Ensemble neural network-based particle filtering for prognostics

Piero Baraldi, Michele Compare, Sergio Sauco, Enrico Zio

Ensemble Neural Network-Based Particle Filtering for Prognostics

P. Baraldi[1], M. Compare[1], S. Sauco[1], E. Zio[1,2]

[1]Politecnico di Milano, Italy

[2]Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy-Electricité de France, Ecole Centrale Paris and Supelec, France

Abstract

Particle Filtering (PF) is used in prognostics applications by reason of its capability of robustly predicting the future behavior of an equipment and, on this basis, its Residual Useful Life (RUL). It is a model-driven approach, as it resorts to analytical models of both the degradation process and the measurement acquisition system. This prevents its applicability to the cases, very common in industry, in which reliable models are lacking. In this work, we propose an original method to extend PF to the case in which an analytical measurement model is not available whereas, instead, a dataset containing pairs «state - measurement» is available. The dataset is used to train a bagged ensemble of Artificial Neural Networks (ANNs) which is, then, embedded in the PF as empirical measurement model.

The novel PF scheme proposed is applied to a case study regarding the prediction of the RUL of a structure, which is degrading according to a stochastic fatigue crack growth model of literature.

1 Introduction

Predictive Maintenance (PM) is an innovative maintenance paradigm, founded on the assessment of

the current health state of an equipment (i.e., state identification) and the prediction of its future evolution (i.e., prognostics, [1] [2]). This allows identifying problems in the equipment at the early stages of development and estimating the RUL (i.e., the residual time span before the degradation state reaches the threshold that leads to a loss of functionality). In principle, an accurate estimate of the RUL enables to run the equipment as long as it is healthy, thus providing additional time to opportunely plan and prepare the maintenance interventions for the most convenient and inexpensive times [2].

The development of prognostic systems capable of reliably predicting the occurrence of a faulty condition in the equipment mainly depends on the quality and quantity of the available information and data on its past, present and future behavior. In this respect, a typical distinction is made between data-driven and model-based approaches. In the former case, modeling of the evolution of the degradation process relies exclusively on process history data. Empirical techniques like Artificial Neural Network (ANN, e.g, [3], [4]), Support Vector Machine (SVM, e.g. [5]), Local Gaussian Regression (LGR, e.g. [6], [7]) are typical examples. In the latter case, a model of the degradation process is used to predict the future evolution of the equipment state and infer the RUL. Physic-based Markov models (e.g., [8], [9]), statistical distributions of failure times (e.g., [10], [11]), empirical degradation models (e.g., [12], [13]) are typical examples. If experimental or field degradation data are available, these can be used to calibrate the parameters of the model or to provide ancillary information related to the degradation state, within the state-observer formulation typical of a filtering problem with given state model; Kalman filtering (KF, e.g., [14], [15]) and Particle Filtering (e.g., [16]-[20]) are typical examples.

This work focuses on Particle Filtering (PF) for prognostics, in recognition of its capability of robustly predicting the future behavior of the equipment degradation, x, which is typically a quantity not directly observable, without requiring the strict hypotheses of the KF ([16]) on the linearity of the system state evolution and noise gaussianity. 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 assess the equipment RUL.

Previous works employing a PF scheme for prognostics (e.g., [16], [18], [19]) assume knowledge of the following information [21]:

1) The knowledge of the degradation model describing the stochastic evolution in time of the equipment degradation $\underline{x}$ (in general a multi-dimensional vector):

$$\underline{x}(t+1) = \underline{g}(\underline{x}(t), \underline{\omega}(t)) \quad (1)$$

where $\underline{g}$ is a possibly non-linear vector function and $\underline{\omega}(t)$ is a possibly non-Gaussian noise.

2) A set of measures $\underline{z}(1),...,\underline{z}(t)$ of past and present values of some physical quantities $\underline{z}$ related to the equipment degradation $\underline{x}$. Although z in general is a multi-dimensional vector, in this work it is considered as a mono-dimensional variable; then, the underline notation is omitted.

3) A probabilistic measurement model which links the measure z with the equipment degradation $\underline{x}$:

$$z(t) = h(\underline{x}(t), \underline{\nu}(\underline{x}(t))) \quad (2)$$

where h is a possibly non-linear vector function and $\nu(\underline{x})$ is the measurement noise vector.

In practical cases where the analytical measurement model is not available, the PF scheme here described would not be directly applicable. To overcome this hurdle, in this work the challenge is notably to combine a data-driven approach for exploiting the measurement data, with a model-based approach, for exploiting degradation modeling. To the authors' best knowledge this is the first time that such issue is addressed in prognostics.

For example, piping of deep water offshore well drilling plants degrades due to a process of scale deposition which may cause a decrease, or even a plug, of the cross sections of the tubular. Giving the inaccessibility of the piping, it is usually impossible to acquire a direct, on line, measure of the scale deposition thickness. On the other side, research efforts are devoted to perform laboratory

tests to investigate the relationships between the scale deposition thickness and other parameters which can be more easily measured during plant operation, such as pressures, temperatures and brine concentrations. The idea is to populate datasets with the values of the measurable parameters for different scale deposition thicknesses, and use the data to build data-driven models for predicting the scale deposition thickness [22].

Another example concerns the crack propagation in bearings of rotating machinery, which often results in damage to the bearings and consequent reduced efficiency, or even severe damage, of the entire motor system. In this respect, several studies have been performed for an analytical description of the crack propagation process in bearing (e.g., [23], [24]). However, a direct measure of the crack depth during online operation is usually not possible, and thus the classical PF scheme is not applicable. On the other hand, since the major tell-tale sign of a bad bearing is an increase in vibration, both in amplitude and complexity ([25], [26]), a possible approach consists in developing laboratory tests to relate the crack depth to the measurements of the vibrations which the crack induces in the assembly. From these tests, an empirical model which links the vibration measurement to the real crack depths may be obtained.

In this context, the specific novelty of the work here presented is the application of PF for prognostics in a case in which the measurement model is not available but a dataset containing a number of pairs made by the state and the corresponding measurement is available for exploitation by a data-driven approach. To do this, a technique based on the use of an ensemble of ANNs [27] is here embedded in the PF scheme. Bagging is used to generate the datasets for training the different ANN predictors whose output are, then, combined to give the ensemble output, which is characterized by a lower variance than the one of the single ANN predictor [28]. The key idea of bagging is to treat the available dataset T as if it were the entire population, and then create alternative versions of the training set, by randomly sampling from it with replacement.

An alternative way to cope with the issue considered in this work consists in the development of a completely data-driven approach. For example, one may use the analytical degradation model to

