

Differential evolution-based multi-objective optimization for the definition of a health indicator for fault diagnostics and prognostics

P. Baraldi ^{a,*}, G. Bonfanti ^a, E. Zio ^{a,b}

^a *Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156 Milan, Italy*

^b *Chair on System Science and the Energetic Challenge, Fondation Électricité de France (EDF), Centrale Supélec, Université Paris Saclay, Grande Voie des Vignes, 92290 Chatenay-Malabry, France*

ABSTRACT

The identification of the current degradation state of an industrial component and the prediction of its future evolution is a fundamental step for the development of condition-based and predictive maintenance approaches. The objective of the present work is to propose a general method for extracting a health indicator to measure the amount of component degradation from a set of signals measured during operation. The proposed method is based on the combined use of feature extraction techniques, such as Empirical Mode Decomposition and Auto-Associative Kernel Regression, and a multi-objective Binary Differential Evolution (BDE) algorithm for selecting the subset of features optimal for the definition of the health indicator. The objectives of the optimization are desired characteristics of the health indicator, such as monotonicity, trendability and prognosability. A case study is considered, concerning the prediction of the remaining useful life of turbofan engines. The obtained results confirm that the method is capable of extracting health indicators suitable for accurate prognostics.

1. Introduction

Condition-based and predictive maintenance approaches can increase safety, minimize downtimes, and ensure mission completion and efficient production [45,9,15,71,5]. To enable this, it is necessary to estimate the current degradation state of the components and systems of interest, and predict the future evolution.

In recent years, diagnostic approaches for the identification of equipment degradation using signal measurements have been proposed and successful industrial applications have been performed [59,58,82]. Furthermore, prognostic approaches have been developed for the prediction of equipment remaining useful life (RUL), i.e. the amount of time that the equipment will continue to perform its function according to specifications [60–62,66,52,71].

Health Indicators (HIs) are introduced in fault diagnostic and prognostic approaches, to measure the amount of equipment degradation. Depending on the equipment and its degradation mechanism, the HI can be a signal directly measured or not.

* *Corresponding author.*

E-mail addresses: piero.baraldi@polimi.it (P. Baraldi), enrico.zio@polimi.it, enrico.zio@ecp.fr (E. Zio).

Examples of directly measurable degradation indicators are the length of a crack in a structure [44,8,55,40], the light output from fluorescent light bulbs or the thickness of a car tyre tread [15]. Measurement noise can affect the raw data obtained from the sensors, possibly obscuring the signal trend; for this reason, filtering techniques are applied to smooth the HI. For example, many approaches rely on Bayesian methods, e.g. Kalman and particle filters, to infer equipment degradation using signal measurements and a physics-based model of the degradation process.

Examples in which a direct measure of the equipment degradation is not available are bearing wear [81,82] and various other mechanisms such as stiction, wear, contamination and degradation of dielectrics in Micro-Electro Mechanical Systems (MEMS) [53]. In these cases, it is necessary to extract a HI from the available measurements, which are not directly related to the component degradation state. Although several HIs have been developed for fault diagnostics and prognostics of different types of industrial components, the solutions already proposed are typically not general, but tailored to the component and the characteristics of the monitored signals.

Typically, HI extraction is performed by (i) pre-processing the raw data to reduce the measurement and process noises and (ii) extracting degradation trends using statistical indicators in the time domain, such as mean and standard deviation, and other indicators in the frequency or time-frequency domains [46,33,4,10].

In [38], Empirical Mode Decomposition (EMD) is used to extract information about the evolution of degradation over time. The main idea is that a time-dependent signal can be described as fast oscillating components superimposed to slow oscillating ones. The residual of this approximation is the trend of the signal, whose variations can show the degradation evolution over time. In [13], an approach based on the use of the Auto-Associative Kernel Regression (AAKR) has been used for the extraction of health indicators. The AAKR empirical model receives in input the current signal value and reproduces in output the value that the signal is expected to have in healthy conditions, before the beginning of the degradation (signal reconstruction). The residual, i.e. the difference between the signal measured value and the AAKR reconstruction, defines the HI.

Once a set of features has been extracted from the raw measurements, it can be useful to combine them into a single degradation indicator, since individual features may contain only partial (and different) information about the equipment degradation state. In [13], a HI is defined as a linear combination of AAKR residuals of different signals. The coefficients of the linear combination are obtained by a genetic algorithm optimization whose objectives are the three metrics of *Monotonicity*, *Trendability* and *Prognosability*, which assess the capability of a HI of properly quantifying the equipment degradation. In [67], an approach based on Principal Component Analysis (PCA) and Hidden Markov Models (HMM) is proposed for the definition of a HI for prognostics.

The objective of the present work is to propose a systematic and general method for the construction of a HI. Since the approach should be applicable to signals obtained from industrial components of different nature, we consider various signal extraction techniques, including statistical indicators, EMD, wavelet and Fourier transforms.

The proposed method differs from previous literature approaches in the technique used for combining the extracted features. Differently from [13], which linearly combines multiple residuals obtained from several AAKR models, each one reconstructing a single measured signal, we define the HI as the residual of an AAKR model which reconstructs in a single step a group of features. The idea behind this approach is that features that individually do not have the characteristics of health indicators may provide useful information when considered jointly with other features. For example, a feature related to the component operating conditions can drive the computation of the residuals of features related to the component degradation and also influenced from the operating conditions.

A large number of features can be extracted from the measured signals and several different combinations (subsets) of them can be selected: then, the definition of a HI can be seen as the problem of selecting the best combination of features to be used. The approach that we explore in this work to address this problem is based on a multi-objective optimization that considers as objectives the metrics of *Monotonicity*, *Trendability* and *Prognosability*. The multi-objective optimization provides a number of Pareto-optimal solutions, which are non-dominated with respect to the considered objectives [10].

We resort to a Binary Differential Evolution (BDE) algorithm for the multi-objective optimization. The choice of BDE is due to the fact that it explores the decision space more efficiently than other multi-objective evolutionary algorithms [57], including Non-dominated Sorting Genetic Algorithm II (NSGA-II) [16], Strength Pareto Evolutionary Algorithm 2 (SPEA2) [72] and Indicator Based Evolutionary Algorithm (IBEA) [74].

The main novelties of the proposed approach are:

- the use of EMD as a feature extraction method and not for directly building a HI, as in [38];
- the use of an AAKR model to combine different features into a HI;
- the BDE multi-objective search for the selection of the optimal HI.

An application is considered, concerning the prediction of the RUL of a fleet of turbofan engines working under variable operating conditions. Data describing the evolution of 21 signals during the engine run-to-failure trajectories have been taken from the NASA Ames Prognostic Data Repository [50]. The obtained results have been compared to those obtained by (1) directly considering one of the measured features as HI, (2) applying the method in [38], where the EMD features are directly used as HIs.

In Section 2, we state the problem and introduce the three metrics of *Monotonicity*, *Trendability* and *Prognosability* used to assess the goodness of the HI. In Section 3, we illustrate the method to extract a HI. Section 4 presents the application of the proposed method on simulated data of aeroplane turbofan engines, while Section 5 concludes the paper.

2. Problem statement

The objective of the present work is the development of a general and systematic method for the definition of a health indicator (HI) that quantifies the degradation of industrial equipment.

We assume to have available the measurements of S signals taken during N run-to-failure trajectories collected from a population of similar equipment. The S signals can be directly or indirectly related to the equipment operating conditions, its internal operational and degradation states.

According to [13,14,23], a HI should have three key properties: monotonicity, trendability and prognosability. Monotonicity is required, assuming that the equipment does not undergo self-healing, which would result in HI non-monotonic trends. Trendability indicates the degree to which the evolution of the health indicator during the degradation of a population of similar components has the same shape and, thus, can be described by the same functional form. Finally, prognosability measures the variance of the HI values at failure time in a population of similar components: one would like that failures occur at the same value of the HI, for all components of the population.

Different metrics have been considered for measuring the degree to which the three properties of monotonicity, trendability and prognosability are satisfied by a given HI [13,35,56]. In this work, we consider the metrics provided by [14], whose values range in $[0, 1]$, where 1 indicates the most satisfactory and 0 the less satisfactory level of the specific HI property.

Monotonicity is defined as:

$$\text{Monotonicity} = \frac{1}{N} \left| \sum_{i=1}^N M_i \right| \quad (1)$$

where M_i is the monotonicity of a single run-to-failure trajectory given by:

$$M_i = \frac{n_i^+}{n_i - 1} - \frac{n_i^-}{n_i - 1}, \quad i = 1, \dots, N \quad (2)$$

with n_i indicating the total number of observations in the i -th run-to-failure trajectory, and n_i^+ (n_i^-) the number of observations characterized by a positive (negative) first derivative. Notice that, according to [14], it is important to properly smooth the noisy raw data to give more accurate estimates of the derivatives. In practice, the HI is typically linearly interpolated considering time windows formed from 5 to 10 observations before computing the derivatives.

As above said, the metric of *Trendability* measures whether the HI trends are modelled by the same underlying functional form in a population of run-to-failure trajectories. In [13], a *Trendability* index based on the comparison of the fractions of positive first and second derivatives in the run-to-failure trajectories has been proposed. Since, however, the proposed index has been shown to be highly susceptible to noise, we define *Trendability* as the minimum of the linear correlation coefficients [15]:

$\text{Trendability} = \min(|\text{corrcoef}_{ij}|), i, j = 1, \dots, N$ (3) where corrcoef_{ij} indicates the linear correlation coefficient between the i th and the j th run-to-failure trajectory, and the minimum is taken over all the possible pairs of trajectories. Notice that the computation of the correlation coefficient between two vectors requires that they are formed by the same number of patterns. Thus, Eq. (3) cannot be directly applied to two run-to-failure trajectories, formed by a different number of patterns due to the different component failure times. This problem is overtaken by defining new HI vectors formed by the same number, L_{HI} , of patterns. In practice, the HI signals are interpolated and the k -th value of the new HI vector, $k = 1, \dots, L_{HI}$ is the interpolated signal value at the fraction k/L_{HI} of the component lifetime.

Finally, *Prognosability* is defined as the standard deviation of the HI value at failure for the available run-to-failure trajectories divided by the average variation of the HI values between the beginning and the end of the trajectories. The obtained value is exponentially weighted to give a metric in the desired 0 to 1 scale:

$$\text{Prognosability} = \exp(-\text{std}(HI_{fail})/\text{mean}|HI_{start} - HI_{fail}) \quad (4)$$

where HI_{start} and HI_{fail} are the HI values at the beginning and end of the run-to-failure trajectories, respectively.

The characteristics of a HI will be represented by the vector:

$$F = [F_1, F_2, F_3] = [\text{Monotonicity}, \text{Trendability}, \text{Prognosability}] \quad (5)$$

The goal is to find the HI characterized by maximum values of the three elements of F . For this multi-objective optimization, we consider Pareto-optimal solutions: a solution $z^* \in \mathcal{F}$ is said to be Pareto-optimal if it is non-dominated with respect to \mathcal{F} , i.e. it does not exist another solution $z' \in \mathcal{F}$ such that $F(z')$ dominates $F(z^*)$,

$$\forall \alpha \in \{1, 2, 3\}, F_x(z') \leq F_x(z^*), \text{ and } \exists \tilde{\alpha} \in \{1, 2, 3\}, \text{ such that } F_{\tilde{\alpha}}(z') < F_{\tilde{\alpha}}(z^*) \quad (6)$$

3. Health indicator selection

The method for the selection of the HI from monitored signals is based on three steps of feature extraction, selection and fusion.

