A Kalman Filter-Based Ensemble Approach With Application to Turbine Creep Prognostics

Piero Baraldi, Francesca Mangili, and Enrico Zio, *Senior Member, IEEE*

***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. 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.**

***Index Terms*—Creep, ensemble, Kalman filter, prognostics and health management.**

ACRONYM

RUL	Remaining useful life.
KF	Kalman filter.
BAGGING	Bootstrapping AGGegatING.
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	ith 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 mth 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 $d(\tau_i)$.
$\mathbf{D}^g$	gth subset of $\mathbf{D}_N$ containing K temporally consecutive measurements; $g = 0, 1, \ldots, G-1$.
G	Number of subsets in which $\mathbf{D}_N$ is partitioned.
$\mathbf{D}^{g,h}$	hth bootstrapped replicate of dataset $\mathbf{D}^g$; $h = 1, 2, \ldots, H$.
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 $\mathbf{D}$.
$f^m(\tau; \mathbf{D}^m)$	mth 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 mth model of the ensemble.
$\hat{RUL}(t)$	Prediction of $\overline{RUL}(t)$ produced by the ensemble.
$w^m(t)$	Weight associated to the mth 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.

Manuscript received August 28, 2011; revised February 24, 2012, June 13, 2012, and July 03, 2012; accepted July 13, 2012. Date of publication October 10, 2012; date of current version November 27, 2012. Associate Editor: M. Xie.

P. Baraldi and F. Mangili are with the Dipartimento di Energia, Politecnico di Milano, 20133 Milano, Italy (e-mail: piero.baraldi@polimi.it; francesca.mangili@mail.polimi.it).

E. Zio is with the Ecole Centrale Paris–Supelec, 91192 Gif-sur-Yvette, France, and also with the Politecnico di Milano, 20133 Milano, Italy.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TR.2012.2221037

$\mathbf{P}(t)$	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.
$r(t)$	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 mth model.
$\sigma^2_{u^m}(t)$	Variance of $u^m(t)$.
$mse^m_{1\rightarrow N_p}$	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^m_o	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.

I. Introduction

PROGNOSTICS AND HEALTH MANAGEMENT (PHM) provides solutions for effective maintenance strategies, to reduce the risk of failure of equipment while supporting 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 equipment undergoing degradation, e.g., time-to-failure data of similar equipment, direct or indirect measures of the degradation path of the monitored equipment or of a set of similar equipment 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, two pieces of information are 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), and
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 some observations indicating the equipment degradation level are available, although affected by uncertainty. This situation is encountered, for example, in rotating equipment where 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, accurate RUL estimate can be obtained [6]. Furthermore, the ensemble is open to accommodate new information by simply adding, to the old models, the new models built on the new information [7]. This approach 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 they are 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 KF 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; and 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 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 II defines the prognostic problem, and the general approach proposed to tackle it. In Section III, the ensemble approach is described, focusing on the BAGGING technique used for the generation of diverse models, and the KF-based aggregation technique. In Section IV, the method is applied to a simulated case study concerning the prediction of the creep evolution in a turbine blade. Finally, Section V states some conclusions, and draws on potential future steps of the work.

II. Degradation Modeling for Prognostics

Let us consider a piece of equipment undergoing a degradation process. When the degradation reaches the threshold value d_{th}, the piece of equipment fails, at τ_f.

At time t, a dataset $\mathbf{D}_N$ of N values $d(\tau_i)$, measured at previous times $\tau_i \leq t, t = 1, 2, \ldots, N$, is available. The objective is to estimate the RUL of the piece of 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., $\hat{\tau}_f = f^{-1}(d_{th}; \mathbf{D}_N)$. Because 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 $d(\tau_i)$, $i = N+1, N+2, \ldots$ 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 modeling 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 the 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], and fuzzy similarity-based methods [16] are typical examples.

III. Ensemble Approach to Prognostics

According to [17], in forecasting, no single method is best for all situations because 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 to be effective solutions because 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], [18]: 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], [19]. To this purpose, different techniques to obtain diversity have been proposed [8], [9], [20], [21], and will be discussed in Section III-A. 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 [22], [7].

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

A. Generation of Multiple Models by BAGGING

Diversity among the individual models of the ensemble can be obtained by different approaches: using different modeling techniques [20], differentiating the training parameters or the models structures [21], projecting the training data into different training spaces [9], [23], [24], or using different datasets to train individual models [8], [25]. 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 the training data using the bootstrapping AGGregatING (BAGGING) method [6], [8], [26], [27]. It consists of generating multiple versions of a predictor by making bootstrap replicates of the dataset, and using these replicates 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].