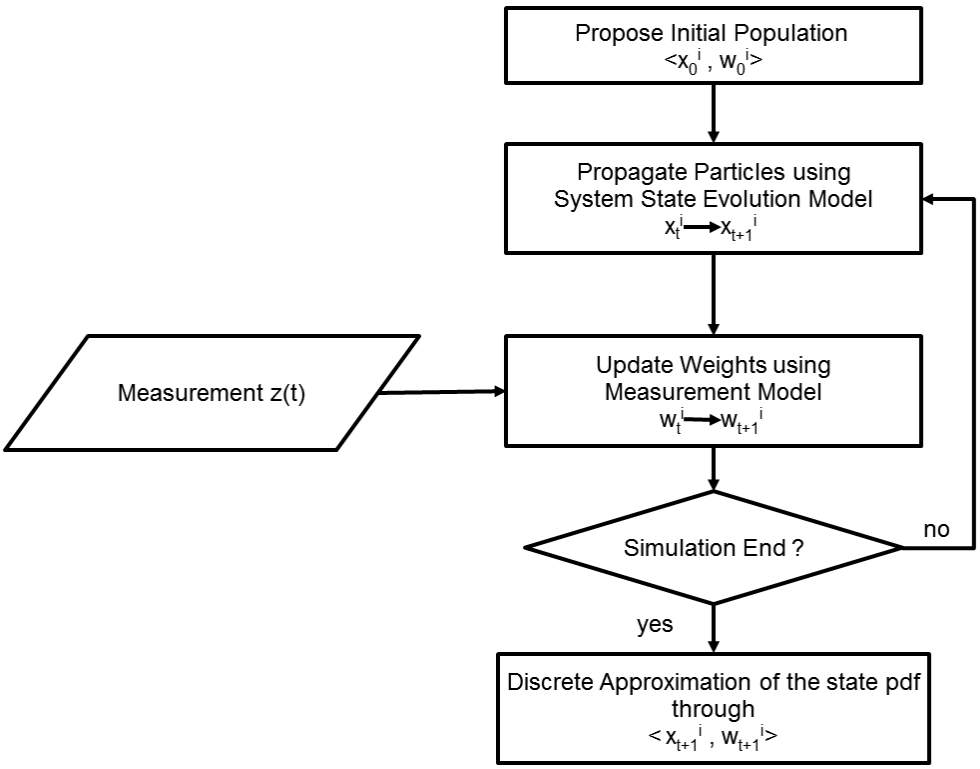
generate samples of degradation evolutions, and, then, utilize them together with the measurement data in a 'traditional' data-driven approach (e.g., a Support Vector Machine) which directly predicts the system future degradation evolution. However, notice that traditional data-driven approaches, differently from the PF approach, are typically not able to continuously update in a Bayesian perspective the uncertainty on the prediction.

The remainder of the work is structured as follows: in Section 2 a brief review of the PF is reported; in Section 3 the problem of the substitution of the measurement model and the technique to overcome this problem are presented. In Section 4 the technique is applied to a case study dealing with the physical phenomenon of the crack growth; Section 5 concludes the work.

2 Particle Filtering for prognostics

Particle Filtering (PF) is a model-based method whose application in prognostics aims at inferring the evolution of an equipment degradation state on the basis of a sequence of noisy measurements; it relies on Bayesian methods to combine a prior distribution of the unknown state with the likelihood of the observations collected, to build a posterior distribution. This technique is widely used in prognostics since it allows to take into account even non-linear systems and/or non-Gaussian noises. The prediction of the equipment degradation state, $\underline{x}$, is performed by considering a set of N_s weighted particles, which evolve independently on each other, according to the probabilistic degradation model of Equation 1. The basic idea is that such set of weighted random samples constitutes a discrete approximation of the true Probability Density Function (pdf) of the system state $\underline{x}$ at time t. Typically, in a PF scheme, when a new measurement is collected, it is used to adjust the predicted pdf through the modification of the weights of the particles. Roughly speaking, the smaller the probability of encountering the acquired measurement value, when the actual component state is that of the particle, the larger the reduction of the particle's weight. On the contrary, a good match between the acquired measure and the particle state results in an increase of the particle importance. This requires the knowledge of the probabilistic law which links the state of

the component to the gathered measure (Equation 2). From this model, the probability distribution $P(z|\underline{x})$ of observing the sensor output z given the true degradation state $\underline{x}$ is derived (measurement distribution). This distribution is then used to update the weights of the particles upon a new measurement collection (for further details see [16], [18], [29]). Schemes of the PF algorithm are presented in Figures 1 and 2.



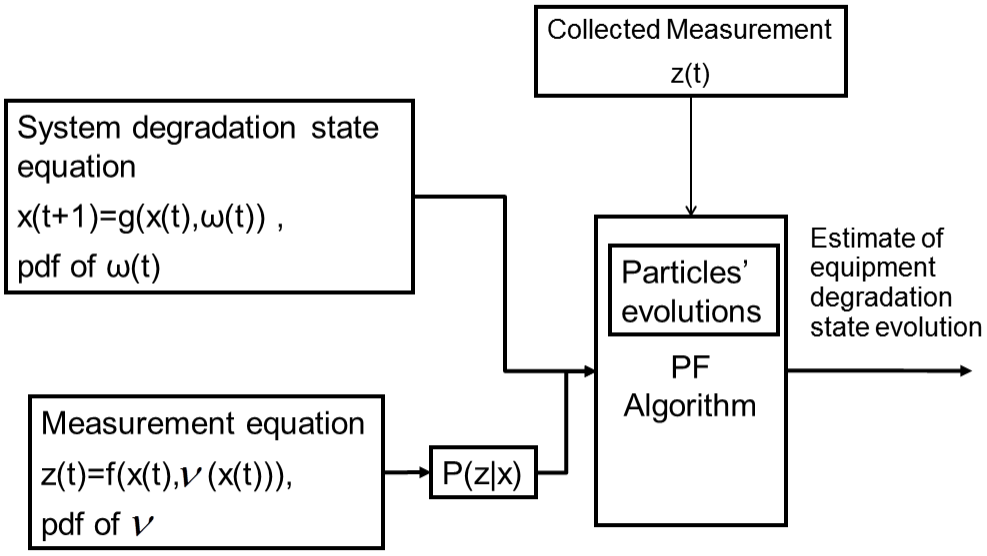


Figure 2: Information flowchart for a PF algorithm

3 Substitution of the measurement model with a bagged ensemble of ANN

In this Section, we propose a method to estimate the pdf $P(z|\underline{x})$ of the measurement z in correspondence of a given equipment degradation state $\underline{x}$, in the case in which the analytical form of the measurement model is not known. The method, derived from [27] and [30], requires the availability of a dataset made of $N_{training}$ couples $(\underline{x}_n, z_n)$. The estimated pdf $\hat{P}(z|\underline{x})$ will be used in the PF scheme according to the scheme of Figure 3.

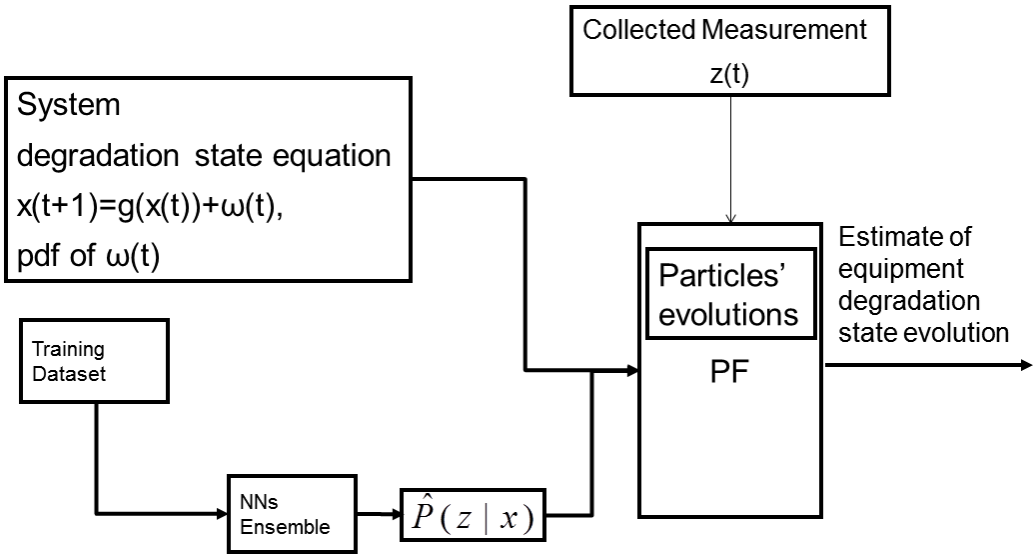


Figure 3: Information flowchart for the modified PF algorithm

Behind the method there is the hypothesis that the measurement model, which is unknown, can be written in the form:

$$z(\underline{x}) = f(\underline{x}) + \nu(\underline{x}) \tag{3}$$

where $f(\underline{x})$ is a biunivocal mathematical function and the measurement noise $\nu(\underline{x})$ is a zero mean Gaussian noise.

