Predictive Maintenance by Risk Sensitive Particle Filtering

Michele Compare[a], Enrico Zio*[a,b], senior member, IEEE

[a]Energy Department, Politecnico di Milano, Milan, Italy
[b]Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy-Electricité de France, Ecole Centrale Paris and Supelec, France
enrico.zio@ecp.fr; enrico.zio@polimi.it; enrico.zio@supelec.fr

Abstract – Predictive Maintenance (PrM) exploits the estimation of the equipment Residual Useful Life (RUL) to identify the optimal time for carrying out the next maintenance action. Particle Filtering (PF) is widely used as a prognostic tool in support of PrM, by reason of its capability of robustly estimating the equipment RUL without requiring strict modeling hypotheses. However, a precise PF estimate of the RUL requires tracing a large number of particles, and thus large computational times, often incompatible with the need of rapidly processing information for making maintenance decisions in due time. This work considers two different Risk Sensitive Particle Filtering (RSPF) schemes proposed in the literature, and investigates their potential for PrM. The computational burden problem of PF is addressed. The effectiveness of the two algorithms is analyzed on a case study concerning a mechanical component affected by fatigue degradation.

Index Terms – Risk sensitive particle filtering, risk function, predictive maintenance.

Acronyms and Abbreviations

PrM   Predictive Maintenance

RUL   Residual Useful Life

PF   Particle Filtering

RSPF   Risk Sensitive Particle Filtering

FSBA   Fuzzy Similarity- Based Approach

RVM   Relevance Vector Machines

SIR   Sampling Importance Re-sampling

SIS   Sequential Importance Sampling

CDF   Cumulative Distribution Function

Notation

d	Failure threshold of the degradation state
D_C	Duration of corrective maintenance actions
D_P	Duration of preventive maintenance actions
$g(x_k, \omega_k)$	Degradation Model
$h(x_k, v_k)$	Measurement model
II	Inspection Interval
N_s	Number of particles traced by PF
P_T	Threshold probability value
$p(z_k \mid x_k^i)$	Probability distribution of the sensor output z_k, given the true degradation state x_k^i
$p(x_k \mid z_1,...,z_k)$	Posterior distribution at the k-th time instant
$p(x_{k+1} \mid x_k, z_1,...,z_k)$	Probability distribution of x_{k+1}, predicted at time k
$q(x_k \mid x_{k-1}, z_1,...,z_{k-1})$	Importance density function
$Q(x_k \mid x_{k-1}^i, z_1,...,z_{k-1})$	CDF of $q(x_k \mid x_{k-1}, z_1,...,z_{k-1})$
$r(x)$	Risk function
t_{PT}	Time instant at which the failure probability reaches P_T
w_k^i	Weight of the i-th particle at the k-th time instant
x_k^i	Position of the i-th particle at the k-th time instant
x_k	Degradation state at the k-th time instant (random variable)

$X_k = \{x_k^i, w_k^i\}$ 　　　　Set of Particles at the k-th time instant

z_k 　　　　Acquired measure at the k-th time instant

γ_k 　　　　Normalization factor

v_k 　　　　Noise on the measurement model

ω_k 　　　　Noise at the k-th time instant

1 Introduction

In recent years, the relative affordability of on-line monitoring technologies has led to a growing interest in new maintenance paradigms such as Predictive Maintenance (PrM). This is founded on the possibility of monitoring equipment to obtain information on its conditions, which is then used to identify problems at an early stage, and predict their changes over time for estimating the equipment Residual Useful Life (RUL). An accurate estimation of the RUL is of great interest, as it would provide lead time to plan, prepare, and execute the repair or the replacement of the equipment, e.g., by delaying the maintenance to the next planned plant outage, by provisioning with spare parts only at the necessary time, by optimizing staff utilization, or while remaining acceptably confident that the system will not fail before maintenance and the equipment lifetime will be fully exploited [1]-[5].

A number of prognostics approaches have been proposed in the literature in support of PrM [6]. Among these, Particle Filtering (PF) is emerging as a powerful model-driven technique, capable of robustly predicting the future behavior of the probability mass distribution that describes the uncertainty in the actual degradation state of the equipment (e.g., the crack depth of a mechanical component, [7]). From the prediction of the future evolution of the degradation and knowledge of the failure threshold (i.e., the degradation value beyond which the equipment loses its function), one can infer the equipment RUL.

Although "PF-based prognostic algorithms have been established as the de facto state of the art in failure prognostics" [8], their application may be impaired in contexts where precise or conservative or both estimations of the RUL are mandatory, because these estimations require large computational times [9]. Indeed, the RUL is related to a failure event, which is generally associated to particles located at the tails of the predicted distributions of the degradation state. This situation entails that the estimation of the failure probability is more sensitive to the imprecision due to the approximate particle representation. The large number of particles that PF needs to trace for providing robust estimates heavily affects its computational burden.

In this work, a solution to this issue is proposed, based on a Risk Sensitive Particle Filtering (RSPF) approach developed in robotics [10], [11], distinctly from the RSPF technique used in control theory and signal processing, whose objective is to minimize the expected exponential of the cumulative quadratic estimation error [12]. A unified view has been proposed in [13].

Our research suggests that the application of RSPF to failure prognostics is new; it has been investigated only in [8], where it is proposed as a method that acts on the model of the evolution of the degradation mechanism to force sampling from the more risky regions of the sample space. Although the scope of this algorithm is the same as that introduced in [10] (i.e., sampling with favor to the particles in the risky regions), there is a fundamental difference between them: the algorithm in [8] tracks the component degradation state, while paying particular care in estimating the tail of the distribution representing the associated uncertainty; on the contrary, the algorithm proposed in [10] factors the risk associated with the particles' positions into the distribution describing the uncertainty in the degradation state: thus, it no longer gives an estimate of the degradation state.

In this work, the two schemes of RSPF are introduced, developed, and compared. To sum up, the contribution of the paper is twofold: on one side, it applies the algorithm in [10] for prognostics purposes and PrM; on the other side, it compares Thrun et al.'s algorithm with that proposed by Orchard et al.

Finally, the issue of reducing the computational times required to estimate the equipment RUL has been tackled in a number of works of literature, which consider techniques different from PF. For example, [9] evaluate and compare the prognostic performance of PF with those of Fuzzy-Similarity Based Approach (FSBA), and Relevance Vector Machine (RVM). From this study, it emerges that FSBA provides predictions of the same accuracy as those of PF in significantly smaller computational times; on the contrary, FSBA does not provide direct indication on the uncertainty in the RUL estimates and does not allow accounting for physical properties. With regards to the comparison of RVM and PF, these techniques have similar pros and cons, and the improvement in computational burden brought by RVM is not significant.

The remainder of the paper is organized as follows. Section 2 briefly describes the main characteristics of PF-based algorithms in PrM applications. Section 3 illustrates the RSPF algorithm proposed in [10], and compares it with the RSPF algorithm proposed in [8]. In Section 4, a case study taken from literature is considered, which concerns the crack growth degradation mechanism in a mechanical component; the application of both RSPF schemes to such an example is also shown and discussed; Section 5 concludes the work.

2 PF-based prognostics for PrM

PF for prognostics is based on [14], [15].

1. A degradation model describing the stochastic evolution of the equipment degradation x in discretized time instants $k = 1, 2, \dots$ is

$$x_{k+1} = g(x_k, \omega_k) \tag{1}$$

where g is a possibly non-linear function, and ω_k is a possibly non-Gaussian noise.

2. A set of measures of past and present values of some physical quantities z related to the equipment degradation x is $z_1, ..., z_k$.

3. A probabilistic measurement model that links z with the equipment degradation x is

$$z_k = h(x_k, v_k) \tag{2}$$

where h is a possibly non-linear function, and v_k is the measurement noise.

The pseudo code of the algorithm follows (adapted from [10]).

1. set $X_k = X_k^{aux} = \varnothing$

2. **for** $i = 1, ..., N_s$

3.   pick the i-th sample $x_{k-1}^i \in X_{k-1}$;

4.   draw $x_k^i \approx p(x_k \mid x_{k-1}^i)$;

5.   set $w_k^i \approx p(z_k \mid x_k^i)$;

6.   add x_k^i to X_k^{aux} with weigh w_k^i

7. **endfor**

8. **for** $i = 1, ..., N_s$ %SIR algorithm

9.   draw x_k^i from X_k^{aux} with probability proportional to w_k^i;

10.   add x_k^i to X_k with $w_k^i = 1 / N_s$

11. **endfor**

Briefly, at every time instant $k = 1, 2, ...$, a set $X_k = \{x_k^i, w_k^i\}$ of particles positioned at x_k^i with weights w_k^i, $i = 1, ..., N_s$, is considered. This set constitutes a discrete approximation of the true probability density function (pdf) representing the uncertainty in the system state x. The particles of X_k evolve statistically independently of each other, according to the probabilistic degradation model of (1). This approach allows predicting the positions of the particles of the set X_{k+1} at time instant $k+1$, given that their current positions are those in X_k.