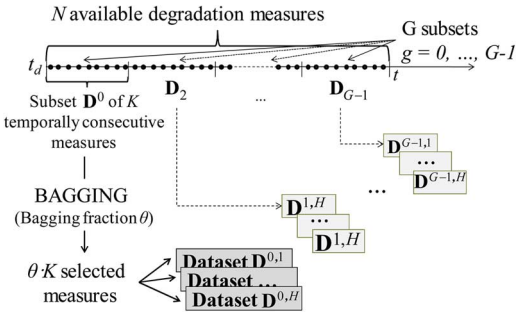

Fig. 1. Procedure for the generation of diverse training datasets.

In this work, to further increase the diversity between the models of the ensemble, the original training dataset $\mathbf{D}_N = [d(\tau_1), \ldots, d(\tau_i), \ldots, d(\tau_N)]$ is partitioned into G subsets, each of which contains K temporally consecutive measurements: $\mathbf{D}^g = [d(gK+1), d(gK+2), \ldots, d(gK+K)]$, $g = 0, 1, \ldots, G-1$. Next, the BAGGING technique is applied to each subset $\mathbf{D}^g$ to generate a number H of bootstrapped replicates $\mathbf{D}^{g,h}$, $h = 1, 2, \ldots, H$, 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 $\mathbf{D}^{g,h}$, $h = 1, 2, \ldots, H$, $g = 0, 1, \ldots, G-1$, 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 into 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 $\hat{RUL}^m = \hat{\tau}_f^m - t$.

B. Aggregation of the Models Outcomes by the KF

Common techniques for the aggregation of multiple model outcomes are based on statistics, e.g., the simple mean (SM) and the median (MD), or on model performance, e.g., the globally weighted average (GWA) [10], [28], [29].

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

$$\hat{RUL} = \frac{\sum_m w^m \cdot \hat{RUL}^m}{\sum_m w^m} \tag{1}$$

According to the SM approach, all weights are equal, i.e., $w^m = (1/(H \cdot G))$, $m = 1, 2, \ldots, H \cdot G$. 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 $\hat{RUL}^{m_C}$ which lies in the center of the distribution of the m model predictions $\hat{RUL}^m$, $m = 1, 2, \ldots, M$.

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

A possibility to improve the accuracy and robustness of the $\hat{RUL}(t)$ predictions in such cases is to properly filter them. In principle, the degradation measurements can be filtered instead of the RUL predictions, but, in the case of non-linear degradation processes, this approach would be a more challenging problem because 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 [30] 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 a discrete time step t. In the discrete KF, the system's dynamics model and the measurement equation take the form

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

The algorithm for the solution of (2) goes through a predict phase, and an update phase. The predict phase (3), and (4) 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 (5)–(7), 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 KF 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 tth time step, and by the term $r(t) = [RUL(t) - RUL(t+\Delta t)]/\Delta t$ representing the RUL rate of modification in time, i.e., $\mathbf{x}(t) = [\overline{RUL}(t)\, r(t)]^{\mathbf{T}}$. 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, because the RUL of a component decreases 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 1 time unit at every time step. For this reason, an element of uncertainty is introduced in (8) through the additive white Gaussian noise $v(t)$.

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

By comparison of (8) with the first equation of the system dynamics model in (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 KF is assumed to be formed by the $RUL(t)$ predictions of the M models in the ensemble, i.e., $\mathbf{z}(t) = [\hat{RUL}^1(t) \ \ldots \ \hat{RUL}^m(t) \ \ldots \ \hat{RUL}^M(t)]^{\mathbf{T}}$.

The observation matrix H has been taken equal to

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

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 the 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. Because, however, the prediction error covariance 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

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

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

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

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 statistical independence between the errors of the different ensemble models f_m, the entries of matrix $\mathbf{R}$ outside its main diagonal are all zero.

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 $\mathbf{Q}$, respectively, the $RUL(t)$ prediction is obtained according to (3)–(7). The drawback of this approach is that the first element of the resulting matrix

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

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.

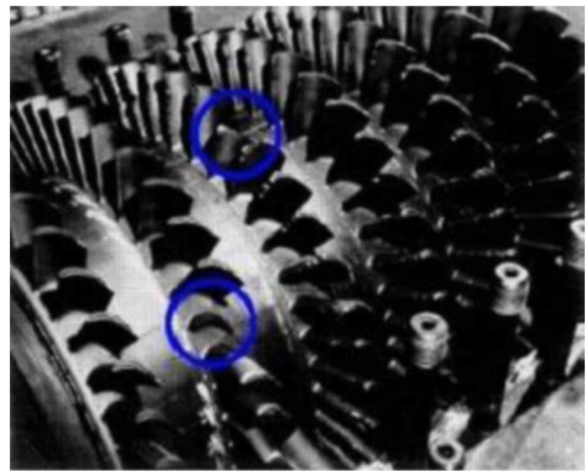

Fig. 2. Blade fracture in a high pressure turbine [31].

In the 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 generated from the previous time estimate $\hat{RUL}(t-1)$, considering that the RUL decreases one time unit for every time unit passed, i.e., $\hat{RUL}^{m+1}(t) = \hat{RUL}(t-1) - 1$. 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 $f(\tau; \mathbf{D})$, because 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.

