbf{z}_{1:i}\}$ of at least i=30 direct creep strain measurements. This dataset has been partitioned into a training dataset $\mathbf{D}^{trn}$ containing the first 75% of the available measurements and a validation dataset D^{val} containing the remaining 25%. An ensemble of H=25 linear least square models $\eta^h(t_j \mid \mathbf{D}^h) = \beta_0^h + \beta_1^h \cdot t_j$ is built and the models $\tilde{\eta}^h[\Delta\mathbf{z}_{j,j'} \mid \mathbf{D}^h] = \Delta\mathbf{z}_{j,j'} / \beta_1^h$ are derived from it. Notice that in a linear process, the time needed to increase the degradation level from ε_j to $\varepsilon_{j'}$ is proportional to the degradation increment $\Delta\varepsilon_{j,j'} = \varepsilon_{j'} - \varepsilon_j$ and does not depend on the initial and final degradation values. The ensemble of models is tested on the validation dataset made of input/output pairs $\mathbf{D}_{i/o}^{val} = \left\{\Delta\mathbf{z}_{j,j'} = \mathbf{z}_{j'} - \mathbf{z}_j ; \Delta t_{j,j'} = t_{j'} - t_j\right\}$, j=N_{trn}+1,…,i-1, j'=j+1,…,i and the prediction residuals $r_{j,j'}$ obtained are used to train the linear model $\chi(\Delta\mathbf{z}_{j,j'}) = \lambda\Delta\mathbf{z}_{j,j'}$ for the variance of Δt. After time t_{30}, each time t_i, i=31,…,87, a new measurement becomes available, a new ensemble of models $\tilde{\eta}^h$ and a new model χ for the prediction error variance are built. The predictions $\hat{rul}_i^h$ and the estimate $\hat{\sigma}_{A+C}^2$ are obtained respectively from the models $\tilde{\eta}^h$ of the ensemble and from model χ in correspondence of the input $\Delta\mathbf{z}_{i,th} = \varepsilon_{th} - \mathbf{z}_i$. Since the data used for training model χ concern creep strain increments which for the first two thirds of the trajectory are smaller than the increment $\Delta\mathbf{z}_{i,th} = \varepsilon_{th} - \mathbf{z}_i$ considered for obtaining the prognostic results, the empirical model $\chi(\Delta\mathbf{z}_{j,j'})$ is used in an input region not described by the training data.

In all three approaches, in correspondence of each prediction $\hat{rul}_i$ we estimate the prediction interval $[C_i^{\inf}(\alpha); C_i^{\sup}(\alpha)]$, i.e., the interval expected to contain the true RUL value rul_i with a probability of 1-α. This interval is obtained as follows:

- In approach 1, $C_i^{\inf}(\alpha)$ and $C_i^{\sup}(\alpha)$ are the $\alpha/2$ and $1-\alpha/2$ percentiles, respectively, of the RUL distribution estimated by Particle Filtering.
- In approaches 2 and 3, assuming that the prediction error has a Gaussian distribution, the value of $C_i^{\inf}(\alpha)$ and $C_i^{\sup}(\alpha)$ can be computed according to the theory of the bootstrap method [24] as:

$$C_i^{\inf}(\alpha) = \hat{rul}_i - c_{conf}^{\alpha}\,\hat{\sigma}_{\hat{rul}_i} \text{ and } C_i^{\sup}(\alpha) = \hat{rul}_i + c_{conf}^{\alpha}\,\hat{\sigma}_{\hat{rul}_i} \tag{17}$$

  where c_{conf}^{α} is the $1-\alpha/2$ percentile of a Student's t-distribution with number of degrees of freedom equal to the number H of bootstrap models.

Figure 5 shows the evolution of the true value rul_i of the blade RUL (continuous thick line), its estimated value $\hat{rul}_i$ (dots) and the corresponding prediction interval for $\alpha = 0.32$ (continuous thin line) obtained during the turbine blade life at times t_i, i=1,…,87, by the three prognostic approaches.

The prediction intervals provided by approach 3 are characterized by large oscillations and low accuracy, especially at the beginning of the trajectory, i.e., when few training data are available. Furthermore, the RUL prediction itself is noisy. This effect can be reduced by properly filtering the predictions. To this purpose, since the time evolution of the RUL is a linear process ($rul(t) = rul(t-1) - 1$), and assuming a Gaussian noise of the prediction, Kalman filtering can be applied [27].

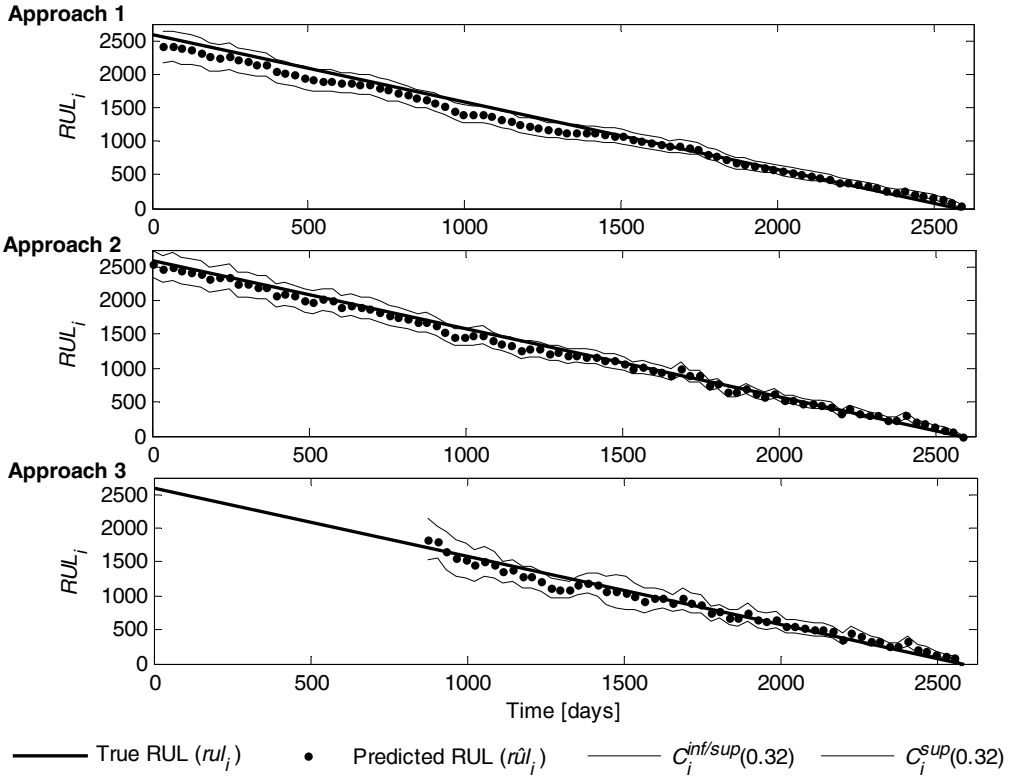

Figure 5: true rul_i (continuous thick line) of a turbine blade, predicted value $\hat{rul}_i$ (dots) and prediction interval $[C_i^{\inf}(0.32); C_i^{\sup}(0.32)]$ (continuous thin line) for the three prognostic approaches.

5.3 RUL distribution and prediction interval

The objective of this Section is to determine whether the estimates of the prediction intervals provided by the three approaches properly describe the uncertainty in the RUL predictions. In practice, we want to know whether the estimates $\hat{\sigma}^2_{\hat{rul}_i}$ are satisfactory approximations of the real $\sigma^2_{\hat{rul}_i}$. According to eq. (4), $\sigma^2_{\hat{rul}_i}$ can be decomposed into the sum of three terms, due, respectively, to the process randomness, σ^2_A, the model error, $\hat{\sigma}^2_B$, and the noise on the observations, σ^2_C. It is also of interest to consider the term $\sigma^2_{A+C} = \sigma^2_A + \sigma^2_C$, since all three approaches proposed do not estimate these two terms separately. The computation of the true value of $\sigma^2_{\hat{rul}_i}$ would ideally require the availability of an infinite number P of equipment degradation trajectories which at time t_i are in the degradation state ε_i. Since in the case study here considered we can artificially generate degradation trajectories, an high number P=1000 of degradation trajectories has been used to numerically approximate the variance $\sigma^2_{\hat{rul}_i}$. For the p-th simulated trajectory, we have computed: 1) its true RUL, $rul_i^p = L^P - t_i$, with L^p being the equipment life duration along the p-th trajectory, 2) the equipment RUL prediction, $\hat{rul}_i^p$, provided by the prognostic model in correspondence of the observations $\mathbf{z}_i^p = \varepsilon_i + \upsilon_i^p$ with υ_i^p a random Gaussian noise with variance σ^2_υ. Then, $\sigma^2_{\hat{rul}_i}$ has been approximated by:

$$\sigma^2_{\hat{rul}_i} = E\left[\left(\hat{rul}_i - RUL_i\right)^2\right] \cong \sum_{p=1}^{P} \frac{\left(\hat{rul}_i^p - rul_i^p\right)^2}{P} \tag{18}$$

Similarly, the computation of the true value of $\sigma_A^2 = E\left[\left(\mu_{RUL_i|\varepsilon_i^p} - RUL_i\right)^2\right]$ is approximated by:

$$\sigma_A^2 \cong \frac{1}{P}\sum_{p=1}^{P}[\mu_{RUL_i|\varepsilon_i} - rul_i^p]^2 \tag{19}$$

where the RUL expected value $\mu_{RUL_i|\varepsilon_i}$ is approximated by $\sum_{p=1}^{P}\frac{L^p - t_i}{P}$.

The real value of $\sigma_{A+C}^2 = E\left[\left(\mu_{RUL_i|\mathbf{z}_i} - RUL_i\right)^2\right]$ has been approximated by considering the P=1000 equipment degradation trajectories which at time t_i are in the degradation state ε_i and for which the observations $\mathbf{z}_i^p$ have been collected, and computing:

$$\sigma_{A+C}^2 \cong \frac{1}{P}\sum_{k=1}^{P}[\mu_{RUL_i|\mathbf{z}_i^p} - rul_i^p]^2 \tag{20}$$

where $\mu_{RUL_i|\mathbf{z}_i^p}$ has been approximated by simulating P'=1000 new degradation trajectories, each one starting from a different degradation state $\varepsilon_i^{p'} = \mathbf{z}_i^p - \upsilon_i^{p'}$. This procedure allows to propagate the uncertainty on the true degradation state given the observation $\mathbf{z}_i^p = \varepsilon_i + \upsilon_i^p$ to the RUL mean value and to approximate $\mu_{RUL_i|\mathbf{z}_i^p}$ by:

$$\mu_{RUL_i|\mathbf{z}_i^p} = \frac{1}{P'}\sum_{p'=1}^{P'} rul_i^{p'} \tag{21}$$

Notice that the terms σ_A^2 and σ_{A+C}^2 do not depend on the approach. In Figure 6 (left), the true RUL distribution $\Pr(RUL_i \mid \varepsilon_i)$, approximated by the distribution of the 1000 simulated rul_i^p, is shown at the times $t_i = t_{18} = 515$ days, $t_{46} = 1355$ days and $t_{73} = 2135$ days, whereas in Figure 6 (right) the true values of the standard deviations σ_A (continuous line) and σ_{A+C} (dots) are reported as a function of the blade creep strain level ε_i.

Differently from the variance terms σ_A^2 and σ_C^2, the model error variance term σ_B^2 depends on the modeling approach used to estimate the RUL. Considerations on σ_B^2 will be done in the following Sections 5.3.1, 5.3.2 and 5.3.3.

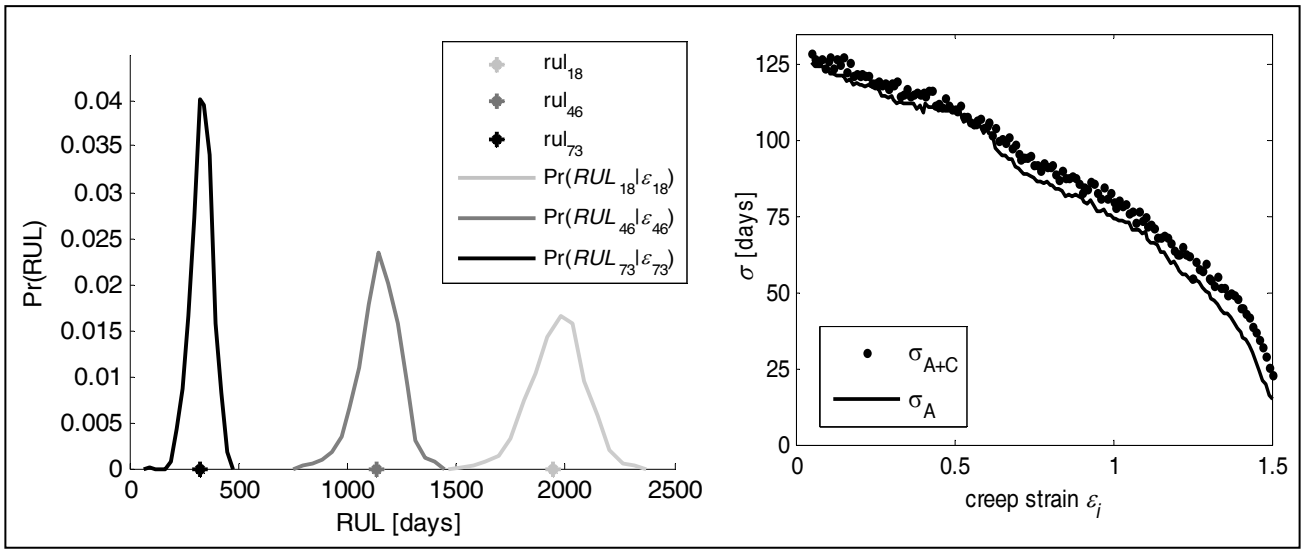

Figure 6: pdf $\Pr(RUL_i \mid \varepsilon_i)$ *of the RUL of a turbine blade at three different instances of the degradation trajectory (left) and evolution of the RUL standard deviations* σ_A *(continuous line) and* σ_{A+C} *(dots) as a function of the creep strain* ε_i *at time* t_i *.*

5.3.1 Prediction interval provided by the PF approach 1

The specific blade undergoing the creep degradation process is characterized by fixed values of the parameters A and n in eq. (15), which in general are not known. In this Section, in order to evaluate the PF performance in the estimate of the prediction uncertainty, the approach is firstly applied assuming to know the exact value of these parameters. In Figure 7, the distribution $\Pr(RUL_i \mid \mathbf{z}_{1:i})$ predicted by the particle filtering method (left, dashed line) and the estimate of the prediction error standard deviation $\hat{\sigma}_{r\hat{u}l_i}$ (right, dots), are compared to the true RUL distribution $\Pr(RUL_i \mid \varepsilon_i)$ (left, continuous line) and standard deviation σ_{A+C} (right, continuous line) of Figure 6. In can be noticed that the method supplies an accurate prediction of the RUL distribution, and correctly estimates the prediction uncertainty for all values of the creep strain ε_i.

The more realistic case where the exact values of parameters A and n are not known has then been considered. In this case, uncertainty in the prognostic model (source of uncertainty B) is introduced. The particle filtering approach 1 handles it by generating particles characterized by different values of A and n randomly sampled from the distributions of Table II. In Figure 8, the true RUL distribution $\Pr(RUL_i \mid \varepsilon_i)$ (left, continuous line) and standard deviation σ_{A+C} (right, continuous line) of Figure 6 are compared to the distribution $\Pr(RUL_i \mid \mathbf{z}_{1:i})$ provided by the method (left, dashed line) and the estimate of the standard deviation $\hat{\sigma}_{r\hat{u}l_i}$ (right, dots).

It can be noticed that the estimated prediction error standard deviation $\hat{\sigma}_{r\hat{u}l_i}$ is larger than the actual RUL standard deviation σ_{A+C}, especially for low values of ε_i, due to the model error variance σ_B^2. There are two main reasons for which the difference between $\hat{\sigma}_{r\hat{u}l_i}$ and σ_{A+C} decreases as the current creep strain gets closer to the failure threshold: *i)* the effect of the variability of the parameters A and n on the RUL distribution is lower if the gap between the degradation level ε_i and the failure threshold ε_{th} is smaller; *ii)* the SIR particle filtering method selects among the large set of particles initially created with random values of A and n those having the values of these parameters closer to those of the specific blade undergoing the creep degradation process.

Figure 9 compares the estimated prediction error standard deviation $\hat{\sigma}_{r\hat{u}l_i}$ with the true prediction error standard deviation $\sigma_{r\hat{u}l_i}$ for the PF approach 1. The results confirm that, as expected, the PF approach 1 supplies accurate estimates of the prediction error variance $\sigma^2_{r\hat{u}l_i}$, combining properly the contribution of the process stochasticity σ^2_A, the noise σ^2_C and the model error σ^2_B, described respectively by the degradation model (eq. (15)), the observation equation (eq. (6)) and the *A* and *n* parameters distributions (Table II).

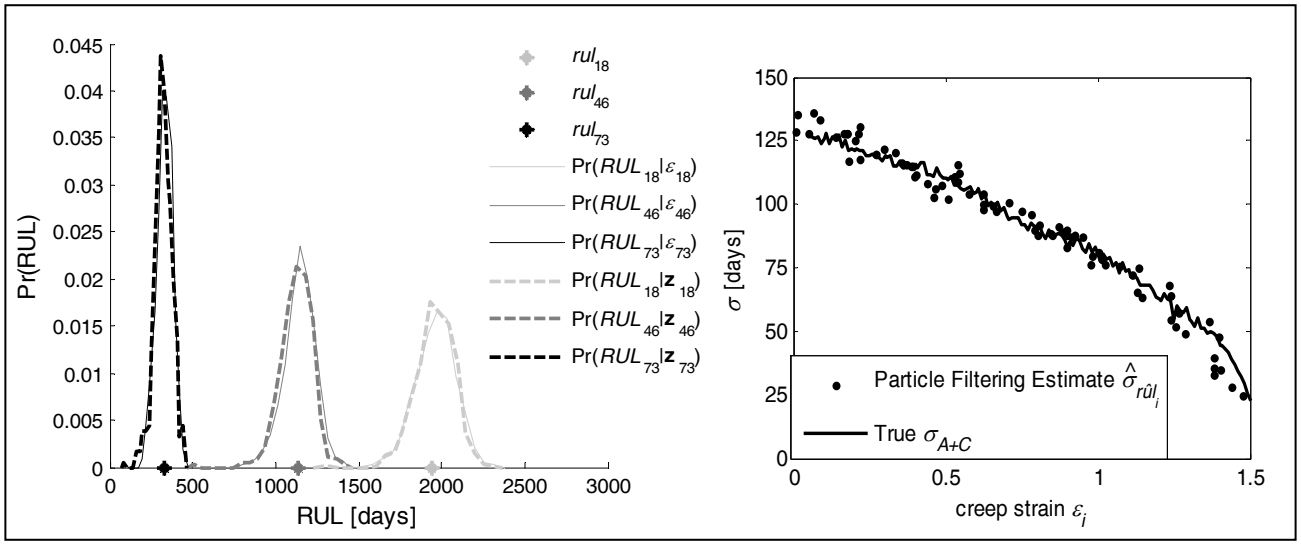

Figure 7: comparison of the pdf $\Pr(RUL_i \mid \varepsilon_i)$ *(left, continuous line) and the standard deviation* σ_{A+C} *(right, continuous line) with, respectively, the pdf* $\Pr(RUL_i \mid \mathbf{z}_{1:i})$ *(left, dashed line) and the estimated standard deviation* $\hat{\sigma}_{r\hat{u}l_i}$ *(right, dots), obtained with the PF approach 1, assuming exact knowledge of the parameters A and n.*

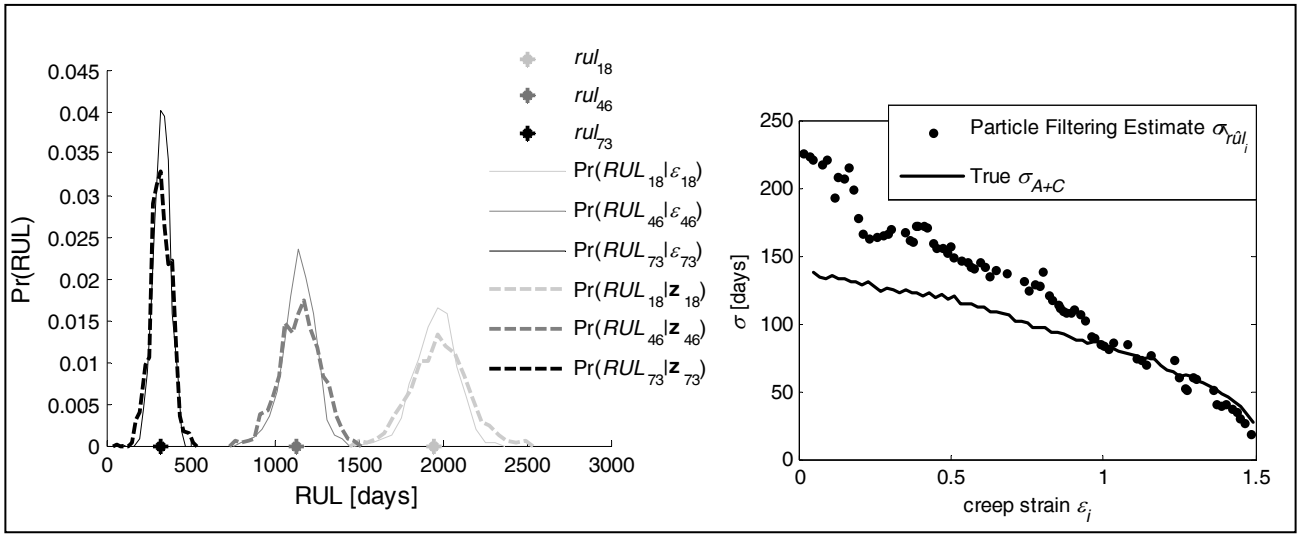

Figure 8: comparison of the pdf $\Pr(RUL_i \mid \varepsilon_i)$ *(left, continuous line) and the standard deviation* σ_{A+C} *(right, continuous line) with, respectively, the pdf* $\Pr(RUL_i \mid \mathbf{z}_{1:i})$ *(left, dashed line) and the estimated standard deviation* $\hat{\sigma}_{r\hat{u}l_i}$ *(right, dots), obtained with the PF approach 1, assuming to know only the distribution of the parameters A and n.*

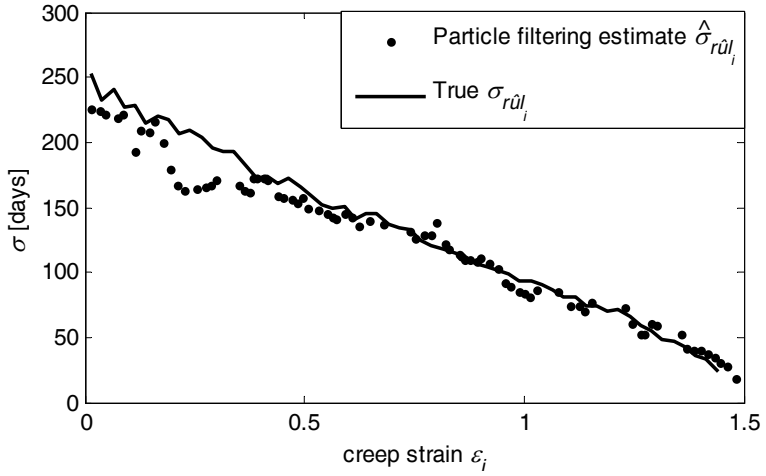

Figure 9: comparison of the estimated prediction error standard deviation $\hat{\sigma}_{r\hat{u}l_i}$ (dots), with the true prediction error standard deviation $\sigma_{r\hat{u}l_i}$ (continuous line).

5.3.2 Prediction interval provided by approach 2

In order to estimate the real uncertainty affecting the RUL prediction of an ensemble of models, it is necessary to add the real model uncertainty σ_B^2 to the noise and process randomness represented by σ_{A+C}^2 (Figure 6, right). The real model uncertainty $\sigma_B^2 = E\left[\left(\mu_{RUL_i|\mathbf{z}_i} - r\hat{u}l_i\right)^2\right]$ can be approximated by following the procedure reported in Appendix A. Basically, M different ensemble models are trained using different sets of creep growth trajectories; then, for P'' test trajectories, the observations $\mathbf{z}_i^{p''} = \varepsilon_i + \upsilon_i^{p''}$, p''=1,…,P'' are simulated and the RUL predictions $r\hat{u}l_i^{p'',m}$ collected for each ensemble model m=1,…,M; finally, the corresponding errors $\mu_{RUL_i|\mathbf{z}_i^{p''}} - r\hat{u}l_i^{p'',m}$ are computed, and the square values of these differences are averaged over the P'' trajectories and the M models to supply the numerical approximation of σ_B^2:

$$\sigma_B^2 \cong \frac{1}{P''}\frac{1}{M}\sum_{p''=1}^{P''}\sum_{m=1}^{M}[\mu_{RUL_i|\mathbf{z}_i^{p''}} - r\hat{u}l_i^{p'',m}]^2 \tag{22}$$

The continuous line in Figure 10 shows the real values of σ_{A+C} (left), σ_B (middle) and $\sigma_{r\hat{u}l_i}$ (right) during the life of a turbine blade as a function of its creep strain ε_i. It can be observed that the term $\sigma_{A+C}(\varepsilon_i)$ dominates the term $\sigma_B(\varepsilon_i)$, except in the very proximity of the failure threshold where σ_{A+C} goes to zero, whereas the model error remains larger due to the uncertainty related to the unknown value of the failure threshold.

The dotted lines in Figure 9 represent the estimates of these quantities provided by the bootstrap ensemble.

The standard deviation σ_{A+C} estimated by the bootstrap ensemble is significantly larger than its real value. As in approach 1, this is due to the fact that the training trajectories have different values of parameters A and n and, thus, the empirical model learns the variance of a population of different blades instead of that of the specific blade with fixed values of A and n.

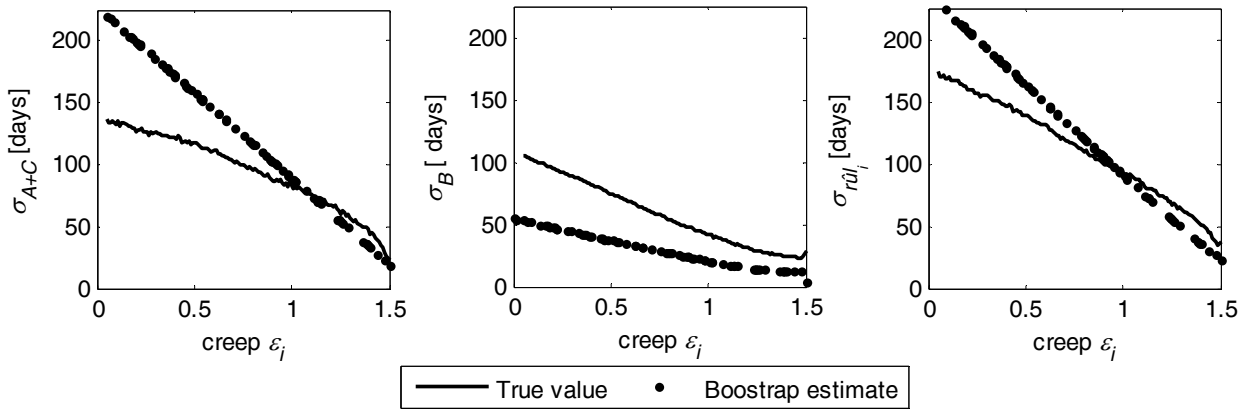

Figure 10: comparison of the bootstrap estimates (dashed line) and true values (continuous line) of $\sigma_B(\varepsilon_i)$ (left), $\sigma_{A+C}(\varepsilon_i)$ (middle) and $\sigma_{r\hat{u}l_i}(\varepsilon_i)$ (right) during the life of a turbine blade for different values of its creep strain ε_i .

In Figure 10, the estimate of the model error variance σ_B^2 appears to be not very accurate. Figure 11 shows that the inaccuracy can be even more remarkable if other test trajectories are considered, characterized by values of A and n far away from the mean value of their respective distributions ($\mu_A = 3 \cdot 10^{-4}$ and $\mu_n = 6$). Notice that the real model uncertainty depends on the test trajectory: the model trained on the historical trajectories tends to learn the 'average' behavior of the general creep growth trajectory; the consequence is that the model makes larger errors when the test trajectory is different from the 'average' trajectory. On the contrary, the estimate, $\hat{\sigma}_B^2$, of the model uncertainty provided by the bootstrap ensemble depends only on the value of ε_i, being independent from the specific values of A and n of the test trajectory.

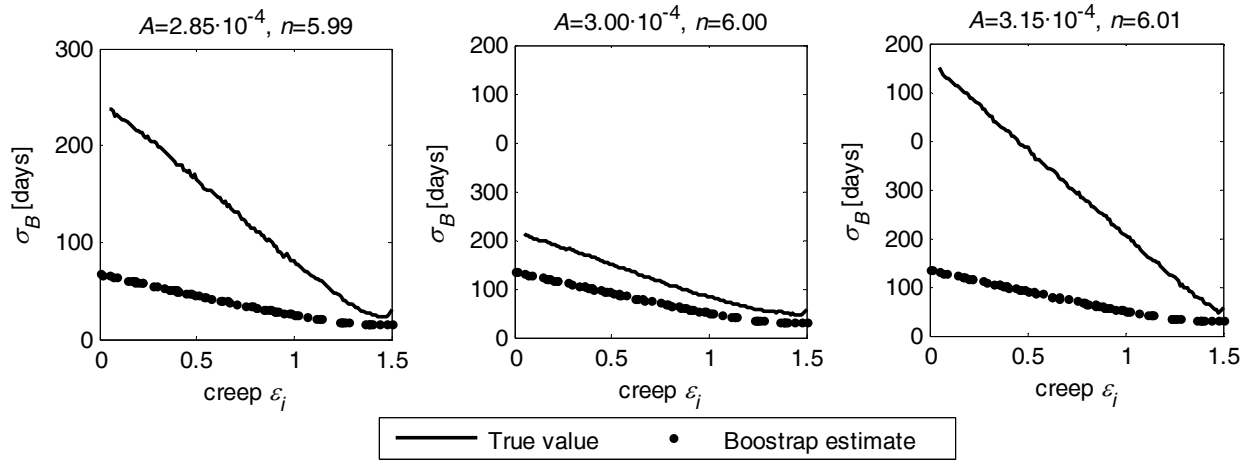

Figure 11: comparison of the bootstrap estimates (dashed line) and true (continuous line) values of σ_B during the life of three turbine blades with different values of parameters A and n .

Differently from approach 1, the ensemble approach 2 is not able to learn the true values of A and n of the current test trajectory. Furthermore, this limitation of the model is not properly described by the prediction interval provided by the ensemble.

The reason for which the proposed bootstrap approach 2 is not able to correctly model the evolution of the error made by the model for a specific test trajectory is that the assumption that the predictive model $f(\mathbf{z}_i)$ is an unbiased estimator of the RUL expected value μ_{RUL_i}, is not fully verified. In fact, if we build several ensemble models trained with different randomly chosen datasets and perform the RUL prediction with each one of them, we notice that the average RUL prediction $\hat{rul}_i$ over the different ensemble models is different from the RUL mean value computed over a set of creep growth trajectories with fixed values of A and n. This is shown in Figure 12 where the distribution $\Pr(\hat{rul}_i \mid \mathbf{z}_i)$ of the prediction $\hat{rul}_i$ obtained at $\varepsilon_i = 0.4$ by several ensemble models is compared to the true RUL distribution $\Pr(RUL_i \mid \varepsilon_i)$ for 3 different values of parameters A and n: 1) $A = 2.85 \cdot 10^{-4}$, $n = 5.99$; 2) $A = 3.00 \cdot 10^{-4}$, $n = 6.00$; 3) $A = 3.15 \cdot 10^{-4}$; $n = 6.01$.

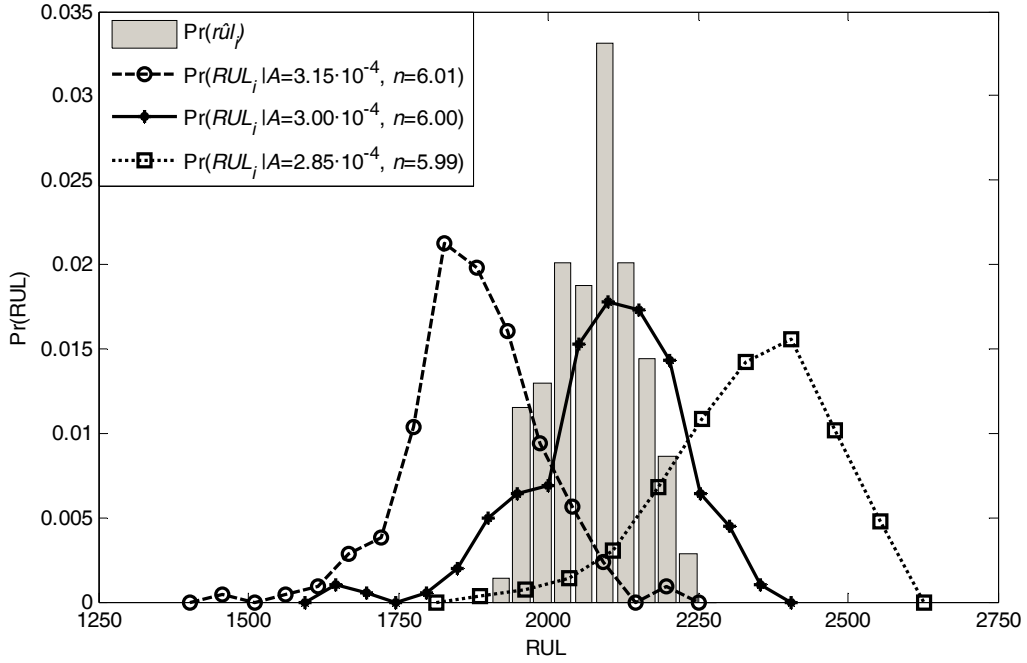

Figure 12: comparison of the distributions of the prediction $\hat{rul}_i$ with the distribution of the actual RUL of three turbine blades with different values of parameters A and n at $\varepsilon_i = 0.4$.

On the other side, the models trained on the historical trajectories are unbiased estimators of the RUL of the generic turbine blade with random values of A and n as it can be seen by comparing the distribution of the RUL prediction $\hat{rul}_i$ with the distribution of the actual RUL of a population of turbine blades with randomly sampled values of A and n (Figure 13).

Figure 14 compares the bootstrap estimates (dots) of $\sigma_{A+C}(\varepsilon_i)$ (left), $\sigma_B(\varepsilon_i)$ (middle) and $\sigma_{\hat{rul}_i}(\varepsilon_i)$ (right) with their true values obtained for a population of different turbine blades with parameters A and n normally distributed.

These results confirm that the bootstrap approach 2 can actually provide a satisfactory estimates of $\sigma_{A+C}(\varepsilon_i)$, $\sigma_B(\varepsilon_i)$ and $\sigma_{\hat{rul}_i}(\varepsilon_i)$ for a population of different turbine blades, and thus the proposed approach correctly quantifies the uncertainty of the prediction produced by the prognostic model for a generic creeping blade.

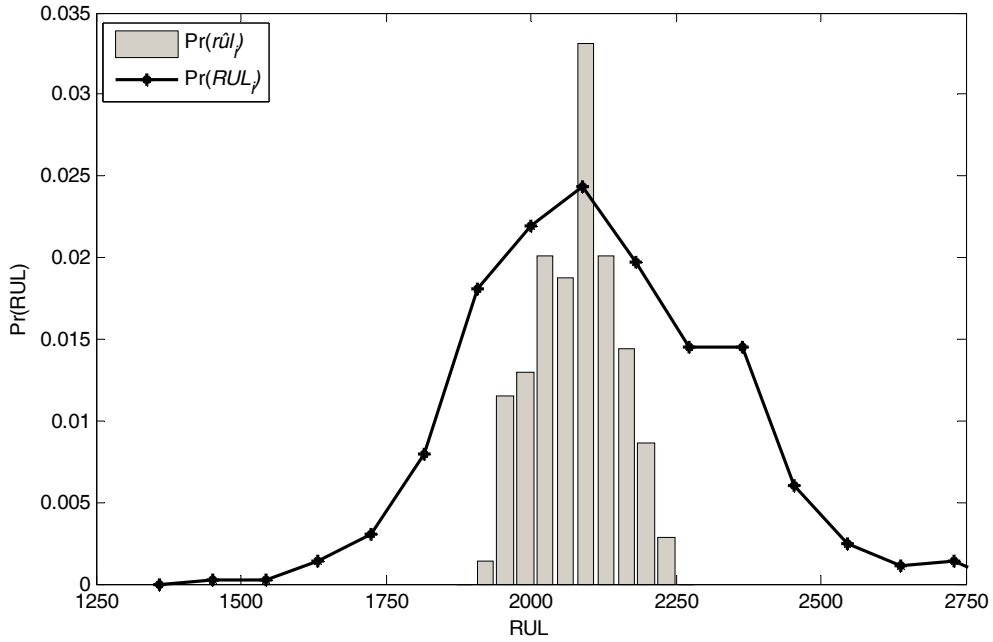

Figure 13: comparison of the distributions of the prediction $r\hat{u}l_i$ with the distribution of the RUL of a population of turbine blades with normally distributed values of A and n at $\varepsilon_i = 0.4$.