The method is based on the use of a bagged ensemble of ANNs, which are employed to build an interpolator $\varphi(\underline{x})$ of the available training patterns $T = \left\{(\underline{x}_n, z_n), n = 1., ..., N_{training}\right\}$. Since the obtained ANNs outputs depend on the training set, ANNs may suffer from instability. To overcome this issue, the use of a bagged ensemble of ANNs has been proposed (e.g., [28], [31]): each base interpolator of the ensemble is trained on different distributions of data according to a bootstrap technique. The key idea is to treat the available dataset as if it were the entire population, and then to create a number B of alternative versions $\{T_b^*\}_{b=1}^B$ of T, by randomly sampling from it with replacement (i.e., every sample is returned to T after sampling so that a pattern could appear multiple times in the same bagged set T_b^*). Using these training sets, the networks $\{\varphi_b(\underline{x}; T_b^*)\}_{b=1}^B$ are built and the output $\varphi_{avg}(\underline{x})$ of the bagged ensemble is obtained by averaging the single ANN output

according to:

$$\varphi_{avg}(\underline{x}) = \frac{1}{B}\sum_{b=1}^{B}\varphi(\underline{x};T_b^*) \tag{4}$$

Empirical studies have established that bagging is a simple and robust method that generally increases the accuracy of a single learner [28], [31].

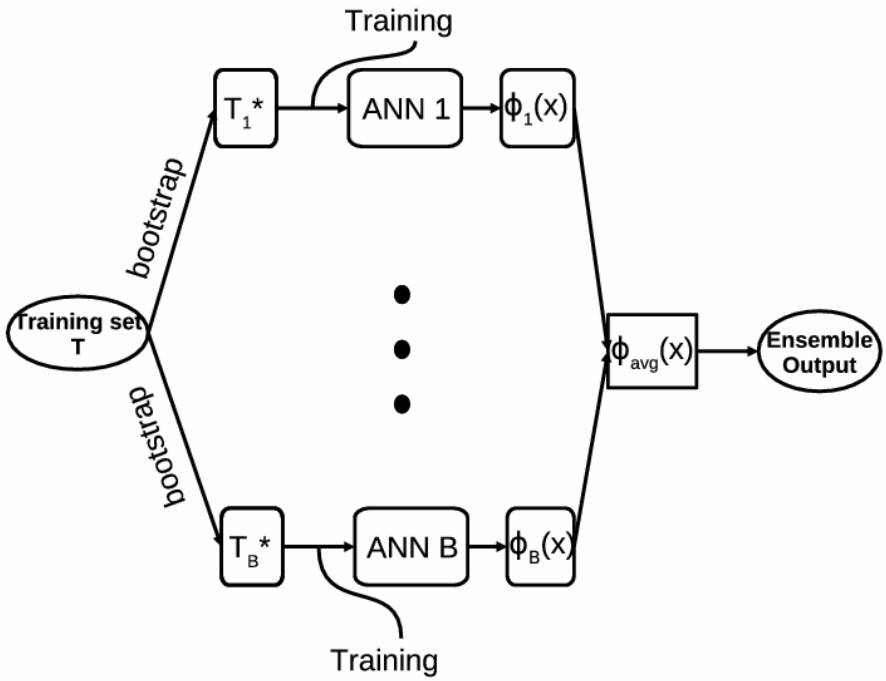


Figure 4: Scheme of a bagging ensemble of ANNs

On the other hand, since PF requires the knowledge of the pdf $P(z|\underline{x})$, the estimate of $f(\underline{x})$ does not suffice to apply PF. In this respect, the procedure proposed in [27] allows to estimate the pdf $P(z|f(\underline{x}))$ from which the pdf $P(z|\underline{x})$ can be obtained, being the function f invertible for hypothesis. The procedure is based on the subtraction of the random quantity $\varphi_{avg}(\underline{x})$ to both sides of Equation 3:

$$z(\underline{x}) - \varphi_{avg}(\underline{x}) = [f(\underline{x}) - \varphi_{avg}(\underline{x})] + \nu(\underline{x}) \tag{5}$$

The left-hand side of Equation 5 is a random variable which represents the error of the ensemble output $\varphi_{avg}(\underline{x})$ with respect to the measurement $z(\underline{x})$. This random error is made up of two contributions (right hand side of Equation 5):

1. The random difference $f(\underline{x}) - \varphi_{avg}(\underline{x})$ between the unknown deterministic quantity $f(\underline{x})$

and the ensemble output $\varphi_{avg}(\underline{x})$. This quantity is a random variable, being $\varphi_{avg}(\underline{x})$ dependent on the random training set T_b, $b=1,...,B$, i.e. different training sets would lead to different ensemble models and thus to different output $\varphi_{avg}(\underline{x})$. Since $f(\underline{x})-\varphi_{avg}(\underline{x})$ can be seen as the model error, its variance will be referred as model error variance and indicated by $\sigma_m^2(\underline{x})$.

2. The intrinsic noise $\nu(\underline{x})$ of the measurement process, whose variance is indicated by $\alpha^2(\underline{x})$.

These two contributions are estimated by means of the procedures described in the following Sections 3.1 and 3.2.

3.1 Distribution of the model error variance

The procedure here used to estimate the distribution $P(\varphi_{avg}(\underline{x})|f(\underline{x}))$ of the ensemble output $\varphi_{avg}(\underline{x})$ given the true value of $f(\underline{x})$, is based on the assumption that the random variable $f(\underline{x})-\varphi_{avg}(\underline{x})$ is gaussian with zero mean and standard deviation $\sigma_m(\underline{x})$, which entails that $P(\varphi_{avg}(\underline{x})|f(\underline{x}))$ is gaussian with mean $f(\underline{x})$. Notice that residual errors in the output of the ANN are usually not caused by variance alone; rather, there may be biases in the output of the ANN, which invalidate the assumption that the mean of the distribution is zero. However, it is generally accepted that the contribution of the variance in the residual error of the ANN dominates that of the bias (see [33] for further details on this). Furthermore, the bias in the output of an ensemble of NNs is expected to be smaller than that of the single ANN.

In order to estimate the model error variance $\sigma_m^2(\underline{x})$, the technique in [27] requires to divide the B networks of the ensemble $\varphi_{avg}(\underline{x})$ into M smaller sub-ensembles, each one containing K networks, and to consider the output $\varphi_{com}^m(\underline{x})$, $m=1,...,M$ of each sub-ensemble as:

$$\varphi_{com}^m(\underline{x})=\frac{1}{K}\sum_{k=1}^{K}\varphi_k(\underline{x}) \qquad (6)$$

The set $\zeta = \{\varphi_{com}^{m}(\underline{x})\}_{m=1}^{M}$ constitutes a sampling of M values from the distribution $P(\varphi_{com}(\underline{x})|\varphi_{avg}(\underline{x}))$ and its sample variance $\hat{\sigma}_{m}^{2}(\underline{x})$ could be used to approximate the unknown variance $\sigma_{m}^{2}(\underline{x})$ of the ensemble output. Notice that the idea behind this procedure is that by estimating $f(\underline{x})$ with $\varphi_{avg}(\underline{x})$, one can approximate $P(\varphi_{avg}(\underline{x})|f(\underline{x}))$ by $P(\varphi_{com}(\underline{x})|\varphi_{avg}(\underline{x}))$. In order to improve the reliability and stability of $\hat{\sigma}_{m}^{2}(\underline{x})$, bagging is also performed on the values of ζ. Thus, P bagging re-sampled sets of ζ are gathered:

$$\Gamma = \{\zeta_{p}^{*}\}_{p=1}^{P} \tag{7}$$

where ζ_{p}^{*} is the p-th subset containing M values of $\varphi_{com}(\underline{x})$, sampled with replacement from ζ. For any subset ζ_{p}^{*}, $p=1,...,P$, the corresponding variance $\sigma_{p}^{2*}(\underline{x})$ is computed; then, the estimate $\hat{\sigma}_{m}^{2}(\underline{x})$ of the variance $\sigma_{m}^{2}(\underline{x})$ is calculated as their average value:

$$\hat{\sigma}_{m}^{2}(\underline{x}) = \frac{1}{P}\sum_{p=1}^{P}\sigma_{p}^{2*}(\underline{x}) \tag{8}$$

Finally, the estimate of the regression distribution proposed by the method is:

$$P(\varphi_{avg}(\underline{x})|f(\underline{x})) \approx N(\varphi_{avg}(\underline{x}), \hat{\sigma}_{m}^{2}(\underline{x})) \tag{9}$$

3.2 Distribution of the measurement noise

In this Section, the technique proposed in [30] is applied to estimate the variance $\alpha^{2}(\underline{x})$ of the Gaussian zero mean noise $\nu(\underline{x})$ affecting the measurement equation (Equation 3).

