Sequential Monte-Carlo sampling based on a committee of artificial neural networks for posterior state estimation and residual lifetime prediction

C. Sbarufatti *, M. Corbetta, A. Manes, M. Giglio

Politecnico di Milano, Dipartimento di Meccanica, via La Masa 1, 20156 Milan, Italy

The application of Bayesian methods to the problem of fatigue crack growth prediction has been growing in recent years. In particular, sequential Monte-Carlo sampling is often presented as an efficient model-based technique to filter the sequential measures of the damage evolution provided as an input to the algorithm. However, a lot of measures are required to reliably identify the system state condition and the underlying model parameters. Many studies rely on the availability of a relatively dense sequence of crack length measures during damage evolution, made in most cases impractical by the consequent maintenance costs. Thus, real-time damage diagnosis is a requirement to enable prognostic health management.

This work focuses on the application of sequential Monte-Carlo sampling to estimate the probabilistic residual life of a structural component subjected to fatigue crack propagation, while real-time estimation of crack length is provided through a committee of artificial neural networks, trained with finite element simulated strain patterns. Multiple crack length observations are available at each discrete time and are provided as the input to the prognostic system, based on a sequential importance resampling algorithm. Each time a new set of measures is available, the algorithm evaluates the posterior distribution of the augmented state vector, including the crack length and a material parameter governing damage evolution. This filtered information is used to numerically update the probability density functions of the residual life of the component. The methodology is applied first to a simulated crack and then to a metallic stiffened panel specimen subject to fatigue crack growth.

Keywords:
Sequential importance resampling
Artificial neural network
Committee
Augmented state vector
Fatigue crack growth

1. Introduction

Fatigue crack growth (FCG) is a primary degradation process in the field of metallic structure reliability, especially for aeronautical applications, and the development of accurate models to describe damage evolution is widely researched (Paris–Erdogan's law [1], Walker's law, Forman's law, NASGRO law [2], to mention a few [3]). The increase of safety requirements and the necessity of advanced maintenance strategies (e.g. the condition based mainte-nance approach) [4] give rise to the adoption of stochastic approaches to improve the reliability of the predictions with respect to design expectations [5]. As a matter of fact, a more precise estimate of the residual life at a specific time can be provided by the observation of the current condition of the component. This gives rise to the conditional residual life (RL) probability density function (PDF) estimate. At a specific point in time and for a particular observation of the condition of a component, a posterior PDF for the residual life (conditional on the observation) can be evaluated based on the availability of a prognostic model. This approach is particularly well described in the Bayesian updating framework [6,7], in which the a priori information on the RL is updated according to the actual observations taken by a diagnostic tool.

Recent advances on sequential Monte-Carlo methods, specifically Sequential Importance Sampling/Resampling (SIS–SIR) algorithms, allowed integrating multiple uncertainties related to the measurement system and the intrinsic randomness of the degradation phenomenon with the mathematical FCG formulations [7–9]. Their suitability for the prediction of evolving phenomena by sequentially updating of the system state estimation has already been demonstrated [8,9]. However, these algorithms require (i) a probabilistic model describing the system evolution, (ii) a stochastic measurement model relating the measure to the system state and (iii) a sufficient number of measures to guarantee the

* Corresponding author. Tel.: +39 022399 8213.
E-mail address: claudio.sbarufatti@polimi.it (C. Sbarufatti).

Published Journal Article available at: http://dx.doi.org/10.1016/j.ijfatigue.2015.05.017

convergence of the algorithm state estimation to the target process evolution [10]. The assumption that a sufficient number of measures are available from any non-destructive testing (NDT) measuring apparatus is widely made [9–11]. This assumption is however only applicable to cases in which maintenance costs are not a primary issue and a maximum exploitation to RL estimation can only be obtained when coupled to a real-time diagnostic system providing automatic crack length measures [12], thereby entering the field of structural health monitoring (SHM) [13]. The aim of this paper is to monitor the damage evolution and to make prognosis on a relatively complex metallic structure with a riveted skin–stringer construction, where diagnosis is made by using an SHM system based on a sensor network and a committee of artificial neural networks (ANNs). This work is an initial step towards the on-line condition monitoring of aeronautical structures, including a prognostic system able to filter multiple information from the SHM diagnostic unit.

Focusing on the diagnostic problem, the algorithm for damage identification requires extensive investigations. In [12], the authors described the optimisation procedure for a diagnostic algorithm based on ANNs, trained on finite element simulated strain patterns. This algorithm is able to generalise to the experimental measures well, providing damage detection, localisation and crack length quantification. A sufficient level of generalisation is especially important for the ANN when using simulation for algorithm training. In addition to some typical regularisation techniques such as cross validation and early stopping, the authors grouped multiple ANN models trained on *Bootstrap* datasets into one committee [14], obtaining further algorithm regularisation. In practice, each time a strain pattern is provided as the input to the diagnostic algo-rithm, the diagnostic output is not a single indication but a series of N_{ANN} outputs, being N_{ANN} the number of ANNs belonging to the committee. Such distribution is related to the uncertainty intrinsic to the diagnostic model training procedure [14]. Various methods studying how to combine the outputs from multiple models are present in the literature [15], the simplest way consists in averaging the prediction of the set of individual models [12,15]. This com-bined observation can be used as the input to the prognostic algo-rithm. However, by simply averaging the observations the information regarding the diagnostic model uncertainty related to a particular measure is lost. A thorough method to combine the multiple outputs from a diagnostic system into the prognostic SIR framework is needed and is the objective of the present study.

In this work, at each discrete time step, the dispersion related to the entire set of multiple observations is combined with the randomness intrinsic to the FCG process, usually described within a dynamic state space (DSS) [7], and a SIR algorithm is adopted to filter the total uncertainty. This filtered information is then used to numerically update the posterior distribution of the residual life. In order to filter also the uncertainty related to the FCG model parameters [16–18], a SIR algorithm with an augmented state vector [19–22] has been implemented. One FCG model parameter typically associated with material properties in linear elastic fracture mechanics is inserted into the state vector and its PDF is sequentially updated in real time. This additional complexity is necessary due to the very high FCG variability one can expect during repeated tests [10,16]. The methodology is applied first to the identification of a simulated crack growth process and then to a real metallic stiffened panel specimen subjected to fatigue crack growth. According to the recent literature, the validation of such a SHM tool on realistic structures (i.e. the portion of a helicopter fuselage) constitutes a novelty in the aeronautics panorama.

The paper is structured as follows: the theory of SIR algorithm is shortly introduced in Section 2, with focus on the extension for diagnostic output filtering as well as for the parameter estimation with an augmented state vector. The critical aspects regarding the SIR implementation for FCG prognosis are detailed in Section 3. The results of the simulated FCG are presented in Section 4. Section 5 is dedicated to the verification of the algorithm performance during one FCG test, showing the overall prognostic system performances. A critical discussion of the work and some possible future extensions of the method are provided in the conclusive section.

2. Theory and methods

Extensive literature is available concerning the mathematical aspects of sequential Monte-Carlo filters (two remarkable examples are [6,7]), therefore this section only summarises the primary aspects of the algorithm procedure and the input requirements. In particular, the basic formulation of system state filtering by SIR algorithm is reported in Section 2.1, followed by its extension for augmented state vector filtering in Section 2.2. The algorithm modification to receive multiple observations as the input is reported in Section 2.3.

2.1. Basics on sequential importance resampling strategy

The definition of a DSS [7], including the model evolution Eq. (1) (consisting of a hidden Markov process of order one) and the observation Eq. (2) (linking the measures with the system state) is considered.

$$x_k = f(x_{k-1}, \vartheta, \omega_{k-1}) \tag{1}$$

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

where x_k is the system state at k-th discrete time, ϑ is a vector collecting the model parameters supposed to be constant during the process evolution, ω_{k-1} is the artificial process noise and η_k is the uncertainty affecting the observation z_k. In a Bayesian updating framework, the objective is to calculate the posterior PDF of the system state at discrete time k, conditioned on the observations, indicated as $p(x_k|z_{1:k})$. This can be evaluated with the well-known prediction and updating steps, respectively performed through the Chapman–Kolmogorov equation and the Bayes' rule [7]. Nevertheless the analytical solution of this problem is only possible if the model is linear and each random process involved is Gaussian. The SIR algorithm is a recursive Bayesian filter, commonly used to approximate the state posterior distribution by a series of samples (often referred to as *particles*). At each k-th discrete time, N_s particles $x_k^{(i)}$ [$i = 1, \ldots, N_s$) approximate the posterior PDF of the system state, $p(x_k|z_{1:k})$. Each of these particles has a linked weight (3) depending on the weight at the previous time step $w_{k-1}^{(i)}$ and the likelihood of that particle given the measure. After a normalisation of the weights is performed in a way that $\sum_{i=1}^{N_S} \tilde{w}_k^{(i)} = 1$, the posterior PDF of the system state can be calculated (4). To limit sample degeneracy, particles are resampled according to the actual posterior distribution, as indicated in Eq. (5).

$$w_k^{(i)} = w_{k-1}^{(i)} p\left(z_k \middle| x_k^{(i)}\right) \tag{3}$$

$$p(x_k|z_{1:k}) \approx \sum_{i}^{N_s} \tilde{w}_k^{(i)} \delta\left(x_k - x_k^{(i)}\right) \tag{4}$$

$$x_k^{(i)} \sim p(x_k|z_{1:k}) \tag{5}$$

The procedure described through Eqs. 3–5 is repeated each time a new observation is provided by a measuring system.

Having applied the SIR algorithm to filter the state vector based on diagnostic observations, the prognosis consists in projecting the filtered particle population in time, up to the failure region that is identified for the specific use case (this may be a critical crack

length in the present study). Similarly to Eq. (4), the posterior PDF of the number of remaining load cycles $N_{r,k}$ can be estimated at discrete time k as reported in Eq. (6) [9].

$$p(N_{r,k}|z_{1:k}) \approx \sum_{i}^{N_s} \tilde{w}_k^{(i)} \delta\left(N_{r,k} - N_{r,k}^{(i)}\right) \tag{6}$$

where $N_{r,k}^{(i)}$ is the number of remaining load cycles associated to the i-th particle trajectory, calculated based on the evolution model f in (1).

