Enabling Health Monitoring Approach Based on Vibration Data for Accurate Prognostics

Kamran Javed, Rafael Gouriveau, *Member, IEEE,* Noureddine Zerhouni, *Member, IEEE*, Patrick Nectoux

***Abstract*—Performances of data-driven prognostics approaches are closely dependent on form, and trend of extracted features. Indeed, features that clearly reflect the machine degradation, should lead to accurate prognostics, which is the global objective of the paper. This paper contributes a new approach for features extraction / selection: the extraction is based on trigonometric functions and cumulative transformation, and the selection is performed by evaluating feature fitness using monotonicity and trendability characteristics. The proposition is applied to time-frequency analysis of non-stationary signals using Discrete Wavelet Transform. The main idea is to map raw vibration data into monotonic features with early trends, which can be easily predicted. To show that, selected features are used to build a model with a data-driven approach namely, the Summation Wavelet-Extreme Learning Machine, that enables a good balance between model accuracy and complexity. For validation and generalization purpose, vibration data from two real applications of Prognostics and Health Management challenges are used: 1) cutting tools from Computer Numerical Control (CNC) machine (2010), and 2) bearings from platform PRONOSTIA (2012). Performances of the proposed approach are thoroughly compared with the classical approach by performing: feature fitness analysis, cutting tool wear "estimation" and bearings "long-term predictions" tasks, which validates the proposition.**

***Index Terms*—Prognostics, monitoring, feature extraction, data-driven.**

I. Introduction

A Prognostics and Health Management system is usually described as the combination of 7 modules that collectively enable linking failure mechanisms with life management (Fig. 1). PHM aims at extending the life cycle of an engineering asset, while reducing exploitation and maintenance costs. For that purpose, prognostics is considered as a key process with future capabilities. In brief, it relies on the usage of condition monitoring (CM) data from operating machinery (acoustic signal, force, vibration, temperature, etc.,) to obtain useful features, to assess the level of degradation, and to predict the evolution of phenomena [1]. The predicted time before a failure is known as the Remaining Useful Life (RUL). The accuracy of RUL estimates is required for further plan of actions to increase safety, minimize downtime, ensure