From Equation 5, one can derive:

$$\begin{aligned} &Var[z-\varphi_{avg}(\underline{x})] = Var[f(\underline{x})-\varphi_{avg}(\underline{x})] + Var[\nu(\underline{x})] + 2E\{[f(\underline{x})-\varphi_{avg}(\underline{x})]\nu(\underline{x})\} \\ &= \sigma_{m}^{2}(\underline{x}) + \alpha^{2}(\underline{x}) \end{aligned} \tag{10}$$

The last equality is due to the independence of the error $[f(\underline{x})-\varphi_{avg}(\underline{x})]$ from the measurement noise $\nu(\underline{x})$. To explain this, notice that $[f(\underline{x})-\varphi_{avg}(\underline{x})]$ depends on the noise values ν_{n} affecting the measures $z_{n} = f(\underline{x}_{n}) + \nu_{n}$, $n=1.,...,N_{training}$, in the training data $T = \{(\underline{x}_{n}, z_{n}), n=1.,...,N_{training}\}$, which are used to build the ensemble model $\varphi_{com}^{m}(\underline{x})$, whereas $\nu(\underline{x})$ is the value of the noise affecting the

measure of the test data $\underline{x}$, not used for training the model. Thus, $\nu_n \quad n=1.,...,N_{training}$, and the values sampled from $\nu(\underline{x})$ in the test data are different, independent realizations of the same random variable. Notice also that $\nu^2(\underline{x})$ obeys a Chi-square $\chi^2(\underline{x})$ distribution with 1 degree of freedom.

The term $\sigma_m{}^2(\underline{x})$ can be estimated according to the procedure illustrated in the previous Section 3.1, whereas, being $z(\underline{x})-\varphi_{avg}(\underline{x})$ a zero mean random variable, its variance is given by:

$$Var[z(\underline{x})-\varphi_{avg}(\underline{x})] = E[(z(\underline{x})-\varphi_{avg}(\underline{x}))^2]$$

Notice that in correspondence of the training couples $(\underline{x}_n, z_n)$, $n=1,...,N_{training}$, one can approximate $E\left[\left(z(\underline{x})-\varphi_{avg}(\underline{x})\right)^2\right]$ by $\left(z(\underline{x})-\varphi_{avg}(\underline{x})\right)^2$ and thus, according to Equation 10, a dataset can be obtained, which is made up of the couples $(x_n, \hat{\alpha}_n^2)$, $n=1,...,N_{training}$, where:

$$\hat{\alpha}_n^2 = max\{(z_n-\varphi_{avg}(\underline{x}_n))^2-\hat{\sigma}^2(\underline{x}_n),0\} \qquad (11)$$

Finally, in order to estimate $\alpha^2(\underline{x})$ for a generic $\underline{x}$, a single ANN is trained using the dataset $(\underline{x}_n, \hat{\alpha}_n^2)$, $n=1,...,N_{training}$.

3.3 Estimate of the measurement distribution *P(z|x)*

Being $\varphi_{avg}(\underline{x})$ an estimate of $f(\underline{x})$, the measurement distribution $P(z\,|\,f(\underline{x}))$ can be approximated by the distribution $P(z\,|\,\varphi_{avg}(\underline{x}))$ which can be derived from the distribution $P(\varphi_{avg}(\underline{x})\,|\,f(\underline{x}))$ and the distribution of the measurement noise $\nu(x)$, according to Equation 5. Since these two distributions are both Gaussian, with means and variances estimated as shown in Sections 3.1 and 3.2, $P(z\,|\,f(\underline{x}))$ is approximated by a Gaussian distribution with mean $\varphi_{avg}(\underline{x})$ and variance $\hat{\sigma}_m^2(\underline{x})+\hat{\alpha}^2(\underline{x})$. Finally, being $f(\underline{x})$ invertible, the distribution $P(z\,|\,\underline{x})$ of the measurement z in correspondence of a given state $\underline{x}$, is given by:

$$P(z\,|\,\underline{x}) \approx P(z\,|\,f(\underline{x})) \approx N(\varphi_{avg}(\underline{x}), \hat{\sigma}_m^2(\underline{x})+\hat{\alpha}^2(\underline{x})) \qquad (12)$$

4 Case study

4.1 Description of the case study

In this Section, we apply the technique described in Section 3 for estimating the measurement distribution $P(z \mid \underline{x})$ to a case study dealing with the crack propagation phenomenon in a component subject to fatigue load. The system state is described by the vector $\underline{x}(t) = (x_1(t), x_2(t))$, whose first element, $x_1(t)$, indicates the crack depth whereas the second element, $x_2(t)$, represents a time-varying model parameter that directly affects the crack growth rate. The evolution of this degradation process is described by the following two equations, which form a markovian system of order one:

$$x_1(t+1) = x_1(t) + 3 \cdot 10^{-4}(0.05 + 0.1 \cdot x_2(t))^3 + \omega_1(t) \quad (13)$$

$$x_2(t+1) = x_2(t) + \omega_2(t) \quad (14)$$

where $\omega_1(t)$ is a Gaussian noise with mean 0.045 and standard deviation 0.116, and $\omega_2(t)$ is a zero mean Gaussian noise with standard deviation 0.010.

In the present case study, it is assumed that a measurement equation is unavailable whereas a dataset formed by the $N_{training}$ pairs $(x_{1,n}, z_n)$ $n = 1, ..., N_{training}$, is available, where the subscript 1 refers to the first element of vector $\underline{x}$(t). With the purpose of showing the feasibility of the proposed approach, the dataset $T = \left\{(x_{1,n}, z_n), n = 1, ..., N_{training}\right\}$ has been artificially generated by simulating the degradation process $\underline{x}(t)$, and sampling the date from the probabilistic measurement model

$$z(t) = f(x_1) + v(x_1) = x_1(t) + 0.25 + v(x_1) \quad (15)$$

where $v(x_1)$ is a zero mean Gaussian noise, whose standard deviation depends on x_1:

$$Std[v(x_1)] = -\frac{1}{120}x_1^2 + \frac{1}{10}x_1 + \frac{1}{2} \quad (16)$$

According to Equation 15, the function $f(\underline{x})$ is given by $x_1 + 0.25$, which is, as required by the method, an invertible function. These degradation and measurement models have been derived from

[18].

Notice that the probabilistic measurement model of Equation 15 has been intentionally taken simple, as the main interest of our work is the quantification of the uncertainty in the RUL prediction and not the ability of the ensemble in reproducing the measurement model. In this respect, the knowledge of the variance of the measurement noise is fundamental, as it determines the amplitude of the prediction intervals of the RUL estimates. Thus, it is the capability of correctly reconstructing the variance that plays a key role in the assessment of the potential of the proposed technique.