IV. 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 turbo-machinery because it is accompanied by abrupt changes in the power conversion system and in the reactor flow conditions [31]. Fig. 2 shows an example of high-pressure turbine blade fracture that occurred in a German power plant [31].

As shown in Fig. 3, the uniaxial creep deformation consists of an augmentation of the original length, and a reduction of the

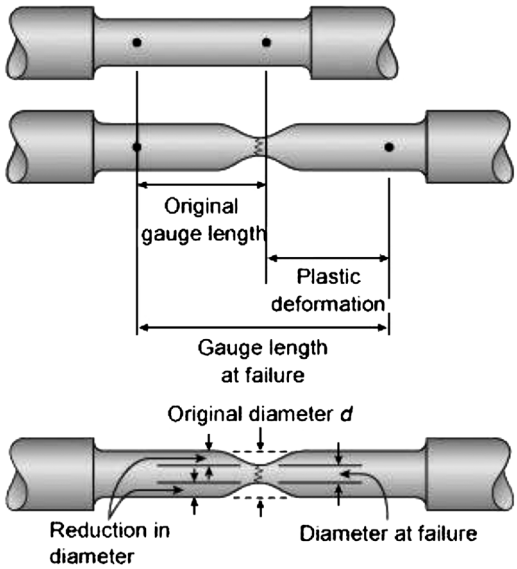

Fig. 3. Schematics of a specimen before and after a creep test [32].

diameter. In this work, the dimensionless 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 a measure of the creep strain. Methods for measuring blades deformation can be found in [33], and [34].

A. Creep Growth Model

The creep evolution has been simulated using the Norton Law, assuming that the dependence from the blade temperature T follows the Arrhenius law [35]:

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

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

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

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, corrosion of the blade, and vibrations of the system or turbulences of the gas flow, are modeled through a random variable δs added to the stress s in (16).

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	$\eta = 2; \xi = 10$

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 286 MWe generation per module [36]; the material inherent characteristics α, n, and ρ are taken assuming that the blade is made of Ni-base cast Superalloy 713LC [36]. The distributions used for the parameters are reported in Table I.

B. Prognostic Problem

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

1) An estimate of the mean ($MTTF = 822.6$ days) and variance ($\sigma^2_{\tau_f} = 5.83 \cdot 10^3$ days^2) of the blades' failure time (τ_f) distribution. These values have been obtained by considering the mean value and the variance of the failure times of a set of 50 simulated degradation paths. Notice that, for these 50 historical trajectories, we assume that no degradation measurement has been taken during the degradation process. This assumption represents a situation in which degradation monitoring is not performed, and thus only the times of a number of failures observed in the past operations of the turbine have been recorded.
2) The failure threshold value $d_{th} = \varepsilon_{th} = 1.5$.
3) The first $N = 50$ direct measurements of the blade creep strain $\mathbf{D} = [d(1), \ldots, d(50)] = [\varepsilon(1), \ldots, \varepsilon(50)]$ (recorded every 20 days). Two different cases will be considered in the following sections: a) $[\varepsilon(1), \ldots, \varepsilon(50)]$ are directly obtained from (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 $[\varepsilon(1), \ldots, \varepsilon(50)]$ to simulate the measurement error (noisy data). Fig. 4 shows an example of simulated creep growth path (upper)

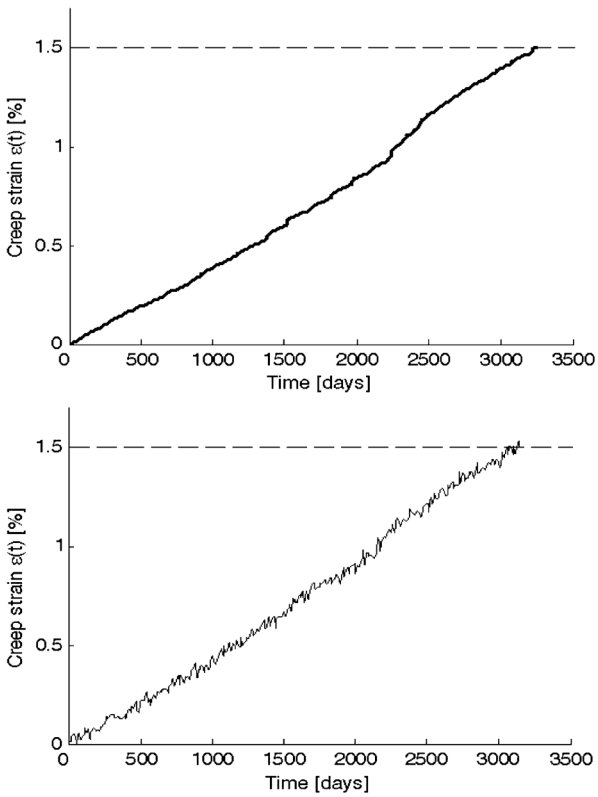

Fig. 4. Example of undisturbed creep growth path (upper) and corresponding noisy measurement sequence (bottom).