2.2. State vector extension for parameter updating

As anticipated in Section 1 and discussed in detail in [10], the prognosis of degradation processes is strictly dependent on the identification of the model parameters for the damage evolution. The SIR algorithm offers a straightforward method to update the PDF of constant model parameters, based on the available sequence of measures. The state vector (x_k) is extended to the *aug-mented state vector* (y_k) which contains both the damage state and the model parameter variables $y_k = [x_k, \vartheta_k]$. The subscript k associ-ated to the model parameter vector ϑ indicates its estimation at a discrete time k and is not a time dependence.

The Chapman–Kolmogorov equation for the augmented state vector prediction and the Bayes' updating step can be expressed through the Eqs. (7) and (8) respectively [22].

$$p(y_k|z_{1:k-1}) = \int_{y_{k-1}} p(y_k|y_{k-1}) p(y_{k-1}|z_{1:k-1}) dy_{k-1} \tag{7a}$$

$$p(y_k|z_{1:k-1}) = \int_{x_{k-1}} \int_{\vartheta_{k-1}} p(x_k|x_{k-1}, \vartheta_{k-1}) p(\vartheta_k|x_{k-1}, \vartheta_{k-1}) p(x_{k-1}, \vartheta_{k-1}|z_{1:k-1}) d\vartheta_{k-1} dx_{k-1} \tag{7b}$$

$$p(y_k|z_{1:k}) \propto p(z_k|[x_k, \vartheta_k]) p(x_k|[x_{k-1}, \vartheta_{k-1}]) p(\vartheta_k|\vartheta_{k-1}) p([x_{k-1}, \vartheta_{k-1}]|z_{1:k-1}) \tag{8}$$

where $p(z_k|[x_k, \vartheta_k])$ is the likelihood of the state vector with respect to the last observation, $p(x_k|[x_{k-1}, \vartheta_{k-1}])$ is the proposal density for the state, $p(\vartheta_k|\vartheta_{k-1})$ is the proposal density for the model parameters (assumed to be independent from the state) and $p([x_{k-1}, \vartheta_{k-1}]|z_{1:k-1})$ is the prior state vector probability.

The sequential prediction–updating–resampling procedure described in Section 2.1 can be reformulated for the combined state-parameter PDF estimation; the weight updating Eq. (3) is modified as shown in Eq. (9).

$$w_k^{(i)} = w_{k-1}^{(i)} p\left(z_k \middle| y_k^{(i)}\right) = w_{k-1}^{(i)} p\left(z_k \middle| x_k^{(i)}, \vartheta_k^{(i)}\right) \tag{9}$$

Eq. (8) clearly shows that a proposal distribution from which samples of the parameter vector ϑ are drawn has to be considered in the SIR algorithm. While some more advanced techniques are available in the literature [19], the approach followed in this work is based on artificial dynamics [22]; the choice is mainly related to its simplicity, nevertheless demonstrating its suitability to the problem under investigation.

According to the theory of artificial dynamics, a random noise ξ is added to the particles to avoid the sample degeneracy, as in Eq. (10). The random noise has a zero mean and a variance that decreases with time [22].

$$\vartheta_k = \vartheta_{k-1} + \xi_{k-1} \tag{10}$$

Regardless of the simplicity of the method, it requires the selection of the initial variance of ξ ($\sigma_{\xi 0}^2$) and the decreasing rate of artificial dynamics variance $f_{\xi}(k)$, as discussed below in Section 3.2.

2.3. Integration of multiple observations

It is widely stated in the literature [14] that the diagnostic performance can be improved by a combination of multiple diagnostic models instead of using a single model in isolation, including the adoption of committees or ensembles. In this paper, the aim is to include a vector of N_{ANN} state observations provided in real-time by N_{ANN} models for the degradation measure (organised in a committee) into the SIR algorithm. Each ANN receives a pattern of real-time strain measures as the input and provides an indication of the crack length as the output. The main algorithm modification to combine the SIR algorithm with multiple outputs from the ANN committee is related to the calculation of particle likelihood $p(z_k|[x_k, \vartheta_k])$ and it has been described in detail in this section.

Likelihood calculations require a probabilistic measurement model which links the measure z with the system degradation state x, as expressed through Eq. (2). Such a model is assumed to be available in most of the studies present in the literature; when a classical NDT is used for damage diagnosis, it can be retrieved by means of dedicated test programs that act to identify sensor calibration curves and their associated uncertainties. A remarkable novel approach has been presented in [23], where the authors use a committee of ANNs to estimate this model based on a series of pairs made by the real state x and the corresponding measurement z; however, a single observation is still assumed to be available at each discrete time from a traditional NDT system.

Moving to multiple observations, suppose N_{ANN} observations of a generic damage feature are available at kth discrete time $z_k^{(j)}$, $j = 1, \ldots, N_{ANN}$ from a committee of N_{ANN} diagnostic ANN models and are organised in the vector z_k. Each ANN acts as a measurement system and is characterised by an error. Two main sources of uncertainty are related to the committee measurement.

(1) The *committee dispersion* (Y_d) is an index of diagnostic precision and is intended here as a non-Gaussian zero-mean PDF, based on the N_{ANN} outputs that are available at each discrete time k, upon removal of their mean value. In the most general case, it is a function of the degradation state, thus $Y_d = Y_d(x)$.
(2) The *committee bias* (Y_b) is intended as the error between the averaged committee output ($\bar{z}_k$) and the target. This is an index of the diagnostic accuracy and has been modelled here as a stochastic Gaussian variable with expectation $g(x)$ and variance $\sigma_b^2(x)$, also described as a function of the degradation state, $Y_b(x) \sim N(g(x), \sigma_b^2(x))$ [23].

Both stochastic variables are combined in the probabilistic measurement model, as in Eq. (11).

$$z_k = x_k + \upsilon_b(x_k) + \upsilon_d(x_k) \tag{11}$$

$\upsilon_b(x_k)$ is a sample of the stochastic variable Y_b and $\upsilon_d(x_k)$ is a vector containing the N_{ANN} samples $\upsilon_d^{(j)}(x_k)$ of the stochastic variable Y_d. A schematic representation of the committee dispersion and bias has been presented in Fig. 1a.

The motivations that lead to the inclusion of both stochastic variables in the measurement model are discussed hereafter, focussing on the related uncertainties.

In this study the committee dispersion is related to the randomness intrinsic to the ANN training. To guarantee a sufficient level of regularisation for the ANN, the database is divided into at least two sets, one for training and one for validation, further *cross-validation* with *early stopping* can be used to avoid over-fitting [14,15]. However, it is difficult to completely avoid over-fitting and different optimal synapse weights are found each time a new training is performed by randomly dividing the input space into two

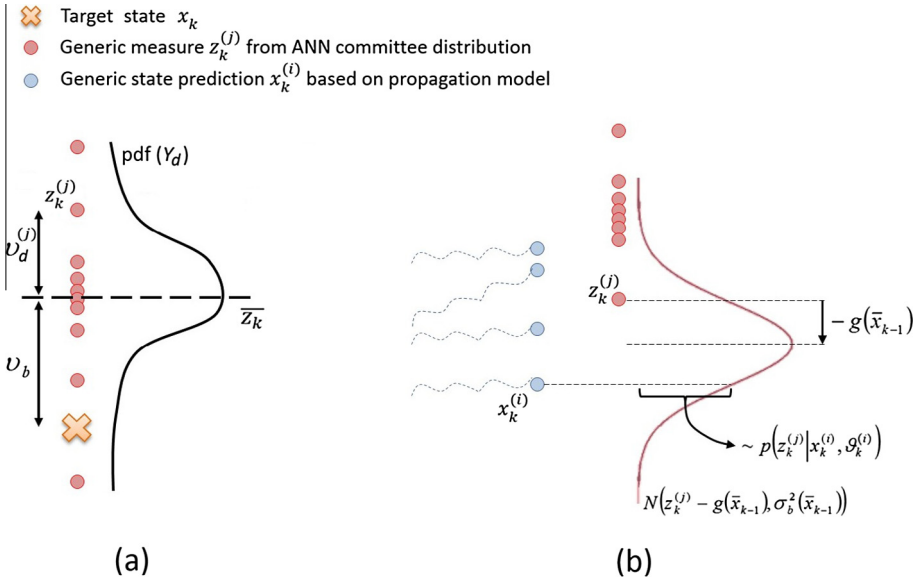

Fig. 1. Schematic representation of (a) committee bias and committee dispersion and (b) the likelihood evaluation.

domains: the training and the validation domain [15]. Thus, it is difficult to define a priori which ANN model has the best diagnostic performance in the real application. The prediction performance of the diagnostic model is increased by grouping multiple models into one committee. While in the most general approach, there are no restrictions for the type and structure of the diagnostic models that are combined together, in this study, different ANN models with the same structure but trained on a different dataset (*Bootstrap* datasets [14]) have been grouped into one committee to provide multiple observation outputs at each discrete time k. Each observation from a different model is considered as a sample of the stochastic variable Y_d. Thus, N_{ANN} samples of Y_d are naturally available at each discrete time as a function of the system state.

The committee bias is related to two main sources of error. The first is associated to a model approximation, which can produce in some cases a persistent bias among the models included in the committee. The second is connected to a bias in the training database. When experimental or operative data are used for diagnostic model training, the environmental or the operative conditions (e.g. the applied load) can be different from those encountered during training. Furthermore, if simulated data are used for model training, any error in the simulation of the input–output variables is reflected in a bias when the diagnostic model is tested with real data. It is very difficult to estimate the committee bias in real applications, especially for real-time diagnostic systems based on distributed sensing like the one described in the Section 5. In a FCG framework, this would require the execution of repeated fatigue tests for the estimation of the committee bias mean $g(x)$ and variance $\sigma_b^2(x)$, as a function of the damage extent, for the specific structure and sensor network. In this work, Y_b is preliminarily assumed to be a known input to the system; in particular $g(x)$ has been fixed to zero and $\sigma_b^2(x)$ has been heuristically selected based on the experience, which constitutes the main simplification for this study. A further extension of the method would be to apply the technique in [23] to provide also an estimation of $Y_b(x)$, based on some sample test cases.

The multiple observations available at a discrete time k from N_{ANN} diagnostic models (dispersed as Y_d around their average) need to be combined, taking into account the bias associated with the average committee output and for this purpose a probabilistic framework for model combination leveraging on Gaussian mixture [14] has been adopted hereafter.