In the feature extraction step, the measured raw signals $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_S(t)]$ are pre-processed in order to reduce the measurement and process noises, and extract degradation trends. In practice, from each measured signal, $x_i(t)$, $i = 1, \dots, S$, we extract M different features $f_{im}(t)$, $m = 1, \dots, M$. The choice of the features to be considered is driven by the characteristics of the monitored component and the measured signals. Feature extraction techniques include the computation of statistical indicators in signal time-windows (e.g., means, standard deviations, etc.) and analytics (e.g., derivatives, elongations, etc.), signal transformation in the frequency domain (e.g., Fourier Transform, Laplace Transform) and in the time-frequency domain (e.g., Short Time Fourier Transforms (STFT) [30], Wigner-Ville Distribution (WVD), Wavelet Transform (WT) [75], and Empirical Mode Decomposition (EMD) [25]). With respect to fault prognostics in mechanical systems, the use of the EMD-based features suggested in [38] and reported in Appendix A.1 has been shown to provide indicators characterized by satisfactory performance with respect to the *Monotonicity*, *Trendability* and *Prognosability* metrics considered in this paper.

The overall number of available features, $S \cdot M$, can be very large: several signals are measured and hundreds of features can be, in principle, extracted. Since several features can be non-informative with respect to the component degradation or contain redundant information, a second step of feature selection is necessary to identify the best subset of the $S \cdot M$ available features from which the HI will be extracted. In practice, among the $2^{S \cdot M} - 1$ possible combinations of feature subsets which can be considered, the objective is to select the one which allows defining the most satisfactory HI in terms of the maximum values of F_1 , F_2 and F_3 (5).

The third step of feature fusion is performed to fuse the information content of the selected features into a single HI. The idea behind the fusion step is that features that individually do not have the characteristics of health indicators may provide useful information when considered jointly with other features. Notice that the second and the third steps are strictly correlated, since the selection of the best feature subset to be used for the extraction of the HI depends on the method used for the fusion of the selected features.

In Section 3.1, we describe the feature fusion method, whereas, in Section 3.2, the identification of the best feature subset from which the HI should be extracted is presented.

3.1. Feature fusion for the construction of a health indicator

In this work, the feature fusion step is performed by an AAKR-based approach. AAKR is a multi-dimensional reconstruction method applied with success in several fault detection applications [24,3,17]. Typically, AAKR reconstructs the values of the features expected during operation of a healthy industrial component as a weighted sum of training patterns taken from historical feature values collected when the component was healthy. Fault detection is, then, based on the quantification of the similarity between the features obtained from the current signal measurements and the reconstructions. Considering a feature subset $\mathbf{f} = [f_1, f_2, \dots, f_K] \subset [f_{im} | i = 1, \dots, S; m = 1, \dots, M]$ with $K \leq S \cdot M$, the similarity is quantified by the residual,

i.e. the Euclidean distance between the vectors of the AAKR reconstructed features, $\hat{\mathbf{f}}(t) = [\hat{f}_1(t), \dots, \hat{f}_K(t)]$, and the features obtained from the currently observed signals, $\mathbf{f}(t) = [f_1(t), \dots, f_K(t)]$:

$$r(t) = \sqrt{\sum_{k=1}^K (\hat{f}_k(t) - f_k(t))^2} \quad (7)$$

When the residual, $r(t)$, is close to zero, the equipment is healthy, whereas it drifts from zero when the component starts degrading.

Since the final objective of prognostics is the prediction of the component failure time, and not of the onset of the degradation process, in this work the AAKR reconstruction model is used to quantify the similarity between the features obtained from the current signal measurements and those expected from a component close to failure. Thus, the AAKR model uses a training set formed by historical feature values collected from components close to failure, i.e. in the last τ_f time instants before the failure of their degradation trajectories. Thus, the residual $r(t)$, which tends to decrease as the component degradation increases and to be close to zero at failure, constitutes a potential HI. A more detailed description of the AAKR algorithm can be found in Appendix A.2.

3.2. Selection of the feature subset

In this subsection, we address the problem of selecting among all the possible $2^{S \cdot M} - 1$ feature subsets, which can potentially be obtained from the available $S \cdot M$ features, the one to be used for the definition of the HI. The performance of a feature subset is evaluated by developing the corresponding feature fusion AAKR-based model and, then, by computing the

three metrics of *Monotonicity*, *Trendability* and *Prognosability* described in Section 2. Thus, the selection of the feature subset to be used for the construction of a HI is treated as a multi-objective optimization problem.

The exhaustive search of all possible combinations of features is in practice not feasible for its computational burden, since for each possible feature subset we have to develop the corresponding AAKR model and evaluate the three metrics F_1 , F_2 and F_3 . For this reason, we use the heuristic Binary Differential Evolution (BDE) algorithm.

3.2.1. BDE

Optimization amounts to finding the values of given parameters (decision variables) that minimize (maximize) one or more objective functions. For complex combinatorial problems, various heuristic optimization algorithms have been developed, such as Ant Colony [1], Genetic Algorithm [54], Binary Particle Swarm optimization [19], Binary Genetic Algorithms [69] and Binary Differential Evolution [22,31]. In this paper, we resort to Binary Differential Evolution (BDE), which has been shown capable of exploring the decision variables space more efficiently than other multi-objective evolutionary algorithms [57].

In BDE, a population of candidate solutions (chromosomes) evolves through three steps: Mutation, Recombination (also called Crossover) and Selection (Fig. 1) [22,31,82].

The decision variables are encoded in the chromosome as binary values of 0 and 1. Thus, each candidate solution $z_{p,G}$, called target vector, of the G th population is encoded by a binary sequence (chromosome) of bits (genes). Each gene, $z_{p,k,G}$, $p = 1 : NP$, $k = 1 : 2 \cdot n$ of each chromosome of the G -th population is conveniently mapped into a continuous variable $\tilde{z}_{p,k,G}$. In practice, the interval $[0, 1]$ is partitioned into two equal subintervals $[0, 0.5]$ and $[0.5, 1]$, such that if the gene $z_{p,k,G} = 0$, $\tilde{z}_{p,k,G}$ belongs to the first sub-interval, whereas if $z_{p,k,G} = 1$, $\tilde{z}_{p,k,G}$ belongs to the second interval. The mapping operator is defined by:

$$\tilde{z}_{p,k,G} = \begin{cases} 0.5 * rand & \text{if } z_{p,k,G} = 0 \\ 0.5 + 0.5 * rand & \text{if } z_{p,k,G} = 1 \end{cases} \quad (8)$$

where *rand* is a random number sampled from a uniform distribution in $[0, 1]$.

In the Mutation phase, for each vector $z_{p,G}$ of the population, a noisy vector v_s is generated by randomly choosing three mutually different vector indices $r_1, r_2, r_3 \in \{1, \dots, NP\}$ with $\{r_1, r_2, r_3\} \neq p$:

$$v_p = z_{r_1,G} + SF(z_{r_2,G} - z_{r_3,G}) \quad (9)$$

where *SF* is called *scaling factor* and $SF \in (0, 2]$ [31].

Then, a sigmoid function is applied to $v_{p,k,G}$ to ensure that the result generated by the mutation operator falls into the interval $[0, 1]$:

$$\frac{1}{1 + e^{-v_{p,k,G}}} \quad (10)$$

An inverse operator is, then, used:

$$v_{p,k,G} = \begin{cases} 0 & \text{if } v_{p,k,G} \in [0, 0.5) \\ 1 & \text{if } v_{p,k,G} \in [0.5, 1] \end{cases} \quad (11)$$

In the Recombination phase, crossover can be introduced in order to increase diversity of the perturbed parameter vectors. To this aim, the trial vector $u_{p,G} = (u_{p,1,G}, \dots, u_{p,k,G}, \dots, u_{p,2 \cdot n,G})$ is defined by:

$$u_{p,k,G} = \begin{cases} z_{p,k,G} & \text{if } U(0, 1) \leq CR \text{ or } k = irand(NP) \\ v_{p,k,G} & \text{if } U(0, 1) > CR \text{ and } k \neq irand(NP) \end{cases} \quad (12)$$

where $U(0, 1)$ is a uniform continuous random value in $[0, 1]$, whereas $irand(NP)$ is a discrete random number in the set $\{1, 2, \dots, NP\}$. The *crossover parameter* $CR \in [0, 1]$ influences the probability that the noisy vector's variables be selected for the mutation process [63].

The purpose of the Selection phase is to avoid stagnation of population in local minima due to the impoverishment of the population. Selection strategies have been deeply investigated in literature [37,48]. According to the MODE-III selection technique [63], each trial vector generated at each iteration by mutation and crossover operations, $u_{p,G}$, is compared only with its



Fig. 1. BDE search engine scheme.

target vector $z_{p,G}$ from which it inherits some variables: if $u_{p,G}$ dominates $z_{p,G}$, it takes its place in the population for the next generation, otherwise $z_{p,G}$, survives [63]. Notice, however, that this approach suffers of a low level of elitism since each trial vector is compared only with its own target vector.

In the present work, we have applied a different technique, called Non-Dominated Sorting Binary Differential Evolution (NSDBE), which combines the robust and effective BDE strategy with the fast non-dominated sorting and ranking selection scheme of NSGA-II [16]. In practice, at the G th generation the combined population of size $2NP$ comprising all $u_{p,G}$ and $z_{p,G}$ is ranked using a fast non-dominated sorting algorithm that identifies the ranked non-dominated solutions of the Pareto optimal set. Then, the first NP candidate solutions are selected according to the crowding distance [16].

3.2.2. BDE for the selection of HI

A candidate solution is a binary sequence (chromosome) of $S \cdot M$ bits (genes), where each bit indicates whether the corresponding feature is present (value set to 1) or discarded (value set to 0) in the feature subset (cf. Fig. 2). The computation of the fitness functions of a candidate solution requires: (1) the development of the AAKR-based feature fusion model, where the model inputs are the features with corresponding value in the chromosome equal to 1 and the model output is the health indicator, (2) the evaluation of the three metrics of *Monotonicity*, *Trendability* and *Prognosability* for the obtained health indicator.

BDE drives the search, generation after generation of candidate features, toward the Pareto-optimal solutions according to the scheme of Fig. 3. Once BDE reaches convergence, the choice of the Pareto front solution for the HI selection is based on the TOPSIS method (Technique for Order Preference by Similarity to an Ideal Solution) [43,4], which is a multiple criteria decision making method whose basic idea is that the chosen solution should have the shortest distance from the ideal solution and the farthest distance from the negative ideal solution. More details on TOPSIS can be found in Appendix A.3.

4. Case study

This section presents the results of the application of the proposed method to a case study concerning the prediction of the RUL of a fleet of turbofan engines working under continuously varying operating conditions. The data used in this work have been taken from the NASA Ames Prognostic Data Repository [50] and consist of $N = 260$ run-to-failure trajectories. Each trajectory is a 24-dimensional time series of different length, formed by $S = 21$ signals measured by sensors and 3 signals referring to the turbofan engines operating conditions (Altitude, Mach Number and Throttle Resolver Angle). These latter signals indicate $C = 6$ different operating conditions influencing the values of the other S measured signals.