4.2 Estimate of the measurement distribution

According to the technique illustrated in Section 3, an ensemble of $B = 200$ ANNs has been built using the available dataset $T = \left\{(x_{1,n}, z_n), n = 1,..., N_{training}\right\}$, where $N_{training} = 1000$. Every ANN has 5 tan-sigmoidal hidden neurons and one linear output neuron. To estimate $\sigma_m^2(\underline{x}) = \sigma_m^2(x_1)$, the ensemble has been divided into $M = 20$ sub-ensembles, and $P = 1000$ bagging resamples of the sub-ensemble outputs $\varphi_{com}^m(\underline{x}) = \varphi_{com}^m(x_1)$ have been considered, $m = 1,..., M$.

The ANN used to estimate the noise variance $\alpha^2(\underline{x})$ is characterized by 5 tansigmoidal hidden neurons and one linear output neuron.

The results are evaluated in terms of the following performance indicators, which are computed by considering a set of $N_{test} = 1000$ couples $(x_{1,i}, z_i)$, $i = 1,..., N_{test}$ obtained from Equations 15 and 16:

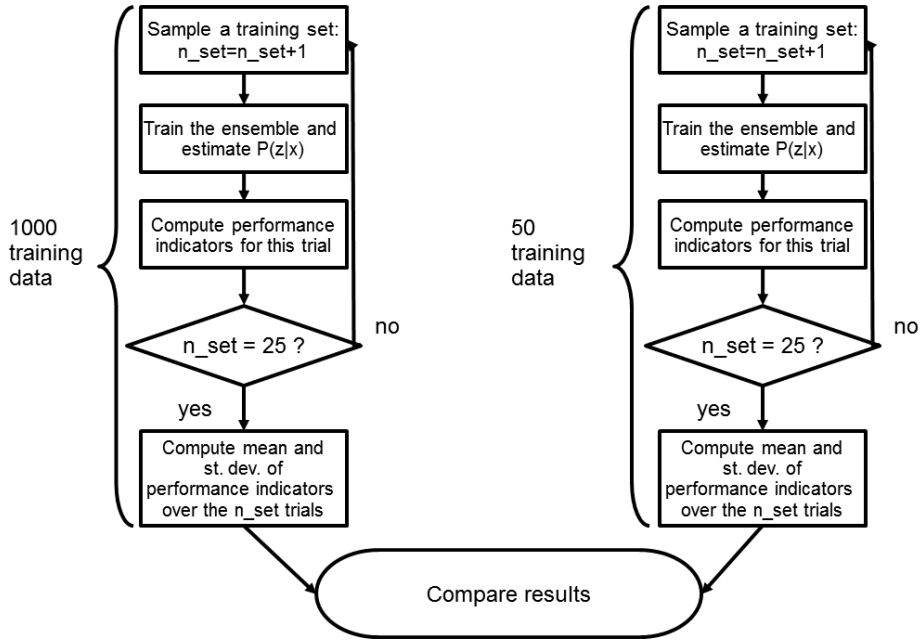


Figure 5: Scheme of the proposed cross-validation

1. The square bias $\bar{b}^2$; i.e., the average quadratic difference between the true value of $f(x_1)$ and the ensemble estimate of this quantity $\varphi_{avg}(x_1)$:

$$\bar{b}^2 = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} (f(x_{1,i}) - \varphi_{avg}(x_{1,i}))^2 \qquad (17)$$

This value gives information on the accuracy of the estimate of $f(\underline{x}) = f(x_1)$ provided by the ensemble. Notice that the computation of this indicator requires the knowledge of the function $f(x_1)$, which is not available if the measurement equation (Equation 15) is not known. Thus, in general one can only compute:

$$MSE = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} (\varphi(x_{1,i}) - z_i)^2 \qquad (18)$$

Small values of *MSE* indicate satisfactory performance of the ensemble.

2. The coverage of the Prediction Interval (PI) with confidence 0.68. This indicator is used to verify the accuracy of the estimate of the distribution $P(z \mid \underline{x})$. A PI with a confidence level γ_p is defined as a random interval in which the observation $z(\underline{x}) = z(x_1)$ falls with probability

γ_p ([27], [32], [34]):

$$P(z(x_1) \in PI_{\gamma_p}(x_1)) = \gamma_p \tag{19}$$

The PI with $\gamma_p = 0.68$ is given by:

$$\varphi_{avg}(x_1) - \sqrt{\hat{\sigma}_m^2(x_1) + \hat{\alpha}^2(x_1)} \le z(x_1) \le \varphi_{avg}(x_1) + \sqrt{\sigma_m^2(x_1) + \alpha^2(x_1)} \tag{20}$$

In order to verify whether the estimate of $P(z \mid x_1)$ provides a satisfactory approximation of the true pdf, we will consider how many times the measurement z_i falls within the $PI_{\gamma_p=0.68}(x_{1,i})$. The closer to γ_p is the portion of data inside the γ_p-confidence interval, the more accurate is the estimation of the parameters of the Gaussian pdf. A graphical representation for the verification procedure is shown in Figure 6.

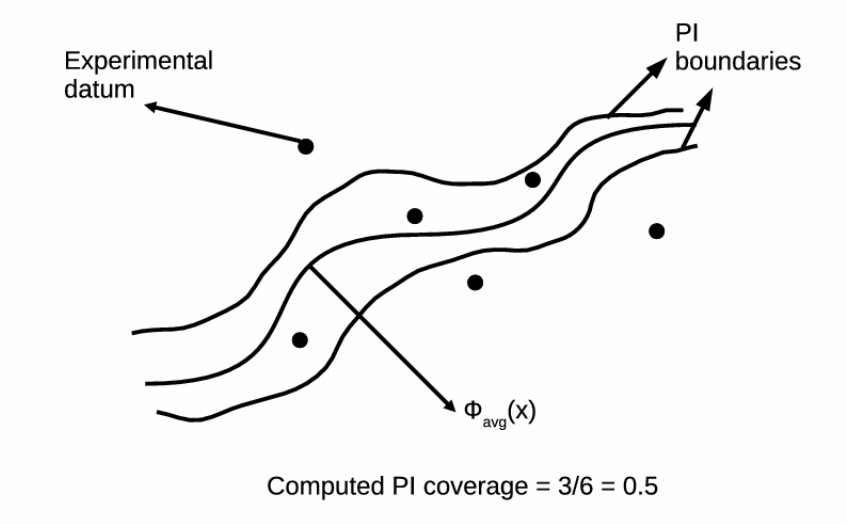


Figure 6: Verification of the coverage for PI

In order to avoid over/under estimating the performance indicators $\bar{b}^2$ and coverage, cross-validation of the results has been done by repeating the computations with $N_{set} = 25$ different, randomly generated training and test sets. Moreover, to verify the sensibility of the methodology to the cardinality $N_{training}$ of the training set, two cases have been considered: case 1 considers $N_{training} = 1000$ couples $(x_{1,n}, z_n)$, $n = 1,...,N_{training}$ to build the ensemble, whereas there are only $N_{training} = 50$ couples for case 2. Table 1 reports means and standard deviations of the performance indicators over the 25 cross-validations in the two cases.

training data	1000	1000	50	50
model	Ensemble	1 ANN	Ensemble	1 ANN
$\bar{b}^2$	0.0040 ± 0.0015	0.0097 ± 0.0060	0.0793 ± 0.0395	0.3877 ± 0.3441
PI coverage	0.6758 ± 0.0366	-	0.5488 ± 0.1031	-

Table 1 Performance indicators in case of 1000 and 50 training data over 25 cross-validations; the mean ± std is reported

Notice that when 1000 training couples $(x_{1,n}, z_n)$, $n = 1,...,N_{training}$ are available, the ensemble output $\varphi_{avg}(x_1)$ is very accurate in the prediction of the function $f(x)$, the bias being very small. Furthermore, notice that the ensemble outperforms a single ANN trained with all the 1000 training patterns. With respect to the estimate of the distribution $P(z|x_1)$, the proposed method provides a satisfactory approximation, being the coverage very close to 0.68.