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

corresponding to the undisturbed case, and the associated noisy creep strain measurements (bottom).

C. Model Development

The procedure proposed in Section III-A has been applied to the 50 available measurements $[\varepsilon(1),\ldots,\varepsilon(50)]$ 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 times, i.e., $RUL(t=0)=MTTF$, and $\sigma^2_{RUL}(t=0)=\sigma^2_{\tau_f}$.

The creep strain measures available, $\varepsilon(\tau_i)$, are not s-uncorrelated, because 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 $\mathbf{D}=[\varepsilon(1),\ldots,\varepsilon(50)]$ 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 $\mathbf{D}'=[\delta\varepsilon_1,\ldots,\delta\varepsilon_{50}]$ instead of $\mathbf{D}=[\varepsilon(1),\ldots,\varepsilon(50)]$, 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 approach is equivalent to assuming 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.

D. Analysis of the KF Ensemble Performance

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 other methods.

- The traditional statistical method based on the estimate of the mean time to failure (MTTF). This value has been computed by averaging the 50 historical failure times τ_f^i, $i=1,\ldots,50$, i.e., $MTTF=\sum_{i=1}^{50}\tau_f^i$. Finally, the RUL prediction is given by $\hat{RUL}(t)=MTTF-t=822.6-t$.
- A single linear predictive model (SM) $f^{SM}(\tau;\mathbf{D})=\varepsilon_o^{SM}+b^{SM}\tau$ whose parameters are estimated by applying the least squares method to all the $N=50$ available creep strain measurements. The RUL prediction is then given by $\hat{RUL}_{SM}(t)=[\varepsilon_{th}-\varepsilon(\tau_{N_p})]/b_{SM}$.
- The same ensemble of models aggregated using two classical aggregation methods such as 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, $\varepsilon(\tau_{51}),\varepsilon(\tau_{52}),\ldots,\varepsilon(\tau_{N_p})$, 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 (11) for updating the computation of the model performances used in matrix $\mathbf{R}$ (10), and in (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.

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

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

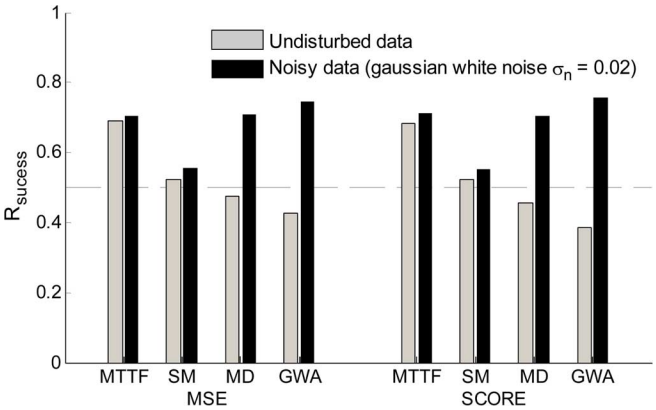

Fig. 5. Fraction of degradation paths in which the KF ensemble outperforms the other methods.

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

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

The SCORE indicator is based on an asymmetric exponential scoring function [37], 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 performance 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

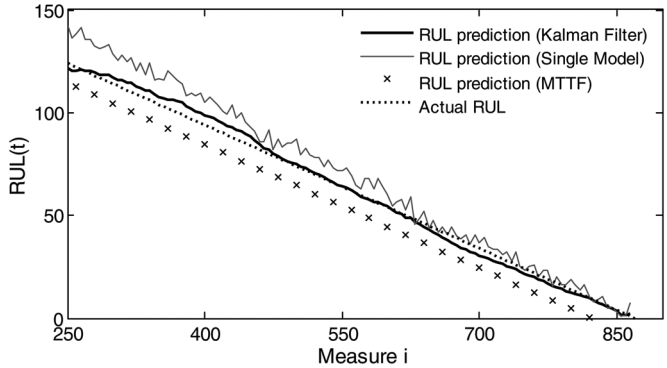

Fig. 6. Comparison of KF ensemble, single model, and MTTF predictions of the RUL for a test degradation path.

the single model and the MTTF methods, both in the case of undisturbed 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 KF. 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.