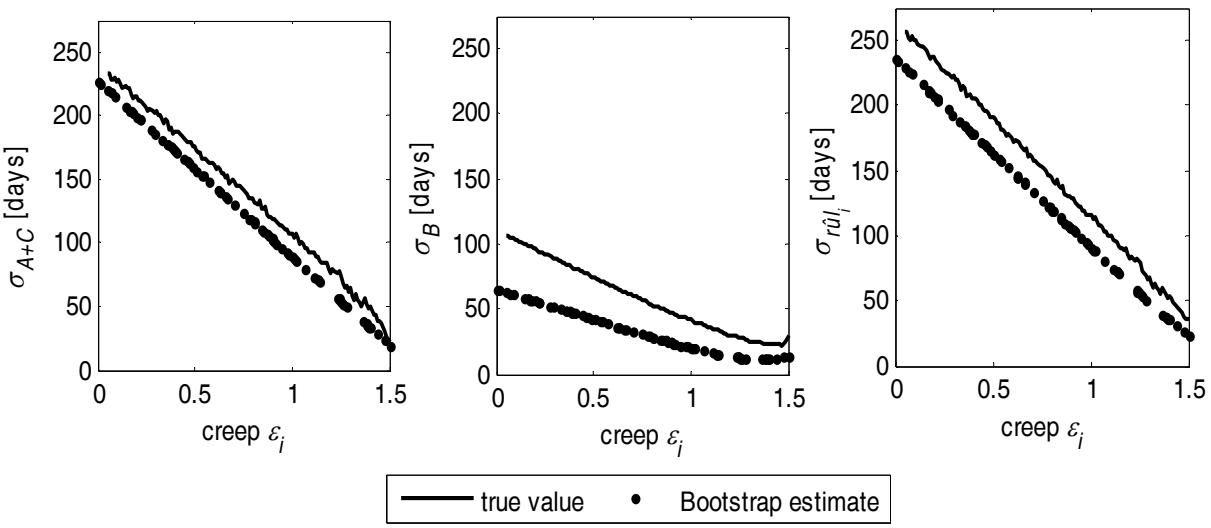

Figure 14: comparison of the bootstrap estimates (dashed line) of $\sigma_{A+C}(\varepsilon_i)$ (left), $\sigma_B(\varepsilon_i)$ (middle) and $\sigma_{r\hat{u}l_i}(\varepsilon_i)$ (right) with their true values (continuous line) for a population of turbine blades with normally distributed values of A and n.

5.3.3 Prediction interval provided by approach 3

Figure 15 shows the analogous results of Figure 9 for approach 3. Notice that good estimates of $\sigma_{A+C}(\varepsilon_i)$ are achieved only for large values of ε_i, i.e., when a large validation dataset is available and the model $\chi(\Delta\varepsilon_{i,i'})$ is used in the same range of input values where it has been trained. We also notice that the value of $\sigma_{A+C}(\varepsilon_i)$ is largely underestimated.

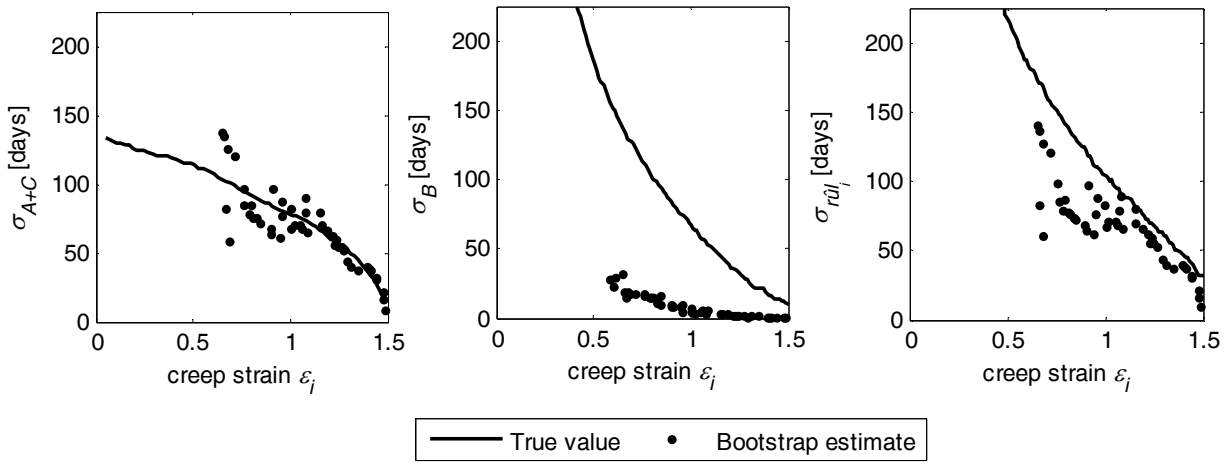

Figure 15: comparison of the bootstrap estimates (dots) and true (continuous line) values of $\sigma_B(\varepsilon_i)$ *(left),* $\sigma_{A+C}(\varepsilon_i)$ *(middle) and* $\sigma_{r\hat{u}l_i}(\varepsilon_i)$ *(right) during the life of a turbine blade for different values of its creep strain* ε_i *.*

To understand the reason for which the proposed bootstrap approach 3 tends to underestimate the model error variance in this case, it must be pointed out that, from a probabilistic point of view, the single available trajectory used for training the models is only one of an infinite number of possible trajectories, which may be drawn from the creep growth process we wish to model. Thus, bootstrap sampling of creep strain measurements from a single degradation trajectory does not account for the variability of all possible degradation trajectories.

In the case of a linear process, it is possible to overcome this limitation, by considering, instead of the sequence of creep strain measurements $z_{1:i}$, the set of independent creep strain increments for time unit $\Delta z_j = (z_{j+1} - z_j)/(t_{j+1} - t_j)$, $j = 1,...,i-1$. In this way, the variability of the training data is increased, and a better representation of the intrinsic variability of the process is provided. An accurate model of the process can still be achieved by estimating the parameters β^h, h=1,…,H, of the models of the ensemble $\tilde{\eta}^h(\Delta\varepsilon_{j,j'}) = \Delta\varepsilon_{j,j'} / \beta^h$ as the average value of the creep strain increments of the bootstrap replicate $\mathbf{D}_{incr}^h$ of the new training dataset $\mathbf{D}_{incr}^{trn} = \{\Delta z_{1:i-1}\}$. Figure 16 shows that a more accurate estimate of σ_B and $\sigma_{r\hat{u}l}$ is achieved using this new ensemble of models.

6 Conclusions

Three prognostic approaches have been investigated, particularly with respect to the treatment of the uncertainty in the predicted equipment RUL. Quantitative considerations have been made with regards to a simulated case study concerning the creep growth process in a high temperature turbine blade. The results show that the particle filtering approach provides a good approximation of the exact distribution of the equipment RUL in the case in which an accurate model reproducing the equipment degradation process is available. A limit of particle filtering is particle impoverishment, which relates to the failure of maintaining the diversity of particles and is caused by the resampling approach adopted to avoid particle degeneracy. Particle impoverishment implies the impossibility of the particles to correctly represent all possible evolutions of the degradation process, which include, for example, changes in time due to the variation of

operating conditions. In this context, different resampling methods, such as the one proposed in [55], which samples particle considering not only the particle weight but also their spatial distribution (state values), can be considered.

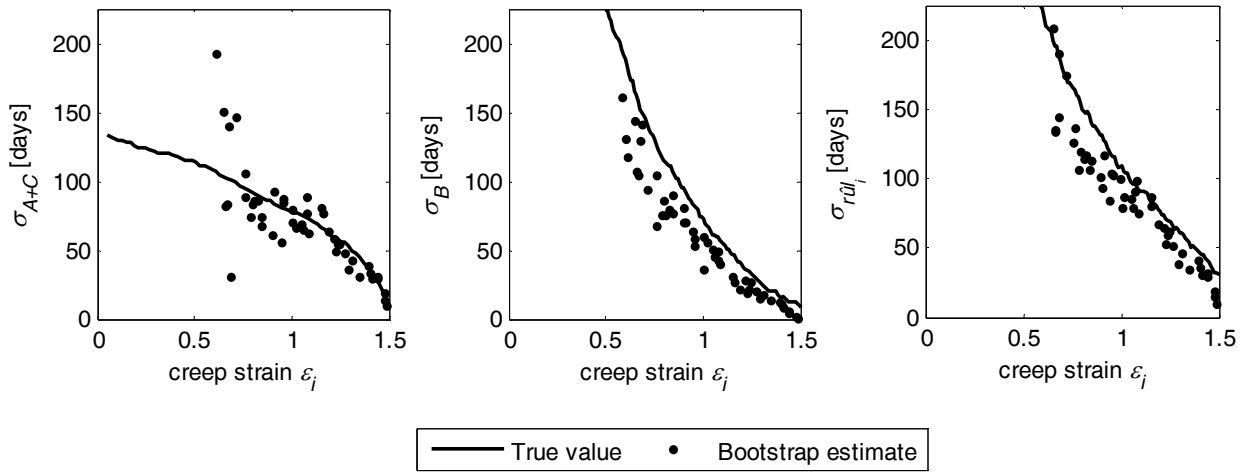

Figure 16: comparison of the bootstrap estimates (dots) and true values (continuous line) of σ_{A+C} (left), σ_B (middle) and $\sigma_{\hat{rul}_i}$ (right) during the life of a turbine blade for different values of its creep strain ε_i obtained using as training data for the ensemble of models the creep strain increments between consecutive observations.

When using model-based approaches, imprecision of the model in the reproduction of the degradation process due to simplifications, incorrect model structure or assumptions on the equipment specific geometries or material properties, etc. can be amplified over time, causing unreliable estimates of the RUL distribution. Using particle filtering, it is possible to include model parameters in the state vector and, thus, perform model adaptation in conjunction with state tracking. In any case, it is very difficult for a physics-based model to account for all aspects of a degradation process; for example, it is common to neglect some of the interactions between different degradation mechanisms or the possible existence of self-healing mechanisms which can reverse the degradation process and are likely to increase the uncertainty of the future degradation evolution. All these non-modeled phenomena can be accounted for by adding further noise to the process model which will result in a larger confidence interval associated to the RUL estimate. Further research is needed to quantify the impact of modeling errors on the final prediction of model-based approaches.

In the bootstrap approaches 2 and 3 considered, it has been shown that a reliable prediction of the equipment RUL with a correct quantification of its uncertainty can be obtained. With respect to the ensemble of bootstrapped models trained with historical sequences of observations in approach 2, the main limitation is that it is not able to learn the peculiar characteristics of the equipment of interest but it tends to reproduce an 'average' behavior. To overtake this problem, a different modeling approach could be used, such as that based on the idea of fuzzy similarity [56], or a procedure for updating the ensemble with the information conveyed by new observations [27].

The application of the bootstrap ensemble in the time series scheme of approach 3, in which only direct measurements of the degradation experienced by the equipment of interest are available, has shown the

importance of injecting diversity into the bootstrapped models by using independent training data, in order to correctly quantify the modeling error. The case study considered is characterized by a linear degradation process, so that independent training data can be obtained by considering the creep strain increments between consecutive measurements; on the contrary, this would not be feasible for non-linear degradation processes. Furthermore, in this case of very little information available, the bootstrap method requires building an empirical model for the RUL variance estimate which is then used outside the region covered by the training data. Although good extrapolations have been obtained in the linear creep growth case study, the feasibility of the approach on more complex models should be verified.

Contrarily to physics-based models, we expect that data-driven methods can automatically learn from data the effects on the equipment RUL of phenomena influencing the degradation process, such as self-healing and interactions between different degradation mechanisms. The capability of data-driven methods of providing correct estimates of the RUL and its uncertain distribution depends on the availability in the training set of examples of the phenomena that we want to represent.

In this work, the problem of detecting the initiation of a degradation process, which is usually achieved by using properly developed diagnostic systems, has not been addressed. Although none of the approaches presented in this work requires knowing the exact time at which degradation has initiated, late detection of an ongoing degradation process will reduce the number of degradation measures available for prognostics; this is expected to reduce the RUL prediction accuracy and increase its uncertainty especially in approaches 1 and 3 which, contrarily to approach 2, generate their RUL prediction on the basis of past degradation measurements. .

Since only artificial data have been used in the case study considered in this work, conclusions about the successful application of these approaches in the field cannot be directly drawn. The analyses performed have shown the potential of these methods in performing RUL prediction with adequate management of its uncertainty; in this sense, they hold promises for future research aimed at confirming this potential in the application to real data.

Appendix A: approximation of the model error and prediction error variances

In the empirical ensemble-based approaches 2 and 3, the true value of the model error variance σ_B^2 has been approximated for different values of the creep strain ε_i, as the mean square value of the model error $\mu_{RUL_i|\mathbf{z}_i} - \hat{rul}_i(\mathbf{z}_i)$ made by different ensembles on 200 creep growth trajectories sampled for a turbine blade with parameters A=3·10^{-4} and n=6. The details of the procedure are sketched in the pseudo-code of Figure 1A for approaches 2 and 3.

1. Sample 200 creep growth trajectories with $A = 3 \cdot 10^{-4}$ and $n = 6$ and simulate a measurement every 30 days according to eq. (16)

FOR p=1:200

2. Sample the current measured value $\mathbf{z}_i^p = \varepsilon_i + \upsilon_i^p$ according to eq. (16);

3. Sample 1000 creep growth trajectories with $A = 3 \cdot 10^{-4}$ and $n = 6$;

FOR p'=1:1000

4. Sample $\varepsilon_i^{p'} = z_i^p - \upsilon_i^{p'}$;

5. Compute $rul_i^{p'}(\varepsilon_i^{p'})$;

END FOR

6. Compute $\mu_{RUL_i|\boldsymbol{\varepsilon}_i^p} = \frac{1}{1000}\sum_{p'} rul_i^{p'}$

Case 2

FOR m=1:50

7. Sample 10 training degradation trajectories with different values of parameters A and n and simulate a measurement every 30 days according to eq. (16);

8. Train the prognostic model $f(\varepsilon)$ using the 10 training trajectories

9. Predict $\hat{rul}_i^{p,m} = f(\mathbf{z}_i^p)$

END FOR

10. Compute the mean square model error

$$\sigma^2_{\hat{B},p}(\varepsilon_i) = \frac{1}{50}\sum_m (\mu_{RUL_i|\mathbf{z}_i^p} - \hat{rul}_i^{p,m})^2$$

Case 3

7-8. Build a prognostic model $\tilde{\eta}(\Delta\varepsilon)$ using the measurements of the p-th trajectory up to ε_i

9. Predict $\hat{rul}_i^p = \tilde{\eta}(\mathbf{z}_{th} - \mathbf{z}_i^p)$

10. Compute the square model error

$$\sigma^2_{\hat{B},p}(\varepsilon_i) = (\mu_{RUL_i|\mathbf{z}_i^p} - \hat{rul}_i^p)^2$$

END FOR

11. Compute the model error variance $\sigma^2_{\hat{B}}(\varepsilon_i) = \frac{1}{200}\sum_p \sigma^2_{\hat{B},p}(\varepsilon_i)$

Figure 1A: procedure for approximating $\sigma^2_{\hat{B}}(\varepsilon_i)$ in approaches 2 and 3.

References

[1] Hess A, Calvello G, Frith P. Challenges, issues, and lessons learned chasing the 'Big P': real predictive prognostics Part 1. Proc IEEE Aerosp Conf, 2006 Mar 5-12; Big Sky, MT.

[2] Tang L, Kacprzynski GJ, Goebel K, Vachtsevanos G. Methodologies for uncertainty management in prognostics. Proc IEEE Aerosp Conf, 2009 Mar 7-14; Big Sky, MT.

[3] Li W, Pham H. An inspection-maintenance model for systems with multiple competing processes. IEEE Trans Rel, 2005;54(2):318–27.

[4] Liu R, Ma L, Kang R, Wang N. The modeling method on failure prognostics uncertainties in maintenance policy decision process. Proc 9th Int Conf on Reliab, Maint and Saf (ICRMS), 2011 Jun 12-15; Guiyang, China.

[5] Saxena A, Celaya J, Saha B, Saha S, Goebel K. Evaluating prognostics performance for algorithms incorporating uncertainty estimates. In Proc. IEEE Aerosp Conf, 2010.

[6] Urbina A, Mahadevan S, Paez TL. Quantification of margins and uncertainties of complex systems in the presence of aleatoric and epistemic uncertainty, Reliab Eng Syst Saf 2011;96(9):1114-25.

[7] Zio E. Prognostics and health management of industrial equipment. In: Kadry S, editor. Diagnostics and Prognostics of Engineering Systems: Methods and Techniques, IGI-Global, 2012 (to appear).

[8] Baraldi P, Cadini F, Mangili F, Zio E. Model-based and data-driven prognostics under different available information. Probab Eng Mech. Submitted for publication, 2012.

[9] Luo J, Pattipati K, Qiao L, Chigusa S. Model-based prognostic techniques applied to a suspension system. IEEE Trans Syst, Man, and Cybern – Part C: Appl and Rev 2008;38(5):1156-68.

[10] Doucet A, de Freitas JFG, Gordon NJ. Sequential Monte Carlo methods in practice. New York: Springer-Verlag; 2001.

[11] Myötyri E, Pulkkinen U, Simola K. Application of stochastic filtering for lifetime Prediction, Reliab Eng Syst Saf 2006;91(2):200-8.

[12] Arulampalam MS, Maskell S, Gordon N, Clapp T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans Sig Process 2002;50(2):174-88.

[13] Anderson BD, Moore JB. Optimal filtering. Englewood Cliffs (NJ): Prentice Hall; 1979.

[14] Kitagawa G. Non-Gaussian state-space modeling of nonstationary time series. J Am Stat Assoc 1987;82:1032-63.

[15] Crisan D, Doucet A, A survey of convergence results on particle filtering methods for practitioners. IEEE Trans on Sig Process, 2002;50(3):736-46.

[16] Djuric PM, Kotecha JH, Zhang J, Huang Y, Ghirmai T, Bugallo MF, Miguez J. Particle filtering. IEEE Sig Process Mag, 2003;20(5):19-38.

[17] Pitt MK, Shephard N. Filtering via simulation: Auxiliary particle filters. J Am Stat Assoc 1999;94(446):590-9.

[18] Marseguerra M, Zio E. Monte Carlo simulation for model-based fault diagnosis in dynamic systems. Reliab Eng Syst Saf 2009;94(2):180-6.

[19] Cadini F, Zio E, Avram D. Model-based Monte Carlo state estimation for condition-based component replacement. Reliab Eng Syst Saf 2009;94(3):752-8.

[20] Cadini F, Zio E, Avram D. Monte Carlo-based filtering for fatigue crack growth estimation. Probab Eng Mech 2009;24:367-73.

[21] Zio E, Peloni G. Particle filtering prognostic estimation of the remaining useful life of nonlinear components, Reliab Eng Syst Saf 2011;96(3):403-9.

[22] Zio E, Compare M. Evaluating maintenance policies by quantitative modeling and analysis. Reliab Eng Syst Saf 2013;109:53−65.

[23] Schwabacher MA. A survey of data-driven prognostic, Proc Infotech@Aerospace, 2005 September 26-29; Arlington, VA.

[24] Heskes T. Practical confidence and prediction intervals in Advances. In: Mozer M, Jordan M, Heskes T, editors. Neural Information Processing Systems 9, Cambridge, MA: MIT Press; 1997, p. 466–72.

[25] Raviv Y, Intrator N. Bootstrapping with noise: An effective regularization technique. Connect Sci 1996;8(3-4):355-72.

[26] Couturier R, Escaravage C. High temperature alloys for the HTGR gas turbine: Required properties and development needs. Proc IAEA Tech Comm Meet on Gas Turbine Power Convers Syst for Modular HTGRs. 2000 Nov 14-16; Palo Alto, CA.

[27] Baraldi P, Mangili F, Zio E. A Kalman filter - based ensemble approach with application to turbine creep prognostics. IEEE Trans on Reliab 2012, to appear.

[28] Saxena A, Celaya J, Saha B, Saha S, Goebel K. Metrics for offline evaluation of prognostic performance. Int. J. of PHM, 2010;1:1-20.

[29] Usynin A, Hines JW, Urmanov A. Uncertain failure thresholds in cumulative damage models. In: Proc. of the Annual Reliability and Maintainability Symposium. IEEE Conference Publications; 2008, p. 334–340.

[30] Wang P, Coit DW. Reliability and degradation modeling with random or uncertain failure threshold. In: Proc. of the Annual Reliability and Maintainability Symposium. IEEE Conference Publications; 2008, p. 334–340.

[31] Orchard M, Kacprzynski G, Goebel K, Saha B, Vachtsevanos G. Advances in uncertainty representation and management for particle filtering applied to prognostics. Proc Int PHM Conf, 2008 Oct 6-9; Denver, CO.

[32] Kotecha JH, Djuric PM. Gaussian sum particle filtering. IEEE Trans Sig process 2003;51(10): 2592-601.

[33] Saha B, Goebel K, Christophersen J. Comparison of prognostic algorithms for estimating remaining useful life of batteries. Trans Inst of Meas & Control 2009;31(3-4):293-308.

[34] Lopes HF, Carvalho CM. Online Bayesian learning in dynamic models: An illustrative introduction to particle methods. In West M, Damien P, Dellaportas P, Polson NG, Stephens DA, editors. Bayesian Dynamic Modelling Bayesian Inference and Markov Chain Monte Carlo: In Honour of Adrian Smith, Clarendon: Oxford University Press; 2012 (to appear).

[35] Wang X, Rabiei M, Hurtado J, Modarres M, Hoffman P. A probabilistic-based airframe integrity management model, Reliab Eng Syst Saf 2009;94(5):932-41.

[36] Zio E. A Study of the bootstrap method for estimating the accuracy of artificial neural networks. Predicting nuclear transient processes. IEEE Trans Nucl Sci 2006;53(3):1460-78.

[37] Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman and Hall; 1993.

[38] Carney JG, Cunningham P, Bhagwan U. Confidence and prediction intervals for neural network ensembles. Proc. Int Joint Conf Neural Netw, 1999 Jul 10-16; Washington, DC.

[39] Gorjian N, Ma L, Mittinty M, Yarlagadda P, Sun Y. Review on degradation models in reliability analysis. Proc 4th World Congr on Eng Asset Manag. 2009 Sept 28-30; Athens, Greece.

[40] Carter TJ, Common failures in gas turbine blades. Eng Failure Analysis, 2005;12:237-47.

[41] Dunn TD, Lommers LJ, Tangira VE. Preliminary safety evaluation of the gas turbine-modular Helium reactor (GT-MHR). Proc Int Top Meet Adv React Saf 1994, Apr 17-21, Pittsburgh, PA.

[42] Saez M, Tauveron N, Chataing T, Geffraye G, Briottet L, Alborghetti N. Analysis of the turbine deblading in an HTGR with the CATHARE code. Nucl Eng and Des 2006;236:574–86.

[43] Made M, Mirmiran A, Walter TA. Local damage assessment of turbine missile impact on composite and multiple barriers. Nucl Eng and Des, 1997;178:145–56.

[44] Goel N, Kumar A, Narasimhan V, Nayak A, Srivastava A. Health risk assessment and prognosis of gas turbine blades by simulation and statistical methods. Can Conf on Electr and Comput Eng. 2008 May 4-7; Niagara Falls, ON.

[45] Flotow A, Mercadal M, Tappert P. Health monitoring and prognostics of blades and disks with blade tip sensors, Proc IEEE Aerosp Conf. 2000 Mar 18-25; Big Sky, MT.

[46] Ye D, Duan F, Guo H, Li Y, Wang K. Turbine blade tip clearance measurement using a skewed dual-beam fiber optic sensor. Opt Eng, 2012;51(8).

[47] Steiner A. Techniques for blade tip clearance measurements with capacitive probes. Meas Sci Technol, 2000;11:865–9.

[48] http://www.capacisense.com/

[49] http://www.thermocoax.com

[50] Dowell M, Sylvester G, Krupp R, Zipfel G. Progress in turbomachinery prognostics and health management via eddy-current sensing, In Proc. IEEE Aerosp Conf, 2000, vol. 6, p. 133-43.

[51] Kwapisz D, Hafner M, Rajamani R. application of microwave sensing to blade health monitoring. In Proc 1st Eur Conf of PHM Soc, 2012, vol. 3, p. 1-8.

[52] Polikar R, Topalis A, Green D, Kounios J, Clark CM. Comparative multiresolution wavelet analysis of ERP spectral bands using an ensemble of classifiers approach for early diagnosis of Alzheimer's disease. Comput in Biol and Med, 2007;37:542–58.

[53] Carter TJ, Common failures in gas turbine blades. Eng Fail Anal 2005;12(2):237-47.

[54] Swindeman RW, Swindeman MJ. A comparison of creep models for nickel base alloy for advanced energy systems, Int J of Press Vessels and Pip 2008;85:72–9.

[55] Li T, Sattar TP, Sun S. Deterministic resampling: Unbiased sampling to avoid sample impoverishment in particle filters. Signal Process, 2012;92:1637–45.

[56] Zio E, Di Maio F. A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system. Reliab Eng and Syst Saf 2010;95(1):49–57.

Paper IV

Equipment Uncertainty in Remaining Useful Life Prediction: a Belief Function Theory Treatment

Piero Baraldi, Francesco Di Maio, Francesca Mangili, Enrico Zio

2013

Equipment Uncertainty in Remaining Useful Life Prediction: a Belief Function Theory Treatment

Piero Baraldi[a], Francesco Di Maio[a], Francesca Mangili[a], Enrico Zio[b,a]

[a]*Dipartimento di Energia, Politecnico di Milano, Italy*

[b]*Chair on Systems Science and the Energetic challenge, European Foundation for New Energy-Electricite' de France, Ecole Centrale Paris and Supelec, France*

Abstract

We develop a prognostic method for estimating the equipment remaining useful time and its uncertainty. The method is based on belief function theory and fuzzy similarity. The maintenance planner defines the maximum acceptable failure probability, and is informed by the prognostic method of the time at which this probability is exceeded.

Keywords: Prognostics, belief function, creep, filter clogging

1 Introduction

Several data-driven methods have been proposed for predicting the Remaining Useful Life (RUL) of degrading equipment [Hines & Usynin, 2008; Zio, 2012, Vachtsevanos , 2006], i.e., the time left before the equipment will stop fulfilling its functions. Data-driven methods rely on the availability of observations collected during the degradation of one or more similar equipments and are usually based on the regression of the future degradation path until the criteria indicating failure is reached [Baraldi 2012a-b,2013a-c]. However, the information available for modeling the future equipment degradation can be scarce and incomplete, e.g., few examples of similar equipment degradation trajectories can be available, the degradation state of the equipment can be not directly measured, and the failure criteria can be not known. Furthermore, the RUL estimate should take into account the intrinsic uncertainty due to the variability of the degradation process (caused, for example, by the micro-structural differences between pieces of the same equipment, or by the unforeseen future loads, operational settings, and external conditions) [Baraldi et al. 2012], which implies that we cannot be sure that two identical pieces of equipment, having walked through the same degradation path up to the present time, will keep following exactly the same path even in the future. Another source of uncertainty in the RUL estimate is the measurement noise.

Thus, given the scarcity of information typically available and the different sources of uncertainty to which the RUL estimate is subject, data-driven models can commit large errors in the RUL estimate [Yan et al., 2004] and uncertainty management becomes a fundamental task in prognostics. In practice, it is necessary to provide the maintenance planner with an assessment of the expected mismatch between the real and

predicted equipment failure times, in order to allow them to confidently plan maintenance actions, according to the maximum acceptable failure probability [Tang et al., 2009].

In this context, the objective of the present work is to provide a measure of confidence in the RUL prediction provided by a data-driven prognostic model. To this purpose, we consider the similarity-based prognostic model proposed in [Zio & Di Maio, 2010] which uses a set of reference degradation trajectories collected in a reference library and performs a data-driven similarity analysis for predicting the RUL of a newly developing degradation trajectory (hereafter called test trajectory). The matching process is based on the evaluation of the distance between the reference and test trajectories [Angstenberger, 2001]. This prognostic model is here extended in order to provide a measure of confidence in the RUL prediction. To address this issue, we adopt a solution based on the belief function theory (BFT) (also called Dempster-Shafer or evidence theory) [Dempster, 1967; Shafer, 1976]. The BFT allows combining different pieces of (uncertain) evidence, based on the assignment of basic belief masses to subsets of the space of all possible events, which are, in this case, the possible values that the equipment RUL can assume. In practice, the proposed method considers each reference trajectory as a piece of evidence regarding the value of the RUL of the test trajectory. These pieces of evidence are discounted based on their similarity to the test trajectory and pooled using the Dempster's rule of combination [Petit-Renaud & Denoeux, 2004]. The result is a basic belief assignment (BBA) that quantifies one's belief about the value of the RUL for the test trajectory given the reference trajectories. From the BBA, the total belief (i.e., the amount of evidence) supporting the hypothesis that the RUL will fall in any specific interval can be computed. In this context, we propose to define a prediction interval as an interval to which a sufficiently large total belief has been assigned.

The method is applied to two case studies considering simulated data generated by a non-linear model of creep growth in ferritic steal and real industrial data concerning the clogging of filters used to clean the sea water pumped in a Boiling Water Reactor (BWR).

The remaining part of the paper is organized as follows: in Section 2 the methodology for the similarity-based prediction of equipment RUL is described and a method for integrating it with the belief function theory to supply a measure of confidence in the similarity-based RUL prediction is proposed; in Section 3 two numerical applications are presented: the first using artificial data about the growth of creep damage in ferritic steel, the second concerning the clogging of sea water filter, for which real data are available; in Section 4 some conclusions are drawn from the results presented.

2 Methodology

We assume the availability of R reference trajectories which contain measurements collected during the process of degradation of R pieces of equipment similar to the one currently monitored (test equipment). Each reference trajectory, r=1,…,R, is made of a sequence $\mathbf{z}_{1:n^r}^r = \left[z_1^r,...,z_i^r,...,z_{n^r}^r\right]$ of n^r observations, $\mathbf{z}_i^r = [z_1^r(\tau_i),...,z_p^r(\tau_i),...,z_P^r(\tau_i)]$, representing the evolution of P relevant parameters z_p^r measured at different time instants τ_i, i=1:n^r up to the last measurement time τ_{n^r} before the failure time, τ_F^r.

With respect to the test equipment, a sequence of observations $\mathbf{z}_{1:I}^{test}$ from τ_1 to the present time τ_I is assumed to be available.

2.1 Similarity-based RUL prediction

The idea underpinning the RUL estimation method is to evaluate the similarity between the test trajectory and the *N* reference trajectories and to use the RULs of the reference equipments to estimate the RUL of the test equipment, taking into account the similarities between the trajectories [Petit-Renaud & Denoeux, 2004; Wang et al. 2008; Zio & Di Maio, 2010].

Trajectory similarity is evaluated considering the pointwise difference between *n*-long sequences of observations. At the present time τ_I, the distance δ_j^r between the sequence of the *n* latest observations $\mathbf{z}_{I-n+1:I}^{test}$ of the test trajectory, and all the *n*-long segments $\mathbf{z}_{j-n+1:j}^r$, $j=n:n^r$ that can be extracted from the reference trajectories $r=1:R$ is computed:

$$\delta_j^r = \sqrt{\sum_{i=1}^{n} d^2(\mathbf{z}_{I-n+i}^{test}, \mathbf{z}_{j-n+i}^r)} \tag{1}$$

where $d^2(\mathbf{x},\mathbf{y})$ is the square Euclidean distance between vectors **x** and **y**.

The similarity s_j^r of the reference trajectory segment $\mathbf{z}_{j-n+1:j}^r$ to the latest test trajectory segment $\mathbf{z}_{I-n+1:I}^{test}$ is defined as a function of the distance measure δ_j^r [Zio & Di Maio, 2010]:

$$s_j^r = \exp\left(-\frac{\left(\delta_j^r\right)^2}{\lambda}\right) \tag{2}$$

The arbitrary parameter λ is set by the analyst: the smaller is the value of λ the stronger the definition of similarity. A strong definition of similarity implies that the two segments under comparison have to be very close in order to receive a similarity value s_j^r significantly larger than zero. In practice, parameter λ is often set to the value that minimizes the error of the similarity-based prediction calculated on a validation dataset.

For the prediction of the test equipment RUL, a value RUL_{j*}^r is assigned to each reference trajectory $r=1:R$ by considering the difference between the trajectory failure time τ_F^r and the last time instant τ_{j*}^r of the trajectory segment $\mathbf{z}_{j*-n+1:j*}^r$ which has the maximum similarity s_{j*}^r with the test trajectory:

$$RUL_{j*}^r = \tau_F^r - \tau_{j*}^r \tag{3}$$

Then, the similarity-based (SB) prediction $\hat{RUL}_I^{SB}$ of the test equipment RUL at time τ_I is given by the weighted sum of the values RUL_{j*}^r:

$$\hat{RUL}_I^{SB} = \frac{\sum_{r=1}^{R} s_{j*}^r RUL_{j*}^r}{\sum_{r=1}^{R} s_{j*}^r} \tag{4}$$

The ideas behind the weighting of the predictions RUL_{j*}^r supplied by the individual trajectories is that: i) all failure trajectories in the reference library can, in principle, bring useful information for determining the RUL of the trajectory currently developing; ii) those segments of the reference trajectories which are most similar to the latest part of the test trajectory should be more informative about the value of its RUL.

2.2 Prediction interval based on belief function theory

Given the uncertainty to which the RUL estimate is subject, maintenance plans cannot usually be based only on the RUL prediction provided by eq. (4). In this Section, we assume that the maintenance planner is able to specify a maximum acceptable failure probability, α, and we proposes a method to identify the latest time at which, according to the available information, we can guarantee that the probability to have a failure is lower than α To this aim, we resort to the Belief Function Theory.

For the ease of clarity, only the notions of BFT necessary for the understanding of the proposed method will be now presented. For further details about the mathematical developments and the possible interpretations of the theory, the interested reader is referred to Dempster (1976), Shafer (1976) and Smets (1998).

The BFT represents the belief of an agent about the value of an uncertain variable Y assuming values y in the frame of discernment Ω_Y, based on the available information, by a basic belief assignments (BBA) made of a set of masses $m_Y(\mathrm{Y}_k)$ assigned to subsets Y_k, k=1,2,… of Ω_Y. The mass $m_Y(\mathrm{Y}_k)$ represents the belief that the value of Y belongs to the subset Y_k. Any subset Y_k with associated a finite mass $m_Y(\mathrm{Y}_k) > 0$, is called focal element; the BBA verifies the condition that the sum of all its masses is 1.

Let us assume that two distinct sources of information induce the two BBAs m_Y^1 and m_Y^2, according to the Dempster's rule of combination the two BBAs can be aggregated into the BBA $m_Y^{1\oplus 2}$:

$$\begin{aligned} &m_Y^{1\oplus 2}(\mathrm{Y}_k) = \frac{1}{K} \sum_{\mathrm{Y}_{k'} \cap \mathrm{Y}_{k''} = \mathrm{Y}_k} m_Y^1(\mathrm{Y}_{k'}) m_Y^2(\mathrm{Y}_{k''}), \quad \forall\, \mathrm{Y}_k \in \Omega_Y, \quad \mathrm{Y}_k \neq \varnothing \\ &m_Y^{1\oplus 2}(\varnothing) = 0 \end{aligned} \tag{5}$$

where

$$K = 1 - \sum_{\mathrm{Y}_{k'} \cap \mathrm{Y}_{k''} = \varnothing} m_Y^1(\mathrm{Y}_{k'}) m_Y^2(\mathrm{Y}_{k''}) \tag{6}$$

is a normalization factor introduced to convert a possibly subnormal BBA (i.e., a BBA assigning a finite mass to the empty set $\varnothing$) into a normal one.

It sometimes occurs that one has some doubts about the reliability of a source of information inducing the BBA m_Y. In this case the discounting operation can be used to reduce by some factor $\chi \in [0,1]$ the belief assigned by m_Y to the evidence conveyed by that information [Petit-Renaud & Denoeux, 2004]:

$$\begin{aligned} &\widetilde{m}_Y(\mathrm{Y}_k) = (1-\chi) m_Y(\mathrm{Y}_k), \quad \forall \mathrm{Y}_k \in \Omega_Y, \quad \mathrm{Y}_k \neq \Omega_Y \\ &\widetilde{m}_Y(\Omega_Y) = \chi + (1-\chi) m_Y(\Omega_Y) \end{aligned} \tag{7}$$

Notice that the mass assigned to the frame of discernment Ω_Y represent the *ignorance* we have about the value of *Y* because it indicates the absence of evidence that the value of *Y* belongs to any subset Y_k of Ω_Y.