4.1. Signal pre-processing

A three steps data pre-processing has been applied to (i) reduce the influence of the operating conditions on the signals behavior, (ii) reduce the signal noise, and (iii) detect the elbow point.

With respect to (i), signal values have been normalized taking into account their ranges in the different operating conditions. Considering N run-to-failure trajectories, S signals and C different operating conditions varying during the whole life of the component, data are normalized by applying the following equation:

$$x_s^{norm}(t) = \frac{x_s^c(t) - \mu_s^c}{\sigma_s^c} \quad (13)$$

where $x_s^{norm}(t)$ represents the s th normalized signal at the time instant t , $x_s^c(t)$ represents the collected s th signal operating in the c th operating condition at the time instant t , μ_s^c and σ_s^c represent the mean value and the standard deviation of the s th signal operating in the c th operating condition.

With respect to (ii), noise has been reduced by applying an exponential filter.

With respect to (iii), since the life of an industrial component is typically divided into healthy and degrading phases, prognostics is typically preceded by the detection of the onset of the degradation process, i.e. the time instant at which the component starts to significantly degrade [47]. In this work, the onset of the degradation process is indicated by the detection of an elbow point in the signals, i.e. a time instant at which the degradation speed drifts from its initial condition, leading to an acceleration of the degradation process. In the remaining part of the paper, every time we refer to a specific trajectory, we will consider only the data subsequent to the elbow point and prognostics will be applied only after the identification of the elbow point.

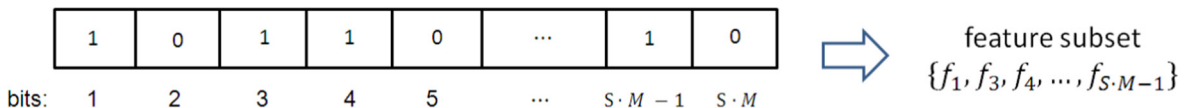


Fig. 2. Example of a chromosome and of the corresponding feature subset.

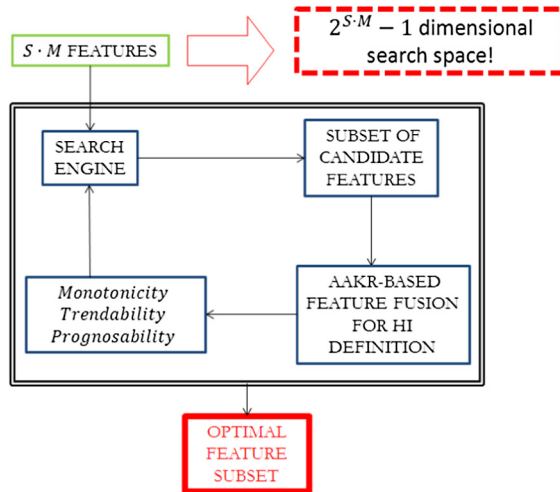


Fig. 3. BDE procedure.

In this case study, according to the analysis in [38], we have considered the feature extraction techniques based on the computation of EMD transform according to which the features to be considered are the EMD trend mean and variance, and the slope and intercept of its linear approximation (see Appendix A.1). These four features are considered together with the raw signal, for a total of $M = 5$ features for each of the $S = 21$ signals.

The obtained set of $S \cdot M = 105$ features has been reduced in order to diminish the computational burden of the BDE analysis for the identification of the best feature subset. To this purpose, we have computed the three metrics of *Monotonicity*, *Trendability* and *Prognosability* of the individual features, and we have selected 25 features belonging or close to the Pareto frontier. This subset of the most promising features, which is formed by 9 raw signals, 6 slopes of the EMD residual linear interpolation and 10 EMD residual variances, has been completed with 3 raw signals measuring the component operating conditions.

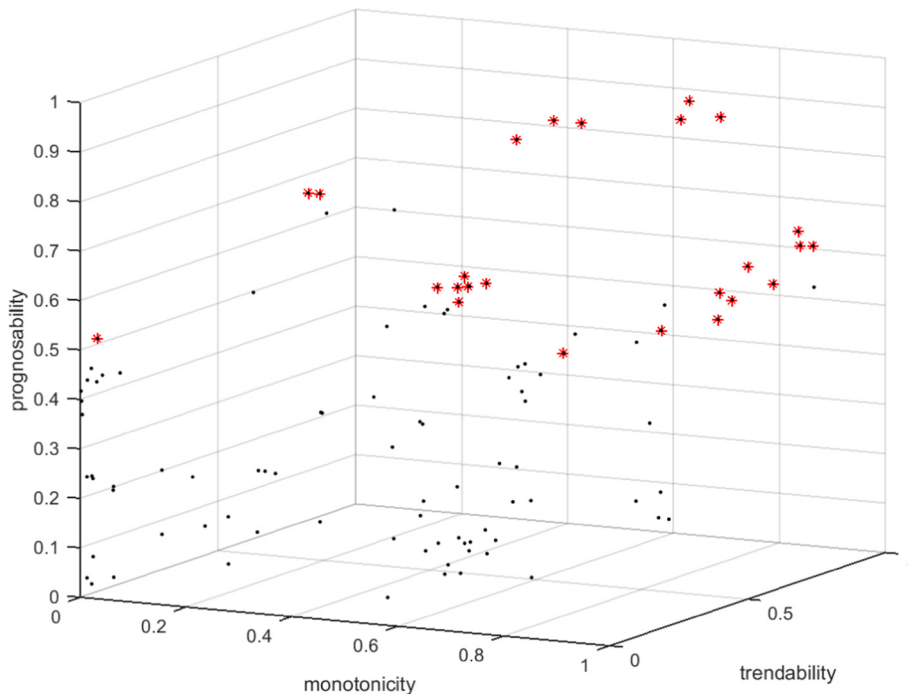


Fig. 4. *Monotonicity*, *Trendability* and *Prognosability* of the 105 extracted features (dots) and of the 25 most promising ones (asterisks).

Fig. 4 shows *Monotonicity*, *Trendability* and *Prognosability* of the 105 extracted features (dots) and of the reduced set of the 25 most promising features (asterisks). These features together with the 3 signals relative to the operating conditions are the 28 features which will be considered for the remaining part of the analysis. Considering that the 28 selected features generate $(2^{28} - 1) \cong 2.7 \cdot 10^8$ possible feature subsets and the construction of a single HI through the application of the AAKR-based procedure of Section 3.1 and the evaluation of the three metrics of monotonicity, trendability and prognosability requires, on average, 3 s on a quad core computer with 2.80 GHz CPUs and 16.0 GB of RAM, an exhaustive search among all the possible solutions of the best performing HI is not feasible.

4.2. BDE application

The $N = 260$ run-to-failure trajectories are divided into three groups:

- (1) a training set made by 100 trajectories, which will be used to train the AAKR reconstruction model;
- (2) a test set made by 100 trajectories, which will be used by the BDE to assess the HI performance;
- (3) a validation set made by 60 trajectories, which will be used to compare the performance of the proposed approach with that of other methods.

Within the BDE search, the evaluation of the fitness of a chromosome, which represents a possible feature subset for the construction of a HI, is performed by developing the corresponding AAKR reconstruction model trained considering the last $\tau_f = 5$ measurements for each one of the 100 training trajectories as historical data representing the component behavior close to failure. Section 4.5 discusses the dependence of the methodology performance from the τ_f parameter.

The BDE crossover parameter, CR , the scale factor parameter, SF , and the number of chromosomes in the population, NP , have been set to 0.60, 0.95 and 200, respectively. The choices of CR and SF allow a gradual and successful exploration of a complex search space and maintain diversity in the population. A large population formed by $NP = 200$ chromosomes has been used to allow a deep exploration of the multidimensional search space [36].

With respect to the BDE search termination criterion, in this work, the decision on when to stop the search is taken by observing the evolution of the Lebesgue-measure of the hyper-volume under the Pareto-optimal set with respect to a lower reference bound usually the ideal worst values of each objective function) [73]. In practice, when two Pareto fronts are compared, higher is the value of such indicator, better is the performance in terms of objective function evaluation and wider is the exploration of the search space. In our case, we set the lower reference point to $(0, 0, 0)$, which are the worst possible value of our objective functions.

Fig. 5 shows the evolution with the number of performed iterations of the value of the Lebesgue integral of the hyper-volume under the Pareto-optimal set. Notice that the most significant increase of the hyper-volume is obtained in the first

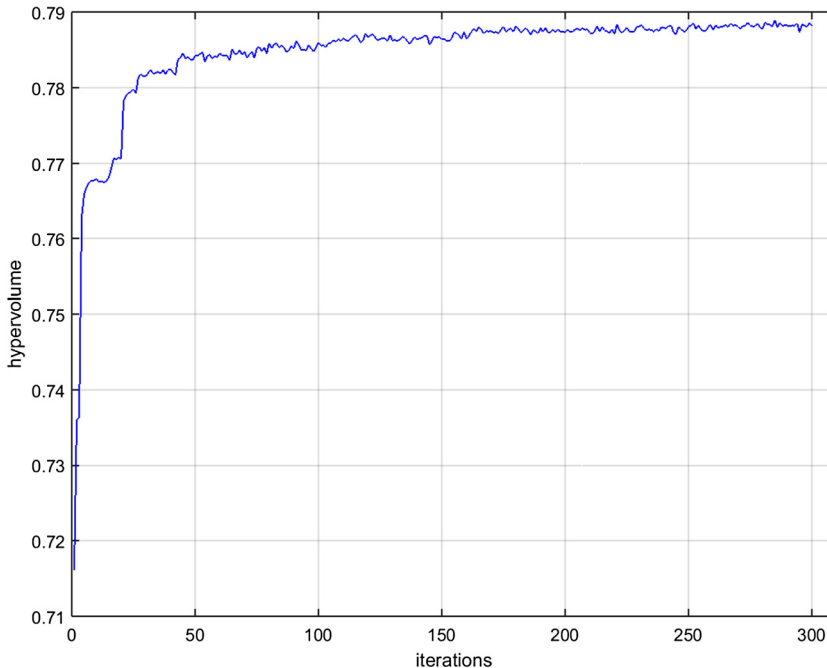


Fig. 5. Evolution of the hyper-volume under the Pareto-optimal frontier with the number of generations.

40 generations, followed by a slower increasing monotonic trend. The search is stopped at the 300th generation, when the hyper-volume reaches a stable value. The total computation time on a quad core computer with 2.80 GHz CPUs and 16.0 GB of RAM has been of 46 h.

4.3. Results

Fig. 6 shows the Pareto-optimal frontier identified by the proposed method at the last generation (circles) and that already reported in Fig. 4, which refers to the 105 single features directly obtained from the 21 measured signals before the feature fusion step (asterisks).

Notice that the HIs found by the proposed method are characterized by large values of *Prognosability* and the Pareto frontier tends to dominate that of the features directly extracted from the raw signals, with the exception of our solutions at the bottom right corner characterized by very large *Monotonicity* and small *Prognosability*.