The obtained 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 because 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 statistically-based aggregation method, such as the median, this outcome 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, showing 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. Notice that the time evolution of the weights assigned to the generic mth 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 relation is due to the fact that the weights are proportional to the model performance, which may vary remarkably because the predicted failure time $\hat{\tau}_i^m$ depends on the noisy creep strain measurement $\varepsilon(\tau_{N_p})$ according to

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

E. 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 pre-

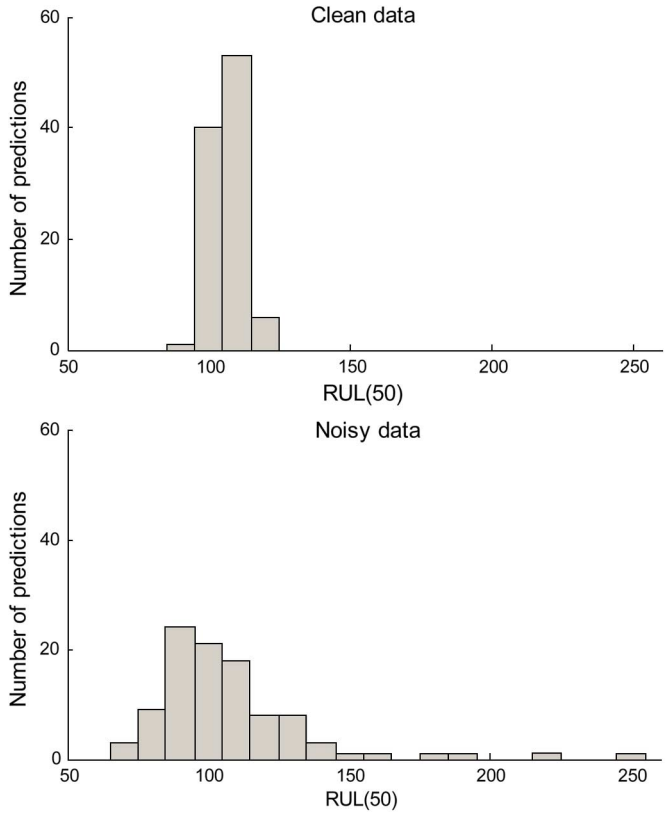

Fig. 7. Histogram of the RUL predictions of 100 different models at $t = 50$ in case of clean (upper) and noisy (bottom) data.

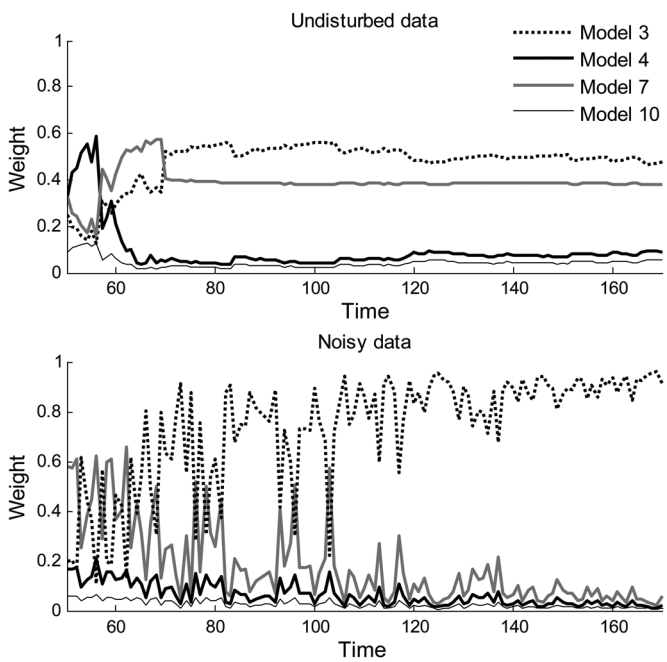

Fig. 8. Weights assigned to 4 different models of the ensemble during a degradation path in case of undisturbed (upper) and noisy (bottom) data.

diction 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. Fig. 9 compares the square root of the MSE (18) made by the single model (continuous line) with that made by the KF ensemble for different

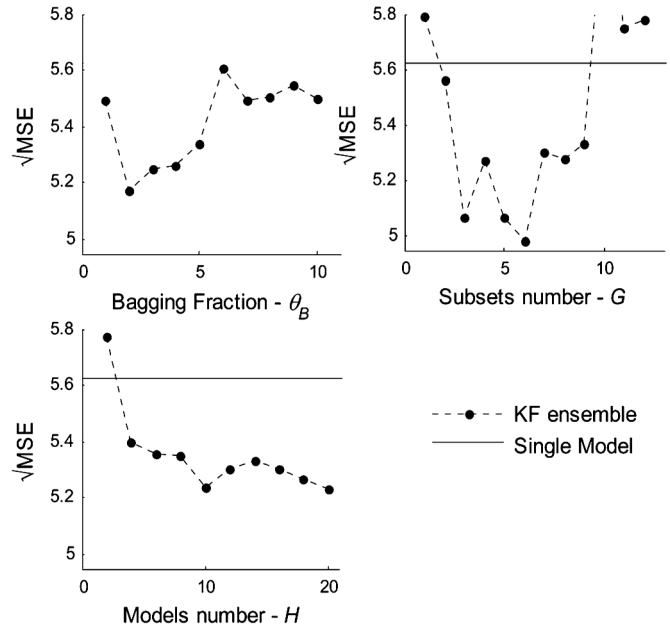

Fig. 9. RMSE of the KF ensemble predictions for different values of ϑ_B (upper, left), K (upper, right), and H (bottom, left) in the case of noisy data.

values of ϑ_B (left), G (middle), and H (right) in the case of noisy data.