When a new measurement z_{k+1} is collected, the predicted pdf $p(x_{k+1} \mid x_k, z_1, ..., z_k)$ is adjusted through the modification of the weights of the particles, thus yielding the posterior distribution $p(x_{k+1} \mid z_1, ..., z_{k+1}) \approx X_{k+1} = \{x_{k+1}^i, w_{k+1}^i\}$. This second, updating step, usually called the filtering step, requires the knowledge of the probabilistic law that links the state of the equipment to the gathered measure (2). From this model, the probability distribution $p(z \mid x)$ of observing the sensor output z given the true degradation state

x is derived (measurement distribution), and used to update the weights of the particles upon a new measurement collection. Roughly speaking, the smaller the probability of encountering the acquired measurement value, when the actual equipment state is that of the particle, the larger the reduction of the particle weight in the posterior distribution. On the contrary, a good match between the acquired measure and the particle state results in an increase of the particle importance (for further details, see [16], [17]).

Notice that the pseudo-code above refers to the Sampling Importance Resampling (SIR) version of PF [16], which builds a new population of particles by sampling with replacement from the set $X_k = \{x_k^i, w_k^i\}$ (lines 8-11), where the chance of a particle being picked is proportional to its weight w_k^i. The final weight assigned to the particles of this new set is $1/N_s$. The SIR algorithm allows avoiding the degeneracy phenomenon (i.e., after few iterations, all but few particles would have negligible weights), which is typical of the standard version of PF (i.e., Sequential Importance Sampling, SIS).

Finally, in the present work, the PF-based prognostic model is supposed to be embedded within the PrM scheme as follows.

- The degradation state is periodically measured, with a period of *II*=20 hours.
- When a measure is acquired, the estimation of the distribution of the current degradation state is updated, and the future evolution of such a distribution is simulated. This approach allows us to estimate the RUL, defined as the difference between time instant t_{PT} at which the failure probability reaches a pre-fixed threshold value P_T (e.g., $P_T = 1\%$) and the current time instant.
- The preventive replacement action is performed either when the estimated RUL is elapsed, or at the acquisition of a measurement if the updated distribution leads to equipment failure with a probability larger than P_T.

For example, Fig. 1. shows the evolution of the estimate of $t_{0.1}$ when the failure time is 753 hours (dashed line), and *II*=20 hours. See that the estimate becomes more accurate with time, and the final value is 744 hours. At this time instant, the component undertakes a preventive replacement action; thus, failure is avoided, and the component life is widely exploited.
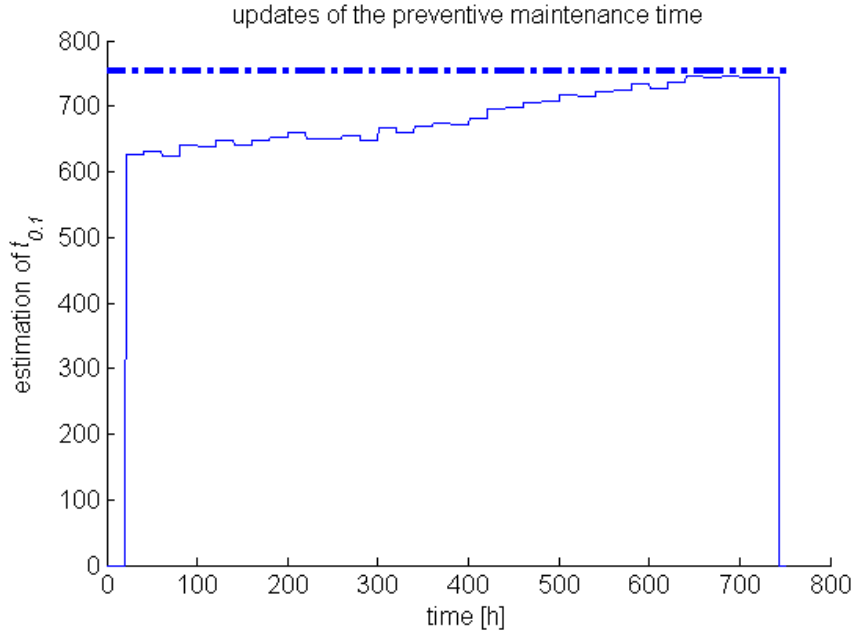
Fig. 1. Evolution of the estimate of $t_{0.1}$.

3 RSPF algorithm: the scope

In the considered PrM setting, we are asked to estimate the equipment failure probability at the successive time instants $k* = k+1, k+2, \ldots$. To this aim, the stochastic behavior of the N_s particles from instant k is simulated according to the model in (1), thus providing an estimation of the distribution $p(x_{k*} | z_1, \ldots, z_k)$. The failure probability P_T is given by the area of the tail of $p(x_{k*} | z_1, \ldots, z_k)$ that crosses the failure threshold d (filled area in Fig. 2). When we are interested in estimating small values of P_T (e.g., $P_T = 1\%$, 5%), PF necessitates handling an appropriate number N_s of particles; otherwise, it gives non conservative values. This necessity is due to the fact that PF provides a particle-based, approximate pdf representing the uncertainty in the state x; the smaller the number of particles N_s, the worse the approximation.

To clarify this issue, let us assume that $P_T = 1\%$ (this may be the case where failures are very high costly); then, PF-SIR needs tracing at least $N_s = 100$ particles to (roughly) estimate $t_{0.01}$ as the time instant at which there is at least 1 particle above d. If N_s is smaller than 100, then the failure probability value is inevitably underestimated. For example, if we consider $N_s = 20$, the time instant $t_{0.01}$ corresponds at least to $t_{0.05}$. Obviously, considering N_s larger than 100 (e.g., 500) would provide a more robust estimation of $t_{0.01}$ because a larger number of particles would exceed the threshold (e.g., 5).
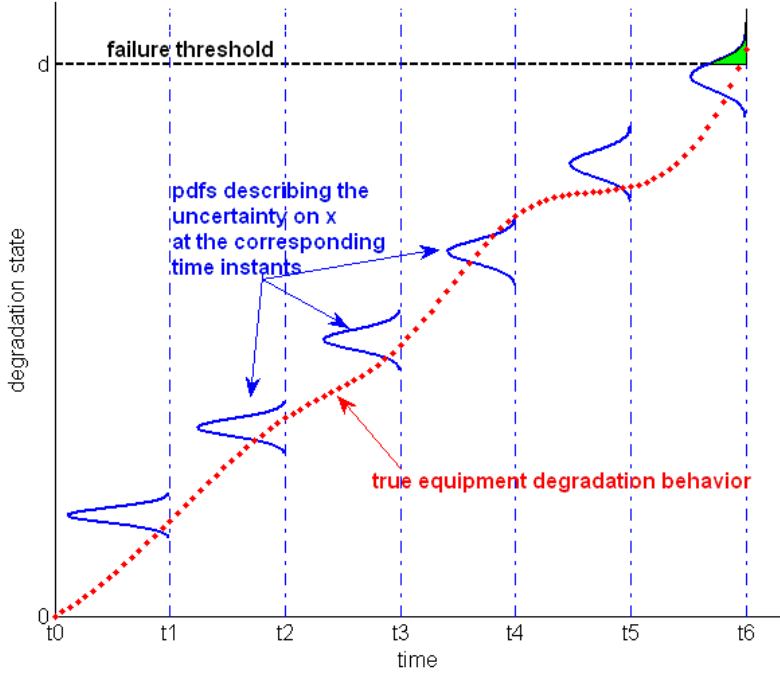
Fig. 2. Estimate of the failure probability.

The larger computational times needed to run the PF algorithm with larger values of N_s may be in contrast with the requirement of estimating the equipment RUL in time to anticipate failures.

RSPF offers a solution to this problem. The idea is to generate particles while giving due account to the risk associated to their positions in the sample space.

In the RSPF scheme proposed in [10], the risk associated to the approximate particle representation is embedded in the representation itself. In details, instead of the posterior distribution $p(x_k \mid z_1,...,z_k)$, the RSPF tracks

$$\gamma_k r(x_k) p(x_k \mid z_1,...,z_k) \tag{3}$$

where $r(x_k)$ is a positive, finite (almost everywhere) risk function, whereas γ_k is a normalization constant ensuring that $\gamma_k r(x_k) p(x_k \mid z_1,...,z_k)$ is a probability distribution. The pseudo-code of the RSPF algorithm proposed in [10] reads as follows.