Consider first a single diagnostic model output $z_k^{(j)}$, where the superscript j indicates the jth ANN model inside the committee. If the bias of the average committee output is described as a Gaussian stochastic variable with the expectation $g(x)$ and the variance $\sigma_b^2(x)$, the likelihood of each ith particle of the evolution model is calculated through Eq. (11),

$$p\left(z_k^{(j)}\middle|y_k^{(i)}\right)\propto L\left(y_k^{(i)}\middle|z_k^{(j)}\right)=N\left(y_k^{(i)}\middle|z_k^{(j)}-g(\bar{x}_{k-1}),\sigma_b^2(\bar{x}_{k-1})\right) \tag{11}$$

where $\bar{x}_{k-1}$ is the expected value of the system state at a previous discrete time and $N\left(y_k^{(i)}\middle|z_k^{(j)}-g(\bar{x}_{k-1}),\sigma_b^2(\bar{x}_{k-1})\right)$ is the probability of the state vector particle $y_k^{(i)}$ according to a Gaussian distribution centred in $z_k^{(j)}-g(\bar{x}_{k-1})$, with the variance $\sigma_b^2(\bar{x}_{k-1})$. A schematic representation of the likelihood of $y_k^{(i)}$ with respect to the observation $z_k^{(j)}$ is shown in Fig. 1b. It has to be noted that, if a systematic bias $g(x)$ is encountered during repeated tests for the bias calibration, this is taken into account to adjust the likelihood calculation.

If the entire set of observations at discrete time k is considered, the expression for the likelihood estimation will depend on a mixture of the N_{ANN} Gaussian components associated to the N_{ANN} observations $z_k^{(j)}$, $j=1,\ldots,N_{ANN}$. A reformulation of the most general expression for the superposition of Gaussian densities [14] specifically addressing the problem at issue is reported in Eq. (12),

$$p\left(z_k\middle|y_k^{(i)}\right)=\sum_{j=1}^{N_{ANN}}\pi_j\cdot p\left(z_k^{(j)}\middle|y_k^{(i)}\right) \tag{12}$$

where the summation is derived from the model combination theory [14] and π_j represent the mixing coefficients, considered as independent from the state and defined in a way that $\sum_{j=1}^{N_{ANN}}\pi_j=1$ and $\pi_j\geq 0$ for all j, thus satisfying the requirement of mixing coefficients to be probabilities. In this study the authors assume the mixing coefficients as equal for all j, thus $\pi_j=\pi=1/N_{ANN}$. Eq. (12) can thus be simplified as Eq. (13),

$$\begin{aligned}p\left(z_k\middle|y_k^{(i)}\right)&\propto\sum_{j=1}^{N_{ANN}}p\left(z_k^{(j)}\middle|y_k^{(i)}\right)\\&\propto\sum_{j=1}^{N_{ANN}}N\left(y_k^{(i)}\middle|z_k^{(j)}-g(\bar{x}_{k-1}),\sigma_b^2(\bar{x}_{k-1})\right)\end{aligned} \tag{13}$$

where the mixing coefficients have been omitted, however considering a proportionality exists between the first and second term. Proportionality is sufficient here because a weight normalisation procedure is operated at each updating iteration as part of the SIR routine. The bias is a property of the committee, thus the same bias expectation $g(x)$ and variance $\sigma_b^2(x)$ are used in each mixture component, independently of j. Finally the weight updating formula in Eq. (9) can be rewritten as Eq. (14) to take multiple observations into account.

$$w_k^{(i)} = w_{k-1}^{(i)} \cdot \sum_{j=1}^{N_{ANN}} N\left(y_k^{(i)} \middle| z_k^{(j)} - g(\bar{x}_{k-1}), \sigma_b^2(\bar{x}_{k-1})\right) \tag{14}$$

3. Algorithm implementation for fatigue crack growth monitoring

3.1. FCG model

The SIR algorithm described in Section 2 is applied in this study to the problem of the fatigue crack growth prediction. The selection of a FCG model is the first requirement of the SIR algorithm. Though much more sophisticated models are available in the literature [3], the Paris–Erdogan law [1] has been selected for the validation of the method. Considering the NASGRO model as an example [2], it allows the FCG description also in the threshold zone and near the limit for unstable crack propagation. However, according to the literature describing the performances of automated real-time diagnostic systems, no reliable crack length measures would be available near the threshold zone. Moreover, safety regulations usually impose a system or a component to be dismissed before the damage becomes critical. For this reasons, in this study the attention is focused on the log-linear zone of the FCG problem, and a stochastic Paris law (15) has been selected as a model for damage evolution.

$$\frac{dx}{dN} = C[\Delta K(x)]^m \cdot \Omega \tag{15}$$

Eq. (13) describes the FCG rate per unit load cycle as a function of the Stress Intensity Factor (SIF) range $\Delta K(x)$ at the crack tips and two empirical constants (C and m) indicated as material-dependent parameters. Ω is a lognormal random process altering the growth rate and representing all the sources of *intra-specimen variability* [18,24]. Under the assumption of a linear damage accumulation, Eq. (15) can be formulated as Eq. (16), which is the core formula used to define the evolution Eq. (1) above.

$$x_k = x_{k-1} + \Delta N \cdot \left.\frac{dx}{dN}\right|_{k-1} = x_{k-1} + \Delta N \cdot C[\Delta K(x_{k-1})]^m \omega_{k-1} \tag{16}$$

where ω_{k-1} is a sample of the random process noise Ω, whose selection is discussed separately in Section 3.2, and ΔN is the discretisation step. While Eqs. (15) and (16) remain generally valid, the calculation of SIFs depends on the load and the geometry of the problem under investigation; thus it is separately described below for each application. Eq. (16) is used as the proposal density for the state prediction, as illustrated in Fig. 2.

A primary problem to be addressed is the large uncertainty associated to the FCG process [16]. This uncertainty is reflected in a large scatter of the FCG data, as was described by the authors in [10,16]. The model parameter updating procedure described in Section 2.2 is implemented in this study to adapt model predictions, by filtering the uncertainties related to model parameters based on the available sequence of measures. The C parameter is inserted in the state vector to be sequentially updated, while m has been considered deterministic. One drawback of such an approach is that, if a bias is present due to any error in the estimation of the SIFs, it is included in the C estimation during the updating step. In practice, a new parameter $\tilde{C}$ is estimated including all the sources of *inter-specimen variability* [18,24]. The augmented state vector described in Section 2.2 can now be formulated as Eq. (17).

$$y_k = [x_k, \vartheta_k] = [x_k, \tilde{C}_k] \tag{17}$$

Some prior information regarding $\tilde{C}$ parameter distribution is necessary to initialise the algorithm. In view of the final validation of the methodology on a panel specimen made of aluminium alloy Al2024-T6, the expected values for C and m have been selected based on literature data [2]. In particular: $\tilde{C} = 2.382e-12$ and $m = 3.2$. $\tilde{C}$ is assumed to have a lognormal distribution with prior variance calculated from Virkler's data [17], keeping the coefficient of variation $CoV = \sigma/|\mu|$ constant (Virkler's data are related to alloy Al2024-T3). The prior probability density function of $\log \tilde{C}$ becomes $\log \tilde{C} \sim N\left(\mu_{\log \tilde{C}}, \sigma^2_{\log \tilde{C}}\right) = N(-26.763,\ 0.9966)$.

As described in Section 2.1, the RL distribution is sequentially updated each time a new observation is available by evaluating the residual life for each particle, then creating the RL posterior probability density function based on Eq. (6). This is a critical aspect of the algorithm due to the computational requirements of the numerical integration. Instead of using a step-by-step particle propagation from x_k state condition to failure (x_{lim}), a method of stochastic integral resolution just used in a SIR framework in [18] is proposed here. The technique consists of the numerical resolution of (13) taking the presence of the stochastic process Ω into account. Further details of the method are not provided for the sake of brevity. The interested reader can refer to [18,25] for a detailed description of the method.

3.2. Algorithm parameters

Two main variables are involved in the algorithm procedure and require the correct setting to make the algorithm work properly. These are the random processes involved in the definition of the proposal PDF, namely Ω and ξ, for the state x and the parameter $\tilde{C}$, respectively.

Focusing on Ω, a lognormal distribution is often selected to define the random process. Mean and variance of the distribution must be linked in order to guarantee an unbiased estimation of the mean crack growth evolution [10]. In addition, the selection of a proper variance strongly influences the algorithm performance [18]. If it is too small, the algorithm is unable to follow the state dynamics, while too large a value will strongly reduce the filtering capability of the algorithm, resulting in a less accurate estimation of both the parameters and the residual life PDF. In the present study, a value based on preliminary sensitivity analyses was selected.

Similar considerations can be provided for the artificial noise ξ. If it is too small, the algorithm suffers from sample impoverishment and is also unable to efficiently adjust the parameter PDF. If it is too large, the loss of information between two consecutive iterations causes a distortion of the actual posterior PDF of parameters. Furthermore, a relatively large artificial dynamic component is needed at the beginning of the sequential filtering process to permit the algorithm to include samples that are outside the prior expectation; nevertheless this artificial term should be reduced to a minimum after the convergence on the correct parameter value is reached. In this study, the process ξ is described by a Gaussian distribution with zero mean and a variance that decreases exponentially during algorithm iterations, as expressed through Eq.(18).