Using the smaller training set made by 50 training patterns, the accuracy of the $P(z|x_1)$ estimate decreases. Notice, however, that the prediction of the $f(\underline{x})$ value provided by the ensemble is still more satisfactory than that of one single ANN.

Table 2 reports the estimates of the two contributions $\bar{\sigma}^2$ and $\bar{\alpha}^2$ of the variance of the estimated measurement distribution $\hat{P}(z|x_1)$. Notice that in this case study, $\bar{\sigma}^2$ is negligible with respect to the variance $\bar{\alpha}^2$ of the measurement noise. Thus, the accuracy of the estimate of the PI is more sensible to the estimate of $\bar{\alpha}^2$.

	1000 training data	*50 training data*	Real value
Var (P(z|x))= $\bar{\sigma}^2 + \bar{\alpha}^2$	0.4489 ±0.1359	0.4955 ±0.0621	-
$\bar{\sigma}^2$	0.0243 ±0.0317	0.0005 ±0.0001	-
$\bar{\alpha}^2$	0.4886 ±0.0276	0.4095 ±0.1642	0.4900

Table 2: Contributions to the $P(z|x)$ variance

In this respect, Figures 7 and 8 show the estimate of $\alpha^2(x_1)$ in the two cases of training sets formed by $N_{training} = 1000$ and $N_{training} = 50$ $(x_{1,n}, z_n)$, $n = 1,...,N_{training}$, pairs, respectively. The estimated values are compared with the true $\bar{\alpha}^2$ value provided by Equation 16. Notice that this comparison, which is done in this work to assess the performance of the methodology, is not possible in real industrial applications if the measurement model (Equations 15 and 16) is not available.

Finally, the larger the cardinality of the training dataset, the better the approximation of $\alpha^2(x_1)$. This is due to the fact that the ANN built to interpolate $\alpha(x_1)$ (reported with dots in Figures 7 and 8) is trained with much more patterns in the first case.

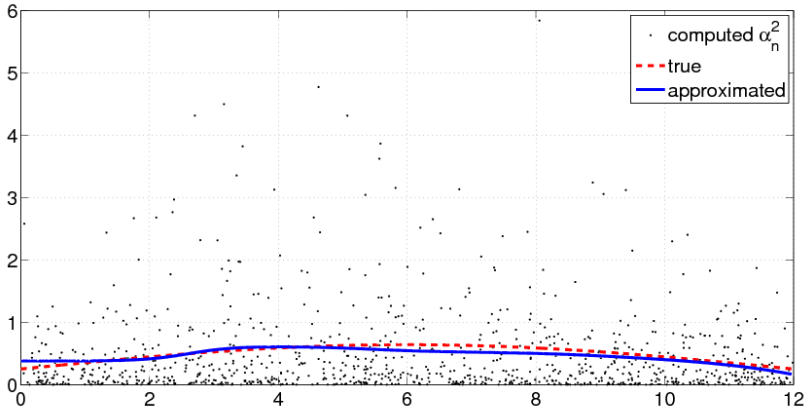


Figure 7: True and approximated measurement noise variance $\alpha^2(x_1)$ (1000 training data)

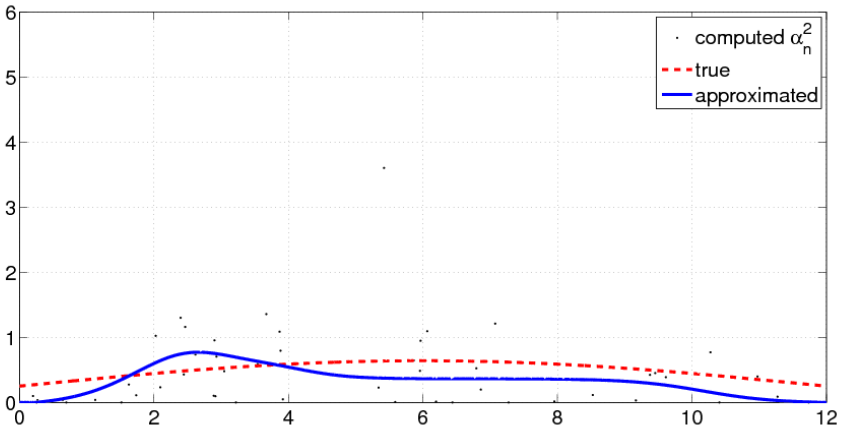


Figure 8: True and approximated measurement noise variance $\alpha^2(x_1)$ (50 training data)

4.3 Crack depth prediction

The objective of this Section is to evaluate the performance of the overall PF scheme in the prediction of the crack depth evolution when the ensemble of ANNs is used to estimate the measurement distribution $P(z \mid x_1)$. To this purpose, the problem tackled consists in predicting the future crack propagation starting from time $t = 80$ in arbitrary units, on the basis of eight measurements of the crack depth taken at time $t_m = 10 \cdot m$, $m = 1,...,8$. The PF has been run and the particles' weights after the collection of the last measurement ($z = 4.6087$ in arbitrary units at $t = 80$

) have been recorded; the particles' weights have been updated by using the estimate of the distribution $P(z|x_1)$ obtained in the previous Section in the two different cases of $N_{training}=50$ and $N_{training}=1000$ training data available. Figure 9 compares the estimates of the pdf $P(z|x_1)$ at time $t=80$ provided by the two ensembles with the distribution which would be obtained by using the true measurement equation for the particle with the highest difference between the weights computed through the analytical measurement model and through the considered technique (with 1000 training data). Table 3 reports the most significant numerical values for the comparison: the means and the standard deviations of the distributions and the weight w assigned to the particle upon the Bayesian update (before this, the weight w is the same in the three cases, since the resampling algorithm has been adopted, see [16]). Notice that the ensemble trained with 50 data provides a distribution characterized by variance lower than the analytical, and thus the weight assigned to the particle ($x_1=5.93$) is higher. This result confirms that the size of the training set impacts on the accuracy of the estimate of $P(z|x_1)$.

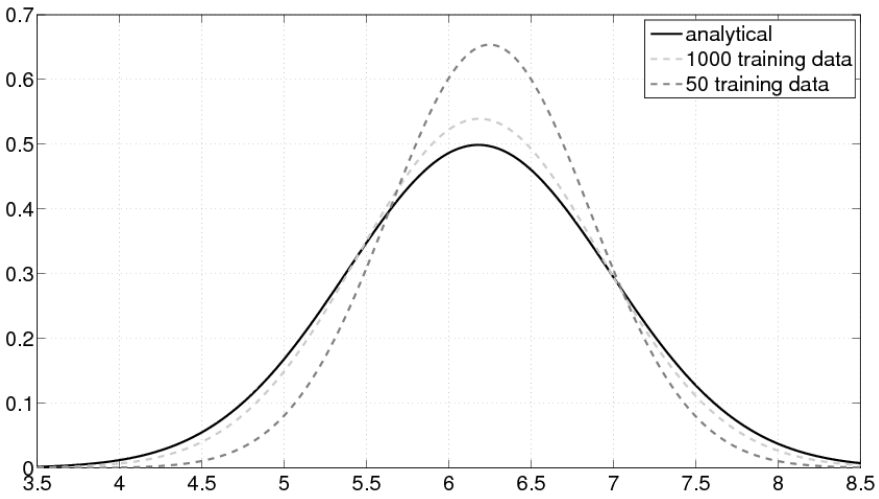


Figure 9: Comparison of the distributions $P(z|x)$ for the three cases under analysis

	analytical	1000 training data	50 training data
mean	6.1800	6.1864	6.2484
std. dev.	0.7997	0.7398	0.6103
w	0.0725	0.0555	0.0177

Table 3: Comparison of the parameters of the distributions $P(z|x_1)$ and of the corresponding weights for the three cases

under analysis

Figure 10 shows the prediction of the crack depth evolution performed at $t=80$, after the last measurement has been acquired, by using the two ensemble models to estimate $P(z|x_1)$. These predictions have been compared to that which would be obtained by directly using the measurement equation in the PF. Since PF provides the estimate of the pdf of the crack depth on the basis of the available measurements, the expected values of the obtained distributions are reported. Notice that, due to the stochastic behavior of the crack, even if the measurement equation were known, the prediction of the crack depth would be different from its true evolution. Furthermore, the accuracy of the PF prediction using the ensemble of ANNs is influenced by the number of available training patterns.