The BFT has been applied to treat uncertain information in classical nonparametric regression by associating to each training pattern of input/output pairs (x_i, y_i) the BBA $m_Y(\mathrm{Y}_i = \{y_i\}) = 1$ having as single focal element the pattern output y_i [Petit-Renaud & Denoeux, 2004].

In the similarity-based approach, it is assumed that the relevance of the information supplied by a training pattern depends on its similarity to the test pattern. Hence, we propose to reduce the belief assigned by the BBA induced by a training pattern proportionally to its dissimilarity to the test pattern by using the discounting operation. In more detail, to apply this procedure to the input/output pairs ($\mathbf{z}^r_{1:n^r}$, RUL^r_{j*}), r=1:R, associating to each reference trajectory its RUL prediction RUL^r_{j*} derived in Section 2.1, we associate to each input/output pair the BBA $m^r_{RUL}(\{RUL^r_{j*}\}) = 1$ and the discounting factor χ, where $\chi = 1 - \gamma \cdot s^r_{j*}$ and $\gamma \in [0,1]$ defines the degree of trust given to the reference trajectories. In particular, the discounting factor χ can be seen as a dissimilarity measure. Thus, the discounted BBAs $\widetilde{m}^r_{RUL}(\{RUL^r_{j*}\})$, r=1:R is obtained from:

$$\begin{aligned} &\widetilde{m}^r_{RUL}(\{RUL^r_{j*}\}) = \gamma \cdot s^r_{j*} \\ &\widetilde{m}^r_{RUL}(\Omega_{RUL}) = 1 - \gamma \cdot s^r_{j*} \end{aligned} \tag{8}$$

The frame of discernment Ω_{RUL} is the domain of RUL_I defined by the interval $[0, \tau_F^{\max} - \tau_I]$, where $\tau_F^{\max}$ is the maximum possible life duration of the equipment provided by an expert. The quantity $\tau_F^{\max} - \tau_I = RUL_I^{\max}$ is the maximum value that can be assumed by the variable RUL at the present time τ_I, whereas 0 is, obviously, the minimum possible value of the equipment RUL.

It is important to notice that if $\gamma < 1$ a part of belief will always be assigned to the *ignorance* represented by Ω_{RUL}, even in the case a reference trajectory were exactly identical to the test one. On the other hand, for $\gamma = 1$, the belief $m^r_{RUL}(\{RUL^r_{j*}\})$ assigned to a prediction RUL^r_{j*} can reach the value of 1 if the reference trajectory r is identical to the test trajectory, so that their similarity is 1: this can produce counterintuitive results (see Section 3.2) and should in general be avoided. In fact, by assigning a belief $m^r_{RUL}(\{RUL^r_{j*}\})$ almost equal to 1 to a single prediction RUL^r_{j*}, generated by a reference trajectory which is, at the present time, almost identical to the test one, we neglect the possibility, indeed relevant, that the two trajectories

diverge in their future evolution due to the intrinsic uncertainty caused by the variability of the degradation process.

Finally, by combining the discounted BBAs $\tilde{m}^r_{RUL}$, r=1:R by the Dempster's rule of combination, we obtain the combined BBA m_{RUL}:

$$\begin{aligned} m_{RUL}(\{RUL^r_{j*}\}) &= \frac{\gamma \cdot s^r_{j*}}{K} \prod_{r' \neq r} (1 - \gamma \cdot s^{r'}_{j*}), \quad r = 1:R \\ m_{RUL}(\Omega_{RUL}) &= \frac{1}{K} \prod_{r=1}^{R} (1 - \gamma \cdot s^r_{j*}) \end{aligned} \tag{9}$$

where

$$K = \prod_{r=1}^{R} (1 - \gamma \cdot s^r_{j*}) + \gamma \cdot \sum_{r=1}^{R} s^r_{j*} \prod_{r' \neq r} (1 - \gamma \cdot s^{r'}_{j*}) \tag{10}$$

Within the framework of BFT, a method for estimating the expected value for the variable RUL_I from the BBA in eq. (9) is given by the pignistic expectation $\hat{RUL}^{PG}_I$ [Smets, 1994; Petit-Renaud & Denoeux, 2004]:

$$\hat{RUL}^{PG}_I = \sum_{r=1}^{R} m_{RUL}(\{RUL^r_{j*}\}) \cdot RUL^r_{j*} + m_{RUL}(\Omega_{RUL}) \frac{RUL^{\max}_I}{2} \tag{11}$$

The main difference between the RUL prediction given by the similarity-based approach $\hat{RUL}^{SB}_I$ in eq. (4) and that given by the pignistic expectation $\hat{RUL}^{PG}_I$ in eq. (11) is that the latter also considers the degree of *ignorance*, i.e., the part of belief assigned indiscriminately to the entire domain Ω_{RUL} represented by the second term of eq. (11). In practice, the pignistic expectation $\hat{RUL}^{PG}_I$ can be seen as a similarity weighted average of the reference trajectories RUL with the addition of a term corresponding to the middle of the RUL domain, which can be interpreted as the *prior* information about the RUL value in the absence of any empirical evidence, weighted by a measure of the overall dissimilarity of the test trajectory with the training trajectories.

Given the BBA in eq. (9), we can also calculate the belief associated to any interval $[RUL^{\inf}_I, RUL^{\sup}_I]$ as the sum of the belief masses associated to all subsets included into $[RUL^{\inf}_I, RUL^{\sup}_I]$. The belief associated to an interval Δ^+_I represents the amount of belief that directly supports the hypothesis $RUL^{test}_I \in [RUL^{\inf}_I, RUL^{\sup}_I]$, where RUL^{test}_I is the true RUL of the test equipment, and it has been interpreted as a lower bound for the probability that $RUL^{test}_I \in [RUL^{\inf}_I, RUL^{\sup}_I]$ or, analogously, as an upper bound for the probability that $RUL^{test}_I \notin [RUL^{\inf}_I, RUL^{\sup}_I]$. Thus, a left bounded interval $\Delta^+_I(\alpha) = [RUL^{\inf}_I(\alpha), +\infty]$ such that a belief $1-\alpha$ is assigned to it provides the following information about the probability distribution of the true equipment RUL, RUL^{test}_I: $P\left(RUL^{test}_I > RUL^{\inf}_I(\alpha)\right) > 1-\alpha$ or, equivalently, $P\left(RUL^{test}_I < RUL^{\inf}_I(\alpha)\right) < \alpha$. This latter

interpretation of $RUL_I^{\inf}(\alpha)$ can be used to plan the maintenance action: performing maintenance before $RUL_I^{\inf}(\alpha)$ guarantees a probability of failure lower than α.

3 Numerical application

In this Section we verify the proposed method on simulated and real data.

Performance are evaluated on N^{tst} test trajectories $\mathbf{z}_{1:I}^q$, q=1:N^{tst}, and N^{trn} different training sets $\{\mathbf{z}_{1:n^r}^r, r=1:R\}_l$, l=1:N^{trn}. Let us define $\hat{RUL}_{I^q(\beta)}^l$ and $RUL_{I^q(\beta)}^{\inf}(\alpha)$, l=1:N^{trn}, q=1:N^{tst}, as the predictions obtained at a time instant $\tau_{I^q(\beta)} = \beta\tau_F^q$ corresponding to the fraction β of the life duration of the q–th test equipment, when training set l is used. Three performance indicators are considered:

- The Mean Square Error (MSE), i.e., the mean value of the square error $(\hat{RUL}_{I^q(\beta)}^l - RUL_{I^q(\beta)}^{true})^2$ made in predicting the true RUL, $RUL_{I^q(\beta)}^{true}$ of the test equipment. The MSE measures the accuracy of the prediction $\hat{RUL}_{I^q(\beta)}^l$ and is desired to be as small as possible.
- The Coverage (cov_α) of the prediction interval $\Delta_I^+(\alpha) = [RUL_{I^q(\beta)}^{\inf}(\alpha), +\infty]$, i.e., the percentage of times the condition $RUL_{I^q(\beta)}^{true} > RUL_{I^q(\beta)}^{\inf}(\alpha)$ is verified, where $1-\alpha$ is the belief associated by the RUL BBA to the interval $\Delta_I^+(\alpha)$. This indicator measures the reliability of the confidence interval; we expect to obtain values of cov_α larger than $1-\alpha$, since the belief $1-\alpha$ associated to interval $\Delta_I^+(\alpha)$ is a lower bound to the possibility that the test equipment RUL belongs to it.
- The mean amplitude (MA_α) of the interval $[RUL_{I^q(\beta)}^{\inf}(\alpha), \hat{RUL}_{I^q(\beta)}^l]$, which gives a measure of the precision of the RUL prediction. In order to have a high precision, we wish the keep the value of MA_α as small as possible.

The performance indicators are calculated for three value of β : $\beta_1 = 25\%$, $\beta_2 = 50\%$ and $\beta_3 = 75\%$.

In the following Section 3.1 the similarity-based method is applied to simulated data concerning the evolution of creep damage in ferritic steel. The influence on the prognostic performance of parameters λ and γ is investigated and some indications about the choice of their values is derived.

On the basis of these results, in Section 3.2, the method is applied to real data taken from a case study about the clogging of filter in a BWR condenser.

3.1 Artificial dataset: Creep growth in ferritic steel

Ferritic steels are widely used for welded steam pipes in the construction of power plant components that operate under high temperature and stress conditions. In such conditions the creep deformation and rupture are important factors in determining the equipment lifetimes.

3.1.1 Creep growth models

We have simulated the evolution of creep damage in ferritic steel exposed to the load σ using the uni-axial form of the non-linear creep constitutive equations proposed within the framework of Continuum Damage Mechanics by Mustata & Hayhurst (2005):

$$\begin{cases} \dot{\varepsilon} = A\sinh\left[\dfrac{B\sigma(1-H)}{(1-\phi)(1-\omega)}\right] \\ \dot{H} = \dfrac{h\dot{\varepsilon}}{\sigma}\left(1-\dfrac{H}{H^*}\right) \\ \dot{\phi} = \dfrac{K_c}{3}(1-\phi)^4 \\ \dot{\omega} = C\dot{\varepsilon} \end{cases} \tag{12}$$

where ε is the creep strain, i.e., the percentage of elongation of the turbine blade in the longitudinal direction with respect to its original length, ϕ and ω are two damage state variable describing, respectively, the coarsening of the carbide precipitates, and the inter-granular creep constrained cavitation damage, H is the hardening state variable, used to represent the strain hardening effect attributed to primary creep, and A, B, H^*, h, K_c, and C are material inherent characteristics. Each characteristic $\varphi_m = A, B, H^*, h, K_c, C$ varies with the temperature according to the Arrhenius law, i.e., $\varphi_m = \varphi_{m0}\exp(-Q_m / T)$, m=1:6, where T is the operating temperature and φ_{m0} and Q_m are parameters whose values have to be determined experimentally.

To generate different trajectories, the intrinsic variability of the degradation process is simulated by sampling the values of the load σ and temperature T to which the steel is exposed at each time step from a normal distributions centered on their mean values, whereas the variability of the degradation process of similar pieces of equipment is simulated by sampling the values of parameters φ_{m0} and Q_m, m=1:6, at the beginning of each new simulated degradation trajectory. Finally, in order to generate the sequence of observations $\mathbf{z}^r_{1:n^r} = \{\varepsilon(\tau_i) + \nu_i\}^r_{i=1:n^r}$, collected with a sampling time of 1000h, a white Gaussian noise ν_i is added to the simulated creep strain $\varepsilon(\tau_i)$ at the observation time τ_i. We assume failure to happen when the limiting creep strain value of 2% is reached. Figure 1 shows an example of simulate creep growth trajectory (upper) and the corresponding sequence of observations $\mathbf{z}^r_{1:n^r}$ (bottom).

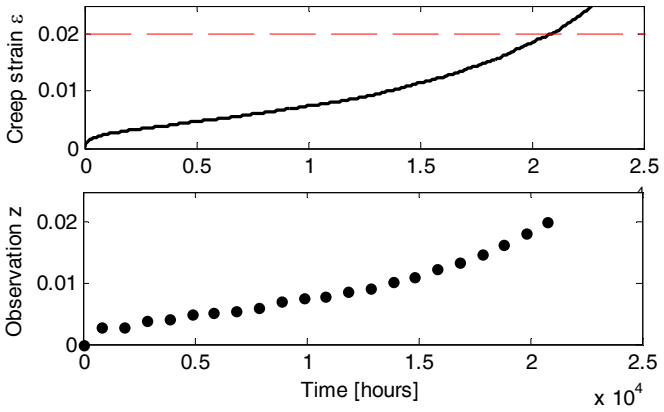

Figure 1: example of simulated creep growth trajectory (upper) with the corresponding sequence of observations (bottom).

3.1.2 Results

The model in eq.(12) is used to generate N^{tst}=50 test trajectories and N^{trn}=50 sets of R=7 reference trajectories. The similarity-based approach proposed in this work is applied to each test trajectory at the three different life fractions $\beta_{1:3}$, using the N^{trn} different sets of reference trajectories simulated. Two RUL predictions are obtained using the similarity-based weighted average in eq. (11), $\hat{RUL}_I^{SB}$, and the pignistic expectation of the BBA in eq.(4). The prediction interval is estimated using the target belief $1-\alpha=0.8$.

In Figure 2 the variation of the square root of the MSE indicator with parameter λ is shown for the three life fractions $\beta_{1:3}$ considered; the performance obtained by the prediction $\hat{RUL}_I^{SB}$ (continuous line), and by the prediction $\hat{RUL}_I^{PG}$ in correspondence of two different values of parameter γ, $\gamma=1$ (thick dashed line) and $\gamma=0.8$ (thin dashed line), are compared. As expected, the prediction error decreases as the life fraction β increases, i.e., as we get closer failure. Results in Figure 2 show that the accuracy of the pignistic prediction $\hat{RUL}_I^{PG}$ in terms of MSE increases with λ and with γ, since the product $\gamma \cdot s_{j*}^r$ decreases with the values of these two parameters (notice that, if λ is low the similarity s_{j*}^r is low for most trajectories) and thus a large belief is assigned to the *ignorance,* i.e., to the domain Ω_{RUL} of RUL_I. This interval is not considered in the similarity-based weighted average $\hat{RUL}_I^{SB}$ whose maximum accuracy obtained for values of parameter λ around 5×10^{-5} remains always higher that the accuracy of $\hat{RUL}_I^{PG}$. The prediction $\hat{RUL}_I^{PG}$ outperforms $\hat{RUL}_I^{SB}$ for value of λ larger than 5×10^{-4} only if parameter γ is close to 1. Due to its higher accuracy, only the similarity-based weighted average $\hat{RUL}_I^{SB}$ will be considered hereafter.

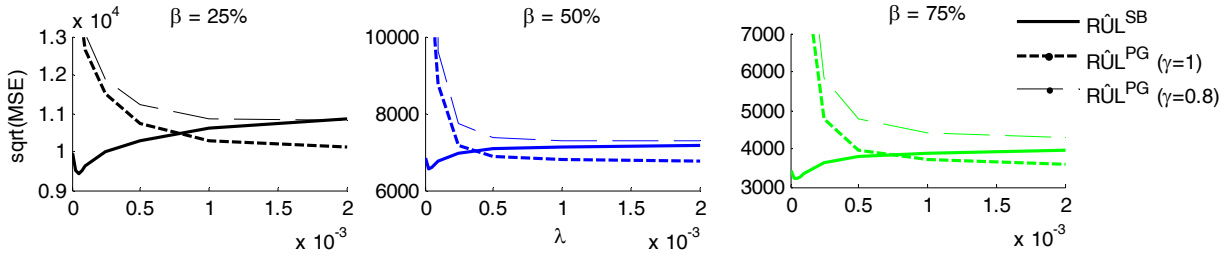

Figure 2: comparison of square root of the MSE made by the prediction $\hat{RUL}_I^{SB}$ obtained using similarity-based weighted average (continuous line) and the prediction $\hat{RUL}_I^{PG}$ obtained as the pignistic expectation from a BBA computed using $\gamma=1$ (thick dashed line) and $\gamma=0.8$ (thin dashed line) as a function of parameter λ.

The precision of the prediction, which is evaluated by the indicator MA_α, is also an important aspect to be considered in the optimization procedure. However, the choice of parameters λ and γ should be subordinated to the verification that the coverage Cov_α is actually larger than $1-\alpha$. Lower values of the coverage would indicate that a too large belief mass has been assigned to the predictions RUL_{j*}^r provided by the reference trajectories most similar to the test trajectory, so that the belief $1-\alpha$ assigned to the prediction interval is not justified by the experimental evidence.

Figure 3 shows the coverage $Cov_{0.2}$ of the left bounded prediction interval $\Delta_I^+(0.2)$ (upper), the square root of the MSE made by the prediction $\hat{RUL}_I^{SB}$ (middle), and the mean amplitude $MA_{0.2}$ of the interval $[RUL_{I^q(\beta)}^{\inf}(0.2), \hat{RUL}_{I^q(\beta)}^{SB,l}]$ (bottom) in correspondence of three different values of parameter λ as a function of the parameter γ.

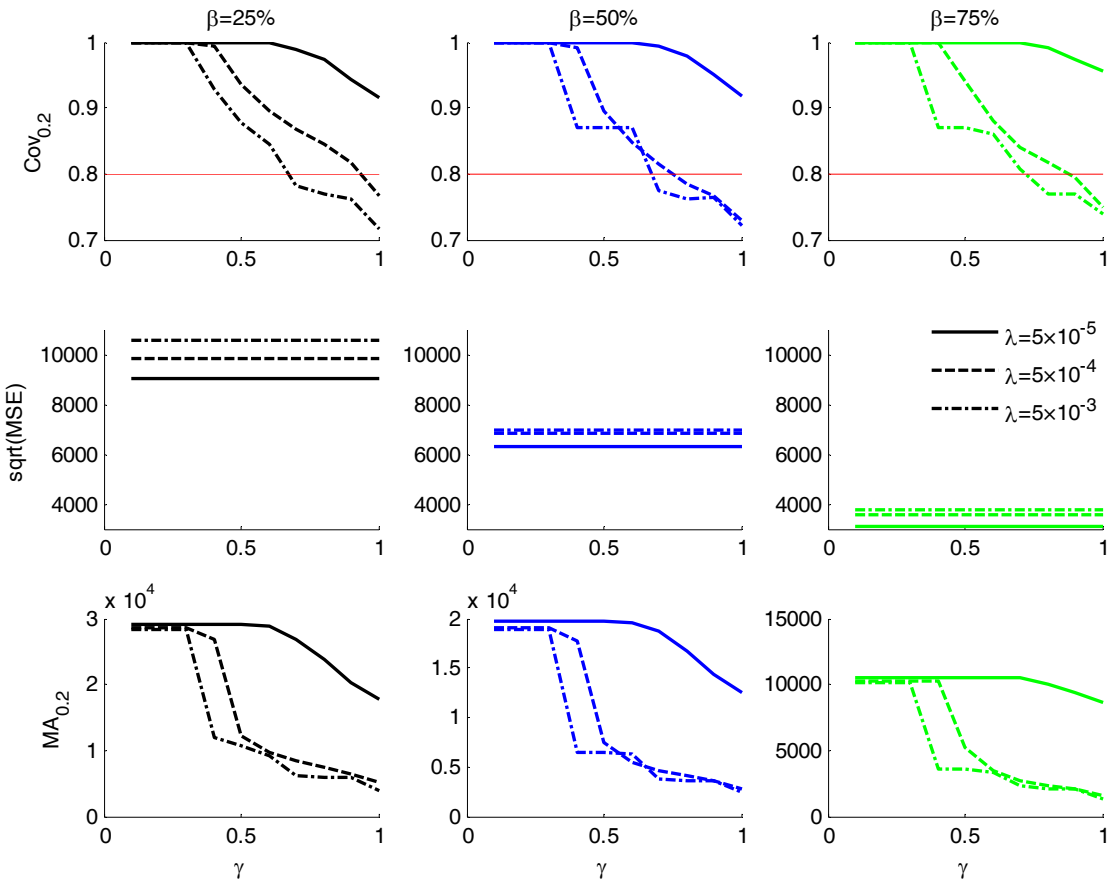

Figure 3: value of the three performance indicators as a function of γ *at three fractions* β *of the trajectory life durations and for three values of* λ.

For the value of $\lambda = 5\times10^{-5}$ that maximizes the accuracy of the prediction $\hat{RUL}_I^{SB}$ the coverage is always larger than the minimum accepted value of $1-\alpha = 0.8$. However, for such a small value of λ the precision, represented by the indicator $MA_{0.2}$, is much lower than for $\lambda = 5\times10^{-4}$ and $\lambda = 5\times10^{-3}$. This is further investigated in Figure 4 which shows the RUL prediction with the relative prediction interval for a specific trajectory (left) in correspondence of two different values of parameter λ: $\lambda = 5\times10^{-5}$ (upper) and $\lambda = 5\times10^{-4}$ (bottom). Notice that for $\lambda = 5\times10^{-5}$ the lower bound of the prediction interval is equal to 0 for large part of the trajectory (Figure 4, upper, left); this does not mean that the evidence of very early failure is high (as demonstrated by the fact that the predicted RUL is far from 0), but only that the evidence drawn from the reference trajectories is not sufficient to assert with the desired belief $1-\alpha = 0.8$ that the RUL value is actually larger than 0. In other words, the prediction $RUL_I^{\inf} = 0$ is a statement of *ignorance* about the value of RUL_I. Contrarily, in the case of $\lambda = 5\times10^{-4}$ (Figure 4, bottom, left) the lower bound of the prediction interval is always higher than 0. Figure 4, right shows the values of the similarity s_{j*}^r assigned to each reference trajectory r=1:7 and the BBA m_{RUL} assigned to the corresponding prediction RUL_{j*}^r and to the RUL domain Ω_{RUL} at time $\tau_{23} = 21811$ hours, which is characterized by a confidence bound equal to 0 using $\lambda = 5\times10^{-5}$. Notice that the similarities s_{j*}^r obtained using $\lambda = 5\times10^{-5}$ are significantly lower than those obtained using $\lambda = 5\times10^{-4}$, and, consequently, the mass $m_{RUL}(\Omega_{RUL})$ assigned to the RUL

domain using $\lambda = 5\times10^{-5}$ is larger than 0.2, so that the total belief assigned to the trajectories predictions RUL^r_{j*} does not reach the required value of 0.8. This is due to the fact that if λ is small, the similarity of a reference trajectory tends to result small except that in the rare case of trajectory very similar to the test trajectory. As a consequence, for very small values of λ, it is often hard to support with sufficient evidence the hypothesis that the RUL value belongs to any subset of the RUL domain Ω_{RUL}.

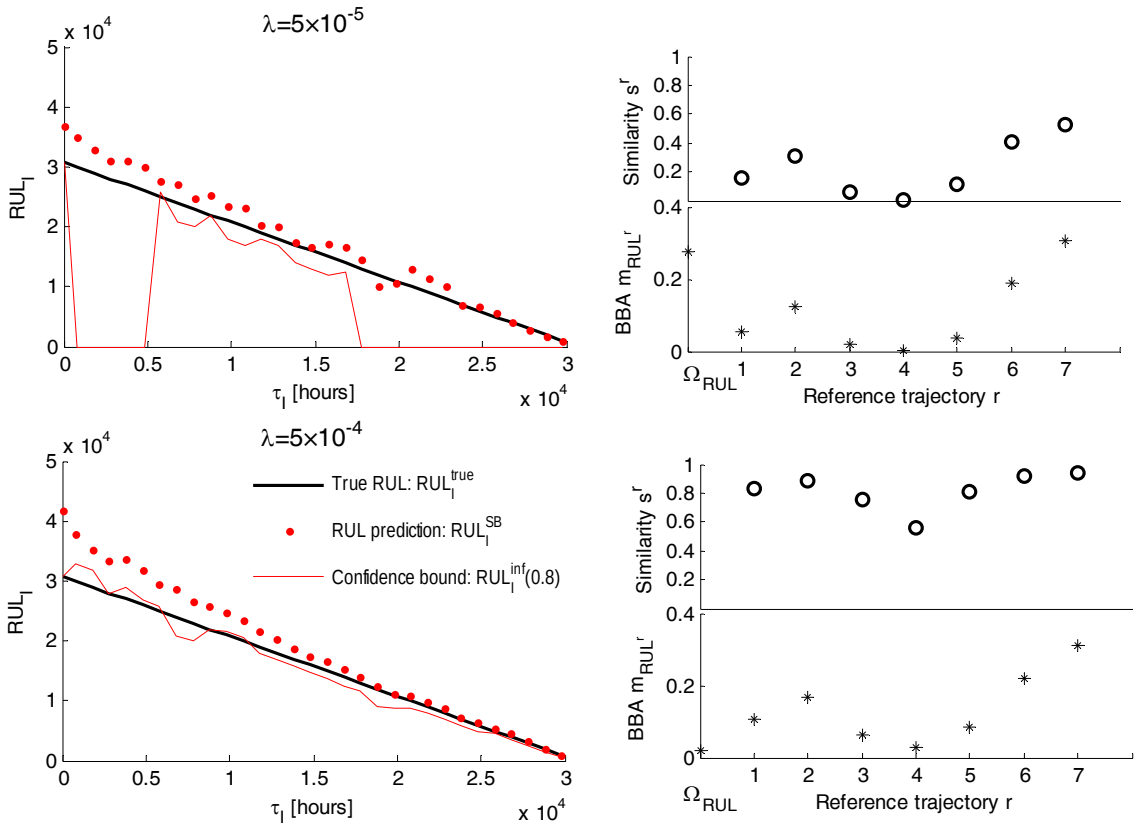

Figure 4: comparison of the RUL prediction with confidence bound (left) and the similarity values and BBAs assigned to the different trajectories at $\tau_{23} = 21811$ *hours (right) for two values of* λ*:* $\lambda = 5\times10^{-5}$ *(upper) and* $\lambda = 5\times10^{-4}$ *(bottom).*

With respect to the optimal choice of the parameters γ and λ, we have adopted the following procedure:

1. we derive conditions on the value of γ as a function of the value of λ by imposing $Cov_{0.2}$>0.8 (for instance, in this case study we have $\gamma <= 0.7$ if $\lambda = 5\times10^{-4}$ and $\gamma <= 0.6$ if $\lambda = 5\times10^{-3}$).
2. For each λ we consider the maximum acceptable value of γ, since in correspondence of it the minimum amplitude $MA_{0.2}$ is obtained and evaluate the square root of the MSE indicator in correspondence of it. The tradeoff between prediction accuracy and precision is considered to set the values of parameters λ and γ.

Table I shows for the three values of lambda considered, the square root of the MSE indicator, the maximum acceptable value of γ, γ_{max}, and the amplitude $MA_{0.2}$ and coverage $Cov_{0.2}$ in correspondence of γ_{max}. Based on these results, we set the parameters to the values $\lambda = 5\times10^{-4}$ and $\gamma = \gamma_{max} = 0.7$, since for this value of λ performance are better with respect to the value $\lambda = 5\times10^{-3}$ both in terms of accuracy and

precision, whereas, compared to $\lambda = 5\times10^{-5}$, a large improvement of the precision is observed, against a small reduction in the accuracy.

Table I: performance indicators for three different value of λ in correspondence of $\gamma_{\max}$.

	λ		
	5×10^{-5}	5×10^{-4}	5×10^{-3}
	$\gamma_{\max}$		
	1.0	0.7	0.6
	$\mathrm{Cov}_{0.8}$		
$\beta = 25\%$	0.915	0.868	0.846
$\beta = 50\%$	0.918	0.814	0.870
$\beta = 75\%$	0.955	0.840	0.861
	$\sqrt{MSE}$ (10^3)		
$\beta = 25\%$	9.065	9.851	10.552
$\beta = 50\%$	6.272	6.807	6.976
$\beta = 75\%$	3.089	3.589	3.793
	$\mathrm{MA}_{0.8}$ (10^3)		
$\beta = 25\%$	17.871	6.217	9.202
$\beta = 50\%$	12.571	4.568	6.233
$\beta = 75\%$	8.646	2.685	3.304

Figure 5 shows the predictions obtained at all measurement time instants τ_I of 4 different test trajectories. Two phenomena can be observed in this Figure: first, some situations of *ignorance* about the value of RUL_I where $RUL_I^{\inf} = 0$, are still encountered. This is due to the fact that the information provided by the reference trajectories is not relevant for a specific test trajectory, e.g., because they are too dissimilar. Another noticeable phenomenon in Figure 5 is the presence of large jumps of the confidence bound $RUL_I^{\inf}$. These jumps occur when the reference trajectory corresponding to the minimum RUL prediction RUL_{j*}^r included in the prediction interval in order to attain the desired belief $\alpha = 0.8$ changes.

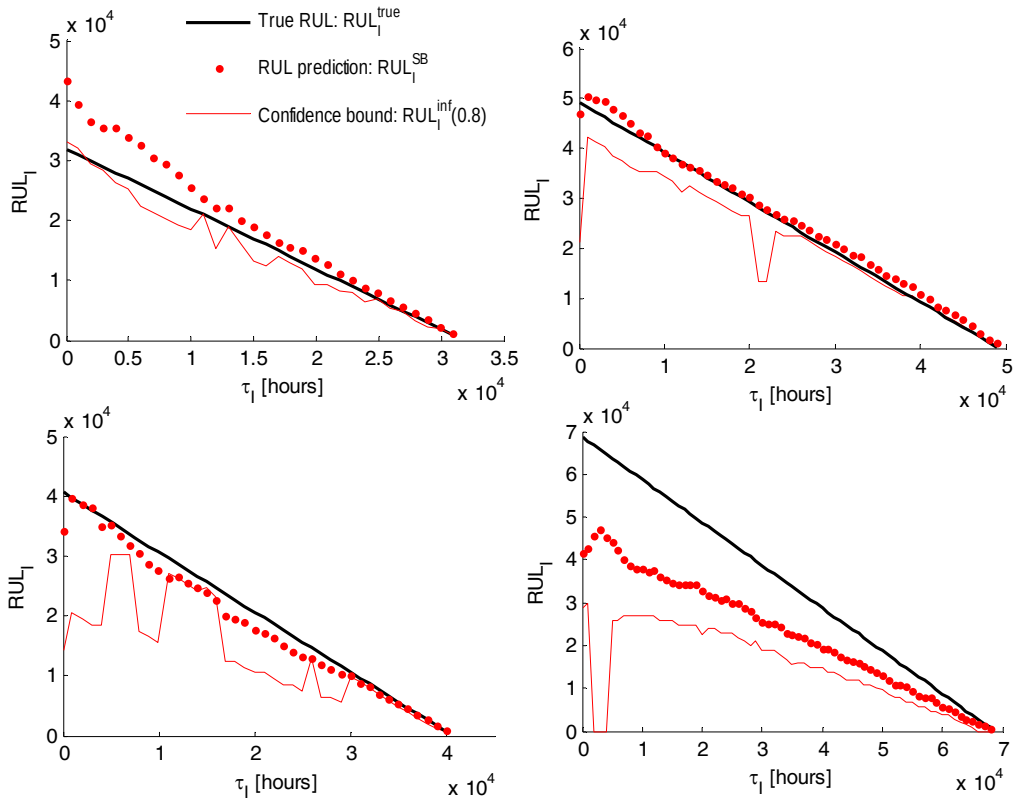

Figure 5: predictions obtained for 4 different test trajectories using $\lambda = 5\times10^{-4}$ *and* $\gamma = 0.7$.

In order to investigate this undesired effect, a new test trajectory (Figure 6, bottom, continuous line) is built by considering parts of different training trajectories (Figure 6, bottom, dots and dashed lines). In particular, the first 7 observations are taken from the reference trajectory r=7 and the last 13 from trajectory r=3. In order to simplify the analysis of the results, no noise is added to the creep strain values simulated, neither for the reference, nor for the test trajectories. RUL predictions are performed during the evolution of such artificial trajectory using $\gamma = 0.7$ and the index r_{inf} of the reference trajectory producing the RUL prediction $RUL_{j*}^{r_{\inf}} = RUL_I^{\inf}$ corresponding to the confidence bound, is recorded and shown in Figure 6 (left, upper). An index r_{inf}=0 indicates that the lower bound is given by the RUL domain Ω_{RUL} , i.e., by the situation of *ignorance* where $RUL_I^{\inf} = 0$. Evident jumps of the confidence bound can be observed when the index r_{inf} changes

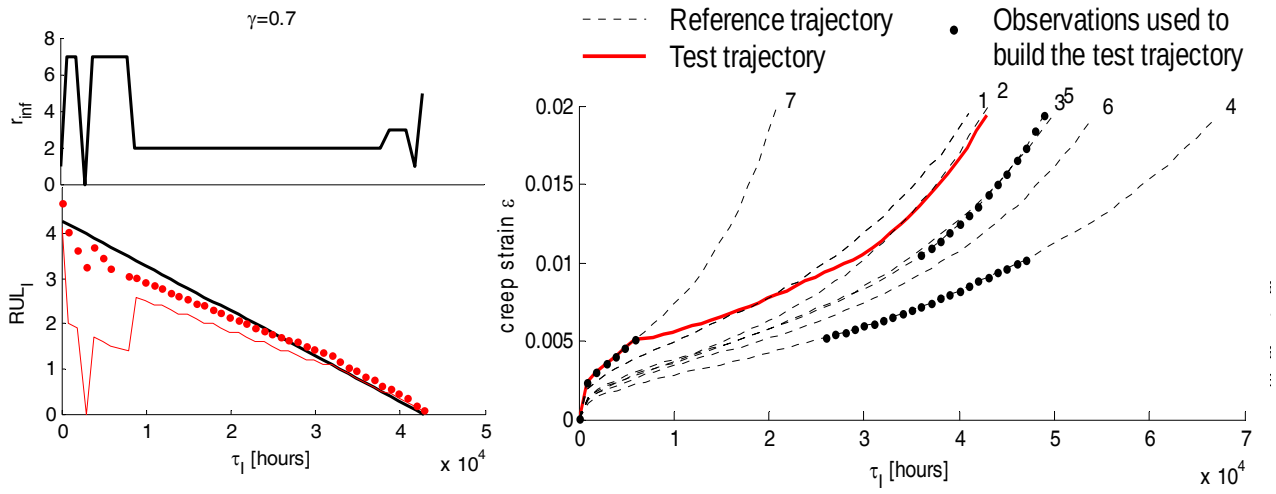

Figure 6: Prediction (left) obtained for an artificial test trajectory (right) using $\gamma = 0.7$.

Although justified by the method, the oscillations of the confidence bound may be confusing for the maintenance planner. A reduction in the oscillations can be obtained by increasing the value of lambda λ or reducing the value of γ, at the price of a lower accuracy and precision.

Figure 7 shows the RUL predictions obtained for the same four trajectories of Figure 5 using for the parameters the values $\lambda = 5\times10^{-3}$ and $\gamma = 0.5$. Table II compares the performance of the prediction computed on N^{tst}=50 test trajectories different from those used for optimizing the parameters, in this case and in the case of Figure 5 where $\lambda = 5\times10^{-4}$ and $\gamma = 0.7$. In the Table, the mean value of the RUL, $\overline{RUL}$, for different values of the life fraction β is also shown, and the performance indicators $\sqrt{MSE}$ and $MA_{0.2}$ are expressed also as a percentage of $\overline{RUL}$.

Table II: RUL prediction performance.