Fig. 7 shows the evolution of the HIs on trajectories of the validation set corresponding to solutions of the Pareto frontier characterized by the largest *Monotonicity* (z_{mono}^*), the largest *Trendability* (z_{trend}^*), the largest *Prognosability* (z_{progno}^*), obtained by applying the TOPSIS method (z_{topsis}^*), and two features directly extracted from the raw signals: one obtained by applying the TOPSIS method to the set of features directly extracted from the signals (z_{11} , since it is the 11th raw signal) and one of those characterized by large *Monotonicity* and small *Prognosability* (z_{emd} , since is one of the EMD-derived features). Notice that this latter HI is not satisfactory for prognostic applications given the very different end-of-life HI values.

As expected, the solution characterized by the largest *Monotonicity* tends to be very smooth and noiseless, but the HI values at the end of the trajectories have a very spread distribution. On the contrary, the solution with the largest *Prognosability* has an extremely narrow distribution of end of life HI values, but it is characterized by a noisy behavior of the HI. The TOPSIS solution provides a trade-off between the previous two solutions, since it is smoother than the solution with the largest *Prognosability* and its final values are more narrow-distributed than those of the solution with the largest *Monotonicity*. The HI corresponding to the solution with the largest *Trendability* is more noisy than that of the TOPSIS solution, whereas that corresponding to the feature directly extracted from the raw signal is characterized by a more spread distribution of the end-of-life HI values.

Table 1 shows the features forming the subset of the selected solutions, i.e. the signals used for the extraction of the corresponding HI. Raw signals tend to provide HIs with a more satisfactory performance in terms of *Prognosability*

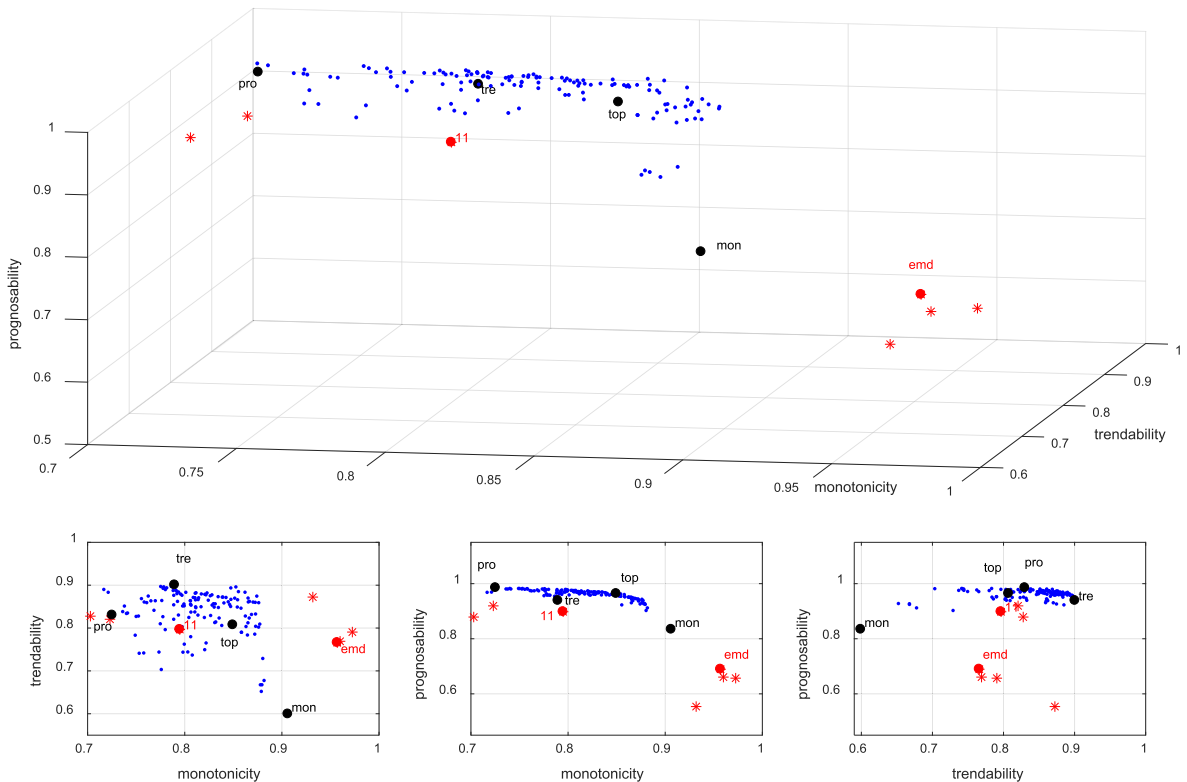


Fig. 6. Pareto frontier of the proposed method (dots) and of features directly extracted from the raw signals (asterisks). The features z_{mono}^* is indicated by "mon", z_{trend}^* by "tre", z_{progno}^* by "pro" and z_{topsis}^* by "top".

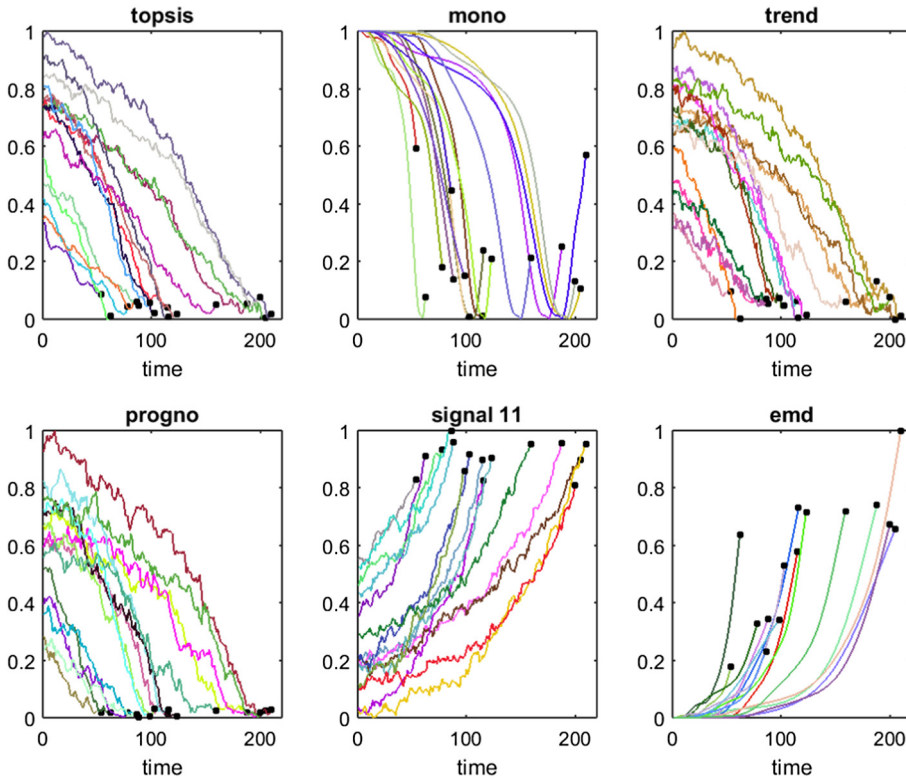


Fig. 7. HIs associated to the selected solutions for some of the trajectories of the validation set.

Table 1

Subsets of features associated to the selected solutions.

		Z_{topsis}^*	Z_{mono}^*	Z_{trend}^*	Z_{progno}^*	Z_{11}
Raw signals	sig.2			✓	✓	
	sig.3			✓		
	sig.4	✓		✓	✓	
	sig.11	✓				✓
	sig.15	✓				
	sig.17			✓		
Operating conditions	cnd.oper.2			✓		
	cnd.oper.3			✓		
EMD-derived features	slope sig. 2	✓				
	slope sig. 3	✓				
	slope sig. 11	✓				
	slope sig. 15	✓				
	variance sig. 2			✓		
	variance sig. 3	✓				
	variance sig. 4	✓				
	variance sig. 7		✓	✓		
	variance sig. 15			✓		
variance sig. 20				✓		

(Z_{topsis}^* , Z_{trend}^* , Z_{progno}^* and Z_{11}), whereas EMD-derived features tend to drive toward more monotonic HIs thanks to their capability of smoothing the signals (Z_{mono}^* and Z_{emd}). Finally, the presence of the operating conditions in the feature subset seems to contribute to the *Trendability* of the HI.

The TOPSIS solution, which is characterized by a satisfactory compromise in terms of *Monotonicity* and *Prognosability*, is extracted from a feature subset formed by raw signals and EMD-derived features. The presence in the Pareto frontier of solutions formed by different numbers of features ranging from 1 (Z_{mono}^*) to 13 confirms that the BDE algorithm has been able to explore thoroughly the search space.

4.4. Analysis of the performance of the prognostic model built with the TOPSIS HI solution

In this section, we verify the performance of a prognostic data-driven model built using as input signal the HI extracted from the features of the TOPSIS solution. The results of the model are compared with those obtained by a similar model whose input is the measured signal with the most satisfactory trade-off among the metrics of *Monotonicity*, *Trendability* and *Prognosability*, i.e. raw signal 11.

The prognostic model is based on the use of Extreme Learning Machine (ELM), which is a single hidden layer feed-forward neural network (SLFN) [28] whose input-hidden layer parameters are randomly sampled. Thus, the only parameters to be optimized are the hidden-output layer parameters which are set by solving a linear system. Therefore, compared to a traditional feed-forward neural network, an ELM is faster to train and it has been shown to provide similar performances [78,26,27,79]. The ELM inputs are the values of the HI at present time and at the previous W consecutive time steps $HI(t), HI(t-1), \dots, HI(t-W)$, whereas the ELM output is the predicted RUL. With respect to the setting of the time-window length W , larger is W , more information are provided to the ELM model on the dynamic evolution of the HI and more accurate are expected to be the predictions. On the other side, too large time-windows can result in ELM models with many input signals, whose training requires a very large number of patterns (curse of dimensionality). Furthermore, larger is the time-window length, larger is the time at which the first RUL prediction can be provided by the ELM model, which requires W measurements after the onset of the degradation process. In this work, a trade-off value of W equal to 11 is used, while Section 4.6 discusses the dependence of the ELM model performance from the time-window length W .

Number of neurons in the hidden layer and type of activation function of the TOPSIS solution have been chosen by adopting a trial-and-error procedure considering as objective the minimization of the *Relative Error (RE)* on the test set [51]. Different ELMs with different combination of number of neurons in the hidden layer and type of activation function have been trained and their performances have been evaluated. The best performances have been obtained by an ELM characterized by a hidden layer of 200 nodes and a sigmoid activation function.

In order to evaluate the performance of the obtained prognostic model, we consider the prognostic metrics of *Cumulative Relative Error (CRE)* [51], *Average $\alpha - \lambda$ Accuracy (AAL)* [51] and *Average Steadiness Index (ASI)* [42].