1.   set $X_k = X_k^{aux} = \varnothing$

2. **for** $i = 1,...,N_s$

3.   pick the i-th sample $x_{k-1}^i \in X_{k-1}$;

4.   draw $x_k^i \approx p(x_k \mid x_{k-1}^i)$;

5.   set $w_k^i \approx p(z_k \mid x_k^i) \cdot r(x_k^i) / r(x_{k-1}^i)$;

6.   add x_k^i to X_k^{aux} with weigh w_k^i

7. **endfor**

8. **for** $i = 1,...,N_s$ %SIR algorithm

9.   draw x_k^i from X_k^{aux} with probability proportional to w_k^i;

10.   add x_k^i to X_k with $w_k^i = 1/N_s$

11. **endfor**

This pseudo-code differs from that of the PF-SIR algorithm given in Section 2 only at line 5.

From (3), it clearly appears that the probability that a particle x_k^i belongs to the set X_k depends not only on the posterior distribution, but also on the risk associated with that sample. In turn, the effect of the risk function is to force the algorithm to sample from the regions of the sample space with higher risk values.

From these considerations, it emerges that not all risk functions will be equally useful to successfully apply RSPF, and identifying a proper shape of $r(x_k)$ is a fundamental issue. In this work, we derive the expression of $r(x_k)$ on the basis of the following consideration: the risk associated with the generic weighted particle x_k^i is strictly related to two main aspects: the closeness of x_k^i to the threshold, and the uncertainty in its next *II* future steps. The larger the distance of x_k^i from the threshold d, the smaller the probability of achieving it within the next *II* steps, whereas the closer the particle to d, the smaller the probability of surviving *II* steps. These two factors influence the RUL estimation, and the corresponding maintenance decision, which comes from the following two rules.

     i)   If RUL $\leq$ *II*, then a preventive replacement of the equipment is done. The risk associated to this choice is that the equipment lifetime is not fully exploited, because it would be removed at least *II* steps before the optimal replacement time. Such preventive action is supposed to last D_P hours.

     ii)   If RUL $>$ *II*, then delay the decision to the next measurement acquisition. Thus, the risk associated with this choice is that the equipment fails before the next *II* steps. In this case, a corrective replacement is performed, which lasts D_C hours.

In this work, we assume that the risk associated with decision i) is negligible with respect to that relevant to ii). That is, the risk function considered accounts for the cost associated with the component failure only, whereas it neglects that of under-exploitation of the component preventively maintained before the optimal time. In this respect, the investigation of more complete risk functions (e.g., [18]) is an issue worthy of investigation in future works. Thus, the aim of the RSPF algorithm proposed in [10] is to generate samples from a distribution that factors-in the risk $r(x_k)$ of not removing the equipment before failure.

The shape of $r(x_k)$ is derived by resorting to the Monte Carlo method. Namely, the sample space $[0,d]$ is discretized into small bins. A large number M of particles are positioned in every bin, and their behaviors are simulated and traced for II steps. The portion of the M particles that cross the threshold provides an estimate of the probability for a particle located at x^i exceeding the threshold d. Such probability value $r(x^i)$ is indicative of the risk associated with the starting degradation state x^i; multiplying that probability value by the failure cost yields the expected cost associated with the decision of leaving the equipment working when its degradation state is x^i.

Finally, notice that accounting for the cost factor in $r(x_k)$ is useless, as the biased sampling is dependent on the ratio $r(x_k^i)/r(x_{k-1}^i)$.

The risk function $r(x_k)$ also enters the RSPF scheme proposed in [8], whose pseudo-code follows.

1.   set $X_k = X_k^{aux} = \varnothing$

2.   **for** $i = 1,...,N_s$

3.   pick the i-th sample $x_{k-1}^i \in X_{k-1}$;

4.   draw $x_k^i \approx \gamma_k r(x_k) p(x_k | x_{k-1}^i)$;

5.   set $w_k^i = w_{k-1}^i \cdot \dfrac{p(z_k | x_k^i) \cdot p(x_k^i | x_{k-1}^i, z_1,...,z_{k-1})}{q(x_k^i | x_{k-1}^i, z_1,...,z_{k-1})}$;

6.   add x_k^i to X_k^{aux} with weigh w_k^i

7.   **endfor**

From this code, it clearly emerges that the risk function enters the algorithm in a different way: it changes the distribution $p(x_k | x_{k-1}, z_1,...,z_{k-1})$ into the importance density function $q(x_k | x_{k-1}, z_1,...,z_{k-1}) = \gamma_k r(x_k) p(x_k | x_{k-1}, z_1,...,z_{k-1})$. This change is used in the prediction step to force the sampling of particles from the risky regions. The weights of the particles are then adjusted to give due account to the introduced bias: the larger the increment in the probability of picking the particle x^i brought by the importance density function $q(x_k | x_{k-1}, z_1,...,z_{k-1})$ with respect to $p(x_k | x_{k-1}, z_1,...,z_{k-1})$, the larger the reduction in its weight. To sum up, this risk sensitive filtering proposed in [8] adjusts the predicted distribution not only on the basis of the distance of the particles from the acquired measurement, but also depending on the extent to which sampling has been forced.

Finally notice that, at a first attempt, the SIS algorithm in support of the RSPF scheme proposed in [8] seems more suitable than the SIR algorithm. In fact, the resampling step (lines 8-11 of the pseudo-codes in Section 2 and Section 3) is expected to cause the biased drawing of the particles from the risky region to be fruitless, as these particles are strongly unlikely to be sampled, especially when the gathered measure is far from this region,

and the number of particles is small. Thus, the particles picked in the risky region cannot give a better estimation of the probability of failure. For the sake of brevity, this algorithm will be referred to as Forced- SIS.

4 Reference Example

4.1 Degradation and measurement models

In this section, the example concerning the fatigue degradation process of a mechanical component is presented. Several models have been developed in the literature to describe the evolution of this complex physical phenomenon (e.g., gamma processes [19], Weibull-based discrete processes [20], etc.). In this work, we adopt the randomized Paris–Erdogan fatigue crack growth model, which is a popular approach in the literature (e.g., [21]).

The justification of this choice is twofold: on one side, such a model has been shown to suitably approximate the degradation evolution of mechanical components subject to fatigue loads [21]; on the other side, it has been successfully applied in applications to prognostics (e.g., [7]).

The most widely used expression of the randomized Paris–Erdogan model is

$$x_k = x_{k-1} + e^{\xi_k} \cdot C \cdot (\beta\sqrt{x_{k-1}})^n \cdot \Delta t \tag{4}$$

where x_k is the crack depth at the (discretized) k^{th} time instant; β, C, and n are constant parameters, which can be estimated from measured crack growth data [22]; $\xi_k \approx N(0,\sigma_\xi)$, $k = 1,\ 2,...$, are statistically independent and identically distributed random variables, whereas Δt is a sufficiently small time interval (in our calculations, $\Delta t = 1$ hour).

The model (4) makes x_k, $k = 1,\ 2,...$, a Markov process with statistically independent increments, which are log-normally distributed [22]. The values of the model parameters, taken from [7], are reported in Table .

4.2 Structure of the maintenance policy

In this work, it is assumed that the crack depth can be observed only through inspections. For the sake of simplicity, these inspections are performed periodically, being non-periodic inspection strategies of difficult implementation in an industrial context. In general, the length *II* of the time interval between two successive inspections may range from $dt \to 0$ to ∞; the former case is generally referred to as continuous monitoring, while the latter corresponds to non-monitored components. In this work, *II* is set to 20 hours.

Generally speaking, the observation z_k of the crack depth that is acquired at the component inspection is just an estimate of its true value x_k because errors and imprecision always affect measurements. In this work, the

uncertainty about the observations is described by a white Gaussian noise $\upsilon \approx N(0, \sigma_\upsilon^2)$, which enters the physical law linking z_k to the depth x_k, leading to the following conditional pdf.

$$p(z_k \mid x_k) = \frac{1}{\sqrt{2\pi\sigma_\upsilon^2}} \cdot e^{-\frac{\left(\ln\left(\frac{z_k}{d-z_k}\right)-\mu_k\right)^2}{2\sigma_\upsilon^2}} \cdot \frac{d}{(d-z_k)z_k} \tag{5}$$

with

$$\mu_k = \beta_0 + \beta_1 \cdot \ln\left(\frac{x_k}{d-x_k}\right) \tag{6}$$

where β_0 and β_1 are parameters to be estimated from experimental data, and d is the component thickness.