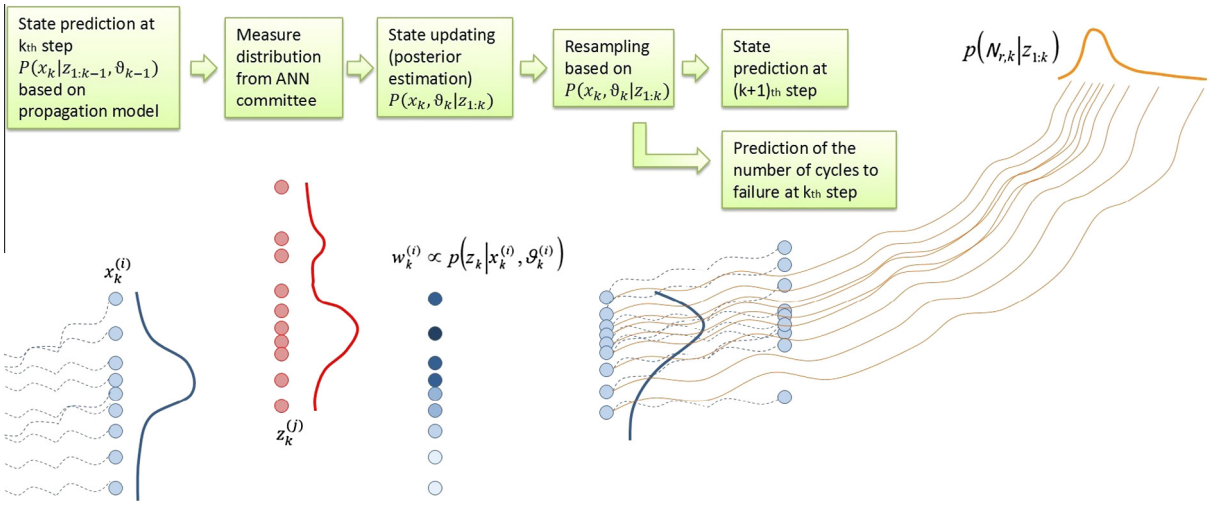

Fig. 2. SIR algorithm filtering procedure for one iteration.

$$\sigma_{\xi}^2(k) = \sigma_{\xi 0}^2 \cdot f_{\xi}(k) = \sigma_{\xi 0}^2 \cdot \frac{1}{k^{\alpha}} \tag{18}$$

where $\sigma_{\xi 0}^2$ is selected as a percentage of the prior parameter variance available from literature data. 25% was selected in this study as a compromise to allow relatively large movements in the parameter domain while limiting the initial explosion of particle parameter variance, thus $\sigma_{\xi 0}^2 = 0.25\sigma_{\log \tilde{C}}^2$. The exponent α has been tuned in order to keep a minimum variance at the end of the system operation, supposing a reasonable number of observations from the ANN dictated by the authors' experience.

The following points summarise the algorithm operations, while specific values of the algorithm parameter settings have been separately provided in the following sections. A schematic representation of the filtering procedure for one step of the SIR algorithm is shown in Fig. 2.

1. Algorithm initialization:

   Collect $z_0^j, j = 1,\ldots,N_{ANN}$ (*committee output at* $k = 0$)

   $\forall i = 1,\ldots,N_s$

   $$\tilde{C}_0^{(i)} \sim p\left(\tilde{C}_0, \sigma_{\log \tilde{C}}^2\right)$$

   $$x_0^{(i)} \sim p(z_0)$$

   $$w_0^{(i)} \sim 1/N_s$$

2. Prediction step:

   $$\sigma_{\xi,k}^2 = \sigma_{\xi 0}^2 \cdot \frac{1}{k^{\alpha}}$$

   $\forall i = 1,\ldots,N_s$

   $$\tilde{C}_k^{(i)} \sim p\left(\tilde{C}_k \middle| \tilde{C}_{k-1}^{(i)}, \sigma_{\xi,k}^2\right)$$

   $$x_k^{(i)} \sim p\left(x_k \middle| \left[x_{k-1}^{(i)}, \tilde{C}_k^{(i)}\right], \sigma_{\Lambda}^2\right)$$

3. Acquire new set of observations
   Collect $z_k^j, j = 1,\ldots,N_{ANN}$

4. Updating step

   $$w_k^{(i)} = \tilde{w}_{k-1}^{(i)} \cdot \sum_{j=1}^{N_{ANN}} N\left(y_k^{(i)} \middle| z_k^{(j)} - g(\bar{x}_{k-1}), \sigma_b^2(\bar{x}_{k-1})\right)$$

   $$\tilde{w}_k^{(i)} = w_k^{(i)} \Big/ \sum_i w_k^{(i)}$$

   $$p(y_k|z_{1:k}) = \sum_i \tilde{w}_k^{(i)} \delta\left(y_k - y_k^{(i)}\right)$$

5. Resampling procedure
   $\forall i = 1,\ldots,N_s$

   $$y_k^{(i)} = [x_k^{(i)}, \tilde{C}_k^{(i)}] \sim p(y_k|z_{1:k})$$

   $$\tilde{w}_k^{(i)} = 1/N_s$$

6. Residual Life posterior PDF updating
   $\forall i = 1,\ldots,N_s$

   – Estimate the RL for each particle $N_{r,k}^{(i)}$ according to the method in [18]
   – Calculate the RL posterior PDF

   $$p(N_{r,k}|z_{1:k}) \propto \sum_i \delta\left(N_{r,k} - N_{r,k}^{(i)}\right)$$

7. Repeat the steps 2–6 at each *k-th* discrete time.

4. Performance analysis on a simulated case

The SHM system with ANN committee-based diagnosis and embedded SIR algorithm is applied to simulated crack propagations. The main features of the FCG simulation and the virtual observations have been described and the performances of the algorithm have been evaluated based on a set of performance indices defined in [26].

4.1. Target FCG and algorithm parameters

A simple mechanical structure is considered in this section to highlight the algorithm performances. The simulated specimen is shown in Fig. 3a. It consists of an aluminium plate with a crack damage positioned on the left side with a crack length *x* propagating under constant amplitude sinusoidal load. Eqs. (15) and (16) allow generating the target damage evolution (reported in

Fig. 3b). The analytical formulation of the stress intensity factor variation (ΔK) within one load cycle is available for this simple test case, as indicated in Eq. (19).

$$\Delta K = F\Delta\sigma\sqrt{\pi x} \tag{19}$$

F is the shape function (assumed constant), $\Delta\sigma$ is the stress variation within one load cycle and x is the crack length, hereafter referred to as the state or health condition. At this stage, no random process noise (Ω) is activated during target FCG generation to better highlight algorithm convergence on the target simulation. Nevertheless the lognormal process noise in Eqs. (15) and (16) has always been used in the SIR routine for the correct approximation of the posterior PDF of the state vector conditioned on the observations. The FCG parameters used to simulate the target crack evolution are reported in Table 1.

If a sequence of virtual real-time observations of the target state is available as indicated in Section 4.2, the system target state (including the crack length and the $\tilde{C}$ parameter) and the residual life are sequentially estimated. A summary of the input parameters for algorithm implementation is reported in Table 2, including the transition density variables and the initialisation parameters. Though the simulated target FCG has been initialised at $x_0 = 3$ mm, the SIR algorithm has been activated after the averaged committee output exceeded a detection threshold (x_{det}), here heuristically set to 5 mm. Moreover, the mean value of prior $\log \tilde{C}$ distribution has been arbitrarily modified with respect to the target in order to verify the convergence of the parameter estimation to the target.

4.2. Virtual real-time crack length observations

The virtual real-time multiple observations have been simu-lated considering the model of uncertainty proposed in Section 2.3 for the diagnostic system based on the ANN committee. First, the *committee bias* (Y_b) has been modelled as a Gaussian distribution with a constant zero mean and a variance that increases as a function of the damage dimension. Concerning the bias expectation $g(x)$, this has been set to zero in order to evaluate the global algorithm performances for an unbiased (on average) measurement system. An increasing variance has been adopted due to the behaviour of the Artificial Neural Networks that are used in Section 5 for the real-time damage diagnosis [12,27]. The uncertainty $\sigma_b^2(x)$ on the definition of the committee bias often increases

Table 1
Features of the crack simulation.

Parameter	Description	Value
C	Empirical constant (mm/cycle · 1/MPa$\sqrt{}$mm)	2.382e−12
m	Empirical constant (–)	3.2
F	Crack shape function (–)	1.12
$\Delta\sigma$	Fatigue load (MPa)	40
x_0	Initial crack length (mm)	3
x_{lim}	Limit crack length (mm)	120
ΔN_s	Load cycle step for FCG simulation (cycles)	100
Ω	Random noise (–)	0

while approaching and exceeding the limit boundaries of the algorithm training domain, due to the inadequacy of the ANNs to perform extrapolation. Nevertheless it is sometimes possible to estimate a priori the limit damage dimensions based on the safety requirements and to design a proper damage database to limit this effect. A variance linearly dependent on the damage state has been adopted, as indicated in the Eq. (20).

$$Y_b(x) \sim N(0, \sigma_b^2(x)) = N\left(0, \sigma_{b,0}^2 \frac{x}{x_0}\right) \tag{20}$$

where $\sigma_{b,0}^2$ is the committee bias variance in correspondence of the initial crack length x_0. A sample of Y_b has been extracted from $N(0, \sigma_b^2(x))$ at each discrete time k, indicating the bias of the average committee output ($\bar{z}_k$) with respect to the target state. It is important to remark that the uncertainty model for the committee bias has to be provided as the input to the algorithm for the calculation of particle likelihood, as specified by Eq. (13). In this preliminary application, the same model for the committee bias is used to both simulate the observations and to calculate the likelihood.

Second, though not strictly Gaussian in general, the *committee dispersion* (Y_d) has also been assumed in this application as a Gaussian distribution with a constant zero mean and a variance that increases with the damage dimension. As a matter of fact, a typical increase of the committee dispersion is encountered during the damage propagation, especially when a model based training is performed as in [12,27], due to the rise of non-linearity and the influence of the damage on the signal of a multitude of sensors, thus increasing the possibility to make errors in the data fusion [27]. Again, a variance linearly dependent on the damage state has been adopted, as indicated in the Eq. (21).

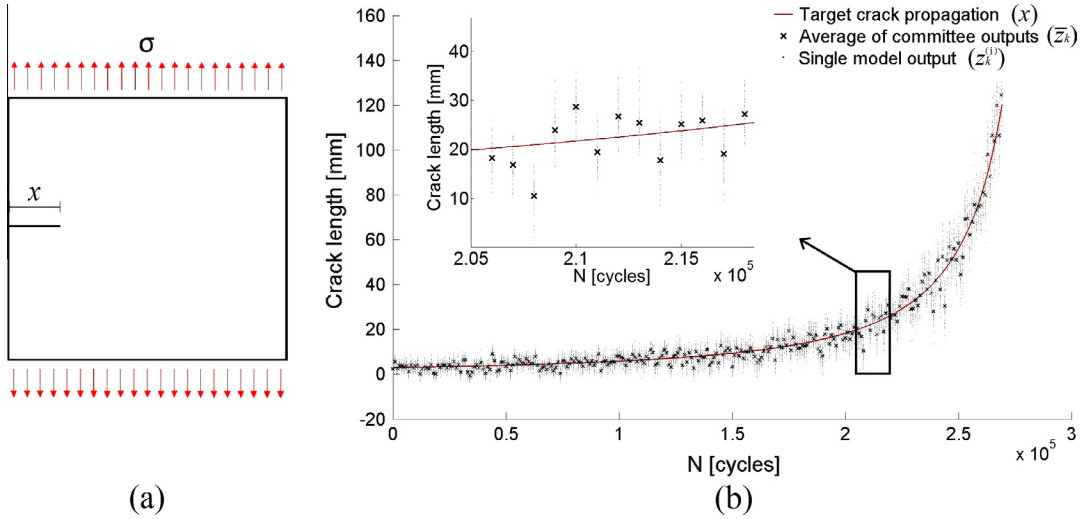

Fig. 3. (a) Virtual specimen for the FCG test and (b) target crack propagation with virtual committee observations.