Considering the generic i th run-to-failure trajectory of the validation set, *Relative Error*, $RE_i(t)$, at time t is defined by:

$$RE_i(t) = \left| \frac{RUL_i(t) - \widehat{RUL}_i(t)}{RUL_i(t)} \right| \quad (14)$$

where $RUL_i(t)$ and $\widehat{RUL}_i(t)$ are the ground truth for RUL and the ELM RUL prediction at time instant t , respectively. *CRE* is the average of $RE_i(t)$ over all the time instants at which the RUL is predicted and over all the N^{val} validation trajectories:

$$CRE = \frac{1}{N^{val}} \sum_{i=1}^{N^{val}} \left(\frac{1}{EOL_i - t_0} \sum_{t=t_0}^{EOL_i} RE_i(t) \right) \quad (15)$$

where EOL_i is the time index of the actual end of life of the i th trajectory, while t_0 is the first time instant at which the RUL is predicted. Small values of *CRE* indicate more accurate prediction.

The *AAL* quantifies prediction quality by determining whether the prediction falls within specified limits at particular times with respect to the ground truth for RUL. The $\alpha - \lambda$ *Accuracy (AL)* is a binary metric that evaluates whether the RUL prediction at specific time instants t_λ falls within specified α -bounds from the ground truth for RUL. The time instants t_λ are fractions of time between the time instant of the first prediction and the end of life, while the α -bounds are expressed as percentage of the ground truth for RUL at each t_λ .

In this work, since we want a point-wise evaluation of the prognostic performance, we do not set the value of the λ parameter, but we calculate *AL* for each time instant of each trajectory:

$$AL_i(t) = \begin{cases} 1, & \text{if } (1 - \alpha) \cdot RUL_i(t) \leq \widehat{RUL}(t) \leq (1 + \alpha) \cdot RUL_i(t) \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

A scalar metric is, then, obtained by computing the average *AL* over all the time instants and all the trajectories:

$$AAL = \frac{1}{N^{val}} \sum_{i=1}^{N^{val}} \left(\frac{1}{EOL_i - t_0} \sum_{t=t_0}^{EOL_i} AL_i(t) \right) \quad (17)$$

The larger is the *AAL* value, the more accurate is the model, since more predictions fall within the corresponding α -bounds.

The third metric is the *ASI*, which quantifies the stability of the failure time predictions. The standard deviation of the difference between the predicted RUL and the ground truth for RUL over a sliding time-window of $tw = 10$ time instants is computed and averaged along the trajectory lengths:

$$SI_i = \frac{1}{EOL_i - t_0 - tw + 1} \sum_{t=t_0}^{EOL_i - tw + 1} std([d(t), d(t+1), \dots, d(t+tw-1)]) \quad (18)$$

where $d(t) = RUL_i(t) - \widehat{RUL}_i(t)$.

Finally, the obtained SI_i values are averaged over all the validation set trajectories:

$$ASI = \frac{1}{N^{val}} \sum_{i=1}^{N^{val}} SI_i \quad (19)$$

The smaller is the ASI value, the more stable is the RUL prediction.

Fig. 8 shows the obtained RUL predictions on two representative trajectories of the validation set. Although the RUL predictions provided by the two models tend to have the same behavior, the TOPSIS HI-based model is more accurate and stable. Table 2 reports the prognostic metrics obtained by the two ELM models, whose input signals are the HI extracted from the feature subset corresponding to the TOPSIS solution and that obtained by directly using signal 11, respectively.

Notice that the TOPSIS HI-based model outperforms the signal 11 HI-based model when the AAL and the ASI metrics are considered, whereas it has a less satisfactory value of the CRE metric.

Fig. 9 shows the average value of the prediction errors as a function of the fraction of component life over all the validation set trajectories. Notice that the predictions provided by the TOPSIS HI-based model are on average more accurate than that of the signal 11 HI-based model, except that for the last few time instants of the run-to-failure trajectory. Notice that, since this last part of the run-to-failure trajectories remarkably influences the CRE metric, being the denominator in Eq. (14) very small, the average CRE of the signal 11 HI-based model is smaller even if in all the other parts of trajectories the TOPSIS HI-based model is more accurate.

A possible cause of the unsatisfactory performance of the TOPSIS HI solution in the last parts of the run-to-failure trajectories is that the residual obtained from the AAKR fusion approach is computed with respect to a training set made by the last $\tau_f = 5$ signal measurements taken before failure from the training trajectories. Therefore, the HI, i.e. the AAKR residual quantifying the similarity between the measurements and the training set, tends to 0 some instants before the component failure, leading to slightly anticipated predictions of the failure times. Section 4.5 discusses the influence of parameter τ_f on the ELM model performance.

4.5. Analysis of the influence of τ_f on the model performance

The number τ_f of measurements that are extracted from the last part of the run-to failure-degradation trajectories and included in the training set is a parameter that needs to be set. One difficulty is that tuning of τ_f by trial-and-error is unfeasible from a computational point of view, since the identification of the optimal health indicator would require the execution of a dedicated time-consuming BDE search for each considered τ_f (Section 4.2).

For guidelines, it is expected that $\tau_f > 5$ would allow obtaining more patterns for training the AAKR and, thus, more complete representations of the failure state and more robust health indicators. On the other side of the coin, increasing τ_f has the undesired effect of including in the AAKR training set patterns collected τ_f time units before the component failure and, thus, of causing systematic underestimation of its RUL.

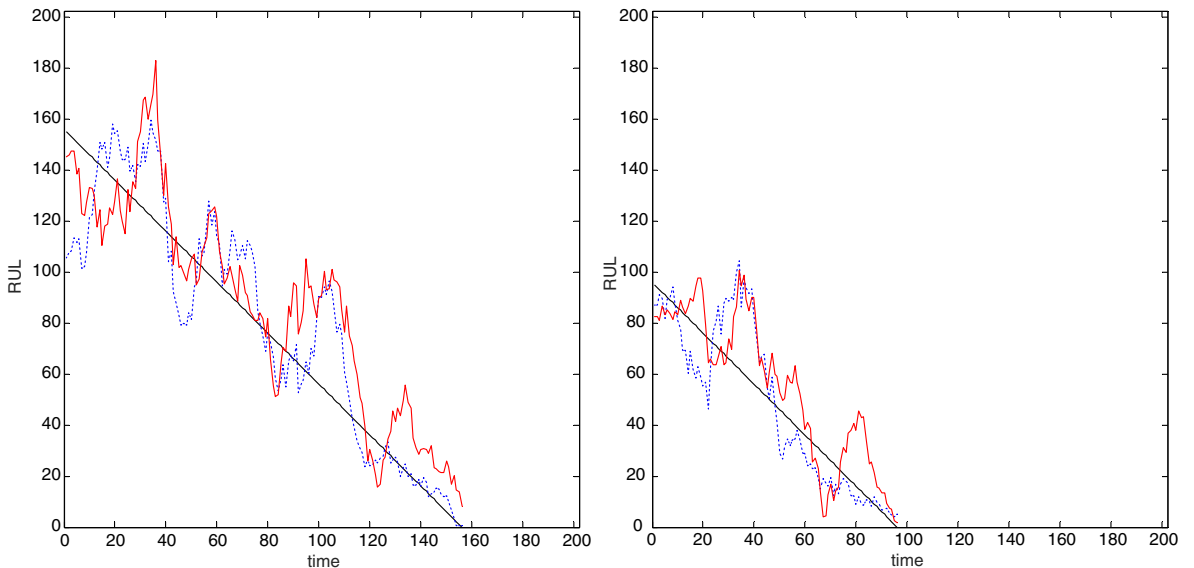


Fig. 8. RUL predictions of the TOPSIS HI-based model (dotted line) and the signal 11 HI-based model (solid line) on two trajectories of the validation set.

Table 2
Metrics for the evaluation of prognostic performance.

feature subset	CRE	AAL	ASI
z_{topsis}	0.598	0.389	6.169
z_{11}	0.582	0.297	7.508

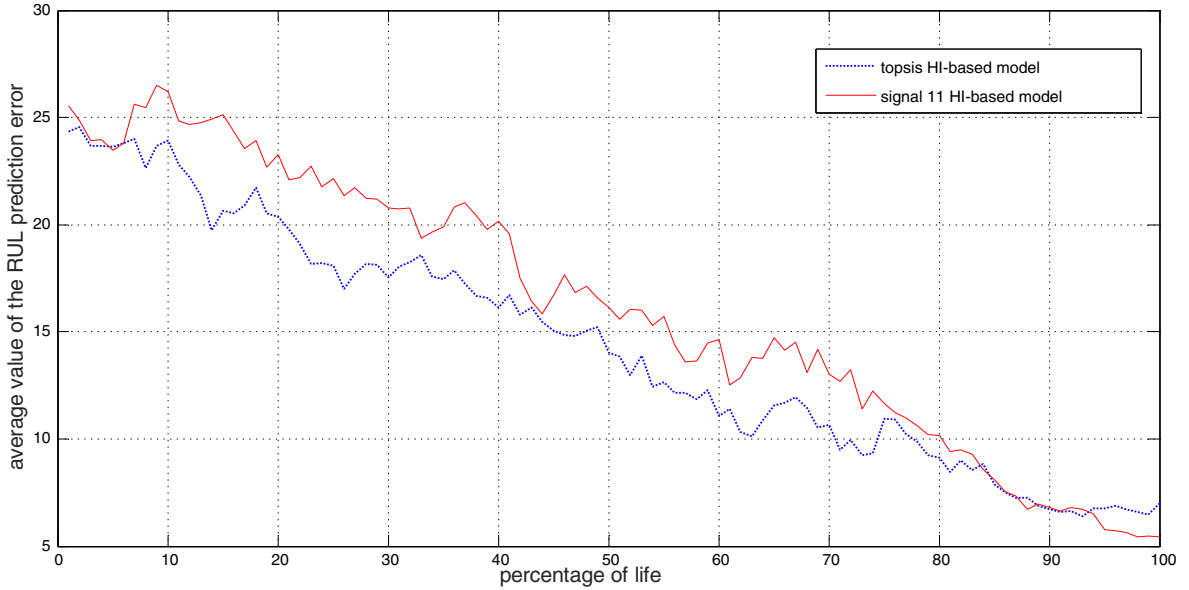


Fig. 9. Average value of the RUL prediction error at different percentages of the component life. The dotted line refers to the TOPSIS HI-based model, the solid line to the signal 11 HI-based model.

To investigate the effect of small τ_f on the overall ELM model performance in this case study, the HI indicator has been developed considering $\tau_f = 1$ and the corresponding ELM has been built. Table 3 reports that the reduction of τ_f causes a remarkable reduction of the overall performances of the ELM model due to the reduction of the AAKR training set, and, thus, the incomplete representation of the component failure state.

In the end, setting τ_f requires a trade-off between the two desiderata of:

- providing a complete representation of the failure state conditions, which would require training the AAKR model using a large number of patterns (large τ_f);
- predicting with precision the failure time, which would require including only the last measurements in the AAKR training set (small τ_f).

The optimal trade-off depends from various factors which are specific of the application under investigation, such as the number of run-to-failure trajectories available, the average lifetime of the component and the consequences of the fault. Future work will be devoted to the development of a systematic methodology for setting the value of τ_f and building the health indicator at the same time. To this aim, the use of coevolutionary search algorithms seems promising [64].

A drawback of the methodology proposed in this work is that unsatisfactory prognostic performances can result due to RUL underestimations in the last part of the component run-to-failure trajectories (Fig. 9). The magnitude of the associated prognostic error is expected to decrease in the case in which many run-to-failure trajectories are available, which would

Table 3
ELM prognostic performance as a function of the number τ_f of measurements extracted from the last part of the run-to failure-degradation trajectories and included in the training set.