The actions performed by the maintenance operators are generally divided into two groups, preventive, and corrective, depending on whether they are performed before, or after the component failure, respectively. In the present reference example, it is assumed that the mechanical component cannot be fixed, but just replaced. Therefore, only two types of actions are possible: preventive replacement, and corrective replacement.

- Preventive replacement is the activity under which the degraded component is substituted by a new one, before its failure. Such action is supposed to last 100 hours.
- Corrective replacement is the activity performed after the component failure. Due to the fact that the failure event is unscheduled, this action brings an additional duration of 100 hours, leading to a total duration of 200 hours. In particular, the additional time may be caused by the supplementary time needed for performing the procedure of replacement after failure, or to the time elapsed between the occurrence of the failure and the start of the replacement actions, due to various logistics and operational delays.

With regards to the costs of the maintenance actions, they are assumed to be proportional to the unavailability of the component. This situation is typical in plants where main maintenance costs are related to business interruption (e.g., energy production plants, [23], [24]). In this setting, the identification of the best maintenance policy, which is in general a multi-objective optimization problem, comes down to the issue of identifying the policy which minimizes the value of the unavailability.

The values of the variables of both the measurement model (taken from [7]) and maintenance model are listed in Table I.

Finally, notice that the reference mission time T does not end with the failure of the component; it is set to T=10,000 hours, which is almost 10 times the expected component life duration [5].

Table I

Parameters of the reference example

Paris Erdogan Model		Measurement Model		Maintenance Model	
Parameters	Values	Parameters	Values	Parameters	Values
C	0.005	β_0	0.06	II	20h
β	1	β_1	1.25	D_P	100h
n	1.3	σ_υ	0.47	D_C	200h
Δt	1	d	100 mm	C_P	$\propto D_P$
σ_ξ	1.7			C_C	$\propto D_C$

4.3 Application of RSPF

In this section, the PrM policy described in Section 2 is applied to the case study at hand, when the risk function $r(x_k)$ is derived according to the procedure proposed in Section 3 (Fig. 3). Notice that, in the first part of the sample space ([0, 100] mm), the estimated probability that the crack mechanism leads the component to failure within *II*=20 hours is zero; this is in contrast with Thrun et al.'s algorithm hypothesis of $r(x_k)$ being strictly positive. To circumvent this issue, the value of $r(x_k)$ is set to a constant, small value up to the crack length value where a significant increasing trend of $r(x_k)$ can be clearly distinguished; by so doing, the risk function does not operate any forcing of the posterior distribution in the first part of the interval [0, 100] mm (see the pseudo-code at Section 3).
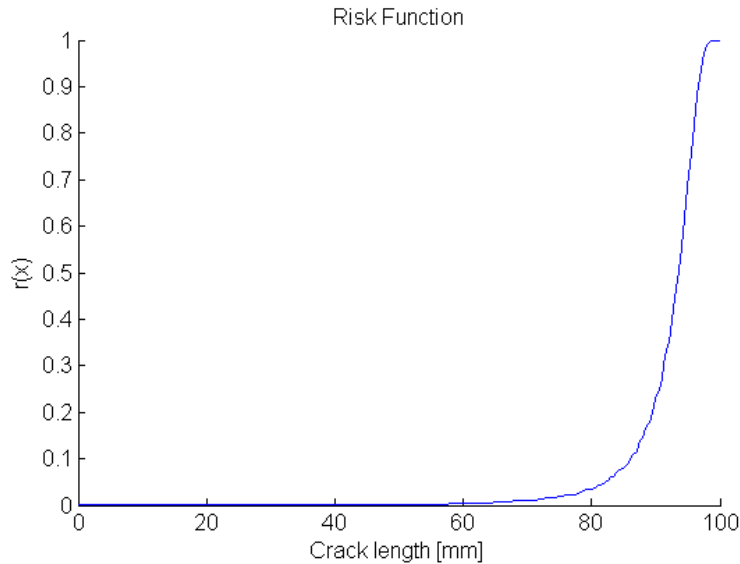

Fig. 3. risk function relevant to the case of the crack growth mechanism.

The effect of factoring the risk related to the particles' positions into the distribution describing the uncertainty in the degradation state can be appreciated by comparing Fig. 4 and Fig. 5. These figures show, respectively, how the classical PF-SIR, and the RSPF algorithms proposed in [10] perform the filtering step at the same time instant (i.e., 621 hours) of the same stochastic trial of the degradation process. In both cases, $N_s = 10$ particles approximate the pdf of the degradation state, which needs to be adjusted on the basis of the acquired measurement value $z_{k=621} = 57.33$ mm, corresponding to the actual crack length of $x_{k=621} = 52.02$ mm.

In these figures, the probability values $p(z_{k=621} | x)$ of acquiring measure $z_{k=621}$ when the degradation state is x are also reported in correspondence of all possible values of x. This curve drives the updating step in the classical SIR scheme: the closer the particle to the measured value, the better its matching with the current degradation evolution, so the larger the particle's chance of being re-sampled, and thus the larger the particle's chance of belonging to the posterior distribution representing the uncertainty in the degradation state (Fig. 4, bottom graph).

When the same situation is handled by the RSPF algorithm proposed in [10] (Fig. 5), then the risk function tends to increase the weights of the particles located at larger values of the sample space. In particular, the larger the increment of the ratio $r(x_k^i) / r(x_{k-1}^i)$ (i.e., the gain in risk of the particle with respect to its value at the previous step, see the pseudo-code at Section 3), the larger the factor multiplying the weights coming from the probability values $p(z_{k=621} | x)$. In turn, the particles sampled from the right part of the crack length axis will have larger weights, and thus will be favored in the re-sampling step (Fig. 5, bottom graph).
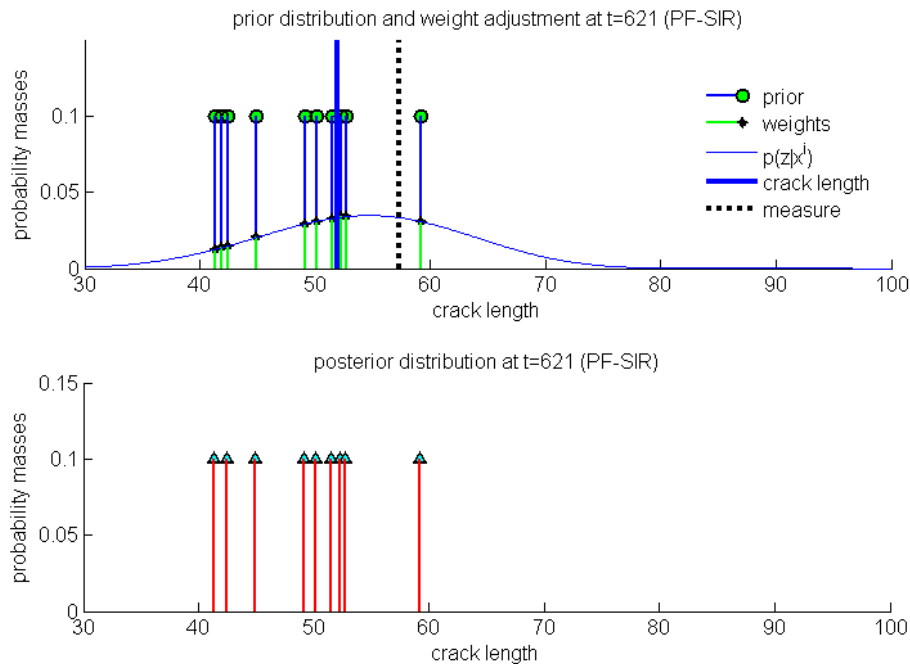

Fig. 4. Example of the filtering step within the SIR PF scheme.

In this situation, the most right particles (i.e., those with the largest crack) traced by the RSPF algorithm reach the value d before the corresponding ones traced by PF-SIR. Then, the estimation of the time instant t_{PT} at which the failure probability reaches a pre-fixed threshold value P_T provided by Thrun et al.'s RSPF algorithm tends to be smaller than that provided by the classical PF.

Finally, the stochastic trial considered above (Fig. 4 and Fig. 5) is given as input to the Forced-SIS algorithm (Fig. 6), in the setting in which the risk function $r(x_k)$ entering the importance density distribution $q(x_k \mid x_{k-1}, z_1, ..., z_{k-1}) = \gamma_k r(x_k) p(x_k \mid x_{k-1}, z_1, ..., z_{k-1})$ is the same as that considered in the RSPF scheme proposed in [10]. This method allows a fair comparison of the results of the algorithms.