Notice also that the linearity of the prediction of the expected value of x_1 can be explained by averaging Equations 13 and 14:

$$E[x_2(t+1)\} = E[x_2(t)] + E[\omega_2(t)] = E[x_2(t)] = \text{constant}$$

$$E[x_1(t+1)] - E[x_1(t)] = 3\cdot 10^{-4}(0.05 + 0.1\cdot E[x_2(t)])^3 + E[\omega_1(t)]$$

$$E[x_1(t+1)] - E[x_1(t)] = \text{constant}$$

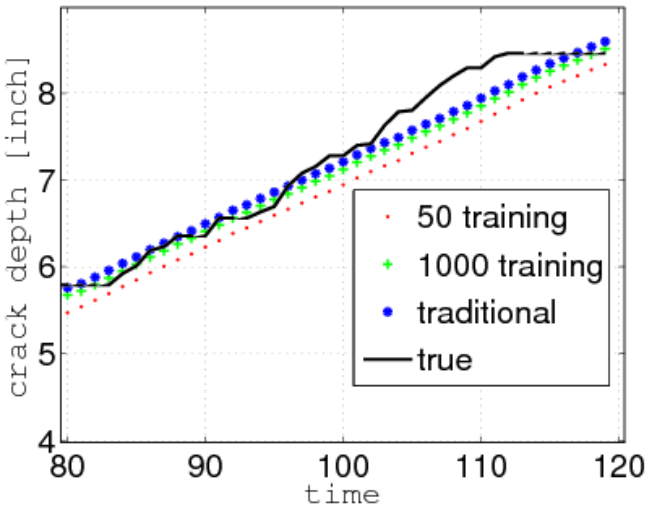


Figure 10: Comparison of the PF predictions with the true state evolution considering different cardinalities of the training set.

To evaluate the impact of replacing the measurement equation with the ensemble of ANNs, $N_{run}=100$ different degradation trajectories have been simulated and the PF predictions of the crack

depth have been performed. Also in this case, three different PF predictions have been performed: by considering for the estimate $P(z|x_1)$ the ensemble of ANNs trained with $N_{training}=1000$ patterns, $N_{training}=50$ patterns and the measurement equation. Each PF run is characterized by the same true trajectory, the same acquired measures and the same state noise vector. The following performance indicators have been computed:

1. The coverage of the PI, with confidence 0.68. In particular, the predictions, provided by the PF, of the crack depth at $t=100$ and $t=120$ have been considered. At each run the boundaries of the PI are computed by considering the 16th and 84th percentiles of the estimate of the pdf of the crack depth. A counter is set to 1 or 0 if the true trajectory belongs or not to the corresponding interval, in analogy with the coverage verification explained in Section 4.2.
2. The average width over the $N_{run}=100$ runs of the PI at $t=100$ and $t=120$.
3. The Mean Square Error (MSE) over the $N_{run}=100$ runs between the prediction of the crack depth provided by the PF and its true value at $t=100$ and $t=120$. Considering for example $t=100$, the MSE_{100} is given by:

$$MSE_{100}=\frac{1}{N_{run}}\sum_{n_{run}=1}^{N_{run}}\left(X_{n_{run}}-o_{n_{run}}\right)^2 \tag{21}$$

where $X_{n_{run}}$ is the true crack depth in the test trajectory at $t=100$ and $o_{n_{run}}$ is the expected value of the crack depth pdf estimated by the PF.

A scheme of the method used for the computation of these performance indicators is shown in Figure 11; the obtained values are reported in Tables 4 and 5.

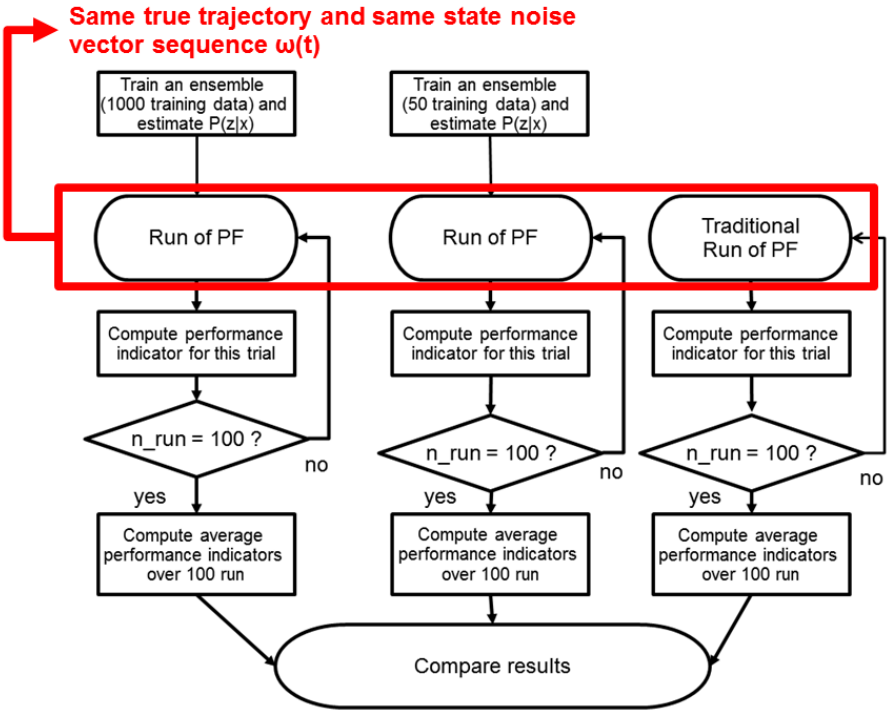


Figure 11: Scheme for the evaluation of the prognostics performance

	traditional	1000 training data	50 training data
coverage	0.6600	0.6200	0.5900
PI width	1.0812	1.0862	0.9605
MSE	0.3125	0.3167	0.3245

Table 4: Performance indicators at *t=100*

	traditional	1000 training data	50 training data
coverage	0.6500	0.7000	0.6000
PI width	1.3058	1.3226	1.2088
MSE	0.3421	0.3464	0.3641

Table 5: Performance indicators at *t=120*

It can be noticed that the coverage of the ensemble trained with 1000 training data is very close to 0.68; furthermore, even the other performance indicators are very close to those which would be obtained by considering the measurement equation. This result confirms that when the size of the training set is sufficiently large, the approximation of the distribution $P(z \mid x_1)$ is accurate and therefore it does not remarkably alter the outcome of the PF. On the contrary, the performance

indicators obtained by considering the ensemble trained with 50 data are less satisfactory; this is due to the worse estimate of the $P(z|x_1)$ provided by the ensemble. In particular, the reduction of the PI width is due to the underestimate of the variance of $P(z|x_1)$ which results in a more peaked particles' distribution when a measurement is collected.

In conclusion, the size of the training dataset has an influence both on the approximation of the measure distribution $P(z|x_1)$ and on the prognostics performance of the PF.

Finally, the performance evaluator proposed in [35] has been computed to evaluate the prediction performance:

$$s = \begin{cases} \sum_{i=1}^{n} e^{-\left(\frac{d_i}{a_1}\right)} - 1 & if\ d_i < 0 \\ \sum_{i=1}^{n} e^{-\left(\frac{d_i}{a_2}\right)} - 1 & otherwise \end{cases}$$

where a_1= 10, a_2=13, n=100 is the number of simulated histories and d is the difference between the estimated RUL and its true value. To compute the value of this performance metric, the following procedure has been adopted:

1. Set the failure threshold to S_T=7.
2. Simulate the evolution of the degradation process; this allows calculating the true value of the true RUL t_{RUL} at t=80 as the difference between the time instant at which the component achieves ST and 80. Moreover, the set of measures sampled according to the measurement model are collected.
3. Use the PF to estimate the component degradation state at t=80 and predict the RUL $\hat{t}_{RUL}$.
4. Calculate the difference $d = \hat{t}_{RUL} - t_{RUL}$.
5. Perform n-1 times the steps 2-4 and compute s.