τ_f	CRE	AAL	ASI
5	0.598	0.389	6.169
1	0.8273	0.2781	7.3903

allow obtaining a complete and satisfactory representation of the component failure state using only the last pattern of the run-to-failure trajectories ($\tau_f = 1$). Notice that the prognostic errors are related to RUL underestimations, which are preferable to RUL overestimations in several industrial applications, from both component availability and safety point of views.

4.6. Analysis of the influence of the ELM time-window length on the model performance

Fig. 10 shows the performance of the TOPSIS HI-based models as a function of the length of the ELM time-window. It is interesting to observe that the relative accuracy of the TOPSIS HI-based model, measured by the metric *CRE*, tends to reach the most satisfactory value for time-windows of length around 12. Time-windows longer than 12 are characterized by slightly more satisfactory performances from the point of view of the accuracy (*AAL* metric) and stability (*ASI* metrics). Increasing the number of ELM inputs has the effect of increasing the smoothness of the predictions and, thus, improving accuracy and stability, whereas it causes larger errors in the last parts of the run to failure trajectories, that negatively influence the *CRE* metric.

5. Conclusions

In this work, we have developed a general and systematic method for selecting a health indicator of the degradation state of industrial equipment, from signals measured during its operation. The developed method is based on the three steps of feature extraction, feature selection and feature fusion. In the first step, the measured raw signals are treated to reduce the measurement and process noises, and extract degradation trends. Since hundreds of features are typically available, a Multidimensional Binary Differential Evolution (BDE) algorithm has been used to select the feature subset to be used for defining the HI, among the several possible combinations of the features. The objectives of the search have been the optimization of statistical properties of HI monotonicity, trendability and prognosability.

The main originality of this work has been to frame the problem of selecting HIs as a feature selection problem. From the methodological point of view, the novelties of the proposed method are:

- using an AAKR model to combine different features into a single HI;
- performing a BDE multi-objective search for the selection of the optimal HI.

The proposed method has been applied for the prediction of the Remaining Useful Life of a fleet of turbofan engines working under variable operating conditions. It has been shown able to provide more satisfactory HIs than those found in literature approaches based on the direct use of features directly extracted from the signals. Furthermore, the performance of an ELM-based prognostic model based on the selected HI has been shown to be satisfactory, in spite of the slight underestimation of the component RUL.

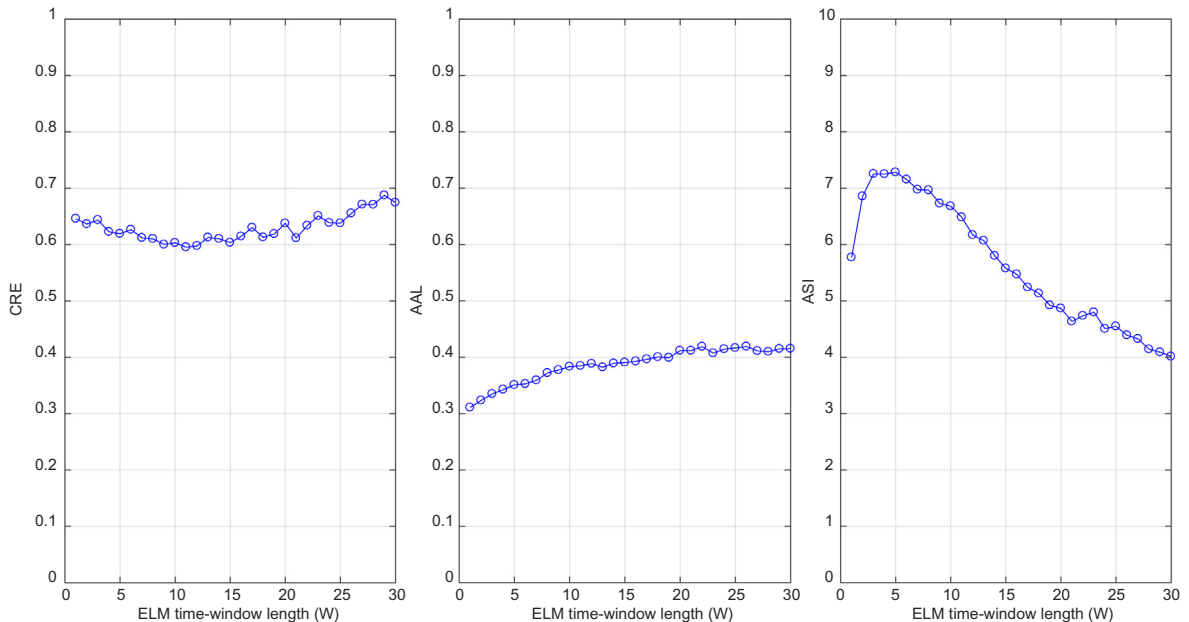


Fig. 10. Variation of the prognostic model performance as a function of the length of the ELM time-window.

Acknowledgements

The participation of Enrico Zio to this research is partially supported by the China NSFC under grant number 71231001.

We would also like to thank Marco Rigamonti and Francesco Cannarile for the help they provided for the analysis of the data and the development of the codes.

Appendix A

A.1. Empirical Mode Decomposition (EMD)

In [38], Empirical Mode Decomposition (EMD) is used to extract information about the effective evolution of the degradation over time of mechanical systems. The main idea of EMD is to locally describe a time dependant signal as the sum of fast oscillating components superimposed to slow oscillating ones. The residual of this approximation is the trend of the signal, whose variations has been shown to catch the degradation evolution over time.

The EMD decomposes each feature into all the possible Intrinsic Mode Functions (IMF) until the residual of the input time series is a monotonic or constant signal. IMF represents a simple oscillatory mode in which amplitude and frequency along the time axis can vary.

A time series is considered an IMF if it satisfies the following requirements:

- the number of local extrema of the time series and the number of its zero-crossings must be equal or differ at most by one;
- at any time the mean value of the upper envelope determined by the local maxima and the lower envelope determined by local minima is zero.

The EMD algorithm consists of the following steps [25]:

1. find all the local maxima and minima of the input signal and compute the corresponding upper and lower envelopes using cubic splines;
2. subtract the mean value of the upper and lower envelopes from the original signal;
3. repeat the previous steps until the signal remains nearly unchanged and an IMF is obtained;
4. remove from the signal the IMF obtained at step 3 and repeat the previous steps if the IMF is neither a constant nor a trend.

This procedure is visually explained in Fig. 11, where $X(t)$ is the input signal.

The generated residual should be constant or monotonic, and it represents the general trend of the feature.

The trend of a feature can represent the severity of the degradation of the component or system; therefore variations in the trend show the evolution of the degradation over time. In order to describe and summarize these variations, we use the following quantifiers: slope and intercept of the linear approximation of the trend, mean and variance of the trend.

At each time instant $t_{current}$ of each history, all the observations, starting from the initial time instant t_0 to the current one $t_{current}$, are processed to measure the proposed quantifiers.

To measure the slope and the intercept, the linear approximation of the trend is calculated, namely the best fitting first degree polynomial of the form $Y = aX + b$ with respect to least squares, from which we obtain the slope a and the intercept b .

The mean and the variance are calculated using their classical expressions:

$$mean = \frac{1}{n} \sum_{j=1}^n x_j, \quad variance = \frac{1}{n-1} \sum_{j=1}^n (x_j - mean)^2 \quad (20)$$

where x_j is the input observation and n is the length of the trend.

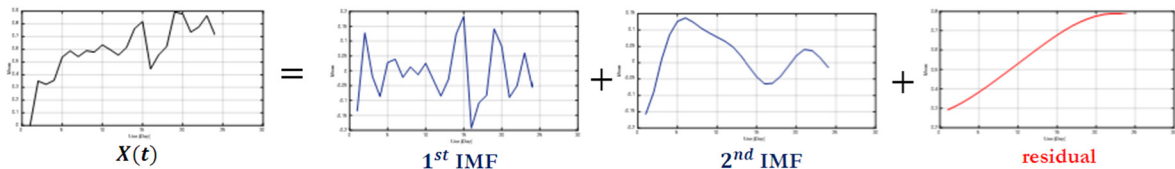


Fig. 11. Signal decomposition in IMFs and residual.

A.2. Auto-Associative Kernel Regression (AAKR)

Different empirical modelling techniques have been applied to the problem of signal reconstruction, such as Auto-Associative Kernel Regression (AAKR) [24,2], Principal Component Analysis (PCA) [18,21], Robust Principal Component Analysis [77], Fault-relevant PCA (FPCA) [68], Partial Least Squares (PLS) [39], Evolving Clustering Method (ECM) [70], Parzen Estimation [32,11], fuzzy-logic based systems [34], AutoAssociative (AA) and Recurrent (R) Neural Networks (NN) [7,20,41,49,65]. In this work, we consider AAKR which has been shown to provide more satisfactory performance than ECM and PCA [12] and is less computationally demanding than AANN, RNN [6] and Parzen Estimation [29]. Furthermore, notice that in many industrial applications of condition monitoring, such as those in energy production plants, it is common to periodically retrain or update the reconstruction models in order to follow the gradual modifications of the signal behavior due to slow degradation processes, maintenance interventions and minor plant reconfigurations. Thus, it is important to develop reconstruction models, such as AAKR, that are easy to train and characterized by few parameters which can be set by plant personnel.

The basic idea behind AAKR is to reconstruct at time t the values of the signals expected in normal conditions, $\hat{\mathbf{x}}^{nc}(t)$, on the basis of a comparison of the currently K observed signal measurements (also referred to as test pattern or test set), $\mathbf{x}^{obs}(t) = [x_1^{obs}(t), x_2^{obs}(t), \dots, x_K^{obs}(t)]$, and of a set of historical signals measurements (also referred to as training set) collected during normal conditions of operation. In practice, AAKR performs a mapping from the space of the measurements of the signals $\mathbf{x}^{obs}(t)$ to the space of the values of the signals expected in normal conditions, $\hat{\mathbf{x}}^{nc}(t)$:

$$\hat{\mathbf{x}}^{nc}(t) = \varphi(\mathbf{x}^{obs}(t) | \bar{\mathbf{X}}^{obs-nc}) : \mathbb{R}^K \rightarrow \mathbb{R}^K \quad (21)$$

where $\bar{\mathbf{X}}^{obs-nc}$ denotes a $L \times K$ matrix containing L historical observations of the K signals performed in normal conditions. Since the mapping is independent from the current time t at which the signals observations are performed, the time t will be omitted for brevity of notation. Thus, x_k^{obs} , $k = 1, \dots, K$, denotes the value of signal k at current time. The reconstruction of the expected values of the signals in normal conditions, $\hat{\mathbf{x}}^{nc} = [\hat{x}_1^{nc}, \hat{x}_2^{nc}, \dots, \hat{x}_K^{nc}]$, is performed as a weighted sum of the available historical observations: for the generic k th of $\hat{\mathbf{x}}^{nc}$, we write:

$$\hat{x}_k^{nc} = \frac{\sum_{l=1}^L w_l \cdot x_{lk}^{obs-nc}}{\sum_{l=1}^L w_l} \quad (22)$$