The analytical expression of the distribution $q(x_k \mid x_{k-1}, z_1, ..., z_{k-1})$ is

$$q(x_k \mid x_{k-1}, z_1, ..., z_{k-1}) = \frac{r(x_k) p(x_k \mid x_{k-1}, z_1, ..., z_{k-1})}{\int_0^d r(x_k) p(x_k \mid x_{k-1}, z_1, ..., z_{k-1}) dx_k} =$$

$$= \begin{cases} \dfrac{r(x_k) \cdot \dfrac{1}{\sqrt{2\pi\sigma_\xi^2}} \cdot (x_k - x_{k-1}) \cdot e^{-\frac{1}{2\sigma_\xi^2}\left(\log\frac{x_k - x_{k-1}}{C \cdot (\beta \cdot \sqrt{x_{k-1}})^n}\right)^2}}{\displaystyle\int_{x_{k-1}}^d r(x_k) \cdot \dfrac{1}{\sqrt{2\pi\sigma_\xi^2}} \cdot (x_k - x_{k-1}) \cdot e^{-\frac{1}{2\sigma_\xi^2}\left(\log\frac{x_k - x_{k-1}}{C \cdot (\beta \cdot \sqrt{x_{k-1}})^n}\right)^2} dx_k} & x_{k-1} \le x_k \le d \\ \\ 0 & otherwise \end{cases}$$

Notice that $r(x_k)$ is not analytically known; then, determining the product $r(x_k) p(x_k \mid x_{k-1}, z_1, ..., z_{k-1})$ at the numerator as well as its integral at the denominator requires a numerical approach. We have to compute, for any particle x^i, the point-wise product of the distribution $p(x_k \mid x_{k-1}^i)$, the risk function $r(x_k)$, and the area below this curve (i.e., the denominator). Once the shape of $q(x_k \mid x_{k-1}^i, z_1, ...z_{k-1})$ has been obtained, it is numerically integrated to get the corresponding Cumulative Distribution Function (CDF) $Q(x_k \mid x_{k-1}^i, z_1, ...z_{k-1})$, and the particles are sampled from this distribution by means of the inverse transform method [25]. That is, a random number R is sampled from a uniform distribution in the right-open interval $[0,1[$, and the point $Q^{-1}(R \mid x_{k-1}^i, z_1, ...z_{k-1})$ is numerically found (Fig. 6). This approach constitutes a sample from $q(x_k \mid x_{k-1}^i, z_1, ...z_{k-1})$.

For example, Fig. 6 compares the CDF used by both Thrun et al.'s RSPF and the classical SIR-PF algorithms to sample the next position of particle x^i given that its current position is $x_{620}^i = 59$ mm, with the CDF used by the algorithm in [8] to sample the particles in the prediction step. From Fig. 6, it appears that the effect of changing

the sampling distribution is to force the drawing of the particles from the more risky region of the crack length axis.

According to the PF algorithm [16], the weights of the sampled particles need to be adjusted as

$$w_k^i = w_{k-1}^i \cdot \frac{p(z_k \mid x_k^i) \cdot p(x_k^i \mid x_{k-1}^i, z_1, ..., z_{k-1})}{q(x_k^i \mid x_{k-1}^i, z_1, ..., z_{k-1})} = w_{k-1}^i \cdot \frac{p(z_k \mid x_k^i) \cdot p(x_k^i \mid x_{k-1}^i, z_1, ..., z_{k-1})}{\gamma_k r(x_k^i) p(x_k^i \mid x_{k-1}^i, z_1, ..., z_{k-1})} = w_{k-1}^i \cdot \frac{p(z_k \mid x_k^i)}{\gamma_k r(x_k^i)} \quad (8)$$

From (8), we can understand why the particles favored by the forced sampling undergo a strong reduction in their weights. In fact, from Fig. 6, it clearly appears that sampling from $q(x_k \mid x_{k-1}^i, z_1, ..., z_{k-1})$ provides points very close to those which would have been sampled from $p(x_k \mid x_{k-1}, z_1, ..., z_{k-1})$, for a large range of values of *R*. This result is due to the shape of the risk function. On the contrary, if *R* is close to 1, the bias introduced is very strong: drawn particles are positioned considerably ahead on the crack length axis ([0, 100]mm). These particles will have a very poor match with the collected measure, and consequently a very small value in the numerator of (8). The strong reduction in the particle weight is only partially counterbalanced by the denominator, which is certainly smaller than 1, being the product of two numbers both smaller than 1.
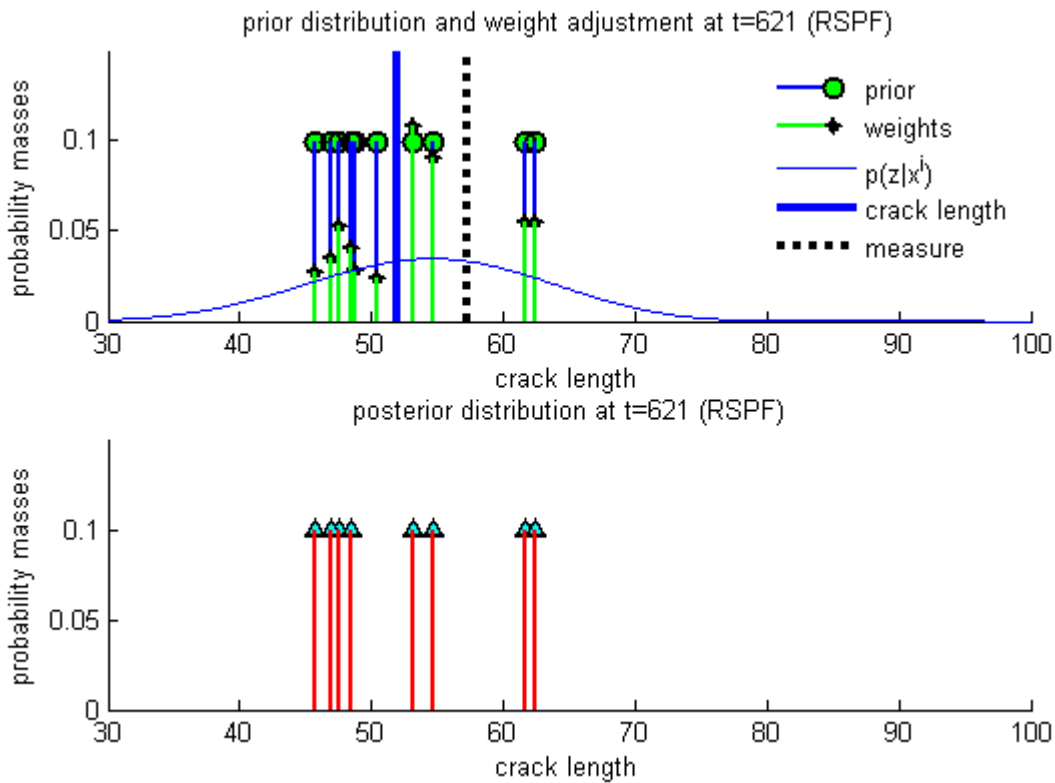

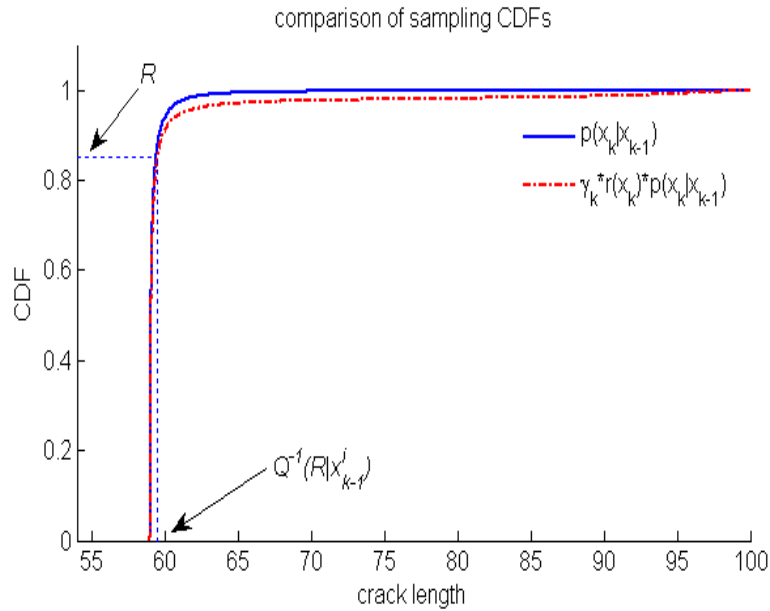Fig. 5. Example of the filtering step within the RSPF scheme.