The values of the performance indicators obtained in the case in which the RUL is predicted by using the 'traditional' PF approach (s=10.30) and the 'data–driven' approach (s=10.65) turn out to be very close to each other.

5 Conclusions

PF is often proposed as prognostic technique for estimating the evolution of the degradation state x of a system; it resorts to analytical models of both degradation state evolution and measurement. This latter is a probabilistic relation between the true degradation state and the corresponding output z of the measurement sensor, and is used to update, within a Bayesian framework, the prediction of the evolution of the degradation state, upon the acquisition of a measure. In practice, the measurement model may not be available in an analytical form; rather, there may be available a set of data which allows, through data-mining techniques, to build the measurement model. In this work, a technique based on an ensemble of ANNs has been investigated to this aim and applied to a case study derived from the literature. The verification conducted on the results shows that, when the training set is sufficiently large, a good approximation of the model may be obtained and its substitution in the PF does not significantly affect its performance.

Additional effort will be dedicated in future works to improve the accuracy of the estimate when only a small training set is available and to extend the applicability of the technique also in those cases in which $f(\underline{x})$ is not biunivocal. Moreover, one more future objective is the substitution also of the model of the evolution of the system state for a data-driven model, like an ensemble of ANNs trained on the basis of an available dataset, with a procedure which may be inspired by the one presented in this work, in order to allow the usage of the PF in those cases where also an analytical model of the evolution of the system is unavailable.

6 References

[1] Jarrell, D., Sisk, D., and Bond, L. (2004). Prognostics and Condition-Based Maintenance: A New Approach to Precursive Metrics. Nuclear Technology 145 (3), 275-286.

[2] Jardine, A., Lin, D., and Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition based maintenance. Mechanical Systems and Signal Processing, 20 (7), 1483-1510.

[3] Bishop, C. (1995). Neural Networks for Pattern Recognition. Oxford University Press.

[4] Worden, K., Staszewski, W., and Hensman, J. (2011). Natural computing for mechanical systems research: A tutorial overview. Mechanical Systems and Signal Processing, 25 (1), 4-111.

[5] Hsu, C., Chang, C., and Lin, C. (2003). A Practical Guide to Support Vector Classification. Technical Report. Department of Computer Science and Information Engineering, National Taiwan University, Taipei.

[6] Rasmussen, C., and Williams, C. (2006). Gaussian Processes for Machine Learning. MIT Press.

[7] Liu, W., Principe, J., and Haykin, S. (2010). Kernel Adaptive Filtering: A Comprehensive Introduction. John Wiley.

[8] Samanta, P., Vesely, W., Hsu, F., and Subudly, M. (1991). Degradation modeling with application to ageing and maintenance effectiveness evaluations. Technical report, NUREG/CR-5612. US Nuclear Regularoty Commission.

[9] Yan, J., Guo, C., and Wang, X. (2011). A dynamic multi-scale Markov model based on methodology for remaining life prediction. Mechanical Systems and Signal Processing, 25 (4), 1364-1376.

[10] Abernethy, R. (1996). The New Weibull Handbook, 2nd edn, North Palm Beach.

[11] Cox, D., and Oakes, D. (1984). Analysis of Survival Data. Chapman and Hall.

[12] Kopnov, V. (1999). Optimal degradation process control by two-level policies. Reliability Engeneering and System Safety, 66 (1), 1-11.

[13] Lam, C., and Yeh, R. (1994). Optimal maintenance policies for deteriorating systems under various maintenance strategies. IEEE Transactions on Reliability, 43 (4), 423-430.

[14] Roweis, S., and Ghahramani, Z. (1999). A unifying review of linear Gaussian models. Neural Computation, 11 (2), 305-345.

[15] Hamilton, J. (1994). Time Series Analysis. Princeton University Press.

[16] Arulampalam, M., Maskell, S., Gordon, N., and Clapp, T. (2002). A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. IEEE Transactions on Signal Processing, 50 (2), 174-188.

[17] Doucet, A., Freitas, J. d., and Gordon, N. (2001). Sequential Monte Carlo methods in practice. Springer-Verlag. Springer Series in Statistics for Engineering and Information Science. New York: Springer-Verlag.

[18] Orchard, M. (2007). A Particle Filtering-based Framework for On-line Fault Diagnosis and Failure Prognosis. PhD thesis, Georgia Institute od Technology.

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

[20] Sikorska, J., Hodkiewicz, M., and Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. Mechanical Systems and Signal Processing , 25 (5), 1803-1836.

[21] Gustaffson, F., and Saha, S. (2010). Particle filtering with dependent noise. In proceedings of the 13th Conference on Information Fusion (FUSION), 26-29 July 2010, Edinburgh, UK.

[22] Moura, M., Lins, I., Ferreira, R., Droguett, E., and Jacinto, C. (2011). Predictive maintenance policy for oil well equipment in case of scaling through support vector machines.Advances in Safety, Reliability and Risk Management - Proceedings of the European Safety and Reliability Conference, ESREL 2011 Troyes, France,pp. 503-507..

[23] Green, I., and Casey, C. (2005). Crack Detection in a Rotor Dynamic System by Vibration Monitoring. Journal of Engineering for Gas Turbines and Power, 127 (2), 425-436.

[24] Huang, R., Xi, L., Li, X., Richard Liu, C., Qiu, H., and Lee, J. (2007). Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. Mechanical Systems and Signal Processing, 21 (1), 193-207.

[25] Harris, T. (1991). Rolling Bearing Analysis (3rd ed.). John Wiley & Son, Inc.

[26] Seker, S., Ayaz, E., and Turkcan, E. (2003). Elman's recurrent neural network applications to condition monitoring in nuclear power plant and rotating machinery. Engineering Applications of Artificial Intelligence, 16, 647-656.

[27] Carney, J., Cunningham, P., and Bhagwan, U. (1999). Confidence and prediction intervals for neural network ensembles. International Joint Conference on Neural Networks, IJCNN '99, 10-16 July 1999, Washington, DC, 1215-1218.

[28] Breiman, L. (1996). Bagging Predictors. Machine Learning, 24, 123-140 .

[29] Cadini, F., Zio, E., and Avram, D. (2009). Model-based Monte Carlo state estimation for condition-based component replacement. Reliability Engineering and System Safety, 94 (3), 752-758.

[30] Nix, D., and Weigend, A. (1994). Estimating the mean and the variance of the target probability distribution. In Proceedins of IEEE International Conference on Neural Networks,World Congres on Computational Intelligence, 27 June-02 July 1994, Orlando, FL, vol. 1, 55-60.

[31] Breiman, L. (1999). Combining predictors. In Sharkey A.J.C., Combining Artificial Neural Nets-Ensemble and Modular Multi-net Systems. Springer, Berlin, 31-50.

[32] Heskes, T. (1997). Practical Confidence and Prediction Intervals. In, Advances in Neural Information Processing Systems 9, 176-182, Cambridge, MIT press.

[33] Stuart, G., Bienenstock, E. and Doursat, R. (1992) Neural Networks and the bias/variance dilemma. *Neural Computation*, Vol. 4(1), pp. 1-58.

[34] Montgomery, D., Runger, G., and Hubele, N. (2011). Engineering Statistics. John Wiley & Sons Ltd.

[35] Saxena, A., Goebel K, Simon, D., Eklund, N. (2008) Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation, in International Conference on Prognostics and Heath Management, PHM2008.