		$\beta = 25\%$		$\beta = 50\%$		$\beta = 75\%$	
$\overline{RUL}$ (10^4)		2.711		1.827		0.975	
$Cov_{0.8}$	$\lambda = 5\times10^{-4}$; $\gamma = 0.7$	0.782		0.814		0.849	
	$\lambda = 5\times10^{-3}$; $\gamma = 0.5$	0.838		0.850		0.853	
$\sqrt{MSE}$ (10^3)	$\lambda = 5\times10^{-4}$; $\gamma = 0.7$	9.152	33.8%	5.965	32.6%	3.191	32.7%
	$\lambda = 5\times10^{-3}$; $\gamma = 0.5$	9.822	36.2%	6.160	33.7%	3.411	35.0%
$MA_{0.8}$ (10^3)	$\lambda = 5\times10^{-4}$; $\gamma = 0.7$	8.445	31.2%	5.594	30.6%	3.228	33.1%
	$\lambda = 5\times10^{-3}$; $\gamma = 0.5$	11.300	41.7%	7.167	39.2%	3.960	40.6%

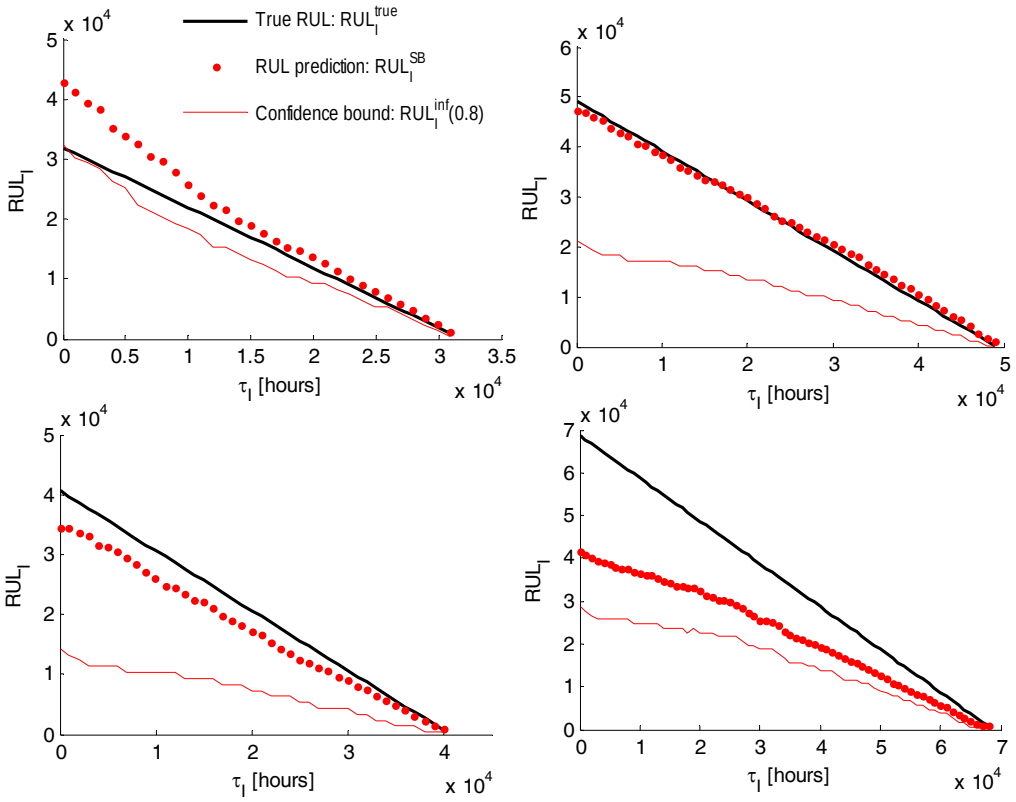

Figure 7: predictions obtained for 4 different test trajectories of Figure 5 using $\lambda = 5\times10^{-3}$ *and* $\gamma = 0.5$.

The results of Figure 7 and Table 2 confirm that the oscillation of the confidence bound can be damped down by increasing the value of lambda λ or reducing the value of γ, but this choice increases the prediction error and the amplitude $MA_{0.2}$. Clearly, to an increased $MA_{0.2}$ corresponds also a higher value of the coverage indicator $Cov_{0.2}$.

When a situation with a larger density of reference trajectories is considered, the oscillations of the lower bound becomes of smaller amplitude, although more frequent. This happens, for example, when a larger number R of reference trajectories is available or when the variability within the degradation trajectories becomes smaller. To show this, we have reduced the variance of the parameters φ_{m0} and Q_m, m=1:6, and of the load σ and temperature T used in the model of eq.(12) to simulate N^{tst}=50 test trajectories and N^{trn}=50 training sets of R=50 reference trajectories. The optimization procedure applied for the case with R=7 has been used to set the parameters to the values $\lambda = 5\times10^{-5}$ and $\gamma = 0.95$. Four examples of the predictions obtained are shown in Figure 8, whereas the values of the performance indicators are presented in Table 3. As expected, with a higher density of training trajectories available, the prediction is both more accurate and precise.

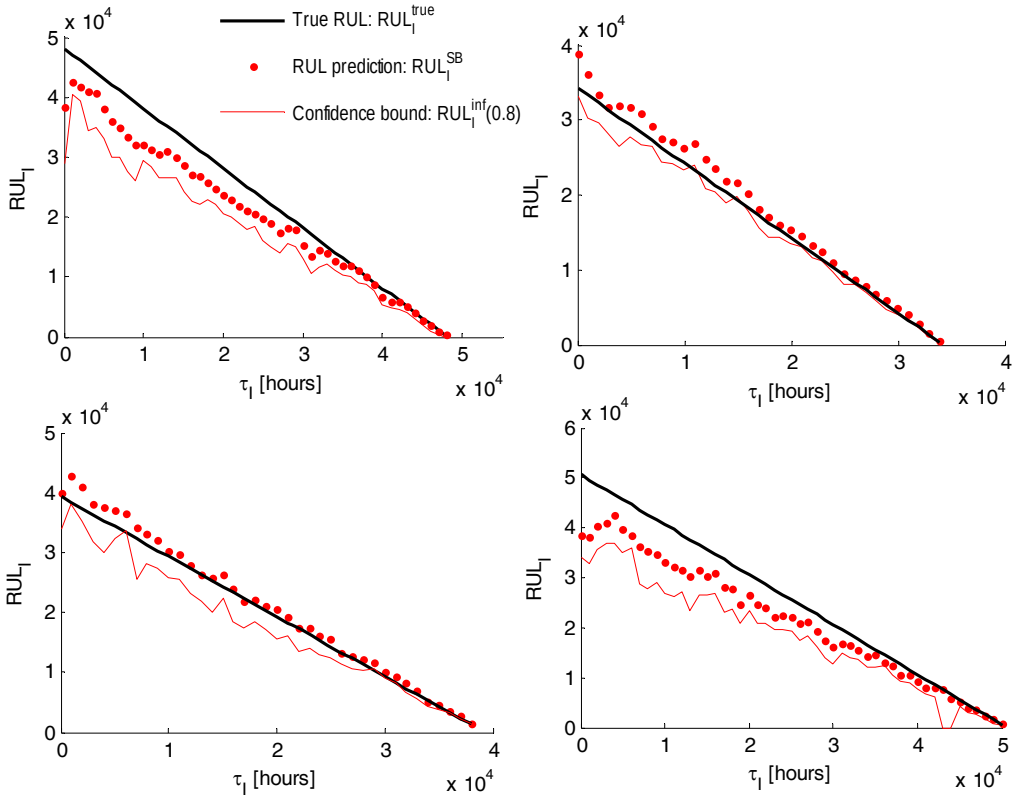

Figure 8: predictions obtained for 4 different test trajectories using R=50 reference trajectories and parameters $\lambda = 5\times10^{-5}$ *and* $\gamma = 0.95$.

Table III: RUL prediction performance with $\lambda = 5\times10^{-5}$ *and* $\gamma = 0.95$

	$\beta = 25\%$		$\beta = 50\%$		$\beta = 75\%$	
$\overline{RUL}$ (10^4)	3.108		2.090		1.102	
$\text{Cov}_{0.8}$	0.814		0.832		0.808	
$\sqrt{MSE}$ (10^3)	5.313	17.1%	3.444	16.5%	1.659	15.1%
$\text{MA}_{0.8}$ (10^3)	4.961	16.0%	3.187	15.2%	1.788	16.2%

3.2 Real dataset: Clogging of BWR condenser filters

In this Section, we consider the heat exchanger filters used to clean the sea water entering the condenser of the BWR reactor of a Swedish nuclear power plant. During operations, filters undergo clogging and, once clogged, can cumulate particles, seaweed, and mussels from the cooling water in the heat exchanger. For this reason, prompt and effective cleaning of the filter is desirable; predictive maintenance can help achieving this result, keeping maintenance costs reasonably low. From data collected on field, we have available sequences of observations $\mathbf{z}^q_{1:n^q}$, q=1:8 taken during the clogging process of Q=8 historical filters. Each observation $\mathbf{z}^q_i = [\Delta P^q_i, \dot{M}^q_i, T^q_i]$ contains the measurements of the pressure drop ΔP^q_i, the flow across the filter $\dot{M}^q_i$, and the sea water temperature T^q_i collected at time τ_i during the clogging process of the q-th filter. It is known that the clogging process is affected by large uncertainties, due to the variable conditions

of the sea water; in this context, the challenge is to provide a sufficiently narrow confidence interval for the RUL prediction.

The prognostic method proposed is applied to each trajectory q at the three life fractions $\beta_{1:3}$, using the remaining R=7 trajectories as reference trajectories. Figure 9 shows how the three performance indicators vary with parameter γ for three values of parameter λ. These results confirm those obtained for the simulated creep growth data: the MSE has a minimum around $\lambda = 0.05$ and the value of the $MA_{0.2}$ indicator decreases with both λ and γ. Notice also that, for the values of λ considered in Figure 9, almost all possible values of γ, are acceptable since the coverage $Cov_{0.2}$ is always larger than 0.8, except for $\lambda = 0.1$ and $\gamma = 1$. The precision obtained for $\lambda = 0.05$ when $\gamma = 1$ is very close to that obtained for $\lambda = 0.1$, whereas the error is lower. Then, we set $\lambda = 0.05$ and $\gamma = 1$ and generate a prognostic prediction in correspondence of each observation available. Results are shown in Figure 10 for all Q=8 test trajectories available.

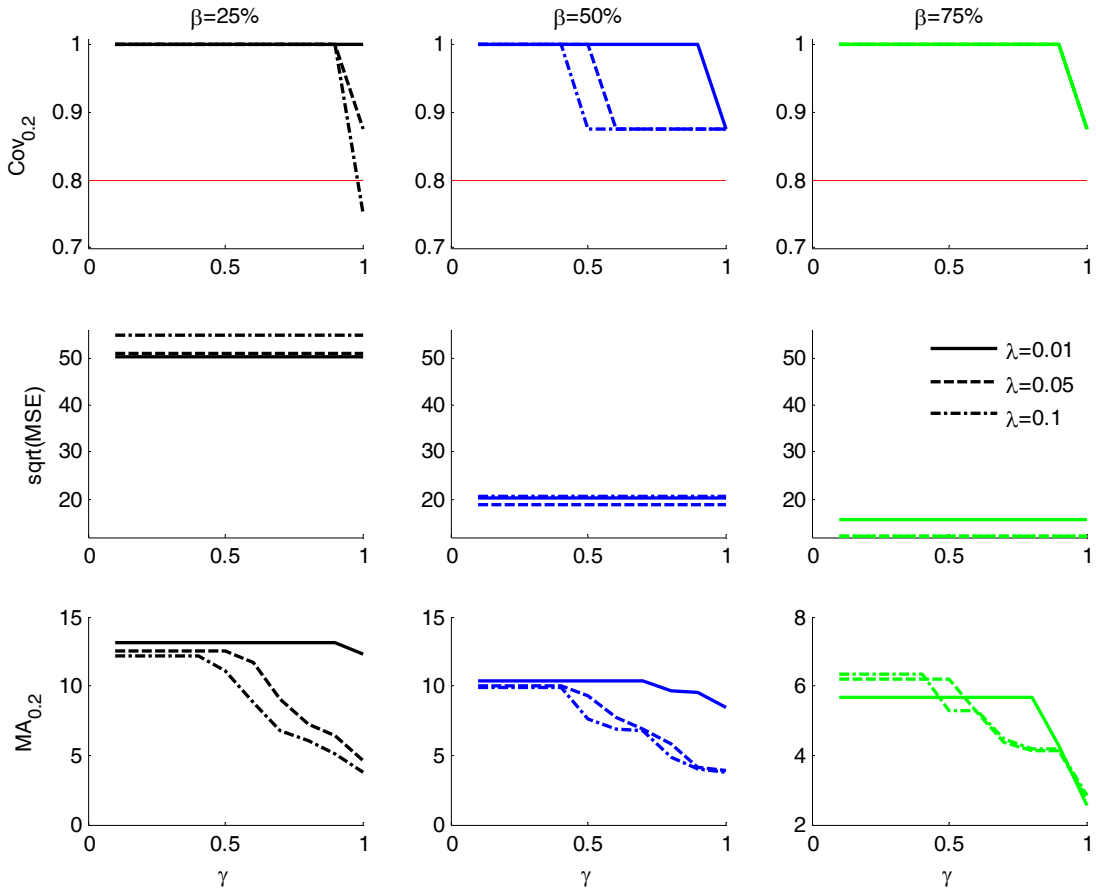

Figure 9: value of the three performance indicators as a function of γ at three fractions β of the trajectory life durations and for three values of λ.

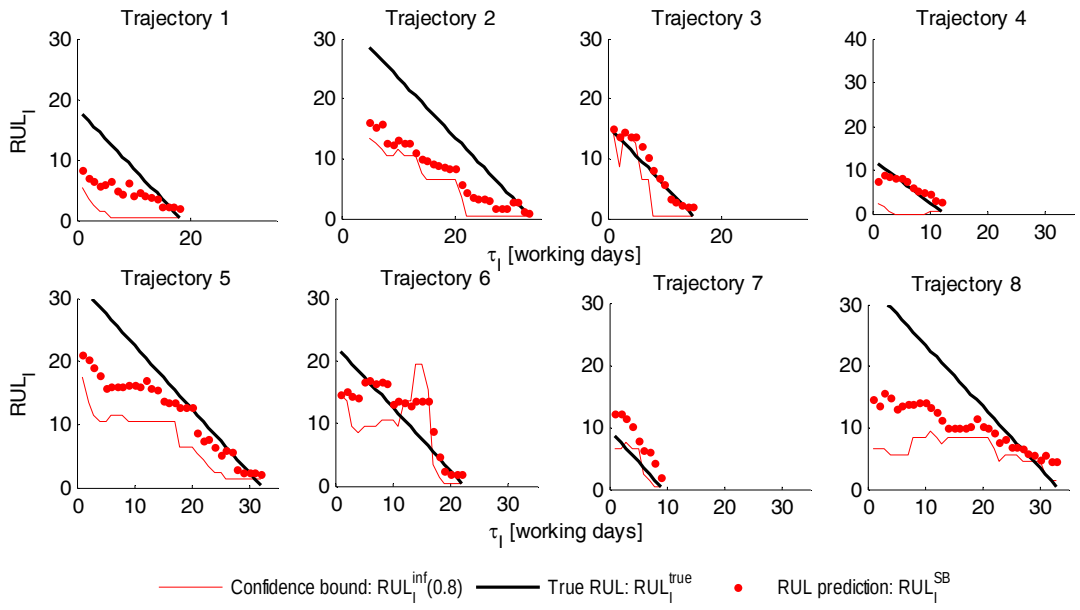

Figure 10: predictions obtained for the Q=8 filter clogging trajectories available using parameters $\lambda = 0.05$ *and* $\gamma = 1$.

In trajectory 4, the confidence bound is for most of the time equal to zero. This means that its similarity with all reference trajectories is rather low, and thus the prediction is very uncertain. In trajectory 6, we notice that the confidence bound is higher than the RUL prediction. This is an example of the counterintuitive results that can be obtained by setting $\gamma = 1$ if two trajectories are very similar. Figure 11 shows the similarities s_{j*}^{r} and the BBAs m_{RUL} assigned to the reference trajectories for the test trajectory 6 at time $\tau_{15} = 15$ working days (upper). We notice that trajectory 8 receives the belief assignment $m_{RUL}(\{RUL_{j*}^{8}\}) = 0.937$. Figure 11 also shows the evolution of the observable parameters ΔP_i^q, $\dot{M}_i^q$, and T_i^q (bottom), for the test trajectory 6 and the reference trajectory 8 receiving the maximum belief assignment. We notice that all three parameters ΔP_i^q, $\dot{M}_i^q$, and T_i^q of the two trajectories are very similar around time $\tau_{15} = 15$ but evolve very differently after that time.

To correct this problem, it is sufficient to reduce the value of parameter γ. Figure 12 shows the predictions obtained for the Q=8 trajectories, when parameters $\lambda = 0.05$ and $\gamma = 0.95$ are used. Results show that in many cases the accuracy is rather low and the prediction interval large. However, due the small number of training trajectories available and the large uncertainties affecting the clogging process, we can be satisfied with this result.

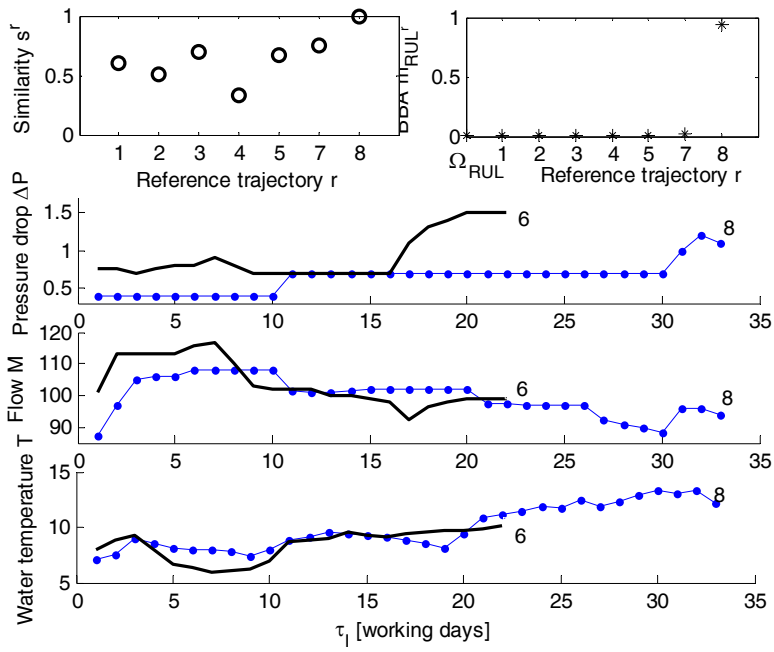

Figure 11: evolution of the three obbservable parameters ΔP_i^q, $\dot{M}_i^q$, *and* T_i^q *(bottom) for trajectories q=6 and a=8, with similarities* s_{j*}^r *and BBAs* m_{RUL} *at time* $\tau_{15} = 15$ *working days (upper).*

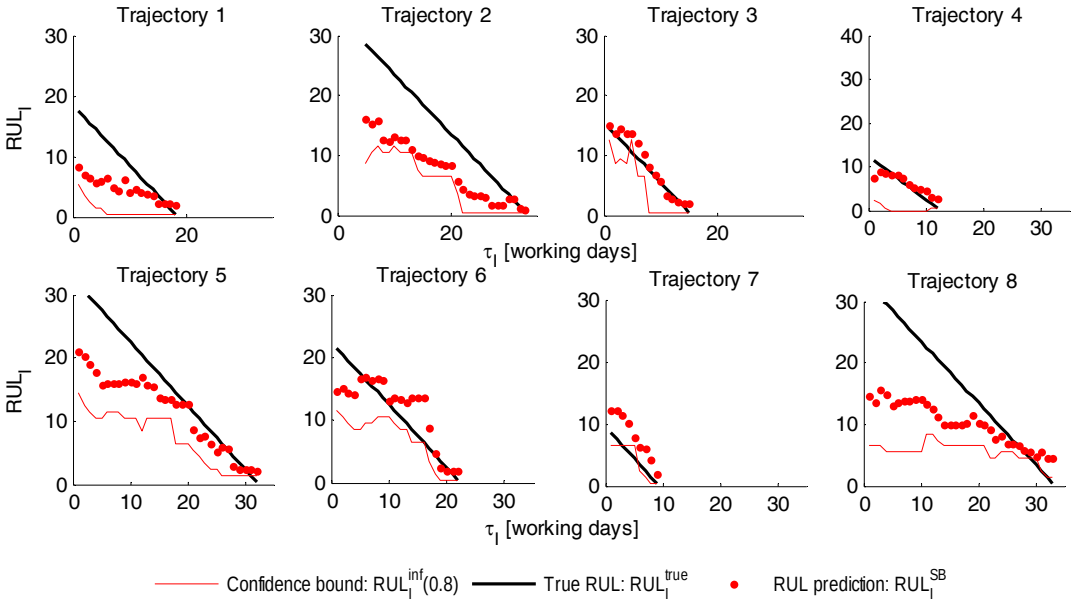

Figure 12: predictions obtained for the Q=8 *filter clogging trajectories available using parameters* $\lambda = 0.05$ *and* $\gamma = 0.95$.

4 Conclusions

In this work, we have considered the problem of directly predicting the RUL of degrading equipment from data and provide a measure of confidence on the prediction, based on a set of reference degradation trajectories followed by similar equipments which have failed in the past. For this we resorted to a similarity-based approach for the prediction of equipment RUL. Since point predictions of the RUL are difficult to use

in practice, without a reliable measure of their uncertainty, we have integrated the similarity-based approach in the framework of belief function theory, and provided a method for supplying a prediction interval for the value of the RUL.

Two key elements in the application of the method are the parameter λ, which defines how strong is the desired interpretation of similarity, and parameter $\gamma \in [0,1]$, which defines the degree of trust given to the reference trajectories. Using artificial data simulated by a non-linear model for creep growth in ferritic steel, we have analyzed how the values of these two parameters influence the performance of the method, and given some indications about how to set them.

Finally, we applied the method to the problem of predicting the RUL of clogging filters, obtaining prediction intervals for the value of the RUL with satisfactorily accuracy, considering the large uncertainties affecting the clogging process.

A limit of the method is the presence of possibly large oscillations in the confidence bound provided, which may be confusing for the maintenance planner. It has been shown that such oscillations can be reduced by conveniently setting the parameter values; however, this would also reduce the accuracy and precision of the prediction, and for this reason we suggest considering filtering as a different solution for smoothing the predictions. Notice however, that the amplitude of the oscillation decreases as the density of the reference trajectories increases.

Finally, the possibility of assigning masses of belief to intervals or fuzzy numbers [Petit-Renaud & Denoeux, 2004] instead of single RUL values should also be investigated.

References

Angstenberger, L., 2001. Dynamic Fuzzy Pattern Recognition, International Series in Intelligent Technologies, 17, Kluwer Academic Publishers.

Baraldi, P., Cadini, F., Mangili, F., Zio, E., 2013a. Model-Based and Data-Driven Prognostics under Different Available Information, Probabilistic Engineering Mechanics, to appear.

Baraldi, P., Cadini, F., Mangili, F., Zio, E., 2013b. Prognostics under different available information. Abstract accepted to the 2013 Prognostics and System Health Management Conference (PHM), Milan, Italy, 8-11 September.

Baraldi, P., Mangili, F., Zio, E., 2012c. A Kalman Filter-based Ensemble Approach for Turbine Creep Prognostics. IEEE Trans. on Reliability.

Baraldi, P., Mangili, F., Zio, E., 2013a. Investigation of uncertainty treatment capability of model-based and data-driven prognostic methods using simulated data. Reliability Engineering and System Safety, Vol 112C, pp. 94-108.

Baraldi, P., Mangili, F., Zio, E., 2012b. Ensemble of Bootstrapped Models for the prediction of the Remaining Useful Life of a Creeping Turbine Blade. IEEE Int. PHM Conf., Denver, Colorado, 18-21 June.

Dempster, A.P., 1976. Upper and lower probabilities induced by a multivariate mapping, Annals of Mathematical Statistics AMS-38, pp. 325-339.

Hines, J.W., Usynin, A., 2008. Current Computational Trends in Equipment Prognostics, Int. J. Comput. Intell. Syst., Vol. 1(1), pp. 94–102.

Mustata, R. Hayhurst, D.R., 2005. Creep constitutive equations for a 0.5Cr 0.5 Mo 0.25V ferritic steel in the temperature range 565 8C–675 8C. International Journal of Pressure Vessels and Piping, Vol 82, pp. 363–372.

Peel, L., 2008. Data Driven Prognostics using a Kalman Filter Ensemble of Neural Network Models, International Conference on Prognostics and Health Management, 6-9 Oct, Denver, CO.

Petit-Renaud, S., Denoeux, T., 2004. Nonparametric regression analysis of uncertain and imprecise data using belief functions. International Journal of Approximate Reasoing, Vol 35, pp.1-28.

Santosh, T.V., Srivastava, A., Sanyasi Rao, V.V.S., Gosh, A. K., Kushwaha, H.S., 2009. Diagnostic System for Identification of Accident Scenarios in Nuclear Power Plants using Artificial Neural Networks, Reliability Engineering and System Safety, Vol. 94, pp. 759-762.

Shafer, G., 1976. A mathematical theory of evidence, Princeton University Press. Princeton, NJ.

Smets, P., 1998. The transferable belief model for quantified belief representation, in: D.M. Gabbay, P. Smets (Eds.), Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, Kluwer Academic Publishers, Dordrecht, pp. 267–301.

Smets, P., 1994. What is Dempster–Shafer_s model?, in: Advances in the Dempster–Shafer Theory of Evidence, Wiley, pp. 5–34.

Tang, L., Kacprzynski, G.J., Goebel, K., Vachtsevanos, G., 2009. Methodologies for Uncertainty Management in Prognostics. Proc IEEE Aerosp Conf, 2009 Mar 7-14; Big Sky, MT.

Vachtsevanos, G., Lewis, F.L., Roemer, M., Hess, A., Wu, B., 2006. Intelligent Fault Diagnosis and Prognosis for Engineering Systems, 1st edition, John Wiley & Sons, Hoboken.

Wang, T., Yu, J., Siegel, D., Lee, J., 2008. A Similarity-Based Prognostics Approach for Remaining Useful Life Estimation of Engineered Systems. International Conference on Prognostics and Health Management, 6-9 Oct, Denver (CO).

Yan, J., Koç, M., Lee, J., 2004. A Prognostic Algorithm for Machine performance Assessment and its Application, Production Planning and Control, 15:8, 796-801.

Zio, E., 2012. Prognostics and Health Management of Industrial Equipment. In: Kadry S, editor. Diagnostics and Prognostics of Engineering Systems: Methods and Techniques, IGI-Global.

Zio, E., Di Maio, F., 2010. A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system. Reliability Engineering & System Safety, Vol. 95(1), pp. 49−57.

Paper V

An approach to Remaining Useful Life prediction based on Gaussian Process Regression for degradation modeling

Piero Baraldi, Francesca Mangili, Enrico Zio

2013

An approach to Remaining Useful Life prediction based on Gaussian Process Regression for degradation modeling

Piero Baraldi[a], Francesca Mangili[a], Enrico Zio[b,a]

[a]*Dipartimento di Energia, Politecnico di Milano, Italy*

[b]*Chair on Systems Science and the Energetic challenge, European Foundation for New Energy-Electricite' de France, Ecole Centrale Paris and Supelec, France*

Abstract

This paper presents a prognostic method that uses Gaussian Process Regression (GPR) to build a stochastic model of the equipment degradation evolution.

GPR is a probabilistic technique for non-linear non-parametric regression that computes the distribution of the future equipment degradation states by constraining a *prior* distribution defined by a Gaussian Process to fit the available training data, based on Bayesian inference. Training data are taken from sequences of degradation measures collected from a set of similar historical equipments which have undergone a similar degradation process in the past. Given new degradation measures from the currently degrading equipment (test trajectory), an estimate of the distribution of the remaining useful life before failure (RUL) is obtained by comparing the future degradation states of the equipment, predicted by the GPR, with a failure criterion.

Different approaches for building the degradation model based on GPR have been developed and tested on simulated data concerning the evolution of creep damage in ferritic steel and real data concerning the clogging of sea water filters placed upstream the heat exchangers of a BWR condenser.

Keywords: Prognostics, Gaussian Process Regression, creep, filter clogging

1 Introduction

Several data-driven and model-based methods have been proposed for predicting the Remaining Useful Life (RUL) of degrading equipment [Hines & Usynin, 2008; Zio, 2012], i.e., the remaining time during which the equipment can continue performing its function in a safe and efficient way, with the final goal of implementing advanced predictive maintenance strategies which have the potential of increasing safety and lowering costs [Muller et al., 2008]. Model-based methods assume that an accurate mathematical model of the degradation process is available from first principle combined with large amount of experimental data. On the other side, data-driven methods rely on the availability of observations collected during the degradation process of one or more similar pieces of equipment, from which the RUL prediction can be directly or indirectly derived. Direct RUL prediction approaches make use of Artificial Intelligence

techniques, e.g., Artificial Neural Networks (ANNs) [Peel, 2008; Santosh et al., 2009], similarity-based regression [Wang et al., 2008; Zio & Di Maio, 2010], etc., to directly map the relation between the observation and the equipment RUL. However, when the observations collected are directly related to the equipment degradation state, the most natural data-driven technique for RUL estimation is to use regression to model the equipment degradation evolution and compare the extrapolated future degradation path to a failure criterion. This approach provides a more informative and transparent output, since it supplies a prediction not only of the current equipment RUL, but of the entire degradation path which can be checked against, e.g., expert intuition to verify its consistency. Moreover, contrarily to direct RUL prediction, this approach can be applied even when no historical failure data are available.

Current research works [Vachtsevanos & Wang, 2001; Goebel et al, 2008] show the effective use of ANNs for degradation-based prognostic systems. A limit of ANNs models is that they do not provide, in general, an explicit quantification of the uncertainty of the predicted states, like do, instead, methods such as Relevance Vector Machine (RVM) [Tipping, 2001] or Gaussian Process Regression (GPR) [MacKay 1998; Rasmussen & Williams, 2006] which make use of Bayesian inference. The prediction uncertainty is caused by the model uncertainty (due, e.g., to the limited amount of data used to build it), the uncertainty on the observation (due, e.g., to sensor noise), and the process uncertainty (due, e.g., to the unforeseen future loads and operating conditions) [Baradi et al., 2012]. The intrinsic ability of RVM and GPR to fit probability distribution functions (pdfs) to the degradation data is desirable for prognostics where uncertainty management is of paramount importance [Tang et al. 2009; Liu et al., 2011]. In practice, the RVM method is actually a special case of a Gaussian Process (GP) [Rasmussen & Williams, 2006]. In this context, this paper proposes the use of the GPR for a prognostic approach [Mohanty et al. 2011] that explicitly models the uncertainty in the future degradation states and provides prediction of the future equipment degradation in the form of a Gaussian pdf. The hypothesis of Gaussianity, required by the GPR approach, allows analytical calculations that make inference simpler. Then, although one cannot always prove that degradation states are normally distributed, in the absence of outstanding evidence to support a different assumption, it is reasonable to adopt this assumption rather than others that would make the inference from data more difficult.

GPR can be used to predict the evolution in time of the distribution of the degradation state or of its rate of growth, which, in turn, can be modeled as a function of time or of degradation. In this work, we proposed a strategy to implement each of these three approaches: 1) modeling the degradation state as a function of time; 2) modeling the degradation rate as a function of time; 3) modeling the degradation rate as a function of the degradation state. In any case, the final output is a prediction of the distribution of the future degradation states. By comparing this distribution at different time instants with the value of a fixed failure threshold, i.e., the maximum value of degradation beyond which the equipment cannot be operated, it is possible to calculate the distribution of the equipment RUL.

The problem of correctly inferring the degradation process of the test equipment from observations coming from different pieces of equipment, which have similar behavior, but have necessarily followed different degradation trajectories than the test equipment, is faced by describing the degradation process as the composition of two elements: a general structure of the degradation trajectory, common to all pieces of equipment, and a variation around this structure which is different for each piece of equipment and thus uncorrelated between trajectories.

Two case studies are considered in this work to validate the method and compare the three approaches proposed. In the first case study, the method is tested on simulated data generated by a non-linear model of creep growth in ferritic steel. In the second case study, we show the results of the application of GPR to real data concerning the clogging of filters used to clean the sea water pumped through the secondary side of a BWR condenser to cool the steam in the primary side. The performance of each proposed approach to GPR-based prognostics is evaluated in terms of the accuracy and precision of the RUL prediction and the reliability of the uncertainty estimates provided.

The remaining part of the paper is organized as follows: Section 2 describes the method for performing RUL predictions based on GPR and details the three approaches proposed for degradation modeling; Section 3 shows the results obtained by applying the GPR to the two case studies considered; finally, in Section 4 we draw some conclusions and suggest potential future work.

2 Methodology

2.1 Gaussian Process Regression

Gaussian process regression is a powerful and flexible approach to performing inference over functions [Rasmussen & Williams, 2006]. In a regression problem which aims to map from an input x to an output $f(x)$, GPR defines the *prior* for the output $f(x)$ in the form of a distribution over functions specified by a Gaussian Process. A GP is a collection of random variables any finite number of which has a joint Gaussian distribution. A real GP $f(x)$ is completely specified by its mean function $m(x)$ and covariance function $k(x,x')$:

$$\begin{aligned} &f(\mathbf{x}) \sim \mathrm{GP}\{m(\mathbf{x}); K(\mathbf{x},\mathbf{x})\} \\ &m(x) = E[f(x)] \\ &k(x,x') = E[(f(x)-m(x))(f(x')-m(x'))] \end{aligned} \tag{1}$$

where $\mathbf{x}$ represents a vector of input values and $K(\mathbf{x},\mathbf{x})$ indicates the co-variance matrix containing the values of $k(x,x')$ evaluated for all possible pairs of inputs in $\mathbf{x}$.

This *prior* is taken to represent our prior beliefs over the kind of functions we expect to observe. Typically the *prior* mean and co-variance functions that we use will have some free parameters, called, usually, hyper-parameters. Although the choice of covariance function must be specified by the user, various methods have been proposed for determining the corresponding hyper-parameters from training data [Rasmussen & Williams, 2006], e.g., the conjugate gradient optimizer that maximizes the marginal likelihood of the training set with respect to the hyper-parameters.

Given the prior information about the GP, the value of the hyper-parameters and a set of training data $\mathbf{D}_{xy} = (\mathbf{x},\mathbf{y}) = \{(x_i, y_i)\}_{i=1:N}$, the posterior distribution over functions is derived by imposing a restriction on the *prior* distribution to contain only those functions that agree with the observed data [Rasmussen &

Williams, 2006]. In other words, we impose the output in correspondence of the test input vector $\mathbf{x}^{tst}$ to be drawn from the same GP as the training data $\mathbf{D}_{xy}$, and thus we have:

$$\begin{bmatrix} \mathbf{y} = f(\mathbf{x}) \\ \mathbf{f}^{tst} = f(\mathbf{x}^{tst}) \end{bmatrix} \sim \mathrm{GP}\left\{ \begin{bmatrix} m(\mathbf{x}) \\ m(\mathbf{x}^{tst}) \end{bmatrix}; \begin{bmatrix} K(\mathbf{x},\mathbf{x}) & K(\mathbf{x},\mathbf{x}^{tst}) \\ K(\mathbf{x}^{tst},\mathbf{x}) & K(\mathbf{x}^{tst},\mathbf{x}^{tst}) \end{bmatrix} \right\} \tag{2}$$

The presence of a white Gaussian noise υ_i with variance σ_υ^2 on the observations y_i can be accounted for by adding the noise variance to the co-variance function:

$$\mathbf{y} = f(\mathbf{x}) + \boldsymbol{\upsilon} \sim \mathrm{GP}\{m(\mathbf{x}); K(\mathbf{x},\mathbf{x}) + \sigma_\upsilon^2 \mathbf{I}\} \tag{3}$$

where $\mathbf{I}$ is the identity matrix.

From eq. (2) the posterior distribution of the output $\mathbf{f}^{tst} \mid D$ in correspondence of the input vector $\mathbf{x}^{tst}$ can be derived [Rasmussen & Williams, 2006]:

$$\mathbf{f}^{tst} \mid \mathrm{D} \sim N(\overline{\mathbf{f}^{tst}}, \mathrm{cov}(\mathbf{f}^{tst})) \tag{4}$$

where

$$\begin{aligned} \overline{\mathbf{f}^{tst}} &= m(\mathbf{x}^{tst}) + K(\mathbf{x},\mathbf{x}^{tst})[K(\mathbf{x},\mathbf{x}) + \sigma_\upsilon^2 \mathbf{I}]^{-1}(\mathbf{y} - m(\mathbf{x})) \\ \mathrm{cov}(\mathbf{f}^{tst}) &= K(\mathbf{x}^{tst},\mathbf{x}^{tst}) - K(\mathbf{x}^{tst},\mathbf{x})[K(\mathbf{x},\mathbf{x}) + \sigma_\upsilon^2 \mathbf{I}]^{-1} K(\mathbf{x},\mathbf{x}^{tst}) \end{aligned} \tag{5}$$

2.2 Prognostic model

It is assumed that R training trajectories are available from measurements collected during the process of degradation of R pieces of equipment similar to the one of interest (test equipment). Each reference trajectory r=1:R is made of a sequence $z_{1:n^r}^r$ of observations z_j^r directly related to the degradation state δ_j^r of the r-th equipment at time τ_j, j=1:n^r, where τ_{n^r} is the last measurement time before failure. Equipment fails when its degradation exceeds the failure threshold value δ^{th}; let τ_F^r indicate the time at which the failure of the r-th piece of equipment occurs. A sequence of observations $z_{1:J}^{test}$ from τ_1 to the present time τ_J is available also for the test equipment.

The goal of the GPR prognostic model is to predict the future degradation states of the equipment of interest and from them compute its RUL. Due to the scatter in the microstructural and manufacturing characteristics, the loading and external conditions variability, etc., the damage state, at any time instant, is better represented by a random variable $\Delta(\tau)$ rather than by a deterministic quantity [Mohanty, 2011]. As a consequence, also the equipment RUL at the present time τ_J should be represented by a random variable $RUL(\tau_J)$ [Baraldi et al., 2012].