Fig. 6. Comparison of the sampling CDF considered within the RSPF and Classical PF algorithms, and that of the scheme proposed in [8]

Fig. 7 shows the effect of the filtering step on the estimate of the posterior distribution in the same situation presented in Fig. 4 and Fig. 5. The sample impoverishment is evident: none but one particle has a weight considerably different from 0. This phenomenon is as strong as the number of particles is small [26]. In addition, the weight adjustment step tends to strongly reduce the weights of the particles forcibly sampled from the right part of the crack length space, which have poor chances of being the particles with significant weights, as explained above. This result leads to the fact that, in general, the very few (even one) important particles are expected to occupy positions not far from the measurement value. Consequently, the algorithm in [8] tends to heavily overestimate the component RUL.
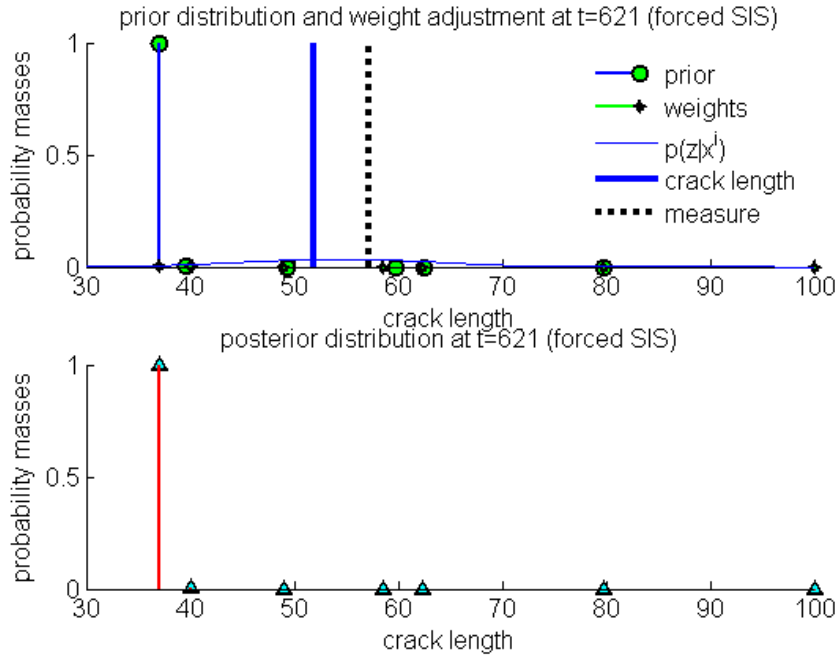
Fig. 7. Example of the filtering step within the SIS scheme proposed in [8].

Fig. 8 summarizes the results of the study. With regards to the performance of Thrun's RSPF algorithm when P_T =0.1 (continuous line), it can be noted that larger values of N_s correspond to smaller values of the component unavailability. This result is due to the more precise approximation of the biased distribution in (3), which allows us to effectively avoid component failures (Fig. 8 (a) and (b)). Such effectiveness is counterbalanced by the poor exploitation of the component life (i.e., the mean crack length of the preventively replaced components decreases (Fig. 8 (c)). For comparison, the performance of the classical SIR-PF algorithm are also reported in Fig. 8 (dotted lines), which are evidently worse than those of the RSPF proposed in [10]. The RUL estimates provided by the SIR algorithm are generally smaller than those provided by Thun et al.'s RSPF algorithm. This difference is that PF-SIR tends to leave the component working longer than Thrun et al.'s algorithm does. Consequently, the number of failures occurring within the mission time under the PF-SIR prognostics setting is on average larger than that of the RSPF scheme proposed in [10]; then, the unavailability is conservatively larger.

Fig. 8 shows also the results relevant to the application of the classical PF-SIR when P_T =0.05. The maintenance performance in this setting is similar to that of the classical PF when N_s =10. This result confirms that, when the P_T value is small, PF-SIR tends to overestimate the RUL, unless a proper number of particles is tracked. Thus, reducing the value of P_T is not a feasible way to reduce the number of failures when a small number of particles is handled by PF.
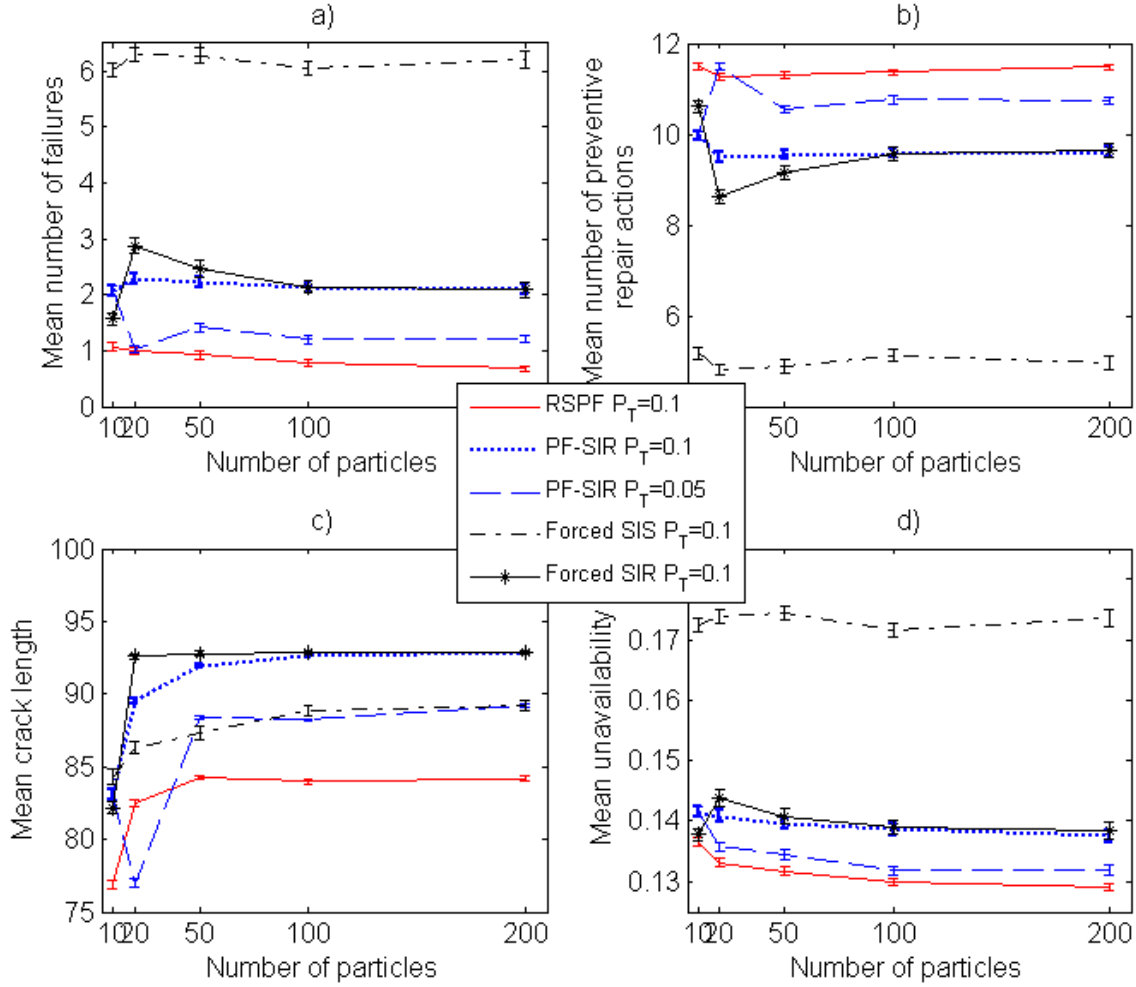
Fig. 8. comparative results of the study.

Finally, the results of the algorithm in [8] are also shown in Fig. 8. It clearly emerges that such an algorithm tends to overestimate the RUL. This means that failures occur before the estimated time, withthe consequent significant worsening of the capability of the PrM approach in avoiding failures, and major increment in the component mean unavailability over the mission time. In particular, the increase in the number of particles traced by this algorithm does not lead to any significant improvement. This result stems from the sample impoverishment phenomenon, which means that the first particles reaching threshold d have, on the whole, a probability mass smaller than P_T. Thus, the estimate of the RUL will coincide with the difference between the time instant at which the first important particle reaches d and the current time instant. In fact, the particles with the largest weights are generally positioned not far from the actual crack length value.

A possible solution to circumvent such a deterioration of the maintenance performance is to apply the re-sampling technique also to the RSPF algorithm proposed in [8], that is, to add lines 8-11 of the pseudo-code given in Section 2 also to the pseudo-code of the algorithm proposed in [8] (Section 3). This algorithm will be referred to as Forced-SIR, to distinguish it from the Forced-SIS. To appreciate the different algorithm dynamics

introduced by re-sampling, we can refer to the same situation investigated above (i.e., time instant $t=621$ hours of the same stochastic trial of the degradation process, with $N_s=10$, acquired measurement value $z_{k=621}=57.33$ mm, and actual crack length $x_{k=621}=52.02$ mm).