Table 2
Input parameters for algorithm implementation.

Parameter	Description	Value
σ^2_{Λ}	Variance of the process noise (proposal variance)	0.1
σ^2_{s0}	Initial value of artificial dynamics' variance	0.2491
α	Decrease exponent for artificial dynamics' variance	1.86
x_{det}	Detection crack length (mm)	5
$\sigma^2_{\log \tilde{C}}$	Variance of prior $\log \tilde{C}$ distribution	0.9966
$\log \tilde{C}_0$	Mean value of prior $\log \tilde{C}$ distribution	−27.63
N_s	Number of particles	2000

$$Y_d(x) \sim N\left(0, \sigma_d^2(x)\right) = N\left(0, \sigma_{d,0}^2 \frac{x}{x_0}\right) \tag{21}$$

where $\sigma^2_{d,0}$ is the committee dispersion variance in correspondence of the initial crack length x_0. N_{ANN} samples of Y_d have been extracted from $N\left(0, \sigma_d^2(x)\right)$ at each discrete time k, indicating the committee dispersion around the average committee output ($\bar{z}_k$).

The vector of the committee measures z_k is thus defined as Eq. (11) at discrete time k. A new committee measure is supposed to be available every ΔN = 1000 load cycles. The simulated real-time observations are reported in Fig. 3b, where the average committee output, affected by bias with respect to the target crack length, and the committee dispersion are shown. The relevant parameters used to simulate the observations are reported in Table 3, heuristically selected based on the author's experience.

4.3. Algorithm performances on a simulated FCG

The SIR output results are illustrated in Fig. 4. The posterior state PDF estimation is shown in (a), the posterior PDF of $\log \tilde{C}$ in (b) and the residual life posterior PDF is reported in (c). Up to approximately $0.8 \cdot 10^5$ cycles the sequential filter is unable to refine the posterior estimation of the RL due to the fact that the evolution trends are entirely hidden in the observation noise. In particular, this is reflected in the large uncertainty affecting the parameter estimation, which produces wrong RL predictions as a consequence. Nevertheless the state posterior estimation appears to be correct after a few iterations. When a significant trend is identified within the observations, after $1 \cdot 10^5$ cycles, the SIR algorithm is able to provide very efficient filtering of the system evolution as well as good estimation of the RL. This is reflected in the convergence of the parameter estimation on the target value, with narrower confidence boundaries. After $1.5 \cdot 10^5$ cycles, a very small error is obtained between the target $\log \tilde{C}$ and the expected value of its posterior PDF. This produces a very accurate estimation of the RL, as visible in Fig. 4c.

Focusing on the state estimation (Fig. 4a), after the SIR algorithm is activated the confidence boundary associated to the estimated state reflects the levels of uncertainty that have been selected for committee bias and dispersion. This uncertainty is strongly reduced after a few iterations of the algorithm, even though it gradually increases with the uncertainties inherent to the observations. As a matter of fact, the increase of the committee dispersion and the bias uncertainty as a function of crack length allows for an enlargement of the particle swarm after resampling, which is reflected in the widening of the 95% confidence boundary of the state prediction. Furthermore, the state prediction converges noticeably towards the target state with very high accuracy after a few iterations. This has been obtained under the hypothesis that the observations have been generated with the exact same measurement model that is embedded into the SIR routine. In a real application, this is only possible if the bias variables, e.g. the bias expectation $g(x)$ and variance $\sigma_b^2(x)$, for particle likelihood calculation are known.

Additional comments arise referring to Fig. 4b. Due to both the absence of the process noise and the consideration of a deterministic value for $\log \tilde{C}$ in the target damage evolution, convergence of

Table 3
Features for the simulation of observations.

Parameter	Description	Value
$\sigma^2_{b,0}$	Initial committee bias variance (mm^2)	2
$\sigma^2_{K,0}$	Initial committee dispersion variance (mm^2)	2
N_{ANN}	Number of ANNs belonging to the committee	100
ΔN	Load cycle step for measure acquisition (cycles)	1000

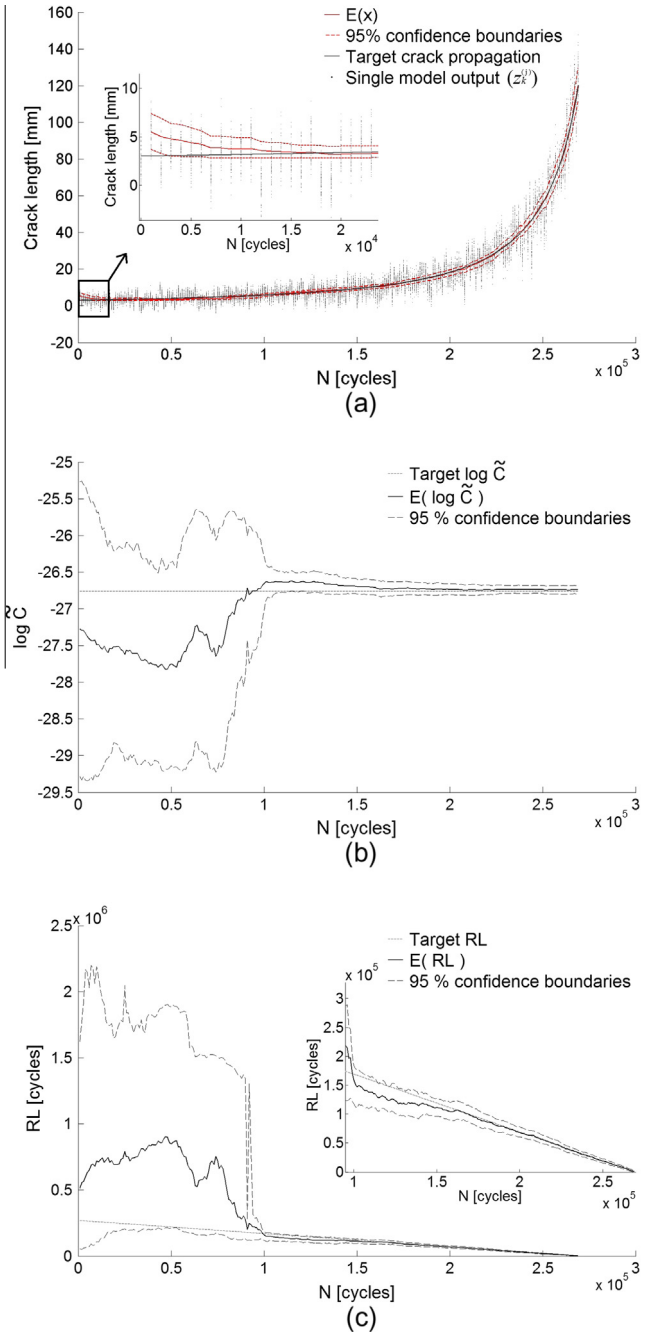

Fig. 4. SIR output in a simulated test. Posterior PDF estimation for (a) the crack length state, (b) the $\log \tilde{C}$ parameter and (c) the residual life.

the parameter estimation on the target with very narrow confidence boundaries has to be expected, if a sufficient number of observations have been provided. After convergence, a very small bias exists for the expected value of log $\tilde{C}$ posterior PDF. This bias depends on many factors. First, the number of observations is obviously limited in practical applications. Second, artificial dynamics cause a small loss of information between two consecutive steps, thus preventing further reduction of the confidence boundaries. Third, the process noise, coupled with artificial dynamics, causes a very large dispersion of the particles in the state-space, thus making the filtering procedure more difficult. However, the latter can be avoided by optimisation of the process noise variances (for both state and parameter(s)).

4.4. Focus on residual life estimation performance

Four metrics developed in [26] to evaluate the algorithm performance in the RL estimation have been used to assess the performance of the SIR algorithm in repeated simulated tests. Only a brief description of each metric with its input parameters is provided below, while the interested reader can refer to [26] for a detailed treatment.

The *Prognostic Horizon* (PH) is defined as the difference between the time index for the end of life and the time index when the RL posterior PDF first meets the specified performance criteria. These performance requirements are specified in terms of an allowable error bound around the target RL (α_{PH}) and a threshold probability mass (β_{PH}) that is required to fall within the α_{PH}-bounds (Fig. 5). It can be calculated in cycle units when a constant amplitude load is considered.

$\alpha - \lambda$ *accuracy* (AL) is a binary metric which evaluates, at a specific time λ, if the RL posterior PDF falls within specified α_{AL}-bounds that shrinks as time passes by (Fig. 5). A *cumulative* $\alpha - \lambda$ *accuracy* (CAL) is used hereafter, averaging the $\alpha - \lambda$ *accuracy* values calculated at each discrete time within the PH interval (Fig. 5). Being $\alpha - \lambda$ *accuracy* a binary variable, its cumulative version can be expressed in terms of a percentage of success. It requires the definition of $\alpha_{AL,0}$ (α_{AL}-bound at the beginning of the PH interval) and β_{AL}.

Relative accuracy is defined as a measure of the error of RL expectation relative to the target RL. *Cumulative relative accuracy* (CRA) is adopted here, defined as a normalised weighted sum of the relative accuracies calculated at discrete times within the PH interval (Fig. 5). Linearly varying weights have been assigned as a function of the distance from the end of life, being p_0 and p_{end} the weights at the beginning and the end of the PH interval, respectively.

Convergence (C_M) is a parameter used to quantify the rate at which a selected metric (M) improves with time. In particular it is calculated as the distance between the origin and the centroid of the area under the curve evaluated for the selected metric. Here, it is calculated based on the relative error between the target and the RL posterior PDF expectation.