In this work, we assume the distributions of the degradation states to be Gaussian with different mean $\overline{\Delta}(\tau)$ and variance $\sigma_{\Delta}^2(\tau)$ at every time instant τ, and use the GPR method to evaluate the conditional probability density function (pdf) $p_{\Delta^{test}(t)}(\delta(t)\,|\,\mathbf{D}_z, z_{1:J}^{test})$ of the future damage state $\Delta^{test}(\tau)$, $\tau > \tau_p$ given the training dataset $\mathbf{D}_z = \{z_{1:n^r}^r\}_{R=1:r}$, and the test trajectory $z_{1:J}^{test}$. Three different approaches are considered to infer the conditional pdf $p_{\Delta^{test}(t)}(\delta(t)\,|\,\mathbf{D}_z, z_{1:J}^{test})$ from data based on GPR:

1) We directly model the degradation state $\Delta(\tau)$ as a function of time by defining the *prior*:

$$\Delta(\boldsymbol{\tau}) \sim \mathrm{GP}\{m_{\Delta}(\boldsymbol{\tau}); K_{\Delta}(\boldsymbol{\tau},\boldsymbol{\tau})\} \tag{6}$$

where $\boldsymbol{\tau}$ is a vector of time instants. The hyper-parameters of $m_{\Delta}(\boldsymbol{\tau})$ and $K_{\Delta}(\boldsymbol{\tau},\boldsymbol{\tau})$ are optimized using the set of training data $\mathbf{D}_{\tau/z}\{(\tau_j; z_j^r\}_{1:n^r;r=1:R}$ derived from the training trajectories. Finally, predictions of the mean $\overline{\Delta}(\tau)$ and the variance $\sigma_{\Delta}^2(\tau)$ of the pdf $p_{\Delta^{test}(t)}(\delta(t)\,|\,\mathbf{D}_z, z_{1:J}^{test})$ of future degradation states are obtained from eq. (5) by conditioning the GP in eq. (6) on the training data in $\mathbf{D}_{\tau/z}$ and those available from the part of test trajectory already observed: $\{(\tau_j; z_j^{test}\}_{1:J}$.

2) We model the degradation rate $d\Delta$ as a function of time by defining the *prior*:

$$d\Delta(\boldsymbol{\tau}) \sim \mathrm{GP}\{m_{d\Delta}(\boldsymbol{\tau}); K_{d\Delta}(\boldsymbol{\tau},\boldsymbol{\tau})\} \tag{7}$$

The hyper-parameters of $m_{d\Delta}(\boldsymbol{\tau})$ and $K_{d\Delta}(\boldsymbol{\tau},\boldsymbol{\tau})$ are optimized using the set of training data $\mathbf{D}_{\tau/dz} = \{(\tau_j; (z_{j+1}^r - z_j^r)/(\tau_{j+1} - \tau_j))\}_{j=1:n^r-1;r=1:R}$ derived from the training trajectories. Prediction of the mean $\overline{d\Delta}(\tau')$ and the variance $\sigma_{d\Delta}^2(\tau')$ of the degradation rate pdf are obtained from eq. (5) by conditioning the GP in eq. (7) on the training data in $\mathbf{D}_{\tau/dz}$ and those available from the part of test trajectory: $\{(\tau_j; (z_{j+1}^{test} - z_j^{test})/(\tau_{j+1} - \tau_j))\}_{j=1:J-1}$. By integrating the mean $\overline{d\Delta}(\tau')$ and the variance $\sigma_{d\Delta}^2(\tau')$ of the degradation rate from the current time τ_J up to time τ it is possible to predict the mean $\overline{\Delta}(\tau)$ and the variance $\sigma_{\Delta}^2(\tau)$ of the pdf of interest $p_{\Delta^{test}(t)}(\delta(t)\,|\,\mathbf{D}_z, z_{1:J}^{test})$:

$$\begin{aligned} \overline{\Delta}(\tau) &= z_J^{test} + \int_{\tau_p}^{\tau} \overline{d\Delta}(\tau')\,\mathrm{d}\,\delta(\tau') \\ \sigma_{\Delta}^2(\tau) &= \int_{\tau_p}^{\tau} \sigma_{d\Delta}^2(\tau')\,\mathrm{d}\,\delta(\tau') \end{aligned} \tag{8}$$

In practice, time can be discretized so that the integrals in eq. (8) become summations over finite increments $\overline{d\Delta}(\tau'_j)(\tau'_{j+1} - \tau'_j)$.

3) We model the degradation rate $d\Delta$ as a function of the degradation state by defining the *prior*:

$$d\Delta(\boldsymbol{\delta}) \sim \mathrm{GP}\{m_{d\Delta}(\boldsymbol{\delta}); K_{d\Delta}(\boldsymbol{\delta},\boldsymbol{\delta})\} \tag{9}$$

where $\boldsymbol{\delta}$ is a vector of degradation states. The hyper-parameters of $m_{d\Delta}(\boldsymbol{\delta})$ and $K_{d\Delta}(\boldsymbol{\delta},\boldsymbol{\delta})$ are optimized using the set of training data $\mathbf{D}_{z/dz} = \{(z_j^r;(z_{j+1}^r - z_j^r)/(\tau_{j+1} - \tau_j))\}_{j=1:n^r-1;r=1:R}$ derived from the training trajectories. Predictions of the mean $\overline{d\Delta}(\delta)$ and the variance $\sigma_{d\Delta}^2(\delta)$ of the degradation rate pdf are obtained from eq. (5) for any value of degradation δ in input by conditioning the GP in eq. (9) on the training data in $\mathbf{D}_{z/dz}$ and on the test data $\{(z_j^{test};(z_{j+1}^{test} - z_j^{test})/(\tau_{j+1} - \tau_j))\}_{j=1:J-1}$. The mean $\overline{\Delta}(\tau)$ and the variance $\sigma_{\Delta}^2(\tau)$ of the pdf of interest $p_{\Delta^{test}(t)}(\delta(t) \mid \mathbf{D}_z, z_{1:J}^{test})$ are obtained as the average and the variance of the degradation states of a large number N^{samp} of degradation trajectories sampled from:

$$\begin{aligned} \boldsymbol{\delta}_{j+1} &= \overline{d\Delta}(\delta_j) + \upsilon_j \\ \upsilon_j &\sim N(0, \sigma_{d\Delta}^2(\delta_j)) \end{aligned} \tag{10}$$

starting from the last observed degradation state z_J^{test} at time τ_J and using the Monte Carlo sampling algorithm.

Notice that, in this particular case where data are available both from historical (training) equipments and from the currently degrading (test) equipment, we aim to learn about the common structure underlying all degradation processes from the test trajectories available, but also to draw from the test trajectory the information about the specific variation around this structure that characterizes the equipment of interest.

To achieve this, the covariance function is built as follows [Mann et al., 2011]:

$$k(x_j^r, x_{j'}^{r'}) = k_1(x_j^r, x_{j'}^{r'}) + k_2(x_j^r, x_{j'}^{r'})\delta(r,r')\delta(j,j') + \sigma_y^2\delta(r,r')\delta(j,j') \tag{11}$$

where x and y are the input and output specific to each approach adopted and the reference index assigned to the test trajectory is $r = R+1$. The first term of the kernel corresponds to the covariance associated with the common structure underlying all degradation trajectories; the second represents the covariance owing to the variation of each trajectory around the common structure of a degradation process. This term assumes a finite value only when x_j^r and $x_{j'}^{r'}$ are taken from the same trajectory, since we assume the variation specific to each trajectory to be uncorrelated across trajectories. Finally, the third term accounts for the observation noise associated with the observation of a specific output y.

The functions chosen to represent our *prior* on the mean and covariance of the GP are fundamental ingredients for the correct application of GPR. In the literature about GPR, the *prior* on the mean is often set to the constant value of zero, after appropriate normalization of the training data [Shi et al., 2005; Rasmussen & Williams, 2006]. However, this does not seem a convenient choice when the function to regress has an evident trend, as for degradation modeling. In these case the *prior* mean function is chosen to accurately represent the trend (linear, power low, etc.) of the function to regress, which can be often guessed from a graphical view of the training data. Notice however that imposing a particular functional form for $m(x)$ does not constrain the updated mean $\overline{\mathbf{f}^{tst}}$ of the output variable to follow that same functional form.

The efficacy of different combination of mean and covariance functions in modeling a set of training data can be evaluated based on the marginal likelihood obtained by each of them.

Given the value of the failure threshold, assumed here to be known, and the conditional distribution of the degradation state $p_{\Delta^{test}(t)}(\delta(t)\,|\,\mathbf{D}_z, z_{1:J}^{test})$, the RUL cumulative distribution function (cdf) $P_{RUL(\tau_J)}(rul_J\,|\,\mathbf{D}_z, z_{1:J}^{test})$ is computed from $p_{\Delta(\tau)}(\delta(\tau)\,|\,\mathbf{D}_z, z_{1:J}^{test})$ as the probability that the degradation $\Delta(\tau)$ at time $\tau = \tau_J + rul_J$ exceeds the failure threshold d_{th} [Baraldi et al., 2013]:

$$\begin{aligned} P_{RUL(\tau_J)}(rul_J \,|\, \mathbf{D}_z, z_{1:J}^{test}) &= \text{Prob}(RUL(\tau_J) < rul_J \,|\, \mathbf{D}_z, z_{1:J}^{test}) \\ &= \int_{d_{th}}^{+\infty} p_{\Delta^{test}(t_J + rul_J)}(\delta(\tau)\,|\,\mathbf{D}_z, z_{1:J}^{test})\,\mathrm{d}\,\delta(\tau) = \\ &= 1 - \Phi\left(\frac{\Delta(t_J + rul_J) - \overline{\Delta}(t_J + rul_J)}{\sigma_\Delta(t_J + rul_J)}\right) \end{aligned} \tag{12}$$

where Φ is the standard normal cdf.

From the RUL cdf one can derive the prediction $\hat{rul}_J$ of the equipment RUL as the mean value of the RUL distribution and the confidence interval $CI(\alpha) = [rul_j^{\inf}(\alpha), rul_j^{\sup}(\alpha)]$ containing the true value of the test equipment RUL, hereafter referred to as rul_J, with probability α.

3 Numerical application

In this Section the three approaches proposed for applying GPR to prognostics are verified on data simulated using a non-linear model of creep growth in ferritic steel (Section 3.1), and on real data taken from a case study about the clogging of filter in a BWR condenser (Section 3.2).

The quality of the prediction depends from the choice of the mean and covariance functions. In this work, different choices for the mean and covariance functions have been compared based on their marginal likelihood. Notice that, finding the optimal choice in absence of any knowledge about the actual process that governs the system can be a hard task. Since, to the best of our knowledge, a widely agreed procedure for the selection of GPR mean and covariance functions is not yet available in literature, we limited this choice to pick out of some reasonable combination of covariance functions available from the literature [Rasmussen & Williams, 2006] the one with the largest marginal likelihood, without assurance that the choice is optimal and that the assumptions about the process are correct. The mean and covariance functions used in this work are listed in Appendix A.

Given N^{tst} test trajectories $\mathbf{z}_{1:J}^q$, q=1:N^{tst}, and N^{trn} different sets of training trajectories $D_z^l = \{\mathbf{z}_{1:n^r}^r, r = 1:R\}$, l=1:N^{trn}, each prognostic approach is used to provide the predictions $\hat{rul}_{J(\beta)}^q \,|\, D_z^l$ and $CI_{J(\beta)}^q(\alpha)\,|\,D_z^l$, in correspondence of the life fraction β, i.e., at time step $\tau_{J(\beta)}^q = \beta \cdot \tau_F^q$. Performances are evaluated by comparing such prediction with the true RUL value $rul_{J(\beta)}^q$ to compute three performance indicators:

- The square root of the Mean Square Error (RMSE), i.e., the average value over all test trajectories q=1:N^{tst} and training sets l=1:N^{trn} of the square error $(\hat{rul}_{J(\beta)}^q \,|\, D_z^l - rul_{J(\beta)}^q)^2$ made in predicting the

true RUL of the test equipment. The MSE measures the accuracy of the prediction and is desired to be as small as possible.

- The Coverage (Cov_α) of the prediction interval $CI^q_{J(\beta)}(\alpha) \mid D^l_z$, i.e., the percentage of times the condition $rul^q_{J(\beta)} \in CI^q_{J(\beta)}(\alpha) \mid D^l_z$ is verified. This indicator measures the reliability of the confidence interval; we want the value of Cov_α to be as close as possible to α.
- The amplitude (MA_α) of the confidence interval $CI^q_{J(\beta)}(\alpha) \mid D^l_z$ averaged over all test trajectories q=1:N^{tst} and training sets l=1:N^{trn}; this indicator gives a measure of the precision of the RUL prediction. In order to have a high precision, we wish the keep the value of MA_α as small as possible.

3.1 Artificial dataset: Creep growth in ferritic steal -

In this Section, GPR is applied to the prediction of the failure time of simulated degradation trajectories representing the accumulation of creep damage in a ferritic steel. Ferritic steels are widely used for welded steam pipes in the construction of power plant components that operate under high temperature and loads. In such conditions the creep deformation and rupture are important factors in determining the equipment lifetimes.

3.1.1 Creep growth models

The evolution of creep damage in ferritic steel exposed to the load σ is simulated using the uni-axial form of the non-linear creep constitutive equations proposed within the framework of Continuum Damage Mechanics by Mustata & Hayhurst (2005):

$$\begin{cases} \dot{\varepsilon} = A \sinh\left[\dfrac{B\sigma(1-H)}{(1-\phi)(1-\omega)}\right] \\ \dot{H} = \dfrac{h\dot{\varepsilon}}{\sigma}\left(1-\dfrac{H}{H^*}\right) \\ \dot{\phi} = \dfrac{K_c}{3}(1-\phi)^4 \\ \dot{\omega} = C\dot{\varepsilon} \end{cases} \tag{13}$$

where ε is the creep strain, i.e., the fraction of elongation of the metallic piece in the longitudinal direction with respect to its original length, ϕ and ω are two damage state variable describing, respectively, the coarsening of the carbide precipitates, and the inter-granular creep constrained cavitation damage, H is the hardening state variable, used to represent the strain hardening effect attributed to primary creep, and A, B, H^*, h, K_c, and C are material inherent characteristics. Each characteristic $\varphi_{1:6} = A, B, H^*, h, K_c, C$ varies with the temperature according to the Arrhenius law, i.e., $\varphi_m = \varphi_{m0} \exp(-Q_m / T)$, m=1:6, where T is the operating temperature and φ_{m0} and Q_m are parameters whose values have to be determined experimentally.

To generate different trajectories, the intrinsic variability of the creep growth process is simulated by sampling the values of the load σ and temperature T to which the steel is exposed at each time step from a normal distribution centered on their mean value, whereas the variability of the creep growth process between similar pieces of equipment is simulated by sampling the value of parameters φ_{m0} and Q_m, m=1:6,

at the beginning of each new simulated trajectory. We assume failure to happen when the limiting creep strain value of 0.02 is reached. Finally, to generate the sequence of observations $\mathbf{z}^r_{1:n^r} = \{\varepsilon(\tau_j) + v_j\}^r_{j=1:n^r}$, collected one every 100 days, a white Gaussian noise v_j with standard deviation $\sigma_v = 2 \times 10^{-4}$ is added to the simulated creep strain $\varepsilon(\tau_j)$ at the observation time τ_j. The time interval between two observations is rather large; in practical applications, this can happen when the costs of an inspection are elevated. Here this choice was done to limit the number of training data, since computational costs of GPR scales typically as $O(N^3)$ with the number N of training data. Figure 1 shows an example of simulate creep growth trajectory (upper) and the corresponding sequence of observations $\mathbf{z}^r_{1:n^r}$ (bottom).

For a better evaluation of the method, we have adopted two different sets of parameters (see Appendix B for their distribution) to simulate trajectories with smaller or larger variability. Figure 2 compares 10 trajectories with low (upper) and high (bottom) variability.

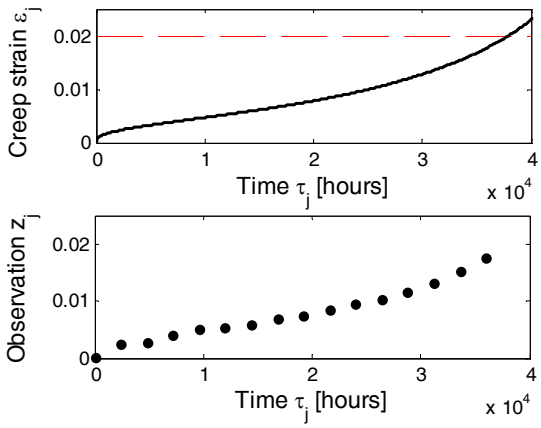

Figure 1: example of simulated creep growth trajectory (upper) with the corresponding sequence of observations (bottom).

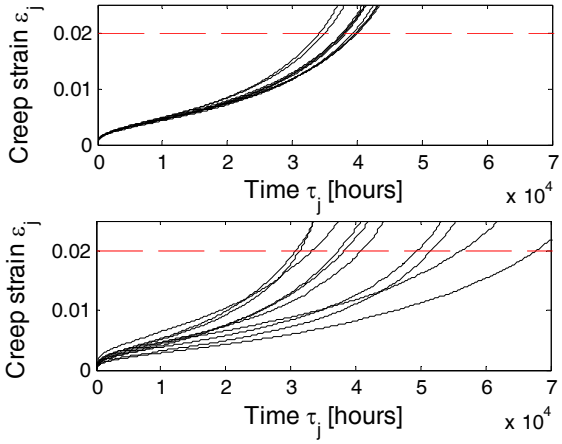

Figure 2: examples of creep growth trajectories with low (upper) and high (bottom) variability.

3.1.2 Results

In this Section we show, first, some results obtained by applying the three approaches proposed in Section 2.2 to a single creep growth trajectory; for this a set of *R*=10 training trajectories is used. In this first part we consider only trajectories with low variability (Figure 2, upper). Then, we compare the performances obtained by the three approaches in correspondence of trajectories with low and high variability.

In Figure 3, the predicted future evolution of degradation obtained by the GPR in approach 1 is compared with the true evolution of the test trajectory at two prediction time instants, $\tau_1 = 1$ hour (left) and $\tau_{12} = 26400$ hours (right). Notice that the prediction accuracy is lower, i.e., the confidence interval is larger, for a prediction done at time $\tau_1 = 1$ since no data are available from the test trajectory, and the GPR can only account for the common structure of the degradation process drawn from the training trajectories; on the other side, at time $\tau_{12} = 26400$ the precision increases, since GPR is able to learn from the data collected during the test trajectory its peculiar behavior thank to the second term $k_2(x^r_j, x^{r'}_{j'})\delta(r,r')$ of the covariance function in eq. (11). See Section 3.1.3 for a further discussion of this aspect.

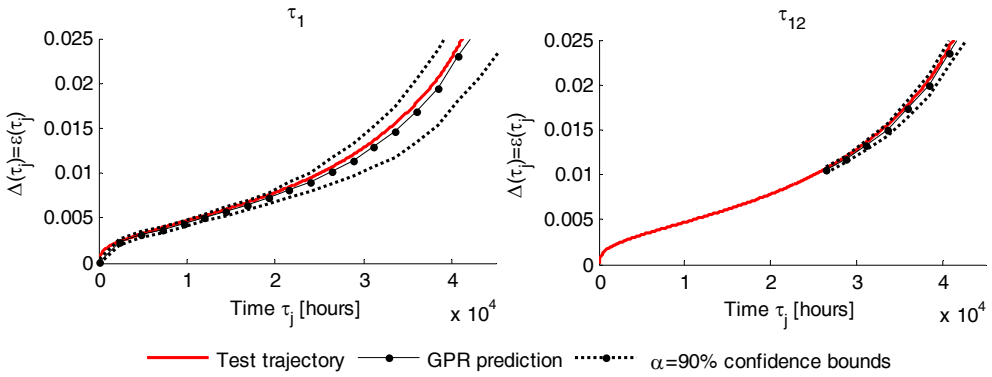

Figure 3: GPR prediction of future degradation states obtained at two time instants $\tau_1 = 1$ *hour (left) and* $\tau_{12} = 26400$ *hours (right) using approach 1.*

In approaches 2 and 3, GPR is used to predict the degradation rate $d\Delta$ as a function of, respectively, the time and the degradation state Δ. In Figure 4 and 5, the predicted degradation rate (upper) and the corresponding prediction of the future degradation evolution (bottom) are shown for approaches 2 and 3, respectively. The predicted value of the degradation rate (continuous line) is compared with its true value (asterisks) observed for the test trajectory.

Results show that these two approaches are less effective in learning the peculiar behavior of the test trajectory as more data are collected during its evolution. This could be due to the fact that the observed degradation rate is affected by a larger noise than the degradation state, since the relative effect of measurement and process noises more evident on the small values of the degradation rates than on those, larger, of the degradation states. For this reason, after a sufficient amount of observations is collected, we expect approach 1 to provide narrower confidence intervals $CI(\alpha)$ than approaches 2 and 3.

Figure 6 compares the RUL prediction obtained by the three approaches at different time instants, with the true RUL value of the test trajectory considered. All three approaches can predict the equipment RUL fairly well, although the prediction of approach 3 appears less accurate in the first part of the trajectory. Approach 2 supplies narrower confidence intervals for the RUL prediction than the other two approaches at the beginning of the trajectory; however, the reliability of such interval has to be verified by checking that their coverage is close enough to the target value of $\alpha = 0.9$.

For a more robust evaluation of the performance of the three approaches proposed, the model in eq. (13) is used to generate N^{tst}=100 test trajectories with low variability (Figure 2, upper) and N^{trn}=10 different sets of R=10 similar training trajectories. For each training dataset simulated, the three approaches are applied to each test trajectory at four different life fractions $\beta_1 = 0.3$, $\beta_2 = 0.6$, $\beta_3 = 0.8$ and $\beta_4 = 0.95$, and the three performance indicators, $Cov_{0.9}$, RMSE, and $MA_{0.9}$ are computed. The same procedure is repeated simulating trajectories with high variability (Figure 2, bottom). Figure 7 and 8 show the values of the three performance indicators in case of, respectively, low and high variability of the trajectories. The average value of the RUL, *mRUL*, of the test trajectories at each life fraction β is also shown for comparison with the values of the indicators RMSE and $MA_{0.9}$. A horizontal (red) line indicates the target value $\alpha = 0.9$ for the coverage (upper, left).

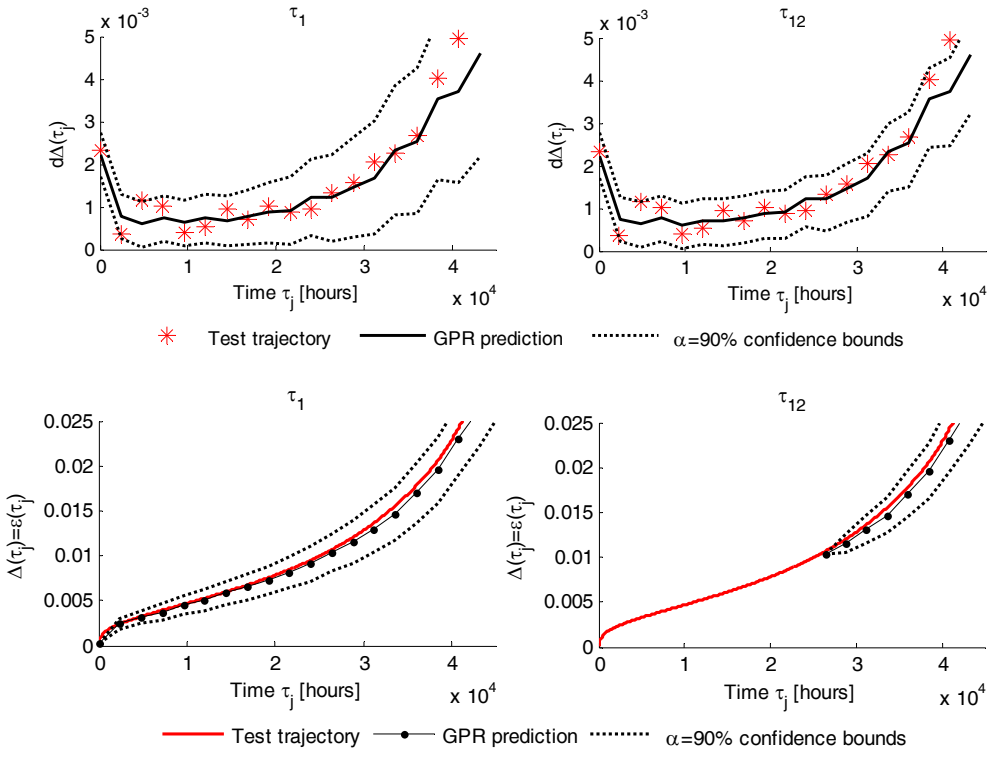

Figure 4: GPR prediction of the degradation rate (upper) and of the future degradation states (bottom) obtained at two time instants $\tau_1 = 1$ *hour (left) and* $\tau_{12} = 26400$ *hours (right) using approach 2.*

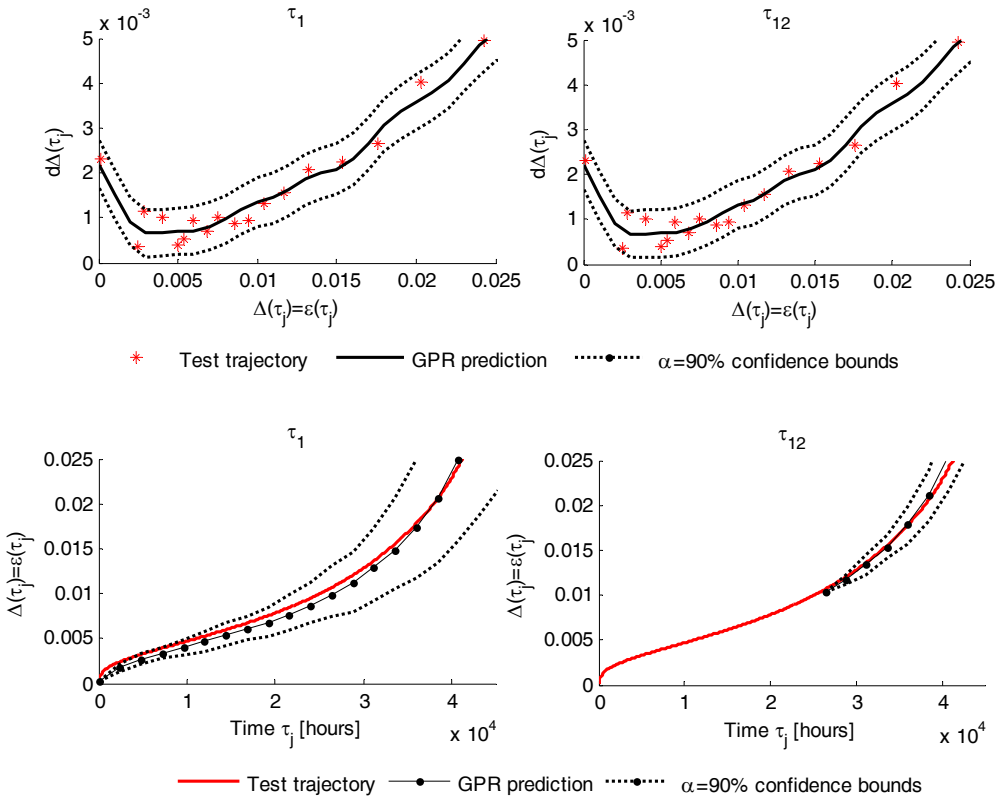

Figure 5: GPR prediction of the degradation rate (upper) and of the future degradation states (bottom) obtained at two time instants $\tau_1 = 1$ *hour (left) and* $\tau_{12} = 26400$ *hours (right) using approach 3.*

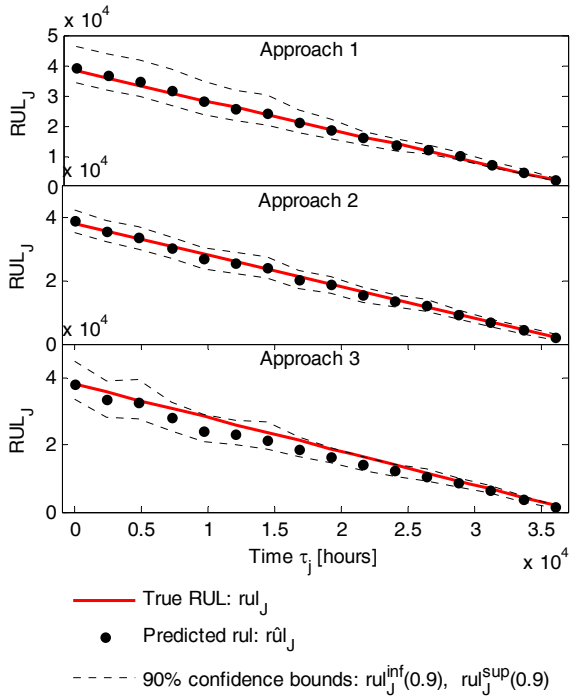

Figure 6: Comparison at different time instants of the RUL prediction with the true RUL value for the three GPR-based approaches.

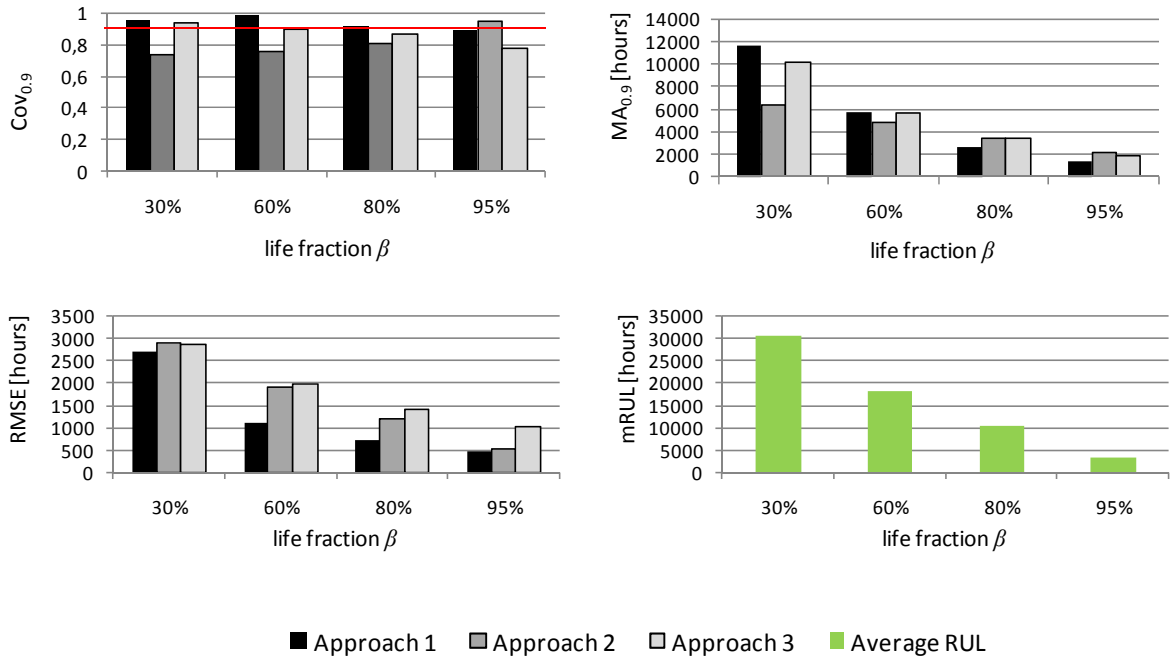

Figure 7: Comparison of the performance of the three approaches for trajectories with low variability. The horizontal (red) line in the $Cov_{0.9}$ figure (upper, left) indicates the target coverage value $\alpha = 0.9$.

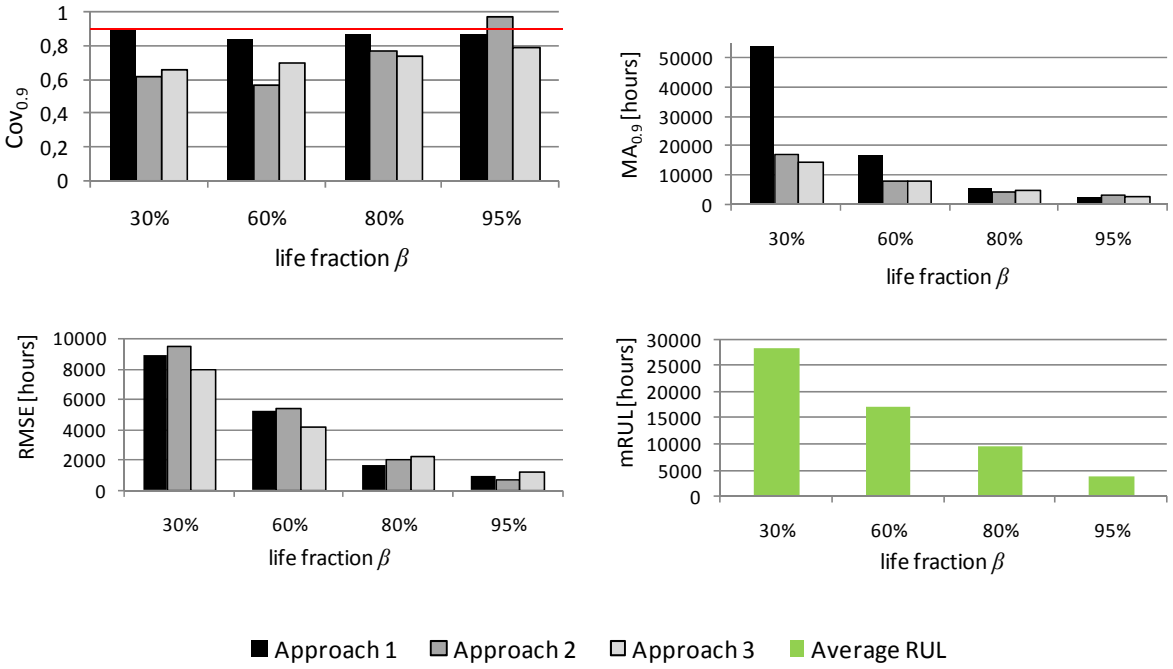

Figure 8: Comparison of the performance of the three approaches for trajectories with high variability. The horizontal (red) line in the $Cov_{0.9}$ figure (upper, left) indicates the target coverage value $\alpha = 0.9$.

In case of creep growth trajectories with low variability, the error RMSE and the amplitude of the confidence interval $MA_{0.9}$ are always much smaller than the mean RUL, being, respectively, around the 10% and the 30% of *mRUL*. As expected, the value of both indicators decreases with the life fraction β, since more data are available and failure is closer, and thus the effects of the process uncertainty on the accuracy and precision of the prediction are reduced. The increase in the accuracy and precision with the life fraction is more evident for approach 1. This approach is able to provide always the most accurate predictions, and, toward the end-of-life of the equipment, even the narrowest confidence interval compared to the other two approaches. Moreover, it is also the only approach obtaining a coverage always close to $\alpha = 0.9$. Approach 2 provides narrower confidence intervals at the beginning of the equipment life, but fails attaining the coverage value of 0.9, so that we cannot rely on such narrow intervals. Finally, approach 3 obtains good coverage except for $\beta_4 = 0.95$, i.e., very close to the equipment failure, but is less accurate (larger RMSE) than approach 1.

Similar results are obtained when trajectories with large variability are considered (Figure 8). Although in this case no approach is able to assure a coverage of 0.9, approach 1 goes very close to this target, whereas approach 2 come up to it only when the equipment is close to failure. The confidence interval supplied by approach 1 is rather large in the first part of the trajectory, with a value that, at $\beta = 0.3$, is almost twice the average RUL value, *mRUL*,. However, due to the large variability of the trajectories, such a large confidence interval may be necessary to correctly represent the intrinsic uncertainty of the future creep evolution. Finally, although in this case we cannot state that approach 1 is the most accurate and precise (at $\beta_1 = 0.3$, $\beta_2 = 0.6$ it is outperformed in accuracy by approach 3 and by approach 2 at $\beta_4 = 0.95$), the RMSE is always comparable with that obtained by the other two approaches and does not exceeds the 30% of the average RUL, *mRUL*.

In this case study, approach 1 has demonstrated to outperform the other two approaches proposed, although all three approaches have shown the capability of tackling the prognostic problem and supply accurate RUL prediction with a measure of the prediction uncertainty. However, the fact that the coverage is often lower than its target value shows that the prediction uncertainty is not correctly quantified. Further research is necessary to identify how to correctly account for all sources of uncertainty affecting the RUL prediction; we suggest, in particular, that model uncertainty could play a significant role, especially in case of high variable trajectories. In this case, resorting to an ensemble of GPR models [Shi et al., 2005], could allow achieving higher reliability and improved prediction performance.

Notice finally that, as already mentioned, the choice of the covariance also impacts the performance of the method; the development of a procedure for the selection of GPR covariance functions is a fundamental requirement for the successful exploitation of GPR in prognostics.