Fig. 9 (top) shows the distribution of the N_s particles at time $t=620$ hours. Sampling from $q(x_k \mid x_{k-1}^i, z_1, ...z_{k-1})$ yields the particle distribution sketched in Fig. 9 (middle). Here, it clearly appears that there has been no forced sampling at $t=621$ hours (i.e., none of the drawn N_s values of R has fallen close to 1). Comparing the situation depicted in Fig. 9 with that in Fig. 7, one notes that the fundamental difference is that particles close to the threshold d are no longer present in Fig. 9. To explain this result, one has to note that, in the scheme considered in Fig. 7, the particles positioned far ahead of the measurement value (or even in d) have undergone a biased sampling at a time instant before $t=621$ hours. Once sampled, their weights are set to values very close to 0, according to (8), and kept up to $t=621$h. In the situation depicted in Fig. 9, particles that have been strongly favored in any of the sampling step before $t=621$ hours have been lost in one of the re-sampling steps (performed every II hours) preceding $t=621$ hours. In turn, in the first part of the sample space, the RSPF scheme proposed in [8] associated with the SIR approach is expected to provide estimates of component RUL not different from those provided by the classical PF-SIR algorithm. When the crack length is deeper, the biased sampling starts providing particle positions which are not incompatible with the measurement values, and thus do not undergo a strong reduction in their weights (8).
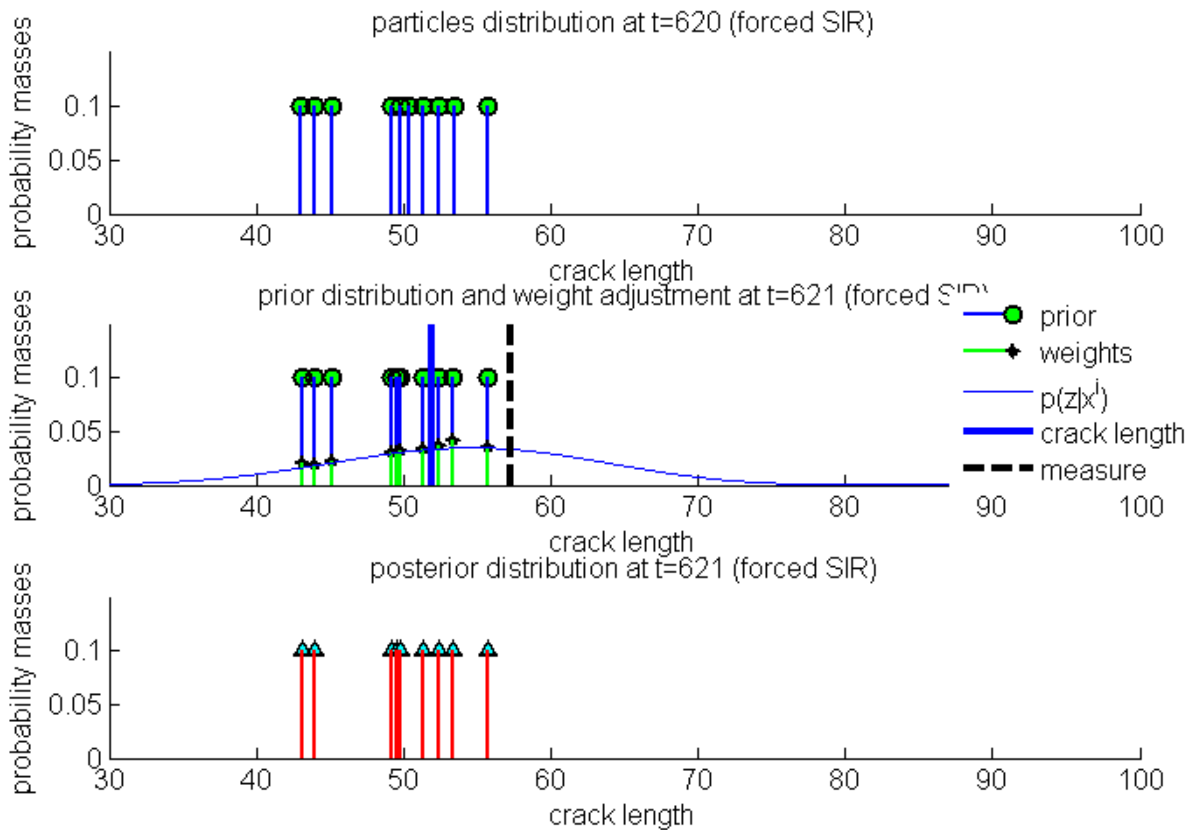
Fig. 9. Example of the filtering step performed by the algorithm proposed in [8], with the SIR scheme.

For example, Fig. 10 shows the particle distribution 60 hours later than that considered in Fig. 9. Notice that forced sampling has yielded 4 particles far ahead of the others, with a total probability mass not negligible; that is, the chance of re-sampling at least one out of these four is not insignificant. In particular, one out of the four most advanced particles assumes a position compatible with the current measure, and is re-sampled in the SIR step. This situation allows the algorithm to underestimate the RUL (i.e., it provides smaller RUL values) in the final part of the component life.
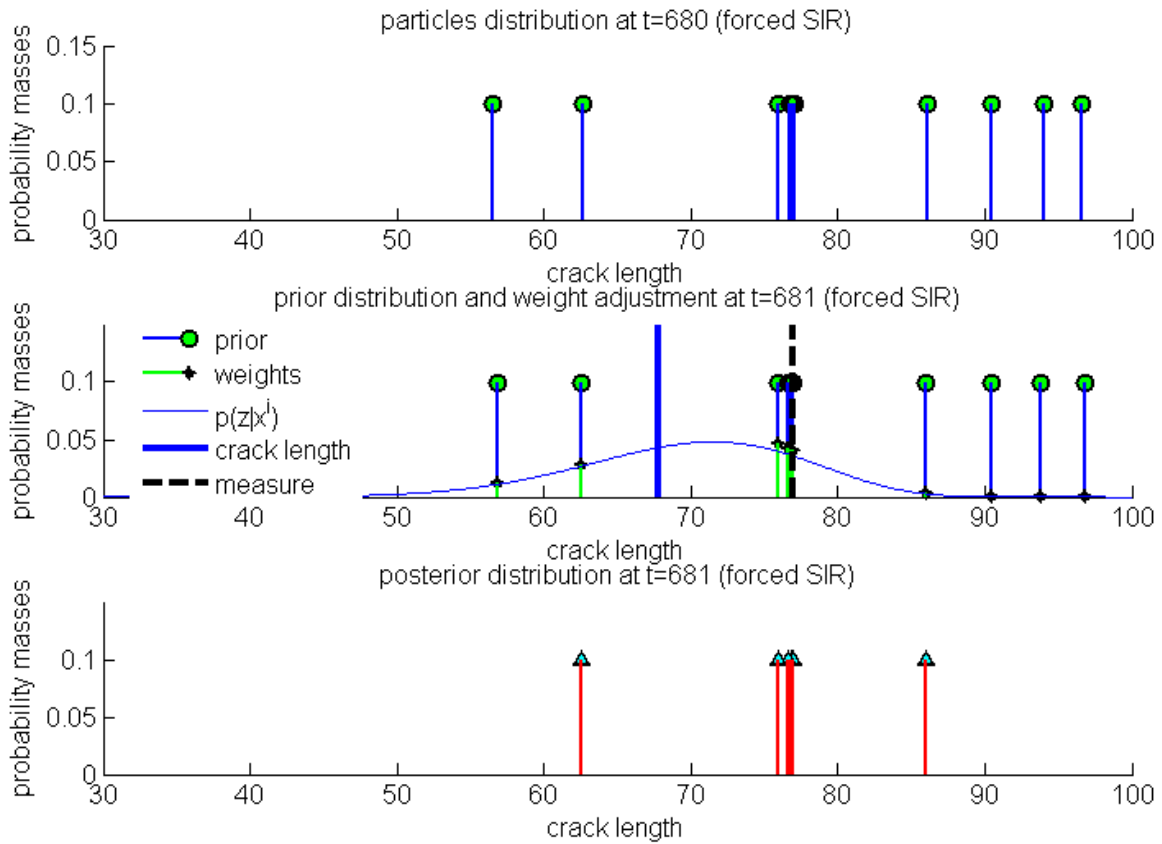
Fig. 10. Example of the filtering step performed by the algorithm proposed in [8], with the SIR scheme.