A summary of the parameters necessary for the calculation of the selected indices is provided in Table 4. Furthermore, *computational time* (CT) is an additional metric that has to be considered for the real-time application of PHM systems. Here it is evaluated as the average CT required to perform one iteration of the algorithm.

The performance indices have been evaluated based on the results in Fig. 4c and are reported in Table 5. The average performance and the standard deviations have been estimated for 39 FCG simulations. Particularly, three different levels of detection threshold have been used (5 mm, 10 mm, and 15 mm); 13 repetitions have been performed at each level. Even though the same target damage evolution has been considered in each repetition, the randomness intrinsic to the measurement model and to the SIR routine induces a slightly different result each time the algorithm is run. The standard deviations reported in Table 5 clearly show the prognostic algorithm robustness to system uncertainties, however for a simulated FCG and specifically for the problem under consideration. These results have not been compared with any other real-time prognostic algorithm in this study, due to the lack of sim-ilar prognostic applications in the literature. However, they become useful for future comparisons with other algorithms for the residual life estimation. The same performance indices are used in Section 5 for the validation of the method on the real scenario.

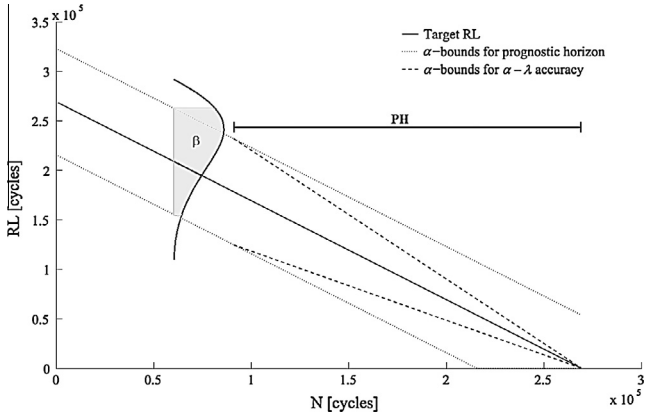

Fig. 5. Representation of the input parameters for the RL performance calculation.

5. System verification on a real application

In this section the SHM system is verified on a relevant aeronautical structure. A metallic panel representative of the rear fuselage of a medium-heavy weight helicopter constitutes the test structure. It is subjected to a real FCG artificially initiated at a rivet hole. The entire PHM system, based on strain pattern acquisition, is described hereafter, including the presentation of the experimental test rig in Section 5.1, the output of the diagnostic system trained on simulated data in Section 5.2, the implementation of the SIR filter to the experimental case in Section 5.3 and the results in Section 5.4. Only a brief overview of the diagnostic system is pro-vided and the interested reader can refer to [12] for a detailed description of the optimisation and feature extraction techniques that have been adopted.

5.1. Experimental test rig

The metallic panel specimen used to verify the applicability of the methodology is shown in Fig. 6. Its dimensions are 600 mm × 500 mm and it is characterised by a typical aeronautical riveted skin–stringer construction, where the skin and the stringers are made of Al2024-T6 and Al7475-T76 respectively. The lower edge of the panel is designed to permit connection to the ground, thereby simulating the skin–stringer-frame connection in real structures. The upper region is designed to receive the load from the actuator. The entire test rig is shown in Fig. 6. A sinusoidal

Table 4
Parameter setup for prognostic performance evaluation.

Parameter	Value
α_{PH}	20%
β_{PH}	50%
$\alpha_{AL,0}$	20%
β_{AL}	50%
p_0	0.1
p_{end}	1

Table 5
Performance of the RL estimation for the SIR algorithm with an augmented state vector (virtual test).

Index	Units	Average performance	Std. deviation of performance
PH	(cycles)	178,000	20,150
CAL	(%)	97.7	7.68
CRA	(%)	82.0	5.7
C_M	(cycles)	52,556	4079
CT	(s)	0.12	/

load with a constant amplitude is applied vertically, with a peak load of 35 kN and a load ratio $R = 0.1$. A similar load is associated to manoeuvres and in real applications occurs at a relatively low frequency. Nevertheless a 12 Hz load frequency is considered in this study to allow a more rapid test execution, without affecting the validity of the final description of algorithm results.

A sensor network for strain pattern acquisition based on fibre Bragg grating technology was installed on the specimen. Five sensors are deployed on each stringer, as shown in Fig. 6, thus 20 strain measures have been simultaneously acquired with an opti-cal interrogator HBM-DI410. Strain acquisition has been automatically activated every 500 load cycles, saving the data relative to the last 30 cycles at a scan rate of 1 kHz. The strain pattern is generated by averaging the 30 strain peaks measured by each of the 20 sensors. The strain patterns are always relative to the 35 kN peak load. Furthermore, a temperature sensor is used to compensate temperature fluctuations.

An artificial notch was created by cutting the panel skin next to a rivet hole, as shown in Fig. 6, to control the location of the fatigue crack initiation. The total width of the artificial notch is 10 mm. After a baseline strain pattern has been acquired from the artificially notched panel, necessary to calibrate the numerical model (in particular eliminating the error of numerical strain prediction for the baseline condition), a fatigue crack is propagated under sinusoidal load with constant amplitude up to a total length of 100 mm. Reference crack length measures (hereafter referred to as target crack propagation) have been taken with a calliper at some discrete times during the FCG. Differently from the case study of Section 4, in which the target FCG was simulated without a process noise addition, here the target crack propagation is naturally affected by its inherent variability. The results are therefore affected by this additional source of uncertainty.

5.2. Real-time crack length observations

A brief overview of the diagnostic system based on strain field pattern acquisition is provided hereafter to clarify how the committee output is generated. The authors stress the fact that no direct measure of crack length is provided as an observation to the SIR algorithm, nevertheless its estimation is generated through a supervised machine learning algorithm trained on simulated data, exploiting the same methodology previously described in [12,27]. Briefly, a numerical finite element model (FEM) has been built to predict the strain variation occurring on the panel speci-men due to the presence of a rivet crack. A database of simulated strain patterns is generated varying the position and the dimension of the damage. These simulated data are then used to train a com-mittee of ANN models for crack length estimation. The description of the FEM used to train the ANN is beyond the scope of this paper and its details have been provided in [12,28]. A brief overview of the diagnostic system based on a committee of ANN is provided hereafter, focussing on the aspects relevant to this study.

Summarising the operation of the diagnostic algorithm, each ANN receives as the input the pattern of the strains collected from the 20 sensors installed on the structure, and gives as the output an estimation of the current crack length affecting the helicopter panel. The strain values used as the inputs of the ANNs are divided by the mean of all the 20 sensors in order to mitigate the load effect. Such normalised strains are defined *damage indices* henceforth, as recently applied in [12]. A multilayer perceptron with single hidden layer is used, with 20 input nodes (corresponding to the 20 damage indices relative to each sensor), one output node (indicating crack length) and 39 hidden nodes. The latter has been optimised as described in [27]. Error back-propagation based on Scaled Conjugate Gradient algorithm is used to adjust the ANN weights

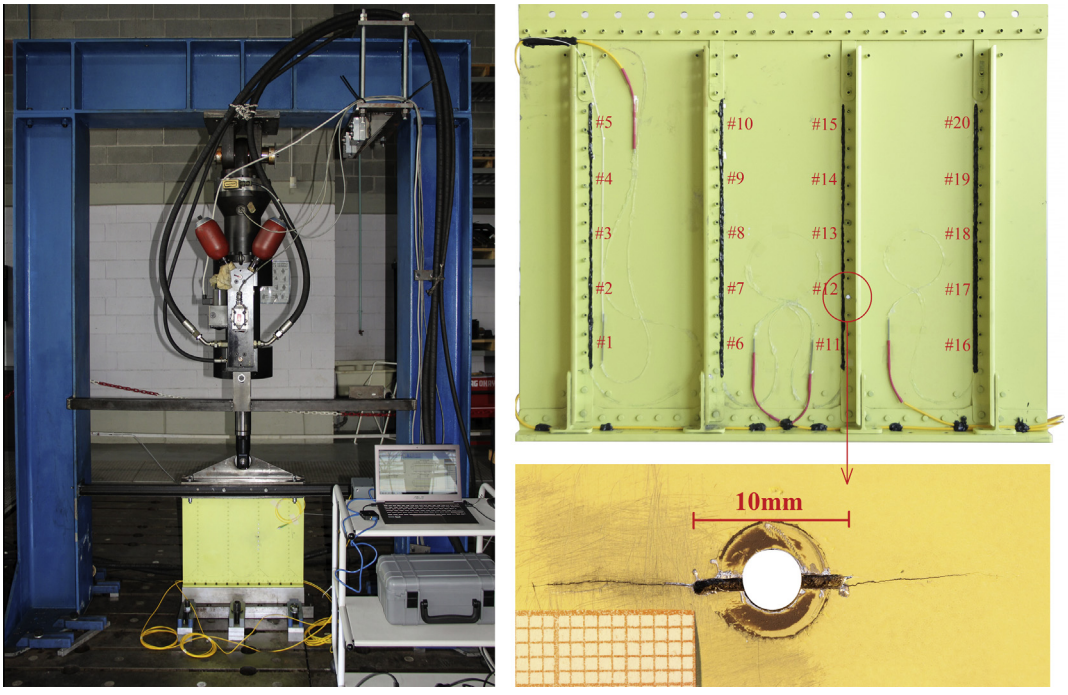

Fig. 6. Experimental setup, with focus on the panel specimen with 20 fibre Bragg grating sensors installed and the artificial notch for rivet crack damage initiation.

during the training procedure. The numerical database has been divided into the training and the validation sets and *cross-validation* with *early stopping* is used to avoid over-fitting [14,15]. This is sufficient in some applications in which the ANN training is based on experimental measures; however, if a model-based training is performed, additional techniques can be adopted to increase the generalisation capability of the ANN. In this study, different ANN models have been grouped into one com-mittee to provide multiple observation outputs. The authors remark that each model has the same structure but is trained on a different *Bootstrap* dataset [14], randomly dividing the input space into the training and validation domains, corresponding to 70% and 30% of the total available cases, respectively. Furthermore, random weight initialisation is assigned at the begin-ning of the training. This results in different optimal weight values each time a new training is performed. In particular, N_{ANN} = 100 is considered in this study, where each ANN is selected as the best among five training trials, based on the root mean square error cal-culated on the validation set.