3.1.3 Analysis of the covariance function

In this Section we study the effects of the two terms $k_1(x_j^r, x'^{r'}_{j'})$ and $k_2(x_j^r, x'^{r'}_{j'})\delta(r,r')$ of the covariance function in eq. (11). For this, approach 1 has been applied using the covariance function in eq. (11) without the first term k_1, first, and then without the second term k_2. The predictions about the future degradation states provided by approach 1 without the first term k_1 (upper) or without the second term k_2 (bottom) of the covariance function are shown in Figure 9 at time instant $\tau_1 = 1$ hour (left) and $\tau_{12} = 26400$ hours (right).

These results show that if the first term k_1 of eq. (11), which refers to the common structure underlying all creep growth trajectories, is not considered (Figure 9, upper), the GPR model predictions at time $\tau_1 = 1$ are very uncertain (Figure 9, upper, left). This is due to the fact that no information is available about the test degradation trajectory, and at the same time the information coming from the training trajectory is not used for conditioning the *prior* GP distribution, but only for optimizing the hyper-parameters of the covariance function k. As more data about the test trajectory become available the prediction improves (Figure 9, upper, right), but remains, however, less accurate and precise than that in Figure 3 (right), where the complete covariance function is used.

On the other side, if we do not consider the second term k_2 of eq. (11) (Figure 9, bottom), which refers to the variation of each trajectory around the common structure of the creep growth process, the GPR model predictions at time $\tau_1 = 1$ are much less uncertain (Figure 9, bottom, left), but no improvement can be observed as more data become available (Figure 9, bottom, right). In other words, the model is not able to learn the peculiar behavior of the test trajectory, since all observations, does not matter which equipment they refer to, are given the same relevance when used for conditioning the *prior* GP distribution.

It is clear that the two terms k_1 and k_2 are complementary, and for this reason their combination, as shown in Figure 3, assures better prediction performance along the entire duration of the degradation trajectory.

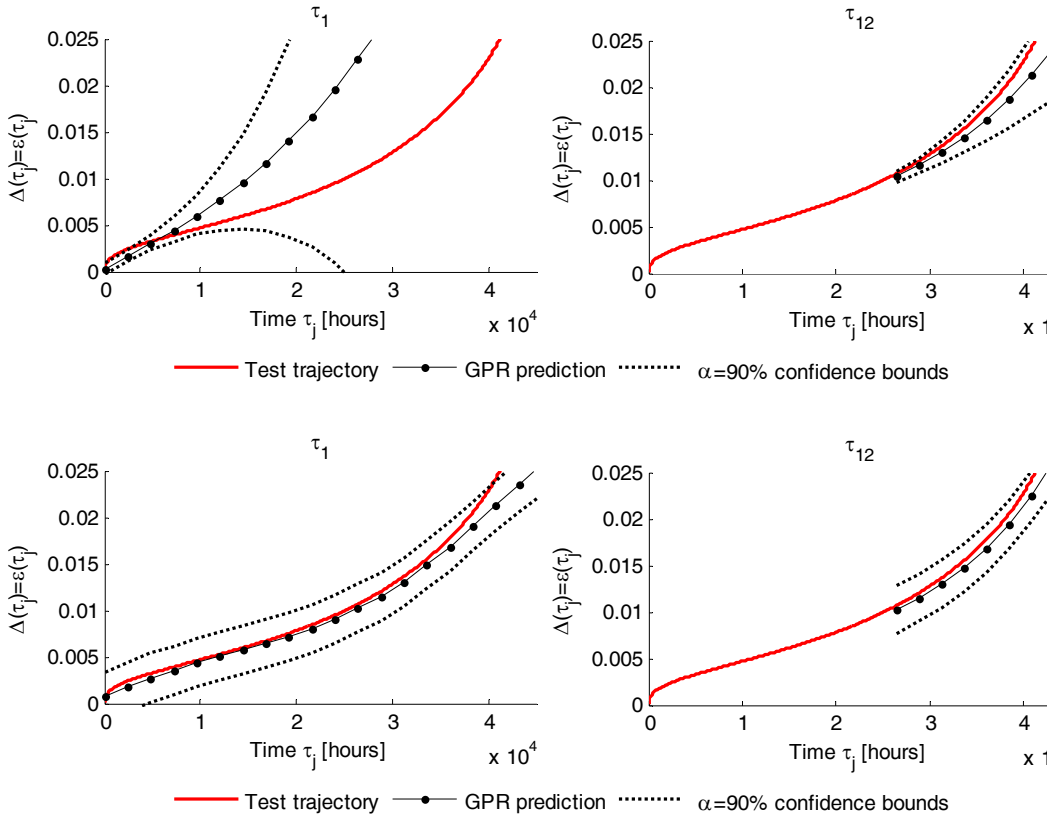

Figure 9:GPR prediction of future degradation states obtained at two time instants $\tau_1 = 1$ *hour (left) and* $\tau_{12} = 26400$ *hours (right) using approach 1 without the first term* k_1 *(upper) or the second term* k_2 *of the covariance function in eq. (11).*

3.2 Real dataset: Clogging of BWR condenser filters

In this Section, we consider the problem of predicting the RUL of filters used to clean the sea water entering the condenser of the BWR reactor of a Swedish nuclear power plant. During operations, filters undergo clogging and, once clogged, can cumulate particles, seaweed, and mussels from the cooling water in the heat exchanger. For this reason, prompt and effective cleaning of the filter is desirable; predictive maintenance can help achieving this result, keeping maintenance costs reasonably low.

An increasing number of articles can be found in the literature concerning the study of filter clogging by solid aerosols [Song et al., 2006] and liquid aerosols [Contal et al., 2004]. Common for these articles is that the results are achieved in a controlled environment: in the experimental setup all degradation quantities, indicators of degradation, and stressors are automatically measured and recorded. This is not the case in this industrial case study where, for a filter q at time τ_j, we only have available the measurements of the pressure drop ΔP_j^q, and the flow across the filter $\dot{M}_j^q$. However, it has been well established that the clogging of a filter medium leads to an increase in pressure drop over the filter as long as the filtration velocity, and thus the flow, is kept constant. It is also known that the pressure drop is proportional to the square of the filtration velocity, and thus we take as an indicator of the state of clogging of filter q at time τ_j the ratio [Nystad, 2009]:

$$z_j^q = \frac{\Delta P_j^q}{(\dot{M}_j^q)^2} \tag{13}$$

Figure 10 shows the sequences of observations $\mathbf{z}_{1:n^q}^q$, q=1:N^{tst} collected on field during the clogging process of N^{tst}=8 filters. We can see from this Figure that the clogging process is affected by large uncertainties, which can be ascribed to the very variable conditions of the sea water; in this context, the challenge is to provide sufficiently narrow confidence interval for the value of the filters RUL.

Due to the absence of physical knowledge about the failure threshold, its value has been arbitrarily set to the value $\delta^{th} = 175$.

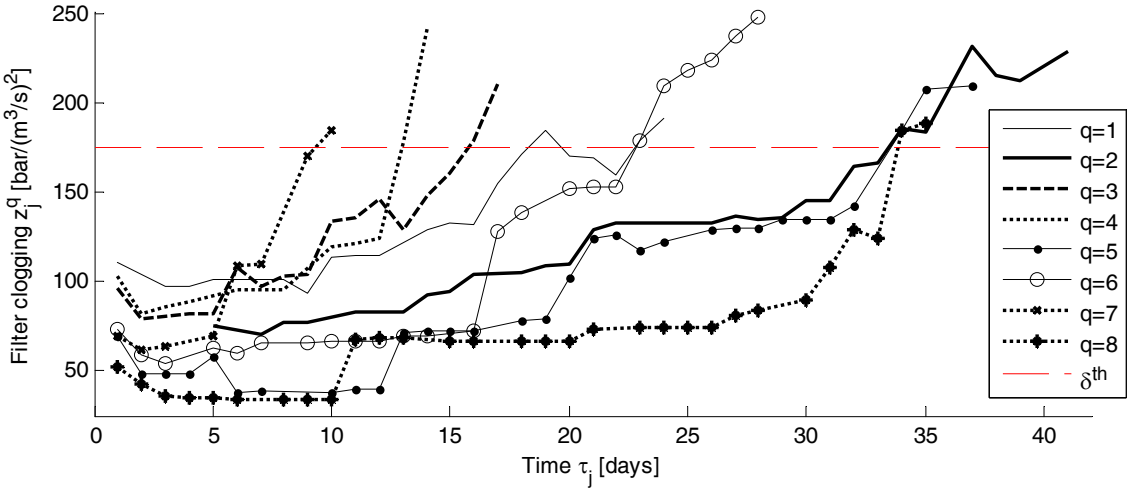

Figure 10: available clogging trajectories $\mathbf{z}_{1:n^q}^q$, *q=1:8.*

The three prognostic approaches proposed are applied at each time instant of each trajectory q=1:8 available, using the remaining R=7 trajectories as training trajectories. Figure 11, 12 and 13 shows the RUL prediction supplied by, respectively, approach 1, 2 and 3.

Due to the large uncertainty of the process, the accuracy of the RUL prediction is always rather low and the confidence intervals are very large in all three approaches. At a first glance, the outcomes of approaches 1 and 2 appear quite similar, unless for trajectories q=5 and q=8 where approach 1 provides more accurate RUL predictions and narrower confidence intervals. In Figure 13 the RUL prediction provided by Approach 3 appear, instead, more accurate, but the confidence intervals are much larger and noisy than for the other two approaches.

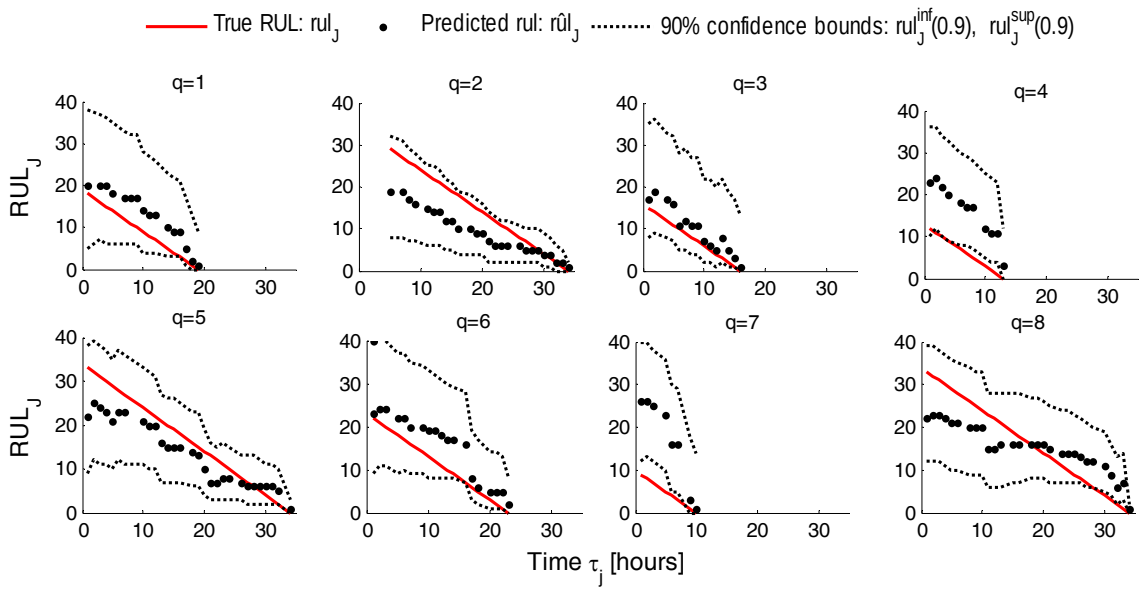

Figure 11: Comparison of the RUL prediction supplied by approaches 1 (dots) and of its confidence bounds (dotted line) with the true RUL value.

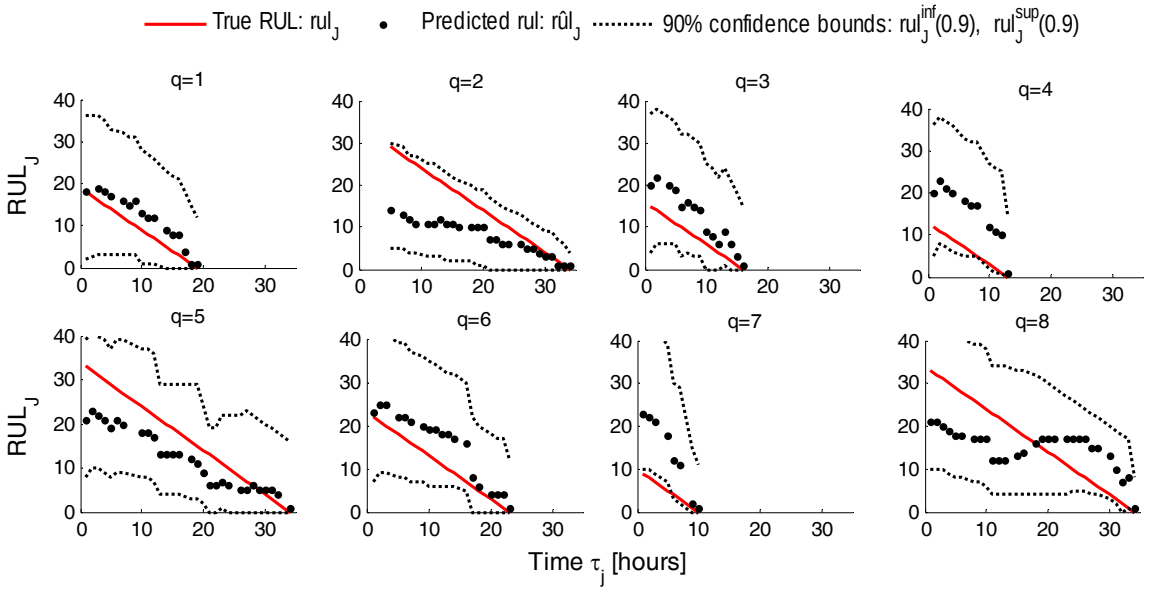

Figure 12: Comparison of the RUL prediction supplied by approaches 2 (dots) and of its confidence bounds (dotted line) with the true RUL value.

To better evaluate the three approaches, the performance indicators RMSE, and $MA_{0.9}$ are also computed in correspondence of the four life fractions $\beta_1 = 0.3$, $\beta_2 = 0.6$, $\beta_3 = 0.8$ and $\beta_4 = 0.95$; since only 8 trajectories are available, to correctly verify the reliability of the method, the coverage indicator $Cov_{0.9}$ is evaluated over all the predictions performed at each time instant τ_j. Figure 14 shows these results and compares the coverage indicator $Cov_{0.9}$ with the target value $\alpha = 0.9$ and the RMSE, and $MA_{0.9}$ indicators with the average value mRUL of the filters RUL at each life fraction β.

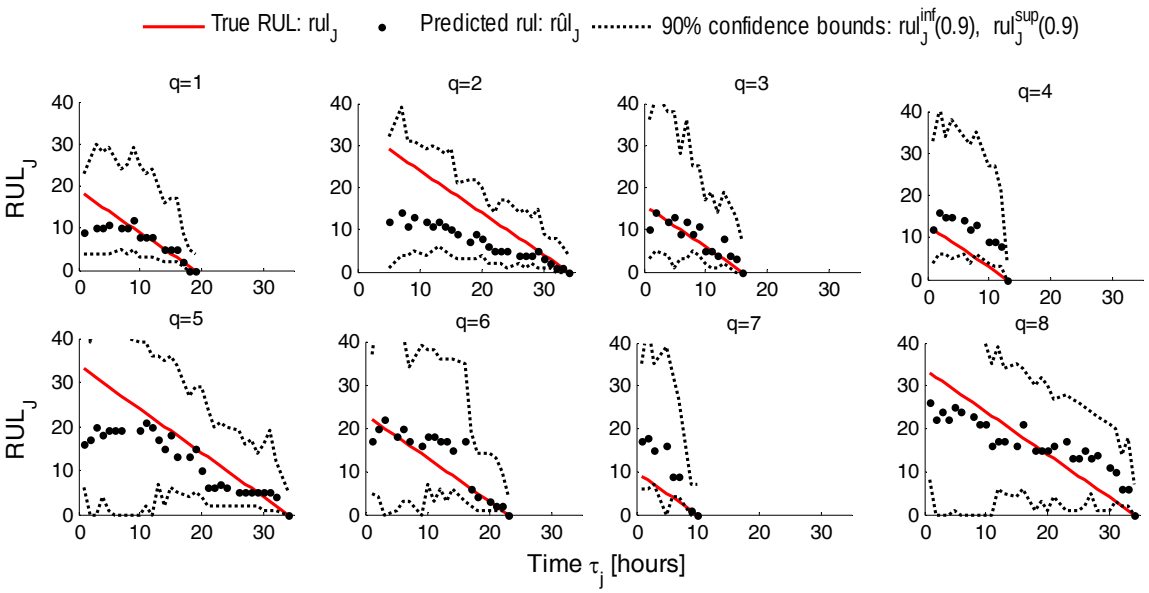

Figure 13: Comparison of the RUL prediction supplied by approaches 3 (dots) and of its confidence bounds (dotted line) with the true RUL value.

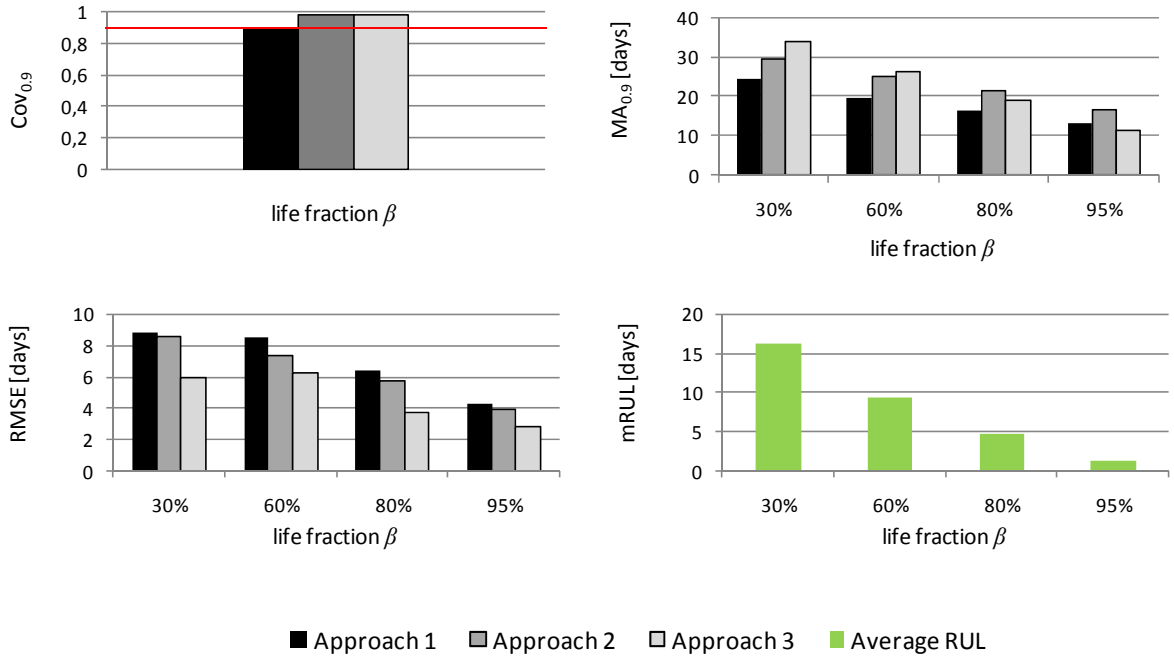

Figure 14: Comparison of the performance of the three approaches on the Ntst=8 filter clogging trajectories. The horizontal (red) line in the $Cov_{0.9}$ figure (upper, left) indicates the target coverage value $\alpha = 0.9$.

Results show that the coverage is very close to the target value $\alpha = 0.9$ for all three approaches. The prediction error (RMSE), however, is always quite large, ranging from being about half of the average RUL value *mRUL* at the beginning of the clogging process ($\beta_1 = 0.3$),up to almost three times (approach 1) this value at $\beta_4 = 0.95$. Approach 3 achieves always the highest accuracy, contrarily to approach 1 that obtains always the largest RMSE. On the other side, approach 1 assures the narrowest confidence intervals for the value of the filters RUL, although their amplitude is still many times the average RUL value *mRUL*.

In the context of filter clogging, due to the large uncertainties involved, we believe that the capability of providing narrower confidence intervals (but still, reliable) of approach 1 should be preferred to the higher prediction accuracy of approaches 2 and 3.

4 Conclusions

In this work, we have considered the problem of predicting the RUL of degrading equipment and providing a measure of confidence in that prediction, based on a set of training degradation trajectories observed in similar equipments which have failed in the past. We have resorted to Gaussian Process Regression to model the evolution of the equipment degradation. GPR treats degradation as a random variable, rather than as a deterministic quantity, and thus is able to model the intrinsic variability of the degradation process and provide predictions about the distribution of future degradation states. By comparing these predictions with a failure threshold, it has been possible to obtain the RUL prediction in the form of a pdf, from which the derivation of confidence intervals for the value of the equipment RUL is straightforward. Moreover, an effective structure of the covariance function has been proposed to allow accounting both for the common structure underlying all degradation trajectories and for the variations specific to each trajectory.

Three different approaches for applying GPR to degradation modeling have been proposed and their performance compared in two case studies concerning, respectively, simulated data about the accumulation of creep damage in ferritic steel, and real data about the clogging of filters used to clean the sea water entering a BWR condenser. In both cases, the first approach proposed, which models the degradation state as a function of time, slightly outperforms the other two, which model the degradation rate as a function of time and, respectively, of the degradation state. However, also approaches 2 and 3 have shown to be capable of providing accurate prediction and could result more suited than approach 1 for treating a different prognostic problem: for example, approach 3 could provide better results if applied to a situation with a weaker observation noise or where the relation between degradation rate and degradation state is more accurately modeled by a GPR, than that between degradation state and time.

When applied to prognostic problems characterized by the high variability of the degradation trajectories, the RUL prediction interval provided by the GPR prognostic approach can fail achieving the desired coverage. We interpreted this as an indication that some sources of uncertainty, most probably the model uncertainty, have not been correctly accounted for. We suggest that resorting to the ensemble techniques [Shi et al., 2005; Baraldi et al., 2012], has the potential of improving these results.

It cannot be ruled out that the results obtained are also affected by the choice of the covariance functions. Here, this choice has been done by a trial and error procedure, with the goal of maximizing the marginal likelihood of the GP with respect to the training data. We believe that the development of a justified procedure for the selection of GPR covariance functions is a fundamental requirement for the successful exploitation of GPR in prognostics.

A further important aspect of GPR is its computational cost which scales typically as $O(N^3)$ with the number of training examples. In our application this was not a problem since we have been working with small training datasets, but it can make the application of the method prohibitive even on modern workstations for larger problems (e.g. $N > 10^4$). Various methods have been suggested in the literature to

reduce this problem by approximating the computations [Rasmussen & Williams, 2006; Shi et al., 2005], and should be considered in future work to make GPR-based prognostics more efficient, especially when used in combination with Monte Carlo sampling (approach 3), which is also a computationally costly procedure.

Appendix A: mean and covariance functions used in the case studies

In the case studies considered in this work, combinations of the covariance functions listed in Table 1A have been used. Table 2A and 3A show the mean function and the terms $k_1(x,x')$ and $k_2(x,x')$ of the covariance function in eq. (11) used in the case studies of Section 3.1 and Section 3.2, respectively.

Table 1A: list of the basic covariance functions used in the case studies of Section 3; symbols b_i, i=1,2,..., are used to indicate the hyper-parameters of the covariance function.

Covariance function	Symbol	Analytical expression of $k(x,x')$
Constant	$Const(x,x')$	b_1
Linear	$\mathrm{Lin}(x,x')$	$b_1(x\cdot x')$
Polynomial	$\mathrm{Poly}_d(x,x')$	$b_1(x\cdot x'+b_2)^d$
Squared exponential	$\mathrm{SE}(x,x')$	$b_1\exp\left(\frac{-(x-x')^2}{b_2}\right)$
Matérn with $\nu=\frac{3}{2}$	$\mathrm{Mat}_3(x,x')$	$b_1\left(1+\frac{\sqrt{3}}{b_2}\lvert x-x'\rvert\right)\exp\left(-\frac{\sqrt{3}}{b_2}\lvert x-x'\rvert\right)$
Rational quadratic	$\mathrm{RQ}(x,x')$	$b_1\left(1+\frac{(x-x')^2}{2b_2b_3}\right)^{-b_2}$
Neural Network	$\mathbf{NN}(x,x')$	$b_1\sin^{-1}\left(\frac{b_2xx'}{\sqrt{(1+b_2x^2)(1+b_2x'^2)}}\right)$

Table 2A: list of functions used to define the prior on the GP mean and covariance in the creep growth case study (Section 3.1); symbols a_i, i=1,2,..., are used to indicate the hyper-parameters of the mean functions.

	Mean function $m(x)=$	Terms of the covariance function in eq. (11)	
		$k_1(x,x')=$	$k_2(x,x')=$
Approach 1	$\sum_{i=0}^{3}a_ix^i$	$\mathrm{Mat}_{3/2}(x,x')$	$\mathrm{RQ}(x,x')+\mathrm{Poly}_3(x,x')$
Approach 2	$\sum_{i=0}^{2}a_ix^i$	$\mathrm{Mat}_{3/2}(x,x')$	$\mathrm{RQ}(x,x')+\mathrm{Poly}_2(x,x')$
Approach 3	a_1x	$\mathbf{NN}(x,x')$	$\mathrm{Poly}_1(x,x')$

Table 3A: list of functions used to define the prior on the GP mean and covariance in the clogging filters case study (Section 3.2); symbols a_i, i=1,2,..., are used to indicate the hyper-parameters of the mean functions.

	Mean function $m(x) =$	Terms of the covariance function in eq. (11)	
		$k_1(x,x') =$	$k_2(x,x') =$
Approach 1	$\sum_{i=0}^{2} a_i x^i$	$\mathrm{SE}(x,x')$	$\mathbf{NN}(x,x') + \mathrm{Poly}_1(x,x')$
Approach 2	$a_0 + a_1 x$	$\mathrm{SE}(x,x')$	$Const(x,x') \times \mathrm{Lin}(x,x') \times \delta(x,x')$
Approach 3	$a_1 x$	$\mathrm{Poly}_1(x,x')$	$\mathrm{Poly}_1(x,x')$

Appendix B: distribution of the parameters in the creep growth model of Section 3.1.1

All parameters φ_{m0} and Q_m, $m = 1:6$ which defines the quantities $\varphi_{1:6} = A, B, H^*, h, K_c, C$ of the creep growth model in eq. (13), the temperature T and the load σ are assumed to have a Gaussian distribution with standard deviation equal to the 0.2% (low variability) or 0.5% (high variability) of their mean values given in Table 1B.

Table 1B: mean values of parameters φ_{m0} and Q_m, $m = 1:6$ of the creep growth model in eq. (13)

Parameter	φ_{10} [h^{-1}]	φ_{20} [MPa^{-1}]	φ_{30}	φ_{40} [MPa]	φ_{50} [h^{-1}]	φ_{60}
Mean value	1.5×10^{-5}	50	5.6	9.5×10^{11}	3.2×10^{6}	3×10^{-5}

Parameter	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6
Mean value	1.9×10^{4}	3450	1450	9800	1.45×10^{4}	6850

Parameter	T [°C]	σ [MPa]
Mean value	565	100

References

Baraldi, P., Cadini, F., Mangili, F., Zio, E., 2013. Model-Based and Data-Driven Prognostics under Different Available Information. accepted for publication in Probabilistic Engineering Mechanics.

Baraldi P., Mangili F., Zio E., 2013. Investigation of uncertainty treatment capability of model-based and data-driven prognostic methods using simulated data. Reliability Engineering and System Safety, Vol 112C, pp. 94-108..

Contal, P., Simao, J., Thomas, D., Frising, T., Callé, S., Appert-Collin, J.C., Bémer, D., 2004. Clogging of fibre filters by submicron droplets. Phenomena and influence of operating conditions. Aerosol Science, Vol. 35, pp. 263-278.

Goebel, K., Saha, B., Saxena, A., 2008. A Comparison of Three Data-Driven Techniques for Prognostics, proc. of the 62nd Meeting of the Society For Machinery Failure Prevention Technology (MFPT), May 6-8, Virginia Beach, Virginia.

Hines, J.W., Usynin, A., 2008. Current Computational Trends in Equipment Prognostics, Int. J. Comput. Intell. Syst., 1(1), 94–102.

Li, W., Pham, H., 2005. An inspection-maintenance model for systems with multiple competing processes. IEEE Transactions on Reliability, vol 54(2), pp. 318–27.

Liu, R., Ma, L., Kang, R., Wang, N., 2011. The modeling method on failure prognostics uncertainties in maintenance policy decision process. Proceedings of the 9th International Conference on Reliability, Maintenance and Safety (ICRMS), Jun 12-15; Guiyang, China.

MacKay, D.J.C., 1998. Introduction to Gaussian processes. In:Bishop CM (ed) Neural networks and machine learning, vol 168. NATO ASI Series, Springer, Berlin, pp 133–165

Mann, R., Freeman, R., Osborne, M., Garnett, R., Armstrong, C., Meade, J., Biro, D., Guilford, T., Roberts, S., 2011. Objectively identifying landmark use and predicting flight trajectories of the homing pigeon using Gaussian processes, Journal of the Royal Society Interface, Vol. 8, pp. 210–219.

Mohanty, S., Chattopadhyay, A., Peralta, P., Das, S., 2011. Bayesian Statistic Based Multivariate Gaussian Process Approach for Offline/Online Fatigue Crack Growth Prediction. Experimental Mechanics, Vol. 51, pp. 833–843.

Muller, A., Sunher, M., Iung, B., 2008. Formalization of a new prognosis model for supporting proactive maintenance implementation on industrial system. Reliability Engineering & system Safety, Vol 93, pp. 234-253.

Mustata, R., Hayhurst, D.R., 2005. Creep constitutive equations for a 0.5Cr 0.5 Mo 0.25V ferritic steel in the temperature range 565 8C–675 8C. International Journal of Pressure Vessels and Piping, Vol 82, pp. 363–372.

Nystad, B.H., 2009. Condition-Based Maintenance (CBM) – filter clogging at OKG 1, a case study, HWR-961, OECD Halden Reactor Project.

Peel, L., 2008. Data Driven Prognostics using a Kalman Filter Ensemble of Neural Network Models, International Conference on Prognostics and Health Management, 06-09 oct, Denver, CO.

Rasmussen, C., Williams, C., 2006. Gaussian processes for machine learning. MIT Press, Cambridge, MA.

Santosh, T.V., Srivastava, A., Sanyasi Rao, V.V.S., Gosh, A. K., Kushwaha, H.S., 2009. Diagnostic System for Identification of Accident Scenarios in Nuclear Power Plants using Artificial Neural Networks, Reliability Engineering and System Safety, 94, 759-762.

Shi, J.Q., Murray-Smith, R., Titterington, D.M., 2005. Hierarchical Gaussian process mixtures for regression, Statistics and Computing, Vol. 15, pp. 31–41.

Song, C.B., Park, H.S., Lee, K.W., 2006. Experimental study of clogging with monodisperse PSL particles. Power Technology, Vol. 163, pp. 152-159.

Tang, L., Kacprzynski, G.J., Goebel, K., Vachtsevanos, G., 2009. Methodologies for uncertainty management in prognostics. Proc IEEE Aerospace Conference, Mar 7-14; Big Sky, MT.

Tipping, M.E., 2001. Sparse Bayesian Learning and the Relevance Vector Machine. Journal of Machine Learning Research, 1:211–244.

Vachtsevanos, G., Wang, P., 2001. Fault prognosis using dynamic wavelet neural networks. Proceedings of IEEE Systems Readiness Technology Conference (AUTEST), pp 857–870.

Wang, T., Yu, J., Siegel, D., Lee, J., 2008. A Similarity-Based Prognostics Approach for Remaining Useful Life Estimation of Engineered Systems. International Conference on Prognostics and Health Management, 6-9 Oct., Denver (CO).

Zio, E., 2012 Prognostics and Health Management of Industrial Equipment. In: Kadry S, editor. Diagnostics and Prognostics of Engineering Systems: Methods and Techniques, IGI-Global.

Zio, E., Di Maio, F., 2010. A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system. Reliability Engineering & System Safety, Vol. 95(1), pp. 49−57.

Paper VI

An Approach to Combining Different Representation of Uncertainty in Data-Driven Prognostics Using Belief Function Theory

Piero Baraldi, Francesca Mangili, Enrico Zio, Giulio Gola, Bent H. Nystad

2013

An Approach to Combining Different Representation of Uncertainty in Data-Driven Prognostics Using Belief Function Theory

Piero Baraldi[a], Francesco Di Maio[a], Francesca Mangili[a], Enrico Zio[b,a]

[a]*Dipartimento di Energia, Politecnico di Milano, Italy*

[b]*Chair on Systems Science and the Energetic challenge, European Foundation for New Energy-Electricite' de France, Ecole Centrale Paris and Supelec, France*

Giulio Gola[c,d], Bent H. Nystad[c]

[c]*Institutt for energiteknikk, Halden, Norway*

[d]*IO Center for Integrated Operations, Trondheim, Norway*

Abstract

We consider two data-driven prognostic approaches for predicting the remaining useful life (RUL) of degrading equipment having available sequences of observations taken during the degradation of similar pieces of equipment. The first degradation-based approach uses Gaussian Process Regression (GPR) to develop a probabilistic model of the evolution of a degradation indicator and provide the probability distribution of the equipment RUL; the second approach uses a similarity-based (SB) method to directly map the relation between the observations and the equipment RUL, and provide a measure of confidence in the RUL prediction by representing it in the framework of belief function theory.

Since the performance of the two approaches are comparable, rather than choosing the best performing one, a method for combining their outcomes and their two different representation of the prediction uncertainty based, respectively, on probabilistic and evidential reasoning, has been proposed. For this, the belief function theory defined on the continuous space is adopted to transform the RUL probability density function supplied by the GPR method into a belief density function based on the least commitment principle. Then, the Dempster's rule is used to aggregate the belief assignments provided by the GPR and the SBR approaches.

The method is applied to the problem of predicting the RUL of filters used to clean the sea water entering the condenser of the boiling water reactor (BWR) in a Swedish nuclear power plant.

Keywords: Prognostics, belief function theory, Gaussian process regression, filter clogging

1 Introduction

For industry, unforeseen equipment failure is costly, both for repair and lost revenue. Recent times have seen the development of predictive maintenance, which is based on the assessment of the actual equipment condition and on the prediction of the optimal time at which performing maintenance. The underlying concept is that of failure prognostics, i.e., predicting the Remaining Useful Life (RUL) of the equipment, defined as the amount of time it will continue to perform its function according to the design specifications [Zio & Compare, 2013].

In general, methods for predicting the equipment RUL can be classified in model-based and data-driven [Zio, 2012]. Model-based methods use an explicit mathematical model of the degradation process to predict the future evolution of the degradation state and, thus, the RUL of the equipment. However, such detailed knowledge is available only for few well studied degradation mechanisms, whereas, most often, equipment degradation depends on the interaction between several degradation mechanisms, which can be very hard to model [Li & Pham, 2005]. Moreover, in order to be used for correctly propagating the uncertainty affecting the RUL predictions, these models should explicitly account for the uncertainty in the degradation process that arises due to scatter in microstructural properties, and to variable loadings and external conditions [Baraldi et al., 2012].

On the other side, data-driven methods are used when an explicit model of the degradation process is not available, but sufficient historical data have been collected. Among data-driven methods one can distinguish between (i) those based on degradation modeling and (ii) those directly predicting the RUL [Wang, 2010]. Degradation-based approaches (i) are based on statistical models that *learn* through regression/trend analysis or stochastic process modeling [Gorjian et al., 2009; Wang, 2010] a model of the degradation evolution from the time series of the observed degradation states, and use it to predict the future degradation states; the predicted degradation is then compared with a failure criteria, such as the maximum value of degradation allowed before the equipment fails performing its functions according to the given specifications (failure threshold). Direct RUL predictions (ii), instead, resort to artificial intelligence techniques to directly map the relation between the observable parameters and the equipment RUL [Schwabacher & Goebel, 2007], without the need of predicting the equipment degradation state and fixing the value of the failure threshold [Peel, 2008, Zio & Di Maio, 2010].