The performance results of the Forced-SIR algorithm are reported in Fig. 8. These results are similar to those of the classical PF-SIR when the number of particles is large. This comparison is expected; the biased sampling is useless when the number of samples is large, because in this case PF is capable of accurately predicting the probability mass of the upper tail of the distribution (Fig. 2).

On the contrary, the performance of the Forced-SIR algorithm is similar to that of the RSPF scheme proposed in [10] with $N_s = 10$ particles. This result is due to the fact that both algorithms tend to underestimate the component RUL, which on one side leads to a reduction of the mean number of failures in the time horizon of interest. On the other side, RUL underestimation reduces the operating life of the component, which leads to a poor exploitation of the components.

Finally, from Fig. 8, see that the performance of the Forced SIR algorithm suffers from instability under small values of N_s. This condition stems from the lack of robustness of the RUL estimates provided by the algorithm; that is, underestimating the RUL requires that at least two particles are sampled with bias, and then re-sampled after the weight adjustment. This scenario is associated with a relatively small occurrence probability.

5 Conclusions

In this work, two different prognostics schemes in support of PrM have been considered, which are based on two different schemes of the RSPF algorithm: one proposed in [10], and the other one in [8]. The former has never been considered for failure prognostics or PrM.

By way of an example related to a component subject to fatigue degradation, we show that the first scheme allows considering a number of particles smaller than that required by the classical PF, leading to better maintenance performance. On the contrary, the algorithm proposed in [8] suffers from the tendency of overestimating the RUL, and leads to maintenance performance considerably worse than those of the algorithm in [10]. An improvement of the algorithm has been proposed, by the introduction of a re-sampling step. Still, the overall setting suffers from a lack of robustness, and provides RUL estimates that are very sensitive to the number of traced particles.

It is reasonable to conclude that the RSPF scheme proposed in [10] is more suitable for prognostics applications with a small number of particles.

In the example considered, the risk function considered accounts for the cost associated with the component failure only, whereas it neglects that of under-exploitation of the component preventively maintained before the optimal time. In this respect, the investigation of more complete risk functions is an issue worth investigating.

Moreover, the performance of the PrM policy has been assessed in the nominal setting only. A further analysis needs to be carried out to understand how and to what extent the performance values are sensitive to the model parameters (e.g., II, the ratio between the cost of the preventive and corrective actions, etc.).

6 Acknowledgements

7 References

[1]. A. Grall, L. Dieulle, C. Bérenguer, and M. Roussignol, "Continuous-Time Predictive-Maintenance Scheduling for a Deteriorating System," *IEEE Transactions On Reliability*, Vol. 51, No. 2, pp.141-150, 2002.

[2]. S. Lu, Y.C. Tu, and H. Lu, "Predictive Condition-based Maintenance for Continuously Deteriorating Systems," *Quality and Reliability Engineering International*, Vol. 23, pp. 71-81, 2007.

[3]. M.Y. You, F. Liu, W. Wang, and G. Meng, "Statistically planned and Individually Improved Predictive Maintenance Management for Continuously Monitored Degrading Systems," *IEEE Transactions on Reliability*, Vol. 59, No. 4, pp. 744-753, 2010.

[4]. H. Fan, C. Hu, M. Chen, and D. Zhou, "Cooperative Predictive Maintenance of Repairable Systems With Dependent Failure Modes and Resource Constraint," *IEEE Transactions on Reliability*, Vol. 60, No. 1, pp. 144-157, 2011.

[5]. E. Zio, and M. Compare, "Evaluating maintenance policies by quantitative modeling and analysis," *Reliability Engineering and System Safety*, Vol. 109, pp. 53-65, 2013.

[6]. E. Zio, "Prognostics and health management of industrial equipment," *Diagnostics and Prognostics of Engineering Systems: methods and techniques*. IGI-Global, pp. 333-356, 2013.

[7]. F. Cadini, E. Zio, and D. Avram, "Model-based Monte Carlo state estimation for condition-based component replacement," *Reliability Engineering and System Safety*, Vol. 94, No. 3, pp. 752-758, 2009.

[8]. M.E. Orchard, L. Tang, K. Goebel, and G. Vachtsevanos, "A novel RSPF approach to prediction of High-Risk, Low-Probability failure events," *Proceedings of the Annual conference of the Prognostics and Health Management Society 2009*, San Diego, CA, 2009

[9]. E. Zio, and F. Di Maio, "Fatigue crack growth estimation by relevance vector machine," *Expert Systems with Applications*, Vol. 39, pp. 10681–10692, 2012.

[10]. S. Thrun, J. Langford, and V. Verma, "Risk Sensitive Particle Filters," *Neural Information Processing Systems (NIPS) 14*, MIT Press, pp. 961-968, Dec. 2001

[11]. V. Verma, G. Gordon, R. Simmons, and S. Thrun, "Real time fault diagnosis," *IEEE Robotics & Automation Megazine,* pp. 56-66, June 2004.

[12]. R.K. Boel, and M.R. James, "Robustness and Risk-sensitive Filtering," *IEEE Transactions on Automatic Control*, Vol. 47, No. 3, pp. 451-461, 2002.

[13]. F. Gustaffson, and S. Saha, "Particle filtering with dependent noise," *Proceedings of the 13th Conference on Information Fusion (FUSION),* Edinburgh, U.K., 2010.

[14]. M.E. Orchard, "A Particle Filtering-based Framework for On-line Fault Diagnosis and Failure Prognosis," PhD dissertation thesis at the Department:of Electrical and Computer Engineering, Georgia Institute of Technology, 2007.

[15]. G. Vachtsevanos, F.L. Lewis, M. Roemer, A. Hess, and B. Wu, "Intelligent Fault Diagnosis and Prognosis for Engineering Systems," John Wiley & Sons, 2006.

[16]. M.S. Arulampalam, S. Maskell, N. Gordon, and T, Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, Vol. 50, No. 2, pp. 174-188, 2002.

[17]. Doucet, J.F.G. de Freitas, and N.J. Gordon, "Sequential Monte Carlo methods in practice," Springer-Verlag, New York, 2001.

[18]. G. Haddad, P.A. Sandborn, M. Pecht, "An Options Approach for Decision Support of Systems With Prognostic Capabilities," *IEEE Transactions on Reliability*, Vol. 61, No. 4, pp. 872-883, 2012.

[19]. J.M. van Noortwijk, "A survey of the application of gamma processes in maintenance," *Reliability Engineering and System Safety*, Vol. 94, pp. 2-21, 2009.

[20]. Wormsen, and G. Härkegård, "A statistical investigation of fatigue behaviour according to Weibull's weakest link theory," *Proceedings of the 15th European Conference on Fracture*, Stockholm, Sweden, 2004.

[21].   F. Kozin, and J.L. Bogdanoff, "Probabilistic models of fatigue crack growth: results and speculations," *Nuclear Engineering and Design*, Vol. 115, pp. 143–171, 1989.

[22].   E. Myötyri, U. Pulkkinen, and K. Simola, "Application of stochastic filtering for lifetime prediction," *Reliability Engineering and System Safety*, Vol. 91, No. 2, pp. 200-208, 2006.

[23].   Amato, M. Compare, M. Gallisto, A. Maccari, M. Paganelli, and E. Zio, "Business interruption and loss of assets risk assessment in support of the design of an innovative concentrating solar power plant," *Renewable Energy*, Vol. 36, No. 5, pp. 1558-1568, 2011.

[24].   K.S. Park, "Condition-Based Predictive Maintenance by Multiple Logistic Function," *IEEE Transactions on Reliability*, Vol. 42, No. 4, pp. 556-560, 1993.

[25].   E. Zio, "The Monte Carlo Simulation Method for System Reliability and Risk Analysis". Springer series in Reliability Engineering, Springer-Verlag, London, 2013.

[26].   N.M. Kwok, G. Fang, and W. Zhou, "Evolutionary Particle Filter: Re-sampling from the Genetic Algorithm Perspective," *IEEE/RSK International Conference on Intelligent Robots and Systems 2005*, Edmonton, Canada, pp. 2935-2940, 2005.

Biographies

Michele Compare (BS in mechanical engng., University of Naples Federico II, 2003, PhD in nuclear engng., Politecnico di Milano, 2011) is currently a post-doc at the Politecnico di Milano. He worked as RAMS engineer, and risk manager. His main research efforts are devoted to the development of methods and techniques in support of the maintenance of complex systems.

Enrico Zio (BS in nuclear engng., Politecnico di Milano, 1991; MSc in mechanical engng., UCLA, 1995; PhD, in nuclear eng'g., Politecnico di Milano, 1995; PhD, in nuclear engng., MIT, 1998) is Director of the Chair in Complex Systems, and 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 and 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 an author or co-author of five international books, and more than 200 papers in international journals.