Diagnostic results have been reported in Fig. 7, where the target crack propagation (corresponding to the measure by the calliper) is compared with the damage assessment performed by the committee of ANNs. The committee output is presented, together with its average. The committee bias for the specific test is the distance between the average of committee outputs ($\bar{z}_k$) and the target, while the scattering of the single ANN outputs at each discrete time indicates the committee dispersion.

Focussing on the bias, the diagnostic system appears to be insensitive to relatively small damage (crack length below 25 mm), which is reflected in the constant average committee output up to $2.2 \cdot 10^5$ cycles. The average committee output remains centred around 10 mm, which is the reference baseline condition used for numerical model calibration. After $2.2 \cdot 10^5$ cycles the average of the committee output approaches the target crack propagation and an efficient damage assessment is performed up to $7 \cdot 10^5$ cycles. After $7 \cdot 10^5$ cycles, a larger bias of the average committee output has been found. Two main reasons are responsible for this behaviour: (i) the crack length approaches the boundaries of the training domain and the ANN approximation is usually less accurate [12,27] and (ii) the strain prediction with the FEM has a bias with respect to the experimental strain when the structure is damaged. The spike around $6.5 \cdot 10^5$ cycles is due to uncontrolled environmental influences.

Concerning the committee dispersion, though not explicitly indicated in the figure for simplicity, the expected value and the mode of the committee output at a specific discrete time practically overlap each other, thus highlighting the Gaussian nature of the committee dispersion for this application [12,27] and justifying the assumption made in Section 4. A gradual increase of the committee dispersion can be appreciated as a function of damage dimension, for the same reasons expressed in Section 4.

The committee output is passed as input to the SIR routine for the posterior estimation of the state vector and the residual life PDFs. Different from Section 4, in this application the committee dispersion is inherent to the N_{ANN} diagnostic models organised in a committee. Nevertheless, a model for the committee bias still has to be provided to integrate the committee observations within the SIR routine, according to the input requirements defined in Section 2.3. The evaluation of error models for real-time systems is absolutely non-trivial and continues to be studied by the research community [23]. It has to be noticed that, if repeated FCG tests had been available and a common bias had been encoun-tered among the tests (e.g. the underestimation of damage dimen-sion when the crack approaches the boundaries of the training domain, as in Fig. 7), it would have been possible to calculate the

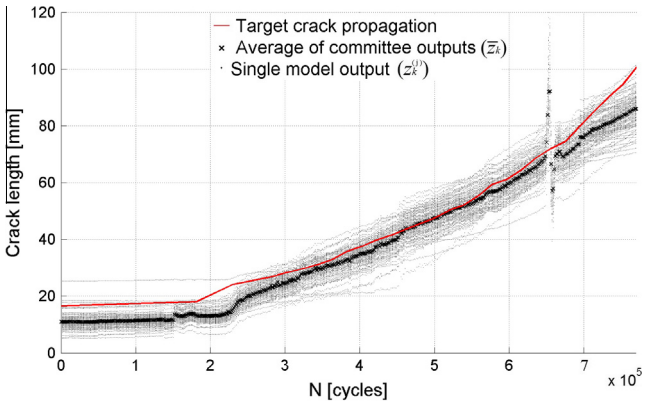

Fig. 7. Output of the diagnostic system for the crack damage quantification. 100 ANNs provide an estimation of the crack length at each discrete time.

average bias $g(x)$ and compensate for it in likelihood calculation, as in Eq. (13). For this single laboratory application, the committee bias model has been heuristically selected and described as Eq. (20). A summary of the parameters relative to the observations is presented in Table 6.

5.3. Application of the SIR filter to the real case

Once the measurement model has been defined in the previous Section 5.2, the attention is directed towards the definition of the SIR algorithm parameters for the application to the experimental use case. The Paris equation is still adopted in its discretised form, as represented in Eq. (16). The additional complexity lies in the calculation of the stress intensity factor, as the crack shape function depends on the crack length itself, i.e. $\Delta K = F(x)\Delta\sigma\sqrt{\pi x}$. Particularly, the crack shape function $F(x)$ has been defined through the simplified analytical formulation proposed in [29] for a skin–stringer structure with a crack extending on both sides from a rivet hole.

The crack length limit x_{lim} for the RL calculation has been set to 100 mm. In a real application, it is reasonable to expect that the anomaly detection is provided after a significant distance from the baseline measure has been recorded [12]. However, since an anomaly detection system is not employed in this application for simplicity, a heuristic detection threshold x_{det} has been set to 25 mm.

A summary of the features selected to define the transition densities (the process noise and the artificial dynamics for state and parameter prediction, respectively) has been provided in Table 7, together with other parameters necessary for the algorithm implementation. In particular, the stress range $\Delta\sigma$, required to calculate SIF variation in one load cycle, has been evaluated as the ratio between the applied load range and the transversal area of the panel.

5.4. Results

The SIR algorithm has been activated after the average committee output exceeded the detection threshold (x_{det}) and the results are reported in Fig. 8a–c for the state, the parameter and the residual life posterior PDF evaluation, respectively.

Focussing on the state (x) filtering (Fig. 8a), the very wide uncertainty associated to the committee output is strongly reduced and a very precise description of the damage evolution is provided. The SIR algorithm is able to filter the strong fluctuations due to environmental influences around $6.5 \cdot 10^5$ cycles, due to the fact that they are not compatible with the crack growth model, the current parameter $\log \tilde{C}$, and the artificial noise(s) describing the DSS.

Table 6
Features for the real-time observations.

Parameter	Description	Value
$\sigma^2_{b,0}$	Initial committee bias variance (mm^2)	9
N_{ANN}	Number of ANNs belonging to the committee	100
ΔN	Load cycle step for measure acquisition (cycles)	500

Furthermore, the expected value of the state posterior PDF after $7 \cdot 10^5$ cycles is noticeably affected by a reduced error if compared with the average committee output (Fig. 7). This is possible (i) if a correct identification of the FCG model parameter is obtained and (ii) if the measurement model takes the possibility of sufficiently large committee biases into account. Thus, an improved accuracy of the state evolution description is obtained. It is however important to remark that if the algorithm is evaluated after $8 \cdot 10^5$ cycles the wrong estimation of the crack length by the committee of ANNs will produce a biased estimation of the system state.

The parameter ($\log \tilde{C}$) posterior PDF approximation is compared in Fig. 8b with the prior $\log \tilde{C}$ expectation and the estimation of $\log \tilde{C}$ through a non-linear regression based on the entire history of the average committee output $\bar{z}_k$. Even though the wide initial uncertainty in the prior $\log \tilde{C}$ is related to literature data [17], it is strongly reduced during the algorithm operation. As a matter of fact, once a trend is identified within the observations, the confidence boundaries associated with $\log \tilde{C}$ posterior PDF are much reduced. However, the algorithm is driven by the ANN committee observations. Thus, if the measure is affected by an error, a wrong $\log \tilde{C}$ posterior expectation is induced, unless this error is included in the model for the observation bias in terms of $g(x)$. It is important to remark that the posterior PDF expectation of $\log \tilde{C}$ calculated by the SIR algorithm approaches the value of $\log \tilde{C}$, evaluated by a non-linear regression on the average committee output, thus validating the SIR output in terms of parameter estimation. The distance between the prior $\log \tilde{C}$ expectation and the expectation of the posterior $\log \tilde{C}$ PDF is associated to (i) sources of *inter-specimen* variability, as anticipated in Section 3.1 and (ii) the fact that the m parameter of the Paris equation is considered as fixed and only $\log \tilde{C}$ is used as fitting variable.

Finally, concerning the RL posterior PDF estimation (Fig. 8c), the initial uncertainty associated with the state vector (crack length and $\log \tilde{C}$) provokes a very wide posterior PDF of the residual life up to $3.3 \cdot 10^5$ cycles. Convergence in the state vector estimation guarantees a more precise indication of the expected RL. As anticipated above, any error in the state vector estimation is reflected in a bias in the RL expectation; the overestimation of the crack length coupled with the overestimation of $\log \tilde{C}$ (with respect to the one calculated by non-linear regression for the entire sequence of observations) provoke the underestimation of the RL expectation

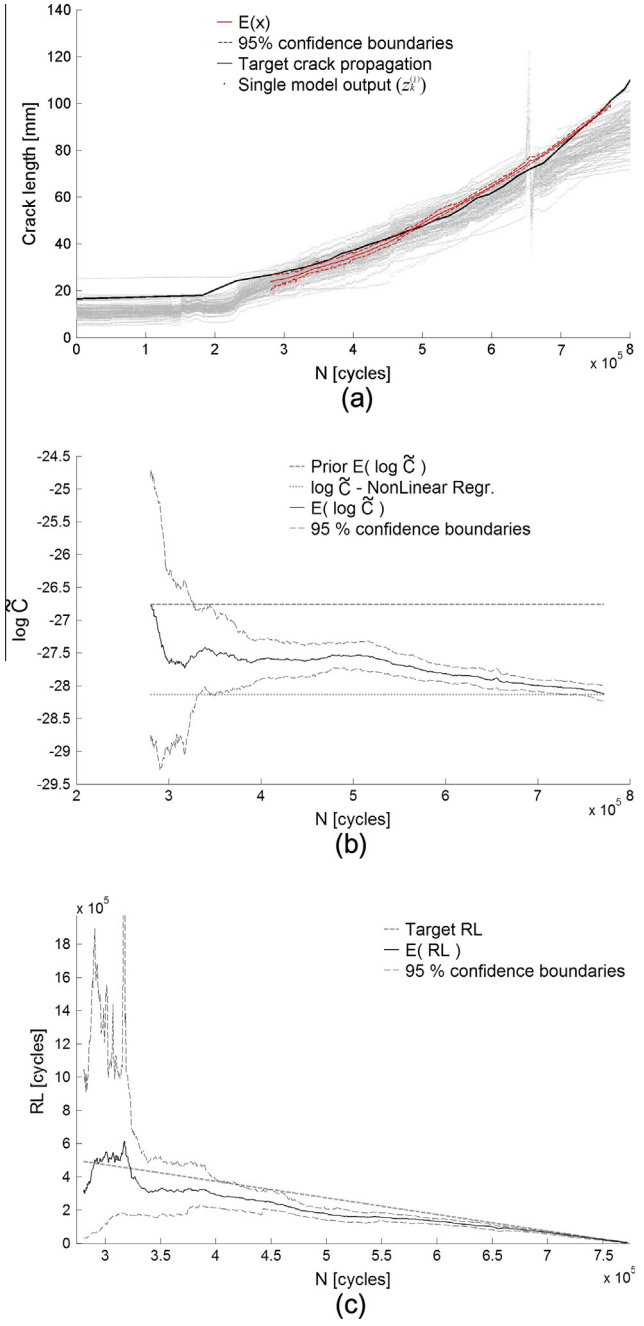

Fig. 8. SIR output in a real application. Posterior PDF estimation for (a) the crack length state, (b) the $\log \tilde{C}$ parameter and (c) the residual life.