Fig. 9 (upper, left) shows that low values of ϑ_B should be preferred because low values of ϑ_B generate more diversity in the bootstrap replicates used to train the ensemble models [38]. With respect to the number of data subsets G, Fig. 9 (upper, right) shows that the ensemble MSE tends to decrease until G reaches the value of 6. Because 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 (bottom, left) shows that the KF ensemble performance remarkably increases until H reaches the value of 10. Because 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.

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 blade's 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 III-B). Fig. 10 shows the variation

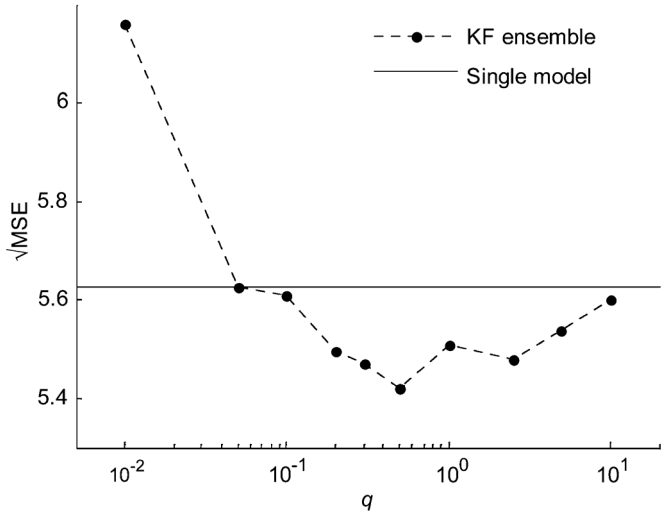

Fig. 10. RMSE of the KF ensemble predictions for different values of q in case of noisy data.

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.

F. Incremental Updating of the Overall Ensemble

In the prognostic scheme previously presented, the information contained in the new measurements $\varepsilon(\tau_i)$, $i = N+1, N+2, \ldots, N_p$, which become available as time passes, is used only for updating the weights in the KF and GWA aggregation methods. This approach allows saving computational time, but may 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 III, 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 method, $H = 5$ 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.

In Table IV, the performance obtained with the updated KF ensemble are compared to those obtained by a single model whose parameters 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.

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

V. CONCLUSIONS

Different forms of information and data may be available for the prognosis of the RUL of 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 equipment. The prognostics have been developed resorting to an ensemble of linear models, individually created by BAGGING, and then aggregated using a KF -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 KF -inspired aggregation strategy allows reducing the noise affecting the prediction. In this respect, the criticality of properly assigning the KF 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 (the higher is q, the 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 levels 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-MODEL KF ENSEMBLE

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

$$\mathbf{x}(t) = \begin{bmatrix} \overline{RUL}(t) \\ r = -1 \end{bmatrix}; \quad \mathbf{z}(t) = \begin{bmatrix} \hat{RUL}^1(t) \\ \hat{RUL}^2(t) \end{bmatrix} \tag{22}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}; \quad \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}; \tag{23}$$

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

$$\mathbf{P}(t) = \begin{bmatrix} p(t) & 0 \\ 0 & 0 \end{bmatrix} \tag{24}$$

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 $\mathbf{P}$ are given by (22)–(24), where $\sigma_{1,2}^2(t) = mse_{1\to N_p}^{1,2}$.

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

$$\hat{\mathbf{x}}^-(t) = \mathbf{A}\hat{\mathbf{x}}(t-1) = \begin{bmatrix} \overline{RUL}(t-1) - 1 \\ -1 \end{bmatrix} \tag{25}$$

$$\begin{aligned} \mathbf{P}^-(t) &= \mathbf{A}\mathbf{P}(t-1)\mathbf{A}^{\mathbf{T}} + \mathbf{Q} \\ &= \begin{bmatrix} p(t-1)+q & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} p^-(t) & 0 \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{26}$$

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

Similarly, expanding (5)–(7), one obtains for the update phase

$$\begin{aligned} \mathbf{K}(t) &= \mathbf{P}^-(t)\mathbf{H}^{\mathbf{T}}\left[\mathbf{H}\mathbf{P}^-(t)\mathbf{H}^{\mathbf{T}} + \mathbf{R}\right]^{-1} \\ &= \begin{bmatrix} p^- & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} p^- + \sigma_1^2 & p^- \\ p^- & p^- + \sigma_2^2 \end{bmatrix}^{-1} \\ &= \frac{1}{p^-\sigma_1^2 + p^-\sigma_2^2 + \sigma_1^2\sigma_2^2}\begin{bmatrix} p^-\sigma_2^2 & p^-\sigma_1^2 \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{27}$$