The weights w_l measure the similarity between the test pattern \mathbf{x}^{obs} and the l th historical observation vector \mathbf{x}_l^{obs-nc} . They are evaluated through a kernel Ker , i.e. a scalar function which can be written as a dot product [76,80]. From the mathematical point of view, a kernel is a function defined in this way:

$$Ker : \mathbb{R}^K \times \mathbb{R}^K \rightarrow \mathbb{R} \text{ s.t. } \exists \phi : \mathbb{R}^K \rightarrow \mathcal{H}, Ker(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle \quad (23)$$

where ϕ is a map from the observation space \mathbb{R}^K in a (possible countable infinite dimensional) Euclidean space \mathcal{H} , and $\langle \cdot, \cdot \rangle$ denotes the dot product. Traditional AAKR adopts as Ker function the Gaussian Radial Basis Function (RBF) with bandwidth parameter h :

$$w_l = Ker(\mathbf{x}^{obs}, \mathbf{x}_l^{obs-nc}) = \frac{1}{\sqrt{2\pi h^2}} e^{-\frac{\|\mathbf{x}^{obs} - \mathbf{x}_l^{obs-nc}\|^2}{2h^2}} \quad (24)$$

Notice that, according to Mercer's theorem (Burges, 1998), the expression of w_l can be seen as a dot product in a countable infinite dimensional Euclidean space:

$$e^{-\frac{1}{2}\|\mathbf{x}^{obs} - \mathbf{x}_l^{obs-nc}\|_2^2} = \sum_{i=0}^{\infty} \frac{(\mathbf{x}^{obs\top} \mathbf{x}_l^{obs-nc})^i}{i!} e^{-\frac{1}{2}\|\mathbf{x}^{obs}\|_2^2} e^{-\frac{1}{2}\|\mathbf{x}_l^{obs-nc}\|_2^2} \quad (25)$$

In this work, in order to account for differences in the scale and variability of the different signals, a Mahalanobis distance is used, defined by the covariance matrix $\bar{\bar{\mathbf{S}}}$, such that:

$$\|\mathbf{x}^{obs} - \mathbf{x}_l^{obs-nc}\|_{mahal}^2 = (\mathbf{x}^{obs} - \mathbf{x}_l^{obs-nc})^T \bar{\bar{\mathbf{S}}}^{-1} (\mathbf{x}^{obs} - \mathbf{x}_l^{obs-nc}) \quad (26)$$

Assuming independence between the signals, $\bar{\bar{\mathbf{S}}}$ is given by:

$$\bar{\bar{\mathbf{S}}} = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_K^2 \end{bmatrix} \quad (27)$$

where σ_k^2 denotes the estimated variance of the k th signal in the historical observations. Alternatively, instead of using the covariance matrix $\bar{\bar{\mathbf{S}}}$, one can obtain the same results by mapping the data into a normalized space:

$$y_k = \frac{x_k - \mu_k}{\sigma_k} \quad (28)$$

where μ_k is the mean value of the k th signal of the historical dataset, and by applying a Gaussian kernel with Euclidean distance in the normalized space.

A novelty we have introduced in the application of AAKR in this work is that the data of the training set are collected in the opposite way with respect to the classical idea of AAKR: we want to reconstruct the values of the signal in abnormal conditions. The ratio behind this is that we want to be sure that the reconstruction of the observed signal is as accurate as possible when it's close to failure rather than when its working conditions are normal, because, obviously, it's much more important to identify a possible failure situation, which may lead to dangerous consequences, rather than a normal situation.

A.3. TOPSIS method for the selection of the best compromise solution

The basic principle of this technique is that the chosen alternative should have the shortest distance from the ideal solution and the farthest distance from the negative ideal solution. Before applying the TOPSIS method [43], we have to normalize the values of the solutions of the Pareto frontier in a scale from 0 to 1, where 0 corresponds to the minimum value of the objective function in the Pareto frontier and 1 to the maximum value.

- P is the number of solutions in the Pareto-optimal set, while dim is the number of objective functions (in our case the objective functions are *Monotonicity*, *Trendability* and *Prognosability*, thus $dim = 3$).
- Compute for each solution $z_p = [z_{p,1}, z_{p,2}, z_{p,3}]$ in the Pareto-optimal set the values $\tilde{z}_{p,\alpha}$:

$$\tilde{z}_{p,\alpha} = \frac{F_\alpha(z_p)}{\sqrt{\sum_{s=1}^P F_\alpha(z_s)^2}} \quad (29)$$

with $p = 1 \div P$ and $\alpha = 1 \div dim$.

- Calculate the weighted values $v_{p,\alpha}$:

$$v_{p,\alpha} = \tilde{z}_{p,\alpha} \cdot w_\alpha \quad (30)$$

with $\sum_{i=1}^{dim} w_i = 1$, $p = 1 \div P$ and $\alpha = 1 \div dim$.

Where w_α denotes the relative importance of the α th objective function, and it's taken equal to $\frac{1}{3}$ for all the objectives.

- Determine the ideal solution and negative ideal solution:

$$A^* = \{v_1^*, v_2^*, v_3^*\} = \max_p(v_{p,\alpha} | \alpha = 1 \div dim) \quad (31)$$

$$A^- = \{v_1^-, v_2^-, v_3^-\} = \min_p(v_{p,\alpha} | \alpha = 1 \div dim) \quad (32)$$

- Compute the separation measures, according to the Euclidean distance. The separation of each candidate solution from the ideal one is given by

$$D_p^* = \sqrt{\sum_{\alpha=1}^{dim} (v_{p,\alpha} - v_\alpha^*)^2} \quad (33)$$

Likewise, the separation from the negative ideal one is given by

$$D_p^- = \sqrt{\sum_{\alpha=1}^{dim} (v_{p,\alpha} - v_\alpha^-)^2} \quad (34)$$

with $p = 1 \div P$.

- Calculate the relative closeness to the ideal solution. For each candidate solution the relative closeness with respect to A^* is defined as

$$C_p^* = \frac{D_p^-}{D_p^* + D_p^-} \quad (35)$$

with $p = 1 \div P$.

- Rank the solutions in increasing order. The solution z_p^* whose value of C_p^* is the highest is the best compromise solution for the TOPSIS method.

References

- [1] A. Al-ani, Feature subset selection using ant colony optimization, *Int. J. Comput. Intell.* (2005) 53–58.
- [2] P. Baraldi, G. Gola, E. Zio, D. Roverso, M. Hoffman, A randomized model ensemble approach for reconstructing signals from faulty sensors, *Exp. Syst. Appl.* 38 (8) (2011).

- [3] P. Baraldi, F. Di Maio, L. Pappalione, E. Zio, R. Seraoui, Condition monitoring of power plant components during operational trajectories, *Proc. Inst. Mech. Eng. Part O: J. Risk Reliab.* 226 (6) (2012) 568–583.
- [4] P. Baraldi, F. Cannarile, F. Di Maio, E. Zio, Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions, *Eng. Appl. Artif. Intell.* 56 (2016) 1–13.
- [5] P. Baraldi, F. Di Maio, P. Turati, E. Zio, Robust signal reconstruction for condition monitoring of industrial components via a modified auto associative kernel regression method, *Mech. Syst. Signal Process.* 60–61 (2015) 29–44.
- [6] P. Baraldi, F. Di Maio, D. Genini, E. Zio, Comparison of data-driven reconstruction methods for fault detection, *IEEE Trans. Reliab.* 64 (3) (2015) 852–860.
- [7] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, NY, USA, 1995.
- [8] V.V. Bolotin, A.A. Shipkov, A model of the environmentally affected growth of fatigue cracks, *J. Appl. Math. Mech.* 62 (313) (1998).
- [9] P.P. Bonissone, F. Xue, R. Subbu, Fast meta-models for local fusion of multiple predictive models, *Appl. Soft Comput. J.* 11 (2) (2011) 1529–1539.
- [10] F. Cannarile, M. Compare, E. Zio, ADAMO: an ADaptive MOdeling approach for prognostics and health management Part I: fault diagnostics, *Mech. Syst. Signal Process.* (2015) (in preparation).
- [11] X. Chen, G. Xu, A self-adaptive alarm method for tool condition monitoring based on Parzen window estimation, *J. Vibroeng.* 15 (3) (2013) 1537–1545.
- [12] R. Chevalier, D. Provost, R. Seraoui, Assessment of statistical and classification models for monitoring EDF's assets, in: *Sixth American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation*, 2009.
- [13] J. Coble, J.W. Hines, Identifying optimal prognostic parameters from data: a genetic algorithm approach, in: *Annual Conference of the Prognostics and Health Management Society*, 2009.
- [14] J. Coble, *Merging Data Sources to Predict Remaining Useful Life – An Automated Method to Identify Prognostic Parameters*, University of Tennessee, 2010 (PhD Diss.).
- [15] J. Coble, J.W. Hines, Applying the general path model to estimation of remaining useful life, *Int. J. Prognostics Health Manage.* 2 (2011) 1–13.
- [16] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.* 6 (2) (2002) 182–197.
- [17] F. Di Maio, P. Baraldi, E. Zio, R. Seraoui, Fault detection in nuclear power plants components by a combination of statistical methods, *IEEE Trans. Reliab.* 62 (4) (2013) 833–845.
- [18] R. Dunia, S.J. Qin, T.F. Edgar, T.J. McAvoy, Identification of faulty sensors using principal component analysis, *AIChE J.* 42 (10) (1996) 2797–2811.
- [19] H.A. Firpi, E. Goodman, Swarmed feature selection, in: *Proceedings of the 33rd Applied Imagery Pattern Recognition Workshop*, 2005, pp. 112–118.
- [20] G. Guglielmi, T. Parisini, G. Rossi, Fault diagnosis and neural networks: a power plant application, *Contr. Eng. Pract.* 3 (5) (1995) 601–620.
- [21] M.F. Harkat, S. Djelal, N. Doghmane, M. Benouaret, Sensor fault detection, isolation and reconstruction using nonlinear principal component analysis, *Int. J. Autom. Comput.* 4 (2) (2007) 149–155.
- [22] X. He, Q. Zhang, N. Sun, Y. Dong, Feature selection with discrete binary differential evolution, in: *International Conference on Artificial Intelligence and Computational Intelligence, AICI 2009*, vol. 4, 2009, pp. 327–330.
- [23] W. He, Q. Miao, M. Azarian, M. Pecht, Health monitoring of cooling fan bearings based on wavelet filter, *Mech. Syst. Signal Process.* 64–65 (2015) 149–161.
- [24] J.W. Hines, D. Garvey, Development and application of fault detectability performance metrics for instrument calibration verification and anomaly detection, *J. Pattern Recogn. Res.* 1 (1) (2006) 2–15.
- [25] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. Roy. Soc. Lond. A: Math. Phys. Eng. Sci.* 454 (1971) (1998) 903–995.
- [26] G.B. Huang, L. Chen, C.K. Siew, Universal approximation using incremental extractive feedforward networks with random hidden nodes, *IEEE Trans. Neural Networks* 17 (4) (2006) 879–892.
- [27] G.B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, *IEEE Trans. Syst. Man Cybern. B Cybern.* 42 (2) (2012) 513–529.
- [28] G. Huang, G.B. Huang, S. Song, K. You, Trends in extreme learning machines: a review, *Neural Netw.* 61 (2015) 32–48.
- [29] H. Hunsop, *Kernel Based Density Estimation, Constrained and Hierarchical Density Estimation for Image Reconstruction and Sensor Networks*, ProQuest UMI Dissertation Publishing, 2011, ISBN 9781243584212.
- [30] T. Kaewkongka, Y. Au, R. Rakowski, B. Jones, A comparative study of short time fourier transform and continuous wavelet transform for bearing condition monitoring, *Int. J. COMADEM* 6 (2003) 41–48.
- [31] R.N. Khushaba, A. Al-ani, A. Al-Jumaily, Feature subset selection using differential evolution and a statistical repair mechanism, *Exp. Syst. Appl.* 38 (9) (2011) 11515–11526.
- [32] S. King, P.R. Bannister, D.A. Clifton, L. Tarassenko, Probabilistic approach to the condition monitoring of aerospace engines, *Proc. Inst. Mech. Eng. Part G: J. Aerosp. Eng.* 223 (5) (2009) 533–541.
- [33] S. Kumar, M. Pecht, Modeling approaches for prognostics and health management of electronics, *Int. J. Perform. Eng.* 6 (5) (2010) 467–476.
- [34] M. Marseguerra, E. Zio, P. Baraldi, A. Oldrini, Fuzzy logic for signal prediction in nuclear systems, *Prog. Nucl. Energy* 43 (1–4 SPEC) (2003) 373–380.
- [35] K. Medjaher, F. Camci, N. Zerhouni, Feature extraction and evaluation for health assessment and failure prognostics, in: *European Conference of the Prognostics and Health Management Society*, Dresden, Germany, 2012.
- [36] R. Mallipeddi, P.N. Suganthan, Empirical study on the effect of population size on differential evolution algorithm, in: *2008 IEEE Congress on Evolutionary Computation, CEC 2008*, art. 4631294, 2008, pp. 3663–3670.
- [37] E. Mezura-Montes, M. Reyes-Sierra, C.A. Coello, Multi-objective optimization using differential evolution: a survey of the state-of-the-art, *Stud. Comput. Intell.* 143 (2008) 173–196.
- [38] A. Mosallam, *Remaining Useful Life Estimation of Critical Components Based on Bayesian Approaches* (Ph.D. Thesis), 2014.
- [39] R. Muradore, P. Fiorini, A PLS-based statistical approach for fault detection and isolation of robotic manipulators, *IEEE Trans. Industr. Electron.* 59 (8) (2012) 3167–3175.
- [40] E. Myötyri, U. Pulkkinen, K. Simola, Application of stochastic filtering for lifetime prediction, *Reliab. Eng. Syst. Saf.* 91 (2) (2006) 200–208.
- [41] K. Nabeshima, T. Suzudu, K. Suzuki, E. Turcan, Real-time nuclear power plant monitoring with neural network, *J. Nucl. Sci. Technol.* 35 (2) (1998) 93–100.
- [42] B.E. Olivares, M.A. Cerda Muñoz, M.E. Orchard, J.F. Silva, Particle-filtering-based prognosis framework for energy storage devices with a statistical characterization of state-of-health regeneration phenomena, *IEEE Trans. Instrum. Meas.* 62 (2) (2013).
- [43] S. Opricovic, G.H. Tzeng, Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS, *Eur. J. Oper. Res.* 156 (2) (2004) 445–455.
- [44] G.F. Oswald, G.I. Schueller, Reliability of deteriorating structures, *Eng. Fract. Mech.* 20 (3) (1984) 479–488.
- [45] M. Pecht, *Prognostics and Health Management of Electronics*, John Wiley & Sons, 2008.
- [46] H. Qiu, J. Lee, J. Lin, G. Yu, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, *J. Sound Vib.* 289 (4–5) (2006) 1066–1090.
- [47] M. Rigamonti, P. Baraldi, E. Zio, I. Roychoudhury, K. Goebel, S. Poll, *Echo State Network for the Remaining Useful Life Prediction of a Turbofan Engine Under Variable Operating Conditions*, 2015.
- [48] A. Salman, A.P. Engelbrecht, M.G.H. Omran, Empirical analysis of self-adaptive differential evolution, *Eur. J. Oper. Res.* 183 (2) (2007) 785–804.
- [49] B. Samanta, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, *Mech. Syst. Signal Process.* 18 (3) (2004) 625–644.
- [50] A. Saxena, K. Goebel, D. Simon, N. Eklund, Damage propagation modeling for aircraft engine run-to-failure simulation, in: *Prognostics and Health Management, PHM 2008, International Conference on, IEEE*, 2008, pp. 1–9.
- [51] A. Saxena, J. Celaya, B. Saha, S. Saha, K. Goebel, Metrics for offline evaluation of prognostic performance, *Int. J. Prognostic Health Manage.* (2010).