Table 7
Summary of input parameters for algorithm implementation in real FCG.

Parameter	Description	Value
x_{det}	Detection crack length (mm)	25
x_{lim}	Limit crack length (mm)	100
σ^2_A	Variance of the process noise	0.1
$\sigma^2_{\tilde{s}0}$	Initial value of artificial dynamics' variance	0.2491
α	Decrease exponent for artificial dynamics	1.32
$\sigma^2_{\log \tilde{C}}$	Variance of prior $\log \tilde{C}$ distribution	0.9966
$\log \tilde{C}_0$	Mean value of prior $\log \tilde{C}$ distribution	−26.76
m	Empirical deterministic constant	3.2
$\Delta\sigma$	Stress range in one load cycle (MPa)	40.7
N_s	Number of particles	2000

Table 8
Performance in the RL estimation through a SIR algorithm with an augmented state vector (experimental test).

Index	Units	Performance
PH	(cycles)	448,500
CAL	(%)	35.2
CRA	(%)	77.8
C_M	(cycles)	208,810
CT	(s)	0.93

between $4 \cdot 10^5$ and $7 \cdot 10^5$. The prognostic performance indices introduced in Section 4.4 are used here to describe the prognostic performances in a real environment and results are reported in Table 8. The same parameter setup presented in Table 4 has been used here for prognostic performance evaluation. The significant reduction in the CAL parameter was expected and is related to the error in the crack length observations by the diagnostic system. A higher CT is required at each iteration to numerically approxi-mate the Paris's integral for the RL calculation.

6. Conclusions

A methodology for the integration of the real-time observations generated by a committee of artificial neural networks in a sequential importance resampling algorithm has been described in the paper. The algorithm, based on Monte-Carlo sampling, is aimed to approximate the posterior PDF of the state vector (including the crack length and a parameter of the FCG model) and the residual life. The availability of multiple models for real-time diagnosis generates multiple observations as the input to the SIR algorithm. The observations are characterised by a dispersion and an average bias, denominated as the committee dispersion and the committee bias, respectively. The likelihood formulation for posterior distribution updating has been modified to consider the availability of multiple observations as the input. The method has been applied first to the identification of a simulated fatigue crack growth, then to a laboratory test during a real fatigue crack growth test on a metallic specimen consisting of a riveted skin–stringer panel representative of the rear fuselage of a medium-heavy weight helicopter.

The results obtained in the preliminary virtual test demonstrate the potentiality of the approach for real-time applications. As a matter of fact, the SIR capabilities to filter the combined uncertainty related to the measure of the crack length and the prior knowledge of FCG model parameters are well-known, as proven by the extensive literature on the method. A robust identification of the residual life posterior PDF is obtained with a limited computational time. The most innovative step consists in the implementation of the entire methodology composed by an ANN-based SHM system with an embedded SIR algorithm in a simplified, though rather complex, laboratory scenario. A machine learning algorithm based on a network of optical strain gauges is trained with simulated strains and can generalise efficiently to the real data, providing real-time structural diagnosis in terms of crack length. Multiple artificial neural network models are combined in a committee thus providing a crack length distribution as the output. These multiple observations are then fused providing the estimation of the required posterior PDFs. Promising results have been obtained also in the laboratory application, still with a limited computational time requirement. The final prognostic performance is evidently strictly dependent on the quality of the diagnostic output; the system can efficiently filter the fluctuations due to environmental influences, even though if the diagnosis is affected by a bias, this will be reflected in a wrong FCG parameter estimation and consequently also in a wrong residual life estimation. However, if the measurement model is correctly determined and assigned as the input to the algorithm, it can mitigate the effect of a wrong damage assessment. Furthermore, as anticipated in [10,16], reliable prognosis is only possible if the correct estimation of the FCG model parameter(s) is obtained. In this study, emphasis is given also to the parameter identification. In particular, a parameter of the FCG model usually associated with material properties is inserted in the state vector and its posterior PDF has been sequentially updated through the SIR algorithm.

Focusing on the introduction of the ANN-based diagnosis within the SIR algorithm, some weaknesses have been identified as topics for future research by the authors. The parameters describing the bias within the measurement model for a committee of diagnostic systems are currently based on heuristic selection, while a rigorous methodology is mandatory to calibrate the committee bias, e.g. the one described in [23], thus allowing the application of the system in real scenarios. However this will require the execution of repeated FCG tests or the definition of model-assisted methods. Drawbacks and unanswered matters specifically related to SIR algorithms are well-known. The tuning of the PDF parameters involved in the transition process remains heuristic so far. A defi-nition a priori of the process noise is often difficult. Moreover, aeronautical structures (like the one presented in this work), as well as civil and mechanical structures, are often subject to vari-able amplitude and random loads, which are in most cases unknown. The applicability of SIR algorithm in the presence of variable and unknown load histories is another mandatory aspect to be considered in future research.

References

[1] Paris PC, Erdogan F. A critical analysis of crack propagation laws. Trans ASME J Basic Eng 1963;85(4):528–34.
[2] NASA J S Centre and Southwest Research Institute. NASGRO reference manual, version 4.02; 2002.
[3] Dowling NE. Mechanical behavior of materials. US: Pearson Prentice Hall; 2007.
[4] Lazzeri L, Mariani U. Application of damage tolerance principles to the design of helicopters. Int J Fatigue 2009;31(6):1039–45.
[5] Cadini F, Zio E, Avram D. Model-based Monte Carlo state estimation for condition-based component replacement. Reliab Eng Syst Saf 2009;94:752–8.
[6] Doucet A, De Freitas JFG, Gordon NJ. Sequential Monte Carlo methods in practice. New York: Springer-Verlag; 2001.
[7] Arulampalam MS, Maskell S, Gordon N, Clapp T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans Signal Process 2002;50(2):174–88.
[8] Orchard ME, Vachtsevanos GJ. A particle filtering approach for on-line fault diagnosis and failure prognosis. Trans Inst Meas Control 2009;31(3–4):221–46.
[9] Cadini F, Zio E, Avram D. Monte Carlo-based filtering for fatigue crack growth estimation. Probabilist Eng Mech 2009;24:367–73.
[10] Corbetta M, Sbarufatti C, Manes A, Giglio M. Real time prognosis of crack growth evolution using sequential Monte Carlo methods and statistical model parameters. Trans Reliab 2014. http://dx.doi.org/10.1109/TR.2014.2366759.
[11] Orchard ME, Vachtsevanos GJ. A particle filtering approach for on-line failure prognosis in a planetary carrier plate. Int J Fuzzy Logic Intell Syst 2007;7(4):221–7.
[12] Sbarufatti C, Manes A, Giglio M. Performance optimization of a diagnostic system based upon a simulated strain field for fatigue damage characterization. Mech Syst Signal Process 2013;40(2):667–90.
[13] Boller C, Chang FK, Fujino Y. Encyclopedia of structural health monitoring. New York, NY: John Wiley and Sons; 2009.
[14] Bishop C. Pattern recognition and machine learning. New York: Springer; 2006.
[15] Bishop C. Neural networks and pattern recognition. Oxford: Oxford University Press; 1995.
[16] Corbetta M, Sbarufatti C, Manes A, Giglio M. On dynamic state-space models for fatigue-induced structural degradation. Int J Fatigue 2014;61:202–19.
[17] Virkler DA, Hillberry BM, Goel PK. The statistical nature of fatigue crack propagation. Technical report AFFDL-TR-78-43; 1978.
[18] Corbetta M, Sbarufatti C, Manes A, Giglio M. Sequential Monte Carlo sampling for crack growth prediction providing for several uncertainties. In: Proc European conference of the Prognostics and Health Management Society, Nantes, France; 2014.
[19] Kantas N, Doucet A, Singh SS, Maciejowski JM. An Overview of sequential Monte Carlo methods for parameter estimation in general state-space models. In: Trans 15th IFAC symposium on system identification; 2009.
[20] Storvik G. Particle filters for state-space models with the presence of unknown static parameters. IEEE Trans Signal Process 2002;50(2):281–9.
[21] Gordon NJ, Salmond DJ, Smith AFM. Novel approach to non-linear/non-Gaussian Bayesian state estimation. IEEE Proc F (Radar Signal Process) 1993;140(2):107–13.
[22] Liu JS, West M. Combined parameter and state estimation in simulation-based filtering. In: Doucet A, De Freitas JFG, Gordon NJ, editors. Sequential Monte Carlo methods in practice. New York: Springer-Verlag; 2001.
[23] Baraldi P, Compare M, Sauco S, Zio E. Ensemble neural network-based particle filtering for prognostics. Mech Syst Signal Process 2013;41:288–300.
[24] Bourinet JM, Lemaire M. Form sensitivities to correlation: application to fatigue crack propagation based on Virkler data. In: Proc 4th international ASRANet colloquium; 2008.

[25] Yang JN, Manning SD. A simple second order approximation for stochastic crack growth analysis. Eng Fract Mech 1996;53(5):667–86.
[26] Saxena A, Celaya J, Saha B, Saha S, Goebel K. Metrics for offline evaluation of prognostic performance. Int J Prognostics Health Manage 2010;1(1).
[27] Sbarufatti C, Manson G, Worden K. A numerically-enhanced machine learning approach to damage diagnosis using a lamb wave sensing network. J Sound Vib 2014;333(19):4499–525.
[28] Sbarufatti C, Manes A, Giglio M. Application of sensor technologies for local and distributed structural health monitoring. Struct Control Health Monitor 2014;21(7):1057–83.
[29] Poe Jr CC. Stress-intensity factor for a cracked sheet with riveted and uniformly spaced stringers. NASA technical report TR R-358 1971, Langley Research Center, Hampton, Va. 23365.