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}^-(t) + \mathbf{K}(t)\cdot\left[\mathbf{z}(t) - \mathbf{H}\cdot\hat{\mathbf{x}}^-(t)\right] = \begin{bmatrix} x_1(t) \\ -1 \end{bmatrix} \tag{28}$$

$$\begin{aligned} x_1(t) &= \overline{RUL}(t-1) - 1 + \frac{1}{p^-\sigma_1^2 + p^-\sigma_2^2 + \sigma_1^2\sigma_2^2} \\ &\quad \cdot \begin{bmatrix} p^-\sigma_2^2 & p^-\sigma_1^2 \end{bmatrix}\begin{bmatrix} \hat{RUL}^1(t) - \overline{RUL}(t-1) + 1 \\ \hat{RUL}^2(t) - \overline{RUL}(t-1) + 1 \end{bmatrix} \\ &= \frac{1}{w^1 + w^2 + w^K}\Big\{w^1\hat{RUL}^1(t) + w^2\hat{RUL}^2(t) \\ &\qquad + w^K\left[\overline{RUL}(t-1) - 1\right]\Big\} \end{aligned} \tag{29}$$

$$\begin{aligned} \mathbf{P}(t) &= [I - \mathbf{K}(t)\mathbf{H}]\cdot\mathbf{P}^-(t) \\ &= \begin{bmatrix} 1 - \frac{(\sigma_1^2+\sigma_2^2)}{\sigma_1^2+\sigma_2^2+\sigma_1^2\sigma_2^2/p^-} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} p^- & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{w^1+w^2+w^K} & 0 \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{30}$$

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 (29) 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.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their constructive comments which have allowed improving the paper.

REFERENCES

[1] G. Niu, B. S. Yang, and M. Pecht, "Development of an optimized condition-based maintenance system by data fusion and reliability-centered maintenance," *Reliability Engineering and System Safety*, vol. 95, no. 7, pp. 786–796, 2010.

[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, pp. 1483–1510, 2006.

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

[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 Trans. Reliability*, vol. 60, no. 1, pp. 3–13, 2011.

[5] J. Coble and J. Hines, "Fusing data sources for optimal prognostic parameter selection," in *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, no. 3, pp. 21–45, 2006.

[7] P. Baraldi, R. Razavi-Fa, and E. Zio, "Classifier-ensemble incremental-learning procedure for nuclear transient identification at different operational conditions," *Reliability Engineering and System Safety*, vol. 96, no. 4, pp. 480–488, 2010.

[8] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140.

[9] L. Peel, "Data driven prognostics using a Kalman filter ensemble of neural network models," in *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, no. 1, pp. 32–41, 2010.

[11] T. Tinga, "Application of physical failure models to enable usage and load based maintenance," *Reliability Engineering and System Safety*, vol. 95, no. 10, pp. 1061–1075, 2010.

[12] D. Jiang and C. Liu, "Machine condition classification using deterioration feature extraction and anomaly determination," *IEEE Trans. Reliability*, vol. 60, no. 1, pp. 41–48, 2011.

[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, no. 5, pp. 4847–4859, 2012.

[14] V. A. Sotiris, P. W. Tse, and M. G. Pecht, "Anomaly detection through a Bayesian support vector machine," *IEEE Trans. Reliability*, vol. 59, no. 2, pp. 277–286, 2010.

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

[16] E. Zio and F. Di Maio, "A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system," *Reliability Engineering and System Safety*, vol. 95, no. 1, pp. 49–57, 2010.

[17] M. Kotanchek, G. Smits, and E. Vladislavleva, "Trustable symbolic regression models: Using ensembles, interval arithmetic and Pareto fronts to develop robust and trust-aware models," in *Genetic Programming Theory and Practice V*, D. E. Goldberg and J. R. Koza, Eds. : Springer, 2008, pp. 201–220.

[18] P. Baraldi, G. Gola, E. Zio, D. Roverso, and M. Hoffmann, "A randomized model ensemble approach for reconstructing signals from faulty sensors," *Expert Systems With Application*, vol. 38, no. 8, pp. 9211–9224, 2011.

[19] P. Baraldi, A. Cammi, F. Mangili, and E. Zio, "An ensemble approach to sensor fault detection and signal reconstruction for nuclear system control," *Annals of Nuclear Energy*, vol. 37, no. 6, pp. 778–790, 2010.