- [52] M. Schwabacher, K. Goebel, A survey of artificial intelligence for prognostics, in: AAAI Fall Symposium Arlington, Virginia, USA, 2007.
- [53] H. Skima, K. Medjaher, C. Varnier, E. Dedu, J. Bourgeois, Hybrid prognostic approach for micro-electro-mechanical systems, in: Aerospace Conference, 2015, IEEE, 2015, pp. 1–8.
- [54] R. Sikora, S. Piramuthu, Framework for efficient feature selection in genetic algorithm based data mining, *Eur. J. Oper. Res.* 180 (2) (2007) 723–737.
- [55] K. Sobczyk, B.F. Spencer Jr, Random fatigue: from data to theory, *J. Eng. Mech.* 119 (2) (1993).
- [56] E. Sutrisno, H. Oh, A. Vasani, M. Pecht, Estimation of remaining useful life of ball bearings using data driven methodologies, in: Prognostics and Health Management, PHM 2012, International Conference on, IEEE, USA, 2012.
- [57] T. Tušar, B. Filipič, Differential evolution versus genetic algorithms in multiobjective optimization, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) LNCS, vol. 4403, 2007, pp. 257–271.
- [58] G. Vachtsevanos, F.L. Lewis, M. Roemer, A. Hess, B. Wu, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*, 1st ed., John Wiley & Sons New Jersey, Hoboken, 2006.
- [59] V. Venkatasubramanian, Prognostic and diagnostic monitoring of complex systems for product lifecycle management: challenges and opportunities, *Comput. Chem. Eng.* 29 (6) (2005) 1253–1263.
- [60] P. Wang, G. Vachtsevanos, Fault prognosis using dynamic wavelet neural networks, in: AUTOTESTCON Proceedings, IEEE Systems Readiness Technology Conference, 2001, pp. 857–870.
- [61] X. Wang, G. Yu, J. Lee, Wavelet neural network for machining performance assessment and its implication to machinery prognostic, in: 5th International Conference on Managing Innovations in Manufacturing (MIM), Milwaukee, Wisconsin, USA, 2002, pp. 150–156.
- [62] W.Q. Wang, M.F. Goldnaraghi, F. Ismail, Prognosis of machine health condition using neuro-fuzzy systems, *Mech. Syst. Signal Process.* 18 (4) (2004) 813–831.
- [63] Y. Wang, Z. Cai, Q. Zhang, Differential evolution with composite trial vector generation strategies and control parameters, *IEEE Trans. Evol. Comput.* 15 (1) (2011) 55–66, art. 5688232.
- [64] B. Wang, K.E. Merrick, H.A. Abbass, Co-operative coevolutionary neural networks for mining functional association rules, *IEEE Trans. Neural Netw. Learn. Syst.* (in press)
- [65] K. Worden, W. Staszewski, J. Hensman, Natural computing for mechanical systems research: a tutorial overview, *Mech. Syst. Signal Process.* 25 (1) (2011) 4–111.
- [66] J. Yan, M. Koç, J. Lee, A prognostic algorithm for machine performance assessment and its application, *Prod. Plann. Contr.* 15 (8) (2004) 796–801.
- [67] X. Zhang, R. Xu, C. Kwan, S.Y. Liang, Q. Xie, L. Haynes, An integrated approach to bearing fault diagnostics and prognostics, in: Proceedings of the American Control Conference, vol. 4, 2005, pp. 2750–2755.
- [68] C. Zhao, F. Gao, Fault-relevant principal component analysis (FPCA) method for multivariate statistical modelling and process monitoring, *Chemom. Intell. Labor. Syst. Syst.* 133 (2014) 1–16.
- [69] E. Zio, P. Baraldi, N. Pedroni, Selecting features for nuclear trajectories classification by means of genetic algorithms, *IEEE Trans. Nucl. Sci.* 53 (3) (2006) 1479–1493.
- [70] E. Zio, P. Baraldi, W. Zhao, Confidence in signal reconstruction by the evolving clustering method, in: Prognostics and System Health Management Conference, PHM-Shenzhen, 2011.
- [71] E. Zio, Prognostics and health management of industrial equipment, *Diagn. Progn. Eng. Syst.: Methods Tech.* 17 (2013) 333–356.
- [72] E. Zitzler, M. Laumanns, L. Thiele, SPEA2: Improving the Performance of the Strength Pareto Evolutionary Algorithm, Technical Report 103, Computer Engineering and Communication Networks Lab (TIK), Swiss Federal Institute of Technology (ETH), Zurich, 2001.
- [73] E. Zitzler, L. Thiele, M. Laumanns, C.M. Fonseca, V.G. Da Fonseca, Performance assessment of multiobjective optimizers: an analysis and review, *IEEE Trans. Evol. Comput.* 7 (2) (2003) 117–132.
- [74] E. Zitzler, S. Künzli, Indicator-based selection in multiobjective search, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 3242, 2004, pp. 832–842.
- [75] S. Abbasion, A. Rafsanjani, A. Farshidianfar, N. Irani, Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine, *Mech. Syst. Signal Process.* 21 (7) (2007) 2933–2945, ISSN 0888-3270.
- [76] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, *Data Min. Knowl. Disc.* 2 (2) (1998) 121–167.
- [77] M. Hubert, P.J. Rousseeuw, K. Vanden Branden, ROBPCA: A new approach to robust principal component analysis, *Technometrics* 47 (1) (2005) 64–79.
- [78] S. Lin, X. Liu, J. Fang, Z. Xu, Is extreme learning machine feasible? A theoretical assessment (Part II), *IEEE T. Neu. Net. Lear.* 26 (1) (2015).
- [79] X. Liu, S. Lin, J. Fang, Z. Xu, Is extreme learning machine feasible? A theoretical assessment (Part I), *IEEE T. Neu. Net. Lear.* 26 (1) (2015).
- [80] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, B. Schölkopf, An introduction to kernel-based learning algorithms, *IEEE T. Neural Networ.* 12 (2) (2001) 181–201.
- [81] Y. Pan, J. Chen, X. Li, Bearing performance degradation assessment based on lifting wavelet packet decomposition and fuzzy c-means, *Mech. Syst. Signal Process.* 24 (2) (2010) 559–566, ISSN 0888-3270.
- [82] M. Rigamonti, P. Baraldi, E. Zio, A. Alessi, D. Astigarraga, A. Galarza, Identification of the degradation state for condition-based maintenance of insulated gate bipolar transistors: A self-organizing map approach, *Microelectron. Reliab.* 60 (2016) 48–61, ISSN 0026-2714.