In this work we resort to data-driven prognostics to tackle the problem of predicting the RUL of filters placed upstream the condenser of the boiling water reactor (BWR) of a Swedish nuclear power plant to clean the sea water entering the secondary side of the cooling system. During operations, particles, seaweed, and mussels from the cooling water can cumulate in the filter medium, causing a clogging process. Thus, to assure correct and efficient operations, which require these wastes to be stopped before entering the condenser, prompt and effective cleaning of the filter is desirable; predictive maintenance can help achieving this result, keeping maintenance costs reasonably low. For this, sequences of observations taken during the clogging process of few historical filters are available (training trajectories). Each observation contains the values of three parameters (pressure drop, flow across the filter, and sea water temperature) which provide indirect indications about the degradation (clogging) state of the filters. It is known that this clogging process is affected by large uncertainties, due to the variable conditions of the sea water; in this context, the challenge is to provide a sufficiently narrow confidence interval for the RUL prediction. In fact, uncertainty

management is of paramount importance in prognostics, and requires providing a measure of the RUL prediction uncertainty, e.g., in the form of a prediction interval for the RUL value. This allows assessing the expected mismatch between the real and predicted equipment failure times, which can be used by the maintenance planner to confidently plan maintenance actions, according to the desired risk tolerance [Tang et al. 2009; Liu et al., 2011].

Two prognostic approaches, one of the degradation-based type (i) and the other of the direct RUL prediction type (ii), are implemented in this work. The degradation-based approach (i) uses Gaussian Process Regression (GPR) [Rasmussen & Williams, 2006; Baraldi, Mangili et al., 2013] to fit probability distribution functions (pdfs) to the training degradation trajectories. This way the uncertainty in the future evolution of the degradation states is explicitly modeled and the predictions about the future degradation state distribution can be provided in the form of a Gaussian pdfs [Mohanty et al. 2011; Baraldi, Mangili et al. 2013]. By comparison of the predicted distribution of the future degradation states with a failure threshold it is possible to estimate the probability distribution of the equipment RUL, and from that derive a prediction interval for the RUL value. On the other side, the direct RUL prediction approach (ii) adopted makes use of a similarity-based (SB) method that performs a data-driven similarity analysis between the training trajectories and a newly developing degradation trajectory (test trajectory) to predict its RUL [Zio & Di Maio, 2010]. To treat the prediction uncertainty, a solution based on the belief function theory (BFT) (also called Dempster-Shafer or evidence theory [Dempster, 1967; Shafer, 1976]) has been adopted. The BFT allows combining different pieces of (uncertain) evidence, based on the assignment of basic belief masses to subsets of the space of all possible events (or 'frame of discernment'), which corresponds, in this case, to the RUL domain. Basically, the prognostic method adopted considers each training trajectory as a piece of evidence regarding the value of the RUL of the test trajectory [Baraldi, Mangili et al., 2013]. These pieces of evidence are discounted based on their similarity to the test trajectory and pooled using the Dempster's rule of combination [Petit-Renaud & Denoeux, 2004]. The result is a basic belief assignment (BBA) that quantifies one's belief about the value of the test trajectory RUL given the evidence provided by the reference trajectories. The calculation of a prediction interval relies on the definition of the total belief assigned by the predicted BBA to an interval, which represents a lower bound for the probability that the test equipment RUL belongs to such interval.

Degradation –based prognostics provides a more informative and transparent output, since it supplies a prediction not only of the current equipment RUL, but of the entire degradation trajectory that the equipment will follow, which can be checked against, e.g., expert intuition to verify its consistency. Moreover, contrarily to direct RUL prediction, this approach can be applied even when no historical failure data are available. However, identifying a degradation indicator and fixing a failure threshold may introduce further uncertainty and sources of errors. Thus, in practice, it is often hard to choose between these two approaches. For this reason, we considered the possibility of aggregating the outcomes of both methods: indeed, the aggregation of multiple models that make use of different pieces of information available has often proven to make the prediction more accurate and robust [Polikar, 2006].

In this context, the main contribution of this work is to propose a technique for aggregating the outcomes provided by two different prognostic approaches making use of different representation of the prediction uncertainty. For this, we have resorted to the extension of the BFT to the continuous real axis $\mathfrak{R}$ [Smets,

2005], which allows the transformation of pfds into belief densities. Then, based on the least commitment principle, the RUL pdf predicted by the GPR method is transformed into a belief density function and the Dempster's rule of combination is used to aggregate it with the BBA provided by the SB approach [Ristic & Smets, 2004].

The remaining part of the paper is organized as follows: in Section 2, we briefly state the prognostic problem of interest for this paper; Section 3 describes the method for performing RUL predictions based on GPR; in Section 4, the methodology for providing prediction intervals for the RUL value based on the SB method in the framework of BFT is described; in Section 5 a BFT-based technique for aggregating the outcomes of the GPR and SB approaches is proposed; Section 6 presents the results of the numerical application of these methods to the prediction of the RUL of clogging filters; finally, in Section 7 we state our conclusions and suggest some potential future work.

2 Problem statement

It is assumed that a set of training trajectories is available from measurements collected during the process of degradation of R pieces of equipment similar to the one of interest (test equipment). Each training trajectory r=1:R is made of a sequence $\mathbf{z}^r_{1:n^r}$ of observations $\mathbf{z}^r_j = [z^r_1(\tau^r_j),...,z^r_P(\tau^r_j),...,z^r_P(\tau^r_j)]$ representing the evolution of P relevant parameters z^r_p measured at different time instants τ^r_j, j=1:n^r during the degradation evolution of the r-th equipment, up to the last measurement time $\tau^r_{n^r}$ before its failure, which occurs at time τ^r_F. Equipment is assumed to fail when its degradation exceeds the a maximum acceptable value δ^{th}, called failure threshold. It is also assumed that an indication $z^r_{\delta,j}$ about the degradation state δ^r_j of the r-th equipment at time τ^r_j can be derived from the values of the observed parameters in $\mathbf{z}^r_j$. A sequence of observations $\mathbf{z}^{test}_{1:J}$ from τ^{test}_1 to the present time τ^{test}_J, and consequently a sequence of indirect degradation observations $z^{test}_{\delta,1:J}$, is available also for the test equipment.

The goal of a prognostics model is to predict the RUL of the test equipment at the present time τ^{test}_J. Due to the scatter in the microstructural and manufacturing characteristics, the loading and external conditions variability, etc., the degradation evolution is better represented by a stochastic process, rather than by a deterministic function [Mohanty, 2011]: as a consequence, the degradation state of an equipment at any future time τ and the equipment RUL at the present time τ^{test}_J should be represented by the random variables $\Delta(\tau)$ and RUL^{test}_J. Then, prognostics is expected to provide with the prediction $\hat{rul}_J$ of the mean value of the variable RUL^{test}_J, a representation of its uncertainty [Baraldi et al., 2012]. In this work, we consider a satisfactorily representation of the prediction uncertainty the left bounded prediction interval $PI(\alpha) = [rul^{\inf}_J(\alpha);+\infty]$ containing the true value of the test equipment RUL, hereafter referred to as rul^{true}_J, with probability at least equal to α. Such definition of the prediction interval do not supply the value of the upper bound for the RUL prediction; however, only the lower bound is of interest for the maintenance planner, since it assures that the risk of a failure happening before a time interval equal to $rul^{\inf}_J(\alpha)$ has passed is lower than $1-\alpha$.

3 Degradation-based prognostics: Gaussian Process Regression

In the context of a degradation-based prognostic approach, we aim to model the evolution of equipment degradation as a stochastic process based on the available observations $z^r_{\delta,1:n^r}$ about the degradation trajectories of similar pieces of equipment. The predicted distribution of the future degradation states is then compared with the failure threshold δ^{th}, whose value is assumed to be known, to predict the distribution of the equipment RUL.

Gaussian process regression is a powerful and flexible approach for performing probabilistic inference over functions [Rasmussen & Williams, 2006] and can be effectively used for modeling degradation as a stochastic process [Baraldi, Mangili et al., 2013]. To do that, it is necessary to assume that the distribution of the degradation states is Gaussian with different mean $\overline{\Delta}(\tau)$ and variance $\sigma^2_\Delta(\tau)$ at every time instant τ. The GPR method is used to evaluate the conditional probability density function (pdf) $p_{\Delta(\tau^{test})}(\delta^{test} \mid \mathbf{D}^+_{\tau/z})$ of the future degradation states $\Delta(\tau^{test})$, $\tau^{test} > \tau^{test}_J$ of the test trajectory, given the observation dataset $\mathbf{D}^+_{\tau/z}$ containing both the dataset $\mathbf{D}^{train}_{\tau/z} = \{(\tau^r_j; z^r_{\delta,j})_{j=1:n^r;r=1:R}\}$ drawn from the training trajectories and the dataset $\mathbf{D}^{test}_{\tau/z} = \{(\tau^{test}_j; z^{test}_{\delta,j})_{j=1:J}\}$ drawn from the test trajectory. For mapping the function $\Delta(\tau)$ given the input τ, the GPR defines the *prior* for it in the form of a distribution over functions specified by a Gaussian Process (GP). A GP is a collection of random variables any finite number of which has a joint Gaussian distribution. A real GP $\Delta(\tau)$ is completely specified by its mean $m_\Delta(\tau)$ and covariance $k_\Delta(\tau,\tau')$ functions:

$$\begin{aligned} &\Delta(\boldsymbol{\tau}) \sim \mathrm{GP}\{m_\Delta(\boldsymbol{\tau}); K_\Delta(\boldsymbol{\tau},\boldsymbol{\tau})\} \\ &m(\tau) = E[\Delta(\tau)] \\ &k(\tau,\tau') = E[(\Delta(\tau) - m_\Delta(\tau))(\Delta(\tau') - m_\Delta(\tau'))] \end{aligned} \tag{1}$$

where $\boldsymbol{\tau}$ represents a vector of input values and $K_\Delta(\boldsymbol{\tau},\boldsymbol{\tau})$ indicates the covariance matrix containing the values of $k(\tau,\tau')$ evaluated for all possible pairs of inputs in $\boldsymbol{\tau}$.

This *prior* is taken to represent our prior beliefs over the kind of functions we expect to observe. Typically the *prior* mean and co-variance functions that we use will have some free parameters, called, usually, hyper-parameters. Although the choice of the covariance function must be specified by the user, various methods have been proposed for determining the corresponding hyper-parameters from training data [Rasmussen & Williams, 2006]. Here, the hyper-parameters are optimized by maximizing with the conjugate gradient method the marginal likelihood of the dataset set $\mathbf{D}^{train}_{\tau/z}$ drawn from the training trajectories.

Given the prior information about the GP, the set of hyper-parameters, and the observation dataset $\mathbf{D}^+_{\tau/z}$, the posterior distribution over functions is derived by imposing a restriction on the *prior* distribution to contain only those functions that agree with the observed data [Rasmussen & Williams, 2006]. In other words, we condition the output in correspondence of the test input vector τ^{tst} to the available observations $\mathbf{D}^+_{\tau/z}$ drawn from the same GP, and thus we have:

$$\begin{bmatrix} \Delta(\boldsymbol{\tau}^{train}) \\ \Delta(\tau^{test}) \end{bmatrix} \sim \mathrm{GP}\left(\begin{bmatrix} m_\Delta(\boldsymbol{\tau}^{train}) \\ m_\Delta(\tau^{test}) \end{bmatrix}; \begin{bmatrix} K_\Delta(\boldsymbol{\tau}^{train}, \boldsymbol{\tau}^{train}) & K_\Delta(\boldsymbol{\tau}^{train}, \tau^{test}) \\ K_\Delta(\tau^{test}, \boldsymbol{\tau}^{train}) & K_\Delta(\tau^{test}, \tau^{test}) \end{bmatrix} \right) \tag{2}$$

where $\boldsymbol{\tau}^{train}$ is the vector of all the inputs in $\mathbf{D}^{+}_{\tau/z}$.

The posterior distribution of the output $\Delta(\tau^{test} \mid \mathbf{D}^{+}_{\tau/z})$ in correspondence of the input τ^{test} is Gaussian with mean $\overline{\Delta}(\tau^{test})$ and variance $\sigma^2_\Delta(\tau^{test})$ which can be derived from eq. (2) [Rasmussen & Williams, 2006]:

$$\Delta(\tau^{test} \mid \mathbf{D}^{+}_{\tau/z}) \sim N(\overline{\Delta}(\tau^{test}), \sigma^2_\Delta(\tau^{test})) \tag{3}$$

where

$$\begin{aligned} \overline{\Delta}(\tau^{test}) &= m_\Delta(\tau^{test}) + K(\boldsymbol{\tau}^{train}, \tau^{test})[K(\boldsymbol{\tau}^{train}, \boldsymbol{\tau}^{train})]^{-1}(\mathbf{z}^{train} - m_\Delta(\boldsymbol{\tau}^{train})) \\ \sigma^2_\Delta(\tau^{test}) &= k_\Delta(\tau^{test}, \tau^{test}) - K_\Delta(\tau^{test}, \boldsymbol{\tau}^{train})[K_\Delta(\boldsymbol{\tau}^{train}, \boldsymbol{\tau}^{train})]^{-1} K_\Delta(\boldsymbol{\tau}^{train}, \tau^{test}) \end{aligned} \tag{4}$$

where $\mathbf{z}^{train}$ is the vector of all the outputs in $\mathbf{D}^{+}_{\tau/z}$.

Since the data $\mathbf{D}^{+}_{\tau/z}$, which are used for conditioning the *prior* GP, come partly from the training equipment and partly from the test equipment, it is possible to make the GPR capable of learning both the structure underlying the degradation processes which is common to all test and training trajectories, and the specific variation around this structure that characterizes the test trajectory. This result is obtained by using a covariance function of the following form [Mann et al., 2011; Baraldi, Mangili et al, 2013]:

$$k(\tau^r_j, \tau^{r'}_{j'}) = k_1(\tau^r_j, \tau^{r'}_{j'}) + k_2(\tau^r_j, \tau^{r'}_{j'})\delta(r, r')\delta(j, j') + \sigma^2_z \delta(r, r')\delta(j, j') \tag{5}$$

where σ^2_z is the variance of the white Gaussian noise affecting the observations $z^r_{\delta,j}$ and the reference index assigned to the test trajectory is $r = R+1$. The first term of the kernel corresponds to the covariance associated with the common structure underlying all degradation trajectories; the second represents the covariance owing to the variation of each trajectory around the common structure of all degradation trajectories. This term assumes a finite value only when τ^r_j and $\tau^{r'}_{j'}$ are taken from the same trajectory, since we assume the variation specific to each trajectory to be uncorrelated across trajectories. Finally, the third term accounts for the observation noise associated with the observation $z^r_{\delta,j}$ of the degradation state δ^r_j.

Given the value of the failure threshold, assumed here to be known, and the conditional distribution of the degradation state $p_{\Delta(\tau^{test})}(\delta \mid \mathbf{D}^{+}_{\tau/z})$, the RUL cumulative distribution function (cdf) $P_{RUL^{test}_J}(rul_J \mid \mathbf{D}^{+}_{\tau/z})$ is computed as the probability that the degradation $\Delta(\tau^{test})$ at the future time $\tau^{test} = \tau^{test}_J + rul_J$ will exceed the failure threshold d_{th} [Baraldi et al., 2013]:

$$
\begin{aligned}
P_{RUL_J^{test}}(rul_J \,|\, \mathbf{D}^+_{\tau/z}) &= \text{Prob}(RUL_J^{test} < rul_J \,|\, \mathbf{D}^+_{\tau/z}) \\
&= \int_{d_{th}}^{+\infty} p_{\Delta(\tau_J^{test}+rul_J)}(\delta \,|\, \mathbf{D}^+_{\tau/z})\,\mathrm{d}\,\delta = \\
&= 1 - \Phi\left(\frac{\Delta(\tau_J^{test} + rul_J) - \overline{\Delta}(\tau_J^{test} + rul_J)}{\sigma_\Delta(\tau_J^{test} + rul_J)}\right)
\end{aligned}
\tag{6}
$$

where Φ is the standard normal cdf.

From the RUL cdf one can derive the prediction $\hat{rul}_J^{GPR}$ of the equipment RUL as the mean value of the RUL distribution and the prediction interval $CI_J^{GPR}(\alpha) = [rul_J^{\inf,GPR}(\alpha), +\infty]$ as the left bounded interval containing with probability α the true value of the test equipment RUL, hereafter referred to as rul_J^{true}.

4 Direct RUL prediction: Similarity-based RUL prediction

Within a direct RUL prediction prognostic approach, the mapping between observations $\mathbf{z}_j^r$ (or sequences of observations) and the corresponding RUL value is derived directly from the training trajectories without modeling the degradation process. In this work, this is done by using the similarity-based regression model presented in Baraldi, Di Maio et al. (2013). The idea underpinning this approach to RUL estimation is to evaluate the similarity between the test trajectory and the *R* reference trajectories available and use the RULs of these latter to estimate the RUL of the former, accounting for how similar they are [Petit-Renaud & Denoeux, 2004; Wang et al. 2008; Zio & Di Maio, 2010]. In this context, the Belief Function Theory is used to treat and quantify the uncertainty affecting the SB RUL prediction by assigning to each training trajectory a mass of belief related to its similarity to the test trajectory. In this paper, only few notions of BFT will be recalled as soon as they become necessary for the understanding of the method. For a general introduction to the BFT and for further details about the mathematical development s and the possible interpretations of the theory, the interested reader is referred to Dempster (1976), Shafer (1976) and Smets (1998).

The first requirement of the approach is to define a measure to evaluate the similarity between trajectories. This is done considering the pointwise difference between *n*-long sequences of observations normalized in the range [0.2;0.8]. At the present time τ_J, the distance d_j^r between the sequence of the *n* latest observations $\mathbf{z}_{J-n+1:J}^{test}$ of the test trajectory, and all *n*-long segment $\mathbf{z}_{j-n+1:j}^r$, $j=n:n^r$ of all reference trajectories $r=1:R$ is computed as:

$$
d_j^r = \sqrt{\sum_{i=1}^{n} \left\| \mathbf{z}_{J-n+i}^{test} - \mathbf{z}_{j-n+i}^r \right\|^2}
\tag{7}
$$

where $\|\mathbf{x} - \mathbf{y}\|^2$ is the square Euclidean distance between vectors $\mathbf{x}$ and $\mathbf{y}$.

The similarity s_j^r of the training trajectory segment $\mathbf{z}_{j-n+1:j}^r$ to the test trajectory is defined as a function of the distance measure d_j^r. In Zio & Di Maio (2010) the following bell-shaped function has turned out to give robust results in SB regression due to its gradual smoothness:

$$s_j^r = \exp\left(-\frac{{d_j^r}^2}{\lambda}\right) \tag{8}$$

The arbitrary parameter λ can be set by the analyst to shape the desired interpretation of similarity: the smaller is the value of λ the stronger the definition of similarity. A strong definition of similarity implies that the two segments under comparison have to be very close in order to receive a similarity value s_j^r significantly larger than zero.

For the prediction of the test equipment RUL, a RUL value $\hat{rul}_{j*}^r$ is assigned to each training trajectory r=1:R by considering the difference between the trajectory failure time τ_F^r and the last time instant τ_{j*}^r of the trajectory segment $\mathbf{z}_{j*-n+1:j*}^r$ which has the maximum similarity s_{j*}^r with the test trajectory:

$$\hat{rul}_{j*}^r = \tau_F^r - \tau_{j*}^r \tag{9}$$

Then, the prediction $\hat{rul}_J^{SB}$ of the test equipment RUL at time τ_J^{test} is given by the similarity weighted sum of the values $\hat{rul}_{j*}^r$:

$$\hat{rul}_J^{SB} = \frac{\sum_{r=1}^{R} s_{j*}^r \hat{rul}_{j*}^r}{\sum_{r=1}^{R} s_{j*}^r} \tag{10}$$

In of belief function theory the belief of an agent about the value of a variable RUL_J^{test}, based on the available information, is represented by a basic belief assignment (BBA) of a mass $m_{RUL_J^{test}}(Y^i)$ to all focal elements Y^i, i=1,2,… of the frame of discernment $\Omega_{RUL_J^{test}}$, i.e., the domain of RUL_J^{test}; all elements assigned with a mass of belief $m_{RUL_J^{test}} > 0$ are the focal elements of the BBA. The BBA verifies the condition that the sum of the masses of all its focal elements is 1. The domain $\Omega_{RUL_J^{test}}$ of RUL_J^{test} is defined as the interval $[0, \tau_F^{\max} - \tau_J^{test}]$, where $\tau_F^{\max}$ is the maximum possible life duration of the equipment; we assumed its value to be indicated by an expert. The quantity $\tau_F^{\max} - \tau_J^{test} = RUL_J^{\max}$ is the maximum value that can be assumed by the variable RUL at the present time τ_J^{test}, whereas 0 is, obviously, the minimum possible value of the equipment RUL. Then, within the BFT framework, the different predictions $\hat{rul}_{j*}^r$ can be taken as the unique focal element of as many BBAs $m_{RUL_J^{test}}^r(\{\hat{rul}_{j*}^r\}) = 1$, r=1:R.

The similarity measure s_j^r defined in eq. (9) is interpreted as a measure about the relevance of the source of information inducing the BBA $m_{RUL_J^{test}}^r(\{\hat{rul}_{j*}^r\}) = 1$ and the discounting operation is used to reduce the

belief assigned by $m^r_{RUL^{test}_J}$ to the evidence conveyed by *r*-th trajectory by some factor $(1-\gamma \cdot s^r_{j*})$, $\gamma \in [0,1]$, representing the dissimilarity between the test and the *r*-th training trajectory. The discounted BBAs $\tilde{m}^r_{RUL^{test}_J}(\{\hat{rul}^r_{j*}\})$, *r*=1:*R*, are thus obtained [Petit-Renaud & Denoeux, 2004]:

$$\begin{aligned} &\tilde{m}^r_{RUL^{test}_J}(\{\hat{rul}^r_{j*}\}) = \gamma \cdot s^r_{j*} \\ &\tilde{m}^r_{RUL^{test}_J}(\Omega_{RUL^{test}_J}) = 1-\gamma \cdot s^r_{j*} \end{aligned} \tag{11}$$

The arbitrary parameter $\gamma \in [0,1]$ defines the degree of trust given to the reference trajectories: in fact, if $\gamma < 1$ a part of belief will always be assigned to the entire RUL domain Ω_{RUL_J}, even in the case a reference trajectory were exactly identical to the test one. Notice that the mass assigned to Ω_{RUL_J} represent the *ignorance* we have about the value of RUL^{test}_J because it indicates the absence of evidence that it belongs to any subset of the RUL domain smaller than the domain itself. For a detailed discussion about the choice of the parameters γ and λ, the interested reader is referred to Baraldi, Di Maio et al. (2013).

According to the Dempster's rule of combination, two distinct sources of information inducing two BBAs, e.g., $m^1_{RUL^{test}_J}$ and $m^2_{RUL^{test}_J}$, can be combined to give the aggregated BBA $m^{1\oplus 2}_{RUL^{test}_J}$ [Petit-Renaud & Denoeux, 2004]:

$$\begin{aligned} &m^{1\oplus 2}_{RUL^{test}_J}(Y^i) = \frac{1}{K} \sum_{Y^{i'} \cap Y^{i''} = Y^i} \tilde{m}^1_{RUL^{test}_J}(Y^{i'}) \tilde{m}^2_{RUL^{test}_J}(Y^{i''}), \quad \forall Y^i \in \Omega_{RUL^{test}_J}, \quad Y^i \neq \varnothing \\ &m^{1\oplus 2}_{RUL^{test}_J}(\varnothing) = 0 \end{aligned} \tag{12}$$

where the mass $m^{1\oplus 2}_{RUL^{test}_J}(\varnothing) = 0$ is imposed to convert a possibly subnormal BBA (i.e., a BBA assigning a finite mass to the empty set $\varnothing$) into a normal one and where K is a normalization factor introduced to make the masses $m^{1\oplus 2}_{RUL^{test}_J}(Y^i)$ assigned to all focal elements sum up to 1. Then, by aggregating through eq. (13) the *R* discounted BBAs $\tilde{m}^r_{RUL^{test}_J}$ in eq. (12) one obtains the aggregated BBA:

$$\begin{aligned} &m_{RUL^{test}_J}(\{\hat{rul}^r_{j*}\}) = \frac{\gamma \cdot s^r_{j*}}{K} \prod_{r' \neq r} (1-\gamma \cdot s^{r'}_{j*}), \quad r = 1:R \\ &m_{RUL^{test}_J}(\Omega_{RUL^{test}_J}) = \frac{1}{K} \prod_{r=1}^{R} (1-\gamma \cdot s^r_{j*}) \end{aligned} \tag{13}$$

where

$$K = \prod_{r=1}^{R} (1-\gamma \cdot s^r_{j*}) + \gamma \cdot \sum_{r=1}^{R} s^r_{j*} \prod_{r' \neq r} (1-\gamma \cdot s^{r'}_{j*}) \tag{14}$$

The information conveyed by a BBA can be represented by a belief $\mathrm{Bel}_f(Y^i)$ or by a plausibility function $\mathrm{Pl}_f(Y^i)$ defined, respectively, as

$$\mathrm{Bel}_f(Y^i) = \sum_{Y^{i'} \subseteq Y^i} m_{RUL_J}(Y^{i'}) \tag{15}$$

and

$$\mathrm{Pl}_f(Y^i) = \sum_{Y^{i'} \cap Y^i \neq \varnothing} m_{RUL_J}(Y^{i'}) \tag{16}$$

The belief associated to an interval $[RUL_J^{\inf}, RUL_J^{\sup}]$ represents the amount of belief that directly supports the hypothesis $RUL_J^{test} \in [rul_J^{\inf}, rul_J^{\sup}]$, whereas the plausibility represent the maximum belief that could be committed to this hypothesis if further information became available. Then, belief and plausibility can be seen as lower and upper bounds on the possibility that the hypothesis $RUL_J^{test} \in [rul_J^{\inf}, rul_J^{\sup}]$ is true. Let us consider a left bounded interval $[rul_J^{\inf}, +\infty]$; the belief assigned to such interval is a lower bound for the evidence that the RUL of the test equipment is larger than the lower bound $rul_J^{\inf}$. Then, we can define the prediction interval for the RUL of the test equipment as the interval $CI_J^{SB}(\alpha) = [rul_J^{\inf,SB}(\alpha), +\infty]$ to which a predefined belief α is assigned.

5 Predictions aggregation based on belief function theory

As shown in Baraldi, Di Maio et al. (2013) and Baraldi, Mangili et al. (2013), both the SB and the GPR approaches are effective prognostic methods which allows providing accurate RUL predictions with reliable measures of the prediction uncertainty. Moreover, they both have some advantages: the SB approach does not require the definition of a degradation indicator and the identification of a failure threshold whereas the GPR approach provides a more informative and transparent output, since it predicts the entire degradation path which can be checked against, e.g., expert intuition to verify its consistency. Thus, it can be hard, in practice, to choose between the two approaches. Moreover, when multiple predictions with a comparable accuracy and reliability are available, an alternative solution to the choice of the best one is their aggregation, which has the potential of increasing the accuracy and robustness of the prediction. For this reason, we undertake in this Section the aggregation of the GPR and SB approaches.

Different techniques for the aggregation of the outcomes of individual models have been proposed in the literature, from statistic methods like the simple mean and the median [Baraldi et al., 2011; Polikar, 2007], to weighed averages of the model outcomes based on the global or local performances of the individual models [Baraldi et al., 2010; Bonissone et al., 2008]. However, the main obstacle in the aggregation of the RUL predictions provided by the SB and the GPR approaches concerns the combination of two different representations of the prediction uncertainty. This work focuses on this second aspect, and with respect to the

choice of a strategy for the combination of the RUL predictions $\hat{rul}_J^{GPR}$ and $\hat{rul}_J^{SB}$, limits to adopt the most straightforward solution, i.e., the simple mean:

$$\hat{rul}_J^{Comb} = \frac{\hat{rul}_J^{GPR} + \hat{rul}_J^{SB}}{2} \tag{17}$$

For the aggregation of the two representation of the RUL uncertainty provided by the SB and the GRP approaches and based, respectively, on evidential and probabilistic reasoning, we use the definition of the BFT on the continuous real axis $\Re$ [Smets, 2005], where masses generalizes into densities so to allow representing also probability density functions. This theory is only briefly introduced here; the interested reader may refer to Smets (2005) and Ristic & Smets (2004) for further details.

To extend the belief function theory to real numbers, it is assumed that masses are only allocated to closed intervals $[\underline{x}, \bar{x}]$. A convenient graphical representation of these intervals is the half-plane highlighted in grey in Figure 1, which contains all the sets of pairs $(\underline{x}, \bar{x}) \in \Re^2 : \underline{x} \le \bar{x}$ representing the an interval $[\underline{x}, \bar{x}]$. A belief density function (BDF) $f(\underline{x}, \bar{x})$ is defined on this half-plane, which assigns to each point $(\underline{x}, \bar{x})$ the mass $m_x([\underline{x}, \bar{x}])$ representing the evidence that $x \in [\underline{x}, \bar{x}]$.

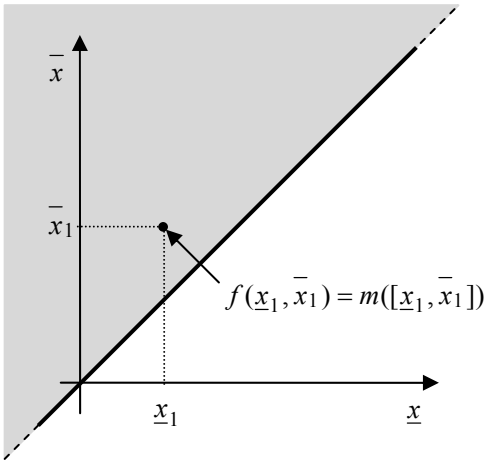

Figure 1: graphical representation of intervals $[\underline{x}, \bar{x}]$.

The total belief $\mathrm{Bel}_f([\underline{x}_1, \bar{x}_1])$ assigned by the BDF f to an interval, e.g., $[\underline{x}_1, \bar{x}_1]$ in Figure 1, is the integral of the BDF $f(\underline{x}, \bar{x})$ over the triangle $(\underline{x}, \bar{x}) \in \Re^2 : \underline{x} \ge \underline{x}_1, \bar{x} \le \bar{x}_1$ highlighted in grey in Figure 2 (left), whereas its plausibility $\mathrm{Pl}_f([\underline{x}_1, \bar{x}_1])$ is the integral of $f(\underline{x}, \bar{x})$ over the half-plane $(\underline{x}, \bar{x}) \in \Re^2 : \underline{x} \le \bar{x}_1, \bar{x} \ge \underline{x}_1$ highlighted in grey in Figure 2 (right).

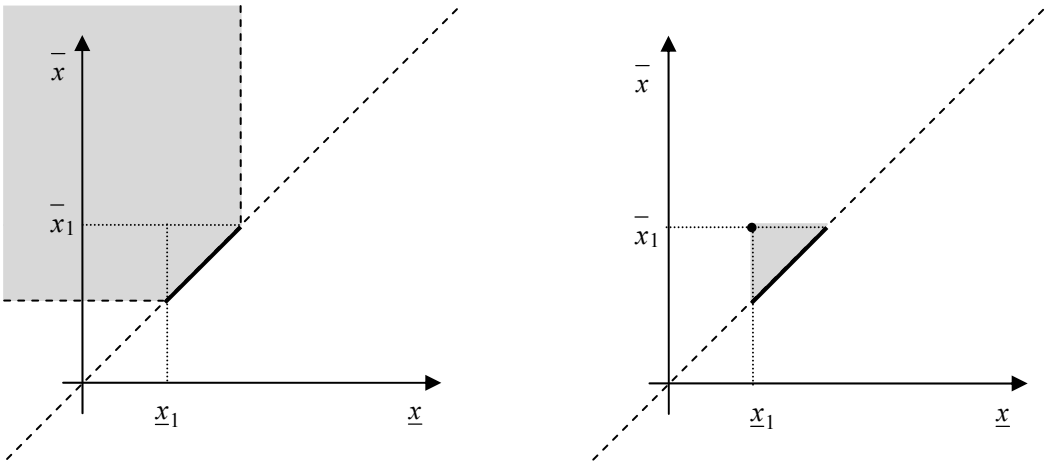

Figure 2: graphical representation of the belief (left) and plausibility (right) associated to the interval $[\underline{x}_1, \bar{x}_1]$.

Within the framework of continuous BFT, the BBA defined by eq. (14) for the RUL value in the similarity based approach can be represented by assigning finite masses only to the degenerated intervals of null dimension $[\hat{rul}^r_{j*}, \hat{rul}^r_{j*}]$, j=1:R, lying on the boundary $\underline{RUL} = \overline{RUL}$ of the half-plane of all possible RUL intervals, and to the RUL domain Ω_{RUL}, i.e., interval $[0, RUL_J^{\max}]$ (see Figure 3). This is represented through the following BDF $f^{SB}_{RUL_J}(\underline{RUL}, \overline{RUL})$:

$$f^{SB}_{RUL_J}(\underline{RUL}, \overline{RUL}) = \sum_{r=1}^{R} m_{RUL_J^{test}}(\{\hat{rul}^r_{j*}\}) \cdot \delta_{\hat{rul}^r_{j*}} + m_{RUL_J^{test}}(\Omega_{RUL_J^{test}}) \cdot \delta_{\Omega_{RUL}} \tag{18}$$

where $\delta_{\hat{rul}^r_{j*}}$ and $\delta_{\Omega_{RUL}}$ are Dirac delta functions which are always zero except when the conditions $\underline{RUL} = \overline{RUL} = RUL^r_{j*}$ and, respectively, $\underline{RUL} = 0, \overline{RUL} = \tau_F^{\max} - \tau_J$ are verified. Notice that, since the equipment RUL assumes only positive values, the half-plane of interest for the RUL prediction (highlighted in grey in Figure 3) is defined by the two constraints $0 \le \underline{RUL} \le \overline{RUL}$.

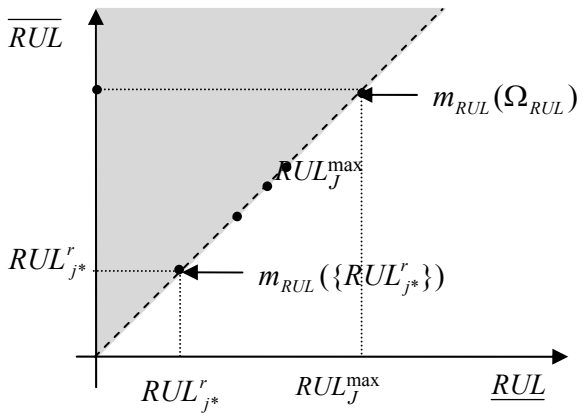

Figure 3: graphical representation of $f_{SB}(\underline{RUL}, \overline{RUL})$.

Concerning the GPR prediction, the straightforward interpretation of a pdf within the framework of continuous BFT would assign a finite value equal to the probability density $p_{RUL_J^{test}}(rul_J|\mathbf{D}^+_{\tau/z})$ to the BDF $f^{GPR1}_{RUL_J}(\underline{RUL},\overline{RUL})$ for all intervals for which $\underline{RUL}=\overline{RUL}=rul_J$ (Figure 4, left). This BDF $f^{GPR1}_{RUL_J}(\underline{RUL},\overline{RUL})$ can be expressed as follows:

$$f^{GPR1}_{RUL_J}(\underline{RUL},\overline{RUL}) = p_{RUL_J^{test}}(\underline{RUL}\,|\,\mathbf{D}^+_{\tau/z})\cdot\delta(\overline{RUL}-\underline{RUL}) \tag{19}$$

where $\delta(\overline{RUL}-\underline{RUL})$ is a Dirac delta function which is always zero except when the condition $\underline{RUL}=\overline{RUL}$ is verified. Such transformation keeps intact the information conveyed by the pdf $p_{RUL_J^{test}}(rul_J|\mathbf{D}^+_{\tau/z})$.

However, our knowledge of the equipment RUL can be considered partial and based on uncertain assumptions. For this reason, a different strategy for representing the GPR prediction in the BFT framework is also considered, which assumes the pdf provided by GPR to be only a representation of some potential betting behaviors, i.e., how an agent would bet about the unknown value of the variable RUL_J^{test}. In the interpretation of the BFT proposed by Smets (1998), this would be called a pignistic pdf $Betf$ induced on $\Re$ by the underlying BDF f whose value is unknown. According to the pignistic transformation [Smets, 2005], many BDF f can induce the same pignistic probability $Betf$. The set of BDF whose related pignistic pdf equals $Betf$ is called the set of isopignistic BDFs induced by $Betf$. In this work, we evoke the least commitment principle (never give more belief than needed) to select the least committed (LC) isopignistic BDF $f^{GPR2}_{RUL_J}(\underline{RUL},\overline{RUL})$ induced by the pignistic pdf $Betf_{GPR}(rul_J)=p_{RUL_J^{test}}(rul_J|\mathbf{D}^+_{\tau/z})$. Other possible choices could be done to select a different isopignistic BDF, however, to limite the scope of this work, only the least committed BDF transformation is considered here.