[20] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
[21] C. Hu, B. D. Youn, P. Wang, and J. T. Yoon, "Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life," *Reliability Engineering and System Safety*, vol. 103, pp. 120–135, 2012.
[22] R. Polikar, L. Udpa, S. S. Udpa, and V. Honavar, "Learn++: An incremental learning algorithm for supervised neural networks," *IEEE Trans. Systems, Man, and Cybernetics—Part C: Applications and Reviews*, vol. 31, no. 4, pp. 497–508, 2001.
[23] C. Lai, M. J. T. Reinders, and L. Wessels, "Random subspace method for multivariate feature selection," *Pattern Recognition Letters*, vol. 27, no. 10, pp. 1067–1076, 2006.
[24] R. Bryll, R. Gutierrez-Osuna, and F. Quek, "Attribute bagging: Improving accuracy of classifiers ensembles by using random feature subsets," *Pattern Recognition*, vol. 36, pp. 1291–1302, 2003.
[25] R. Polikar, "Bootstrap-inspired techniques in computational intelligence," *IEEE Signal Processing magazine*, vol. 24, no. 4, pp. 59–72, 2007.
[26] P. Baraldi, R. Razavi-Far, and E. Zio, "Bagged ensemble of FCM classifiers for nuclear transient identification," *Annals of Nuclear Energy*, vol. 38, no. 5, pp. 1161–1171, 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. Nuclear Science*, vol. 53, no. 3, pp. 1460–1478, 2006.
[28] P. Baraldi, E. Zio, G. Gola, D. Roverso, and 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, no. 2/3, pp. 212–220, 2011.
[29] P. Baraldi, A. Cammi, F. Mangili, and E. Zio, "Local fusion of an ensemble of models for the reconstruction of faulty signals," *IEEE Trans. Nuclear Science*, vol. 57, no. 2, pp. 793–806, 2010.
[30] S. K. Yang and T. S. Liu, "State estimation for predictive maintenance using Kalman filter," *Reliability Engineering and System Safety*, vol. 66, no. 1, pp. 29–39, 1999.
[31] 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, pp. 574–586, 2006.
[32] [Online]. Available: http://www.twi.co.uk/content/jk69.html
[33] J. Ferleger, "Method and Apparatus for Determining Turbine Blade Deformation," U.S. Patent 5238366, August 24, 1993, issued.
[34] S. Hoyte, E. Gebhardt, E. Bauknight, S. Draper, and J. Grant, "Method and Systems for Measuring Blade Deformation in Turbines," U.S. Patent 7762153, July 27, 2010, issued.
[35] 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, pp. 72–79, 2008.
[36] R. Couturier and C. Escaravage, "High temperature alloys for the HTGR gas turbine: Required properties and development needs," in *IAEA Technical Committee Meeting on Gas Turbine Power Conversion Systems for Modular HTGRs*, Palo Alto, CA, November 14–16, 2000.
[37] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *International Conference on Prognostics and Health Management*, Denver, CO, October 6–9, 2008.
[38] G. Martıacute;nez-Muñoz and A. Suárez, "Out-of-bag estimation of the optimal sample size in bagging," *Pattern Recognition*, vol. 43, pp. 143–152, 2009.

Piero Baraldi (BS in nuclear engng., Politecnico di Milano, 2002; PhD in nuclear engng., Politecnico di Milano, 2006) is an assistant professor of Nuclear Engineering at the Department of Energy at the Politecnico di Milano. He is the current chairman of the European Safety and Reliability Association, ESRA, Technical Committee on Fault Diagnosis. His main research efforts are currently devoted to the development of methods and techniques (neural networks, fuzzy and neuro-fuzzy logic systems, ensemble system, kernel regression methods, clustering techniques, genetic algorithms, differential evolution) for system health monitoring, fault diagnosis, prognosis, and maintenance optimization. He is co-author of 42 papers on international journals, and 38 papers on proceedings of international conferences.

Francesca Mangili (MSc in nuclear engng, Politecnico di Milano, 2009) is currently attending the doctoral program in Energy and Nuclear Science Technology at Politecnico di Milano, Department of Energy, supported by a PhD grant of the Institutt For Energiteknikk (IFE), OECD Halden Reactor Project. Her research topics deal with the development of soft computing techniques, system health monitoring, and fault diagnosis and prognosis. She is co-author of two papers in international journals, and two papers in proceedings of international conferences.

Enrico Zio (BS in nucl. engng., Politecnico di Milano, 1991; MSc in mechanical engng., UCLA, 1995; PhD, in nucl. engng., Politecnico di Milano, 1995; PhD, in nucl. engng., MIT, 1998) is Director of the Chair in Complex Systems, the Energetic Challenge of Ecole Centrale Paris and Supelec, full professor, Rector's delegate for the Alumni Association, and past-Director of the Graduate School at Politecnico di Milano, adjunct professor at University of Stavanger. He is the Chairman of the European Safety and Reliability Association ESRA, member of the Korean Nuclear society, member of China Prognostics and Health Management society, and past-Chairman of the Italian Chapter of the IEEE Reliability Society. He is serving as Associate Editor of IEEE TRANSACTIONS ON RELIABILITY, and as editorial board member in various international scientific journals. He has functioned as Scientific Chairman of three International Conferences, and as Associate General Chairman of two others. His research topics include analysis of the reliability, safety and security of complex systems under stationary and dynamic conditions, particularly by Monte Carlo simulation methods; and development of soft computing techniques for safety, reliability and maintenance applications, system monitoring, fault diagnosis and prognosis. He is author or co-author of five international books and more than 170 papers on international journals.