It can be shown [Smets, 2005] that for a unimodal pignistic pdfs with mode $\mu=\arg\max[Betf^{GPR}_{RUL_J}(rul_J)]$ the focal sets of LC BDF $f^{GPR2}_{RUL_J}(\underline{RUL},\overline{RUL})$ are intervals $[\underline{RUL},\overline{RUL}]$ which satisfy

$$Betf^{GPR}_{RUL_J}(\underline{RUL}) = Betf^{GPR}_{RUL_J}(\overline{RUL}) \tag{20}$$

Such intervals form a line in the half plane of all possible intervals which starts from point (μ,μ) (see Figure 4, right). Since $Betf_{GPR}$ is a bell-shaped density, $\underline{RUL}$ is uniquely defined by a function of $\overline{RUL}$, $\phi(\overline{RUL})$ and the LC BDF is defined by:

$$f^{GPR2}_{RUL_J}(\underline{RUL},\overline{RUL}) = [\phi(\overline{RUL})-\overline{RUL}]\partial Betf^{GPR}_{RUL_J}(\overline{RUL})\delta[\underline{RUL}-\phi(\overline{RUL})] \tag{21}$$

where $\partial Betf^{GPR}_{RUL_J}(\overline{RUL})$ is the derivative of $Betf^{GPR}_{RUL_J}$ calculated at $\overline{RUL}$ and $\overline{RUL}\geq\mu$.

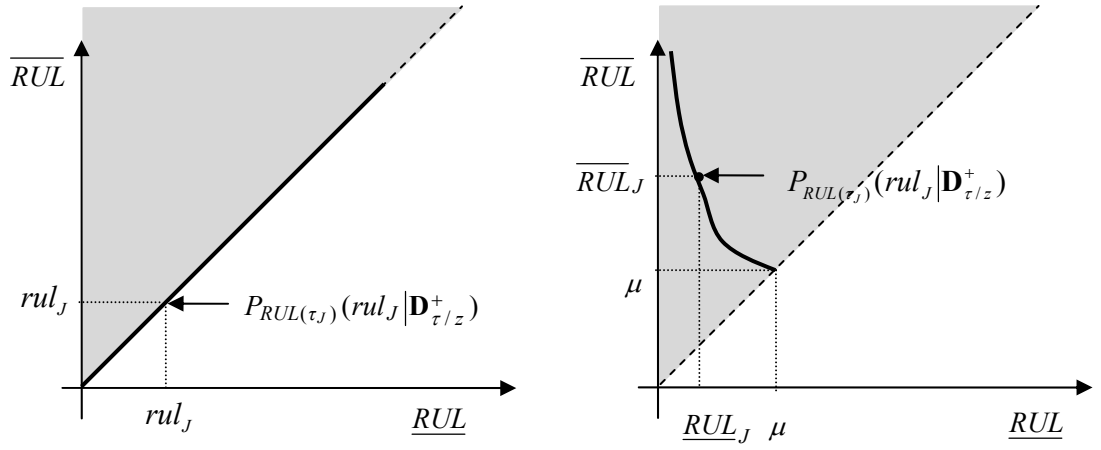

Figure 4: graphical representation of the two possible transformations $f_{GPR1}(\underline{RUL},\overline{RUL})$ and $f_{GPR2}(\underline{RUL},\overline{RUL})$ of the RUL probability density function supplied by the GPR into a belief density function.

Once the RUL pdf predicted by the GPR approach is expressed in the form of a BDF function, either $f^{GPR1}_{RUL_J}(\underline{RUL},\overline{RUL})$ or $f^{GPR2}_{RUL_J}(\underline{RUL},\overline{RUL})$, it can be combined with the SB prediction expressed by the BDF $f^{SB}_{RUL_J}(\underline{RUL},\overline{RUL})$ using the Dempster's rule of combination. For a definition of the conjunctive combination rule in the continuous frame of the real axis $\Re$, please refer to Smets (2005). The two combination strategies, which considers, respectively, $f^{GPR1}_{RUL_J}(\underline{RUL},\overline{RUL})$ or $f^{GPR2}_{RUL_J}(\underline{RUL},\overline{RUL})$, are hereafter referred to as *Comb*1 and *Comb*2. The combined BDFs obtained from *Comb*1 and *Comb*2 are, respectively,

$$f^{Comb1}_{RUL_J}(\underline{RUL},\overline{RUL}) = \frac{1}{K_1} m^{\Omega}_{RUL_J} \cdot f^{GPR1}_{RUL_J}(\underline{RUL},\overline{RUL}) + \sum_{r=1}^{R} m^{r}_{RUL_J} \cdot p_{RUL^{test}_J}(RUL^r_{j*}|\mathbf{D}^+_{\tau/z}) \cdot \delta_{RUL^r_{j*}} \tag{22}$$

and

$$f^{Comb2}_{RUL_J}(\underline{RUL},\overline{RUL}) = \begin{cases} \frac{1}{K_2} m^{\Omega} f^{GPR2}_{RUL_J}(\underline{RUL},\overline{RUL}) & \overline{RUL} \le \overline{\Omega}, \underline{RUL} \ne \overline{RUL} \\ \frac{1}{K_2} m^{\Omega}[\underline{RUL} - \phi^{-1}(\underline{RUL})] \partial Bet f^{GPR2}_{RUL_J}(\phi^{-1}(\underline{RUL})) & \underline{RUL} < \phi(\overline{\Omega}), \overline{RUL} = (\overline{\Omega}) \\ \frac{1}{K_2} m^{r}\ \mathrm{Pl}^{GPR2}_{RUL_J}([\underline{RUL},\overline{RUL}]) & \underline{RUL} = \overline{RUL} = RUL^r_{j*} \end{cases} \tag{23}$$

where K_1 and K_2 are normalization constants that make $f^{Comb1}_{RUL_J}(\underline{RUL},\overline{RUL})$ and $f^{Comb2}_{RUL_J}(\underline{RUL},\overline{RUL})$ integrate to 1 and $\overline{\Omega}$, m^r and m^{Ω} are used instead of $RUL^{\max}_J$, $m_{RUL_J}(\{RUL^r_{j*}\})\cdot$ and $m_{RUL_J}(\Omega_{RUL})$, respectively, for ease of notation.

Given the combined BDF $f_{RUL_J}^{Comb1}(\underline{RUL}, \overline{RUL})$ or $f_{RUL_J}^{Comb2}(\underline{RUL}, \overline{RUL})$, a credible left-bounded interval for the value of the RUL can be estimated as in Section 4 for the SB approach, by taking the intervals $CI_J^{Comb1}(\alpha) = [rul_J^{\inf,Comb1}(\alpha), +\infty]$ and $CI_J^{Comb2}(\alpha) = [rul_J^{\inf,Comb2}(\alpha), +\infty]$ to which a predefined belief α is assigned by the relative BDF.

6 Filter clogging prognostics – a case study

In this Section, we consider the problem of predicting the RUL of filters used to clean the sea water entering the condenser of the BWR reactor of a Swedish nuclear power plant. During operations, filters undergo clogging and, once clogged, can cumulate particles, seaweed, and mussels from the cooling water in the heat exchanger. For this reason, prompt and effective cleaning of the filters is desirable; predictive maintenance can help achieving this result, keeping maintenance costs reasonably low.

From data collected on field, we have available N^{tst}=8 sequences of observations $\mathbf{z}_{1:n^q}^q$, q=1: N^{tst} taken during the clogging process of Q=8 historical filters. Each observation $\mathbf{z}_j^q = [\Delta P_j^q, \dot{M}_j^q, T_j^q]$ contains the measurements of the pressure drop ΔP_j^q, the flow across the filter $\dot{M}_j^q$, and the sea water temperature T_j^q collected at time τ_j^q during the clogging process of the q-th filter. To apply degradation-based prognostics, it is necessary to derive from the observations $\mathbf{z}_j^q$ and indication of the state of clogging affecting the filter at time τ_j. An increasing number of articles can be found in the literature concerning the study of filter clogging by solid aerosols [Song et al., 2006] and liquid aerosols [Contal et al., 2004]. Common for these articles is that the results are achieved in a controlled environment: in the experimental setup all degradation quantities, indicators of degradation, and stressors are automatically measured and recorded. This is not the case in this industrial case study; however, it has been well established that the clogging of a filter medium leads to an increase in pressure drop over the filter as long as the filtration velocity, and thus the flow, is kept constant. It is also known that the pressure drop is proportional to the square of the filtration velocity. Given these results, we take as an indicator of the state of clogging of filter q at time τ_j the ratio [Nystad, 2009]:

$$z_{\delta,j}^q = \frac{\Delta P_j^q}{(\dot{M}_j^q)^2} \tag{24}$$

Also, due to the absence of physical knowledge about the failure threshold, it has been arbitrarily set to the value $\delta^{th} = 175$. Figure 5 shows the sequences of observations $z_{\delta,1:n^q}^q$, q=1:N^{tst} collected on field during the clogging process of the filters. We can see from this Figure that the clogging process is affected by large uncertainties, which can be ascribed to the very variable conditions of the sea water; in this context, the challenge is to provide sufficiently narrow confidence interval for the value of the filters RUL.

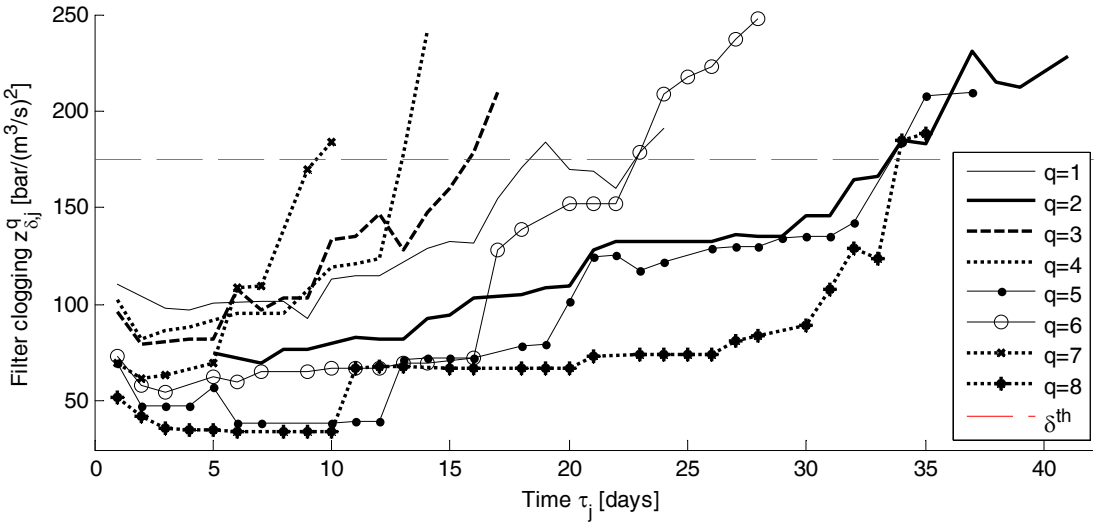

Figure 5: available clogging trajectories $\mathbf{z}^q_{1:n^q}$, *q=1:8.*

6.1 Results

The two prognostic approaches proposed in Sections 3 and 4 and the combination strategy proposed in Section 5 are applied at all time instants of each trajectory q, using the remaining R=7 trajectories $r \neq q$ as training trajectories.

The *priors* on the mean and covariance function of the GP used to model the clogging process in the GPR approach are set as in Baraldi, Mangili et al. (2013) and are:

$$
\begin{aligned}
& m_{\Delta}(\tau) = \sum_{i=0}^{3} a_i \tau^i \\
& k_1(\tau,\tau') = b_1 \exp\left(\frac{-(\tau-\tau')^2}{b_2}\right) \\
& k_2(\tau,\tau') = c_1 \tau \cdot \tau' + c_2 + c_3 \sin^{-1}\left(\frac{c_4 \tau \tau'}{\sqrt{(1+c_4\tau^2)(1+c_4\tau'^2)}}\right) +
\end{aligned}
\tag{25}
$$

where $a_{0:3}$, $b_{1:2}$, and $c_{1:4}$ are the hyper-parameters optimized by maximizing the marginal likelihood of the dataset $\mathbf{D}^{train}_{\tau/z}$ derived from the training trajectories.

Notice that the application of the method proposed in Section 5 to transform the RUL pdf predicted by the GPR approach into the BDF $f^{GPR2}_{RUL_J}$ only applies for bell-shaped distributions. However, the GPR can generate even multimodal pdf. In this work, we impose artificially a bell-shape to the pdfs predicted by the GPR. In practice, if for some values rul_J lower than the coordinate μ of the maximum of the predicted pdf $Betf_{GPR}(rul_J)$, its derivative is negative (or, vice versa, is positive for some values rul_J larger μ), the value of $Betf_{GPR}(rul_J)$ is forced to be constant instead of decreasing (increasing), and the new unimodal pdf thus obtained is normalized so to integrate to one. This device is here justified by the fact that a single failure

mode, clogging, determines the filters RUL, and thus we have no reasons to believe that the RUL distribution should be multi-modal. As a consequence, we attribute the possibly multimodal form of the predicted RUL pdf, to errors made by the GPR model in propagating far in the future the distribution of the degradation states. However, in general, this device should be avoided by extending the LC inverse pignistic transformation to the case of multimodal pdfs.

Concerning the SB approach, we used parameters $\lambda = 0.05$ and $\gamma = 0.95$ according to the results obtained in Baraldi, Di Maio et al. (2013).

Figures 6 and 7 show the RUL prediction and the lower bound of the prediction interval for $\alpha = 0.9$ supplied by, respectively, the GPR and the SB approaches. Due to the large uncertainty of the process, for both approaches the accuracy of the RUL prediction is always rather low and the confidence intervals large. Notice, however, that the GPR approach supplies, in general, smaller prediction intervals than the SB approach. This latter in fact predicts a lower bound which is very often equal to zero; as pointed out by Baraldi, Di Maio et al. (2013) this does not means that the evidence of very early failure is high (as demonstrated by the fact that the predicted RUL can be much larger than zero), but only that the evidence drawn from the reference trajectories is not sufficient to assert with the desired belief of $\alpha = 0.9$ that the RUL value belongs to any subset of the RUL domain $\Omega_{RUL_J^{test}}$. In other words, the prediction $rul_J^{\inf,SB} = 0$ is a statement of *ignorance* about the value of RUL_J^{test}. Although the large intervals predicted by the SB approach cannot be used efficiently by the maintenance planner, they can provide a correct indication that the information conveyed by the training trajectories is not relevant for a specific test trajectory, e.g., because they are too dissimilar. This can be seen for trajectories q=4 and q=7 where the GPR approach provides narrower confidence intervals, but such intervals do not include the true RUL value. Notice also that, for these two trajectories, the prediction $\hat{rul}_J^{SB}$ is more accurate than $\hat{rul}_J^{GPR}$.

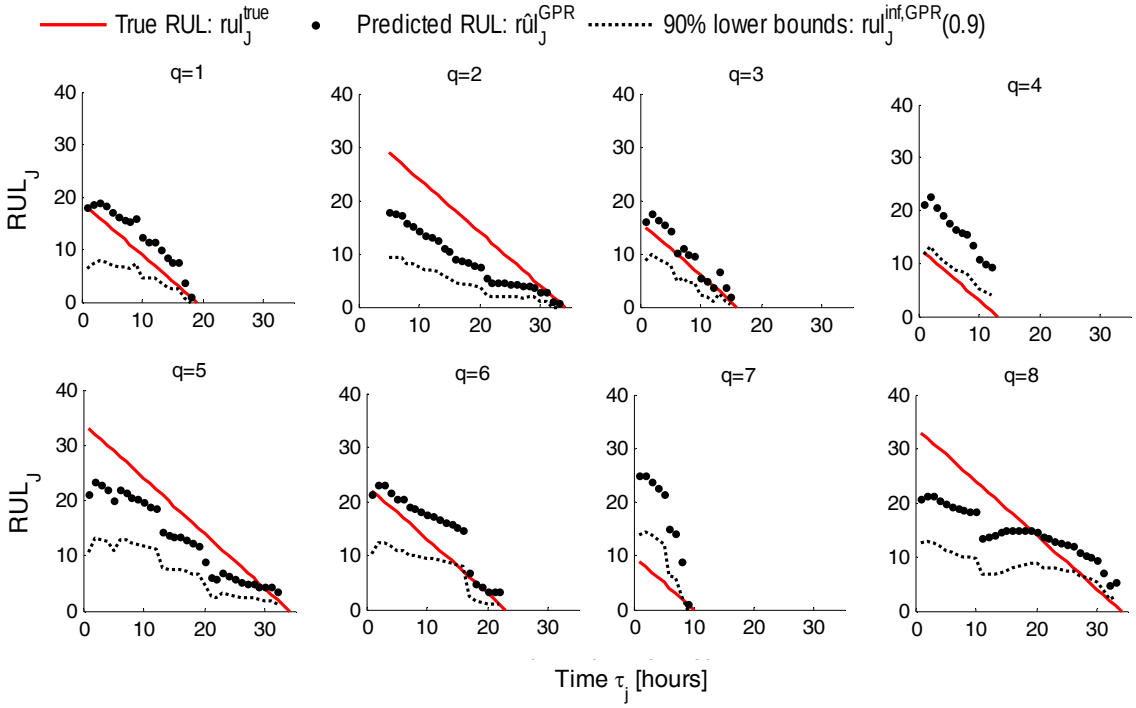

Figure 6: Predictions $\hat{rul}_J^{GPR}$ and $rul_J^{\inf,GPR}(0.9)$ supplied by the GPR approach.

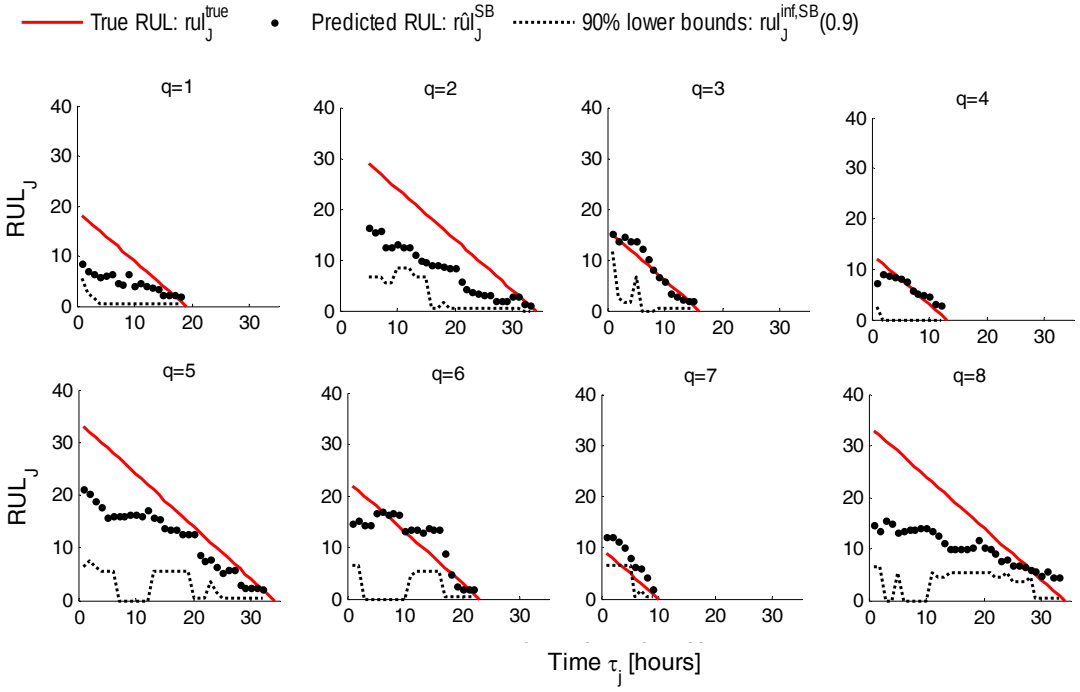

Figure 7: Predictions $\hat{rul}_J^{SB}$ and $rul_J^{\inf,SB}(0.9)$ supplied by the SB approach.

Figure 8 shows the RUL prediction $\hat{rul}_J^{Comb}$ (dots) obtained by averaging the GPR and the SB approaches and the lower bounds of the prediction intervals for $\alpha = 0.9$ obtained using the combination strategies *Comb*1 (dotted line) and *Comb*2 (continuous line). When *Comb*1 is used, the predicted confidence bounds are very similar to those of the GPR approach. This is due to the fact that *Comb*1 keeps intact the information conveyed by the pdf $p_{RUL_J^{test}}(rul_J|\mathbf{D}_{\tau/z}^+)$, and thus provides much stronger information about the RUL value with respect to that provided by the SB BBA. When, instead, the information conveyed by the pdf $p_{RUL_J^{test}}(rul_J|\mathbf{D}_{\tau/z}^+)$ is relaxed through the LC transformation performed by *Comb*2, the resulting BDF $f_{RUL_J}^{Comb2}(\underline{RUL}, \overline{RUL})$ is influence also by the information conveyed by the SB BBA. This can be observed for trajectories q=4 and q=7, where *Comb*2 provides larger confidence bounds than the *Comb*1 and the GPR approaches, thus appearing to be a more robust approach, since its prediction intervals include the true RUL value, and, at the same time are much narrower than those provided by the SB approach.

To better evaluate the different approaches considered, they are applied to each trajectory q in correspondence of the life fraction β, i.e., at time step $\tau_{J(\beta)}^q = \beta \cdot \tau_F^q$ to provide the predictions $\hat{rul}_{J(\beta)}^q$ and $CI_{J(\beta)}^q(0.9) = [rul_{J(\beta)}^{\inf,q}(0.9), +\infty]$. Four life fractions $\beta_1 = 0.3$, $\beta_2 = 0.6$, $\beta_3 = 0.8$ and $\beta_4 = 0.95$ are considered. Performances are evaluated by comparing such prediction with the true RUL value $rul_{J(\beta)}^{q,true}$ to compute two performance indicators:

- The square root of the Mean Square Error (RMSE), i.e., the average value over all clogging trajectories q=1:N^{tst} of the square error $(\hat{rul}_{J(\beta)}^q - rul_{J(\beta)}^{q,true})^2$ made in predicting the true RUL of the test equipment. The MSE measures the accuracy of the prediction and is desired to be as small as possible.

- The amplitude ($MA_{0.9}$) of the interval $[rul_{J(\beta)}^{\inf,q}(0.9), RUL_{J(\beta)}^{\max}]$, averaged over all clogging trajectories q=1:N^{tst}; this indicator gives a measure of the precision of the RUL prediction. In order to have a high precision, we wish the keep the value of $MA_{0.9}$ as small as possible.

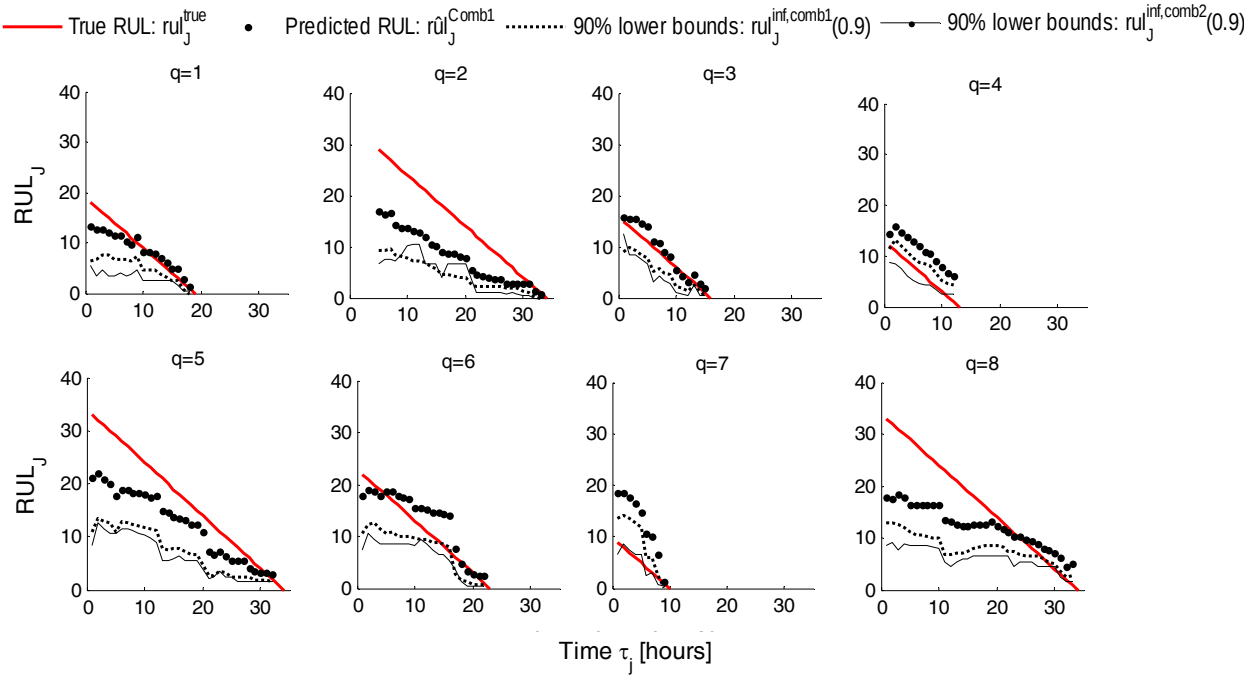

Figure 8: Prediction $\hat{rul}_J^{Comb}$ *(dots) and prediction bounds* $rul_J^{\inf,Comb1}(0.9)$ *(dotted line) and* $rul_J^{\inf,Comb2}(0.9)$ *(continuous line) supplied by Comb1 and Comb2, respectively.*

A third indicator, the coverage $Cov_{0.9}$, is evaluated over all the predictions performed at each time instant τ_J^q along the clogging trajectories by considering the percentage of times the condition $rul_{J(\beta)}^{q,true} \in CI_{J(\beta)}^{q}(0.9)$ is verified. This indicator measures the reliability of the confidence interval; we want the value of $Cov_{0.9}$ to be larger than α=0.9.

Figure 14 shows the value of these indicators obtained for the GPR and the SB approaches and their combination *Comb*2. *Comb*1 is not considered since its RUL prediction is equal to that of *Comb*2 whereas its prediction interval is almost equal to that of the GPR approach. The average value of the RUL, *mRUL*, of the test trajectories at each life fraction β is also shown (bottom, right) for comparison with the value of the RMSE indicator. A horizontal (red) line indicates the target value $\alpha = 0.9$ for the coverage (upper, left).

The results in Figure 9 show that the error and the amplitude of the prediction interval decrease with the equipment life fraction and that the error is always comparable to the average RUL value *mRUL*. These results confirm what already seen in Figures 6, 7 and 8: the GPR approach provides prediction intervals which are narrower that those of the SB approach but cannot assure the desired coverage of 0.9; on the other side, the SB approach makes a smaller error than the other methods, and has a coverage very close to one, but at the price of providing the largest prediction intervals. The proper combination of the two methods allows obtaining the desired coverage with smaller prediction intervals than the SB approach; however, the accuracy is reduced in comparison with the similarity based approach. We expect that different strategies for

the aggregation of $\hat{rul}_J^{GPR}$ and $\hat{rul}_J^{SB}$, accounting, e.g., for the historical performance of the two methods, have the potential of improving the accuracy of the prediction $\hat{rul}_J^{Comb}$.

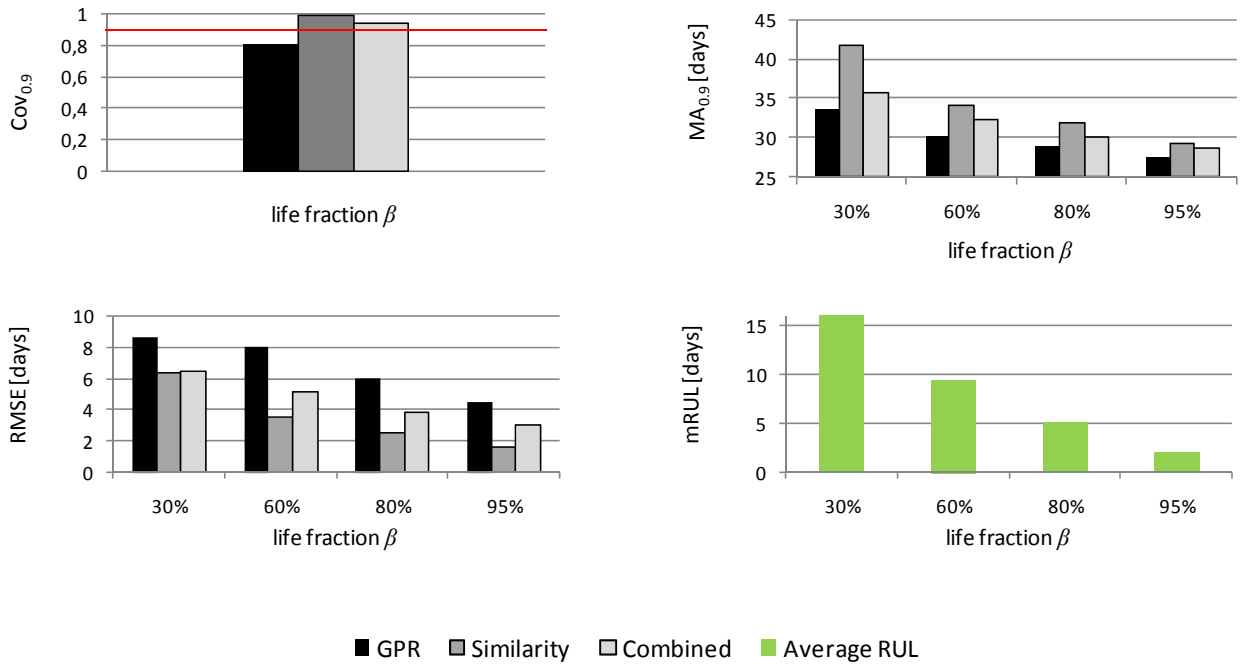

Figure 9: Comparison of the performance indicators $Cov_{0.9}$ (upper, left), $MA_{0.9}$ (upper, right) and RMSE (bottom, left) for the three approaches. The fourth graph (bottom, right) shows the average RUL, mRUL. The horizontal (red) line in the $Cov_{0.9}$ graph (upper, left) indicates the target coverage value $\alpha = 0.9$.

7 Conclusions

In this work, we have considered the problem of predicting the RUL of degrading equipment and providing a measure of its uncertainty, based on sequences of observations collected during the degradation trajectories of a set of similar equipments which have failed in the past. Two different prognostic approaches have been considered: a degradation-based approach resorting to Gaussian process regression to model the evolution of the equipment degradation and predict the probability distribution of the equipment RUL, and a direct RUL prediction approach which exploits SB regression and belief function theory for inferring a basic belief assignment for the value of the test equipment RUL.

In the application of the two methods to the real data concerning the clogging of filters used in a BWR condenser, it is observed that the prediction intervals provided by the SB approach have good coverage, but are extremely large, whereas those provided by the GPR are narrower but do not achieve the desired coverage.

A third approach has, then, been proposed, which is based on the aggregation of the outcomes of these two complementary methods. The main difficulty in performing such aggregation has been the necessity of combining two different representations of uncertainty, based, respectively, on probabilistic and evidential reasoning. By resorting to the belief function theory on continuous variables, it has been possible to translate

both representations of the uncertainty variable RUL_J within the same framework and subsequently aggregate them using the Dempster's rule.

The results obtained by applying the method to the filter clogging case study, have shown that the RUL predictions obtained by aggregating the GPR and the SB approaches represents a good compromise since they reach the desired coverage, contrarily to the GPR predictions, keeping the prediction intervals narrower than those provided by the SB approach alone.

Notice that, the aggregation method proposed in this paper with reference to the aggregation of two specific prognostic methods, can be applied to any situation requiring the combination of multiple uncertain predictions represented in the different frameworks of probabilistic and evidential reasoning.

Concerning the accuracy of the combined prediction $\hat{rul}_J^{Comb}$, future research should consider aggregation strategies different than the simple mean adopted in this work, e.g., performance-based aggregation strategies, which, by accounting for the prediction error made by each approach on historical validation data, have the potential of improving the accuracy of the aggregated prediction.

Notice also that, in a situation where one had reasons to be more confident about the accuracy and reliability of the available RUL pdf, isopignistic transformations different than that based on the least commitment principle, should also be considered, since they have the potential of reducing the amplitude of the prediction intervals.

References

Angstenberger, L., 2001. Dynamic Fuzzy Pattern Recognition, International Series in Intelligent Technologies, 17, Kluwer Academic Publishers.

Baraldi, P., Cammi, A., Mangili, F., Zio, E., 2010. Local Fusion of an Ensemble of Models for the Reconstruction of Faulty Signals - IEEE Transactions on Nuclear Science, vol. 57, pp. 793- 806.

Baraldi, P., Di Maio, F., Mangili, F., Zio, E., Gola, G., Nystad, B.H., 2013. Similarity-based Prediction of Remaining Useful Life using belief function theory for uncertainty treatment.

Baraldi, P., Mangili, F., Zio, E., 2013. Investigation of uncertainty treatment capability of model-based and data-driven prognostic methods using simulated data. Reliability Engineering and System Safety, Vol 112C, pp. 94-108.

Baraldi, P., Zio, E., Gola, G., Roverso, D., Hoffmann, M., 2011. Two novel procedures for aggregating randomized model ensemble outcomes for robust signal reconstruction in nuclear power plants monitoring systems, Annals of Nuclear Energy, Vol. 38 (2-3), pp. 212-220.

Baraldi, P., Mangili, F., Zio, E., Gola, G., Nystad, B.H., 2013. A prognostics approach based on Gaussian Process Regression for degradation modeling.

Bonissone, P., Xue, F., Subbu, R., 2008. Fast Meta-models for Local Fusion of Multiple Predictive Models. Applied Soft Computing Journal 11(2), 1529–1539.

Dempster, A.P., 1976. Upper and lower probabilities induced by a multivariate mapping, Annals of Mathematical Statistics AMS-38 325-339.

Gorjian, N., Ma, L., Mittinty, M., Yarlagadda, P., Sun, Y., 2009. Review on Degradation Models in Reliability Analysis, Proceedings of the 4th World Congress on Engineering Asset Management, 28-30 Sept, Athens.

Li, W., Pham, H., 2005. An inspection-maintenance model for systems with multiple competing processes. IEEE Trans Rel, Vol. 54(2), pp. 318–27.

Liu, R., Ma, L., Kang, R., Wang, N., 2011. The modeling method on failure prognostics uncertainties in maintenance policy decision process. Proc 9th Int Conf on Reliab, Maint and Saf (ICRMS), Jun 12-15; Guiyang, China.

Mohanty, S., Chattopadhyay, A., Peralta, P., Das, S., 2011. Bayesian Statistic Based Multivariate Gaussian Process Approach for Offline/Online Fatigue Crack Growth Prediction. Experimental Mechanics, Vol. 51, pp. 833–843.

Peel, L., 2008. Data Driven Prognostics using a Kalman Filter Ensemble of Neural Network Models, International Conference on Prognostics and Health Management.

Petit-Renaud, S., Denoeux, T., 2004. Nonparametric regression analysis of uncertain and imprecise data using belief functions. International Journal of Approximate Reasoing, Vol 35, pp.1-28.

Polikar, R., 2006. Ensemble Based Systems in Decision Making, IEEE Circuits and Systems Magazine, Vol. 6(3), pp. 21-45.

Polikar, R., 2007. Bootstrap Inspired Techniques in Computational Intelligence, IEEE Signal Processing Magazine, Vol. 24(4), pp. 56-72.

Rasmussen, C., Williams, C., 2006. Gaussian processes for machine learning. MIT Press, Cambridge, MA.

Ristic, B., Smets, P., 2004. Belief function theory on the continuous space with an application to model based classification. In: IPMU (Ed.), Information Processing and Management of Uncertainty, 2004, pp. 1119–1126

Schwabacher, M., Goebel, K., 2007. A Survey of Artificial Intelligence for Prognostics. Association for the Advancement of Artificial Intelligence Fall Symposium, 9-11 Nov, Arlington (VA).

Shafer, G., 1976. A mathematical theory of evidence, Princeton University Press. Princeton, NJ.

Smets, P., 1998. The transferable belief model for quantified belief representation, in: D.M. Gabbay, P. Smets (Eds.), Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, Kluwer Academic Publishers, Dordrecht, pp. 267–301.

Smets, P., 2005. Belief functions on real numbers. International Journal of Approximate Reasoning, Vol 40 (3), pp. 181-223.

Tang, L., Kacprzynski, G.J., Goebel, K., 2009. Vachtsevanos G. Methodologies for Uncertainty Management in Prognostics. Proc IEEE Aerosp Conf, Mar 7-14; Big Sky, MT.

Wang, T., Yu, J., Siegel, D., Lee, J., 2008. A Similarity-Based Prognostics Approach for Remaining Useful Life Estimation of Engineered Systems. International Conference on Prognostics and Health Management, 2008, 6-9 Oct., Denver (CO).

Zio, E., 2012. Prognostics and Health Management of Industrial Equipment. In: Kadry S, editor. Diagnostics and Prognostics of Engineering Systems: Methods and Techniques, IGI-Global.

Zio, E., Compare, M., 2010. Evaluating maintenance policies by quantitative modeling and analysis. Reliability Engineering System Safety, Vol. 109, pp. 53–65.

Zio, E., Di Maio, F., 2010. A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system. Reliability Engineering & System Safety, Vol. 95(1), pp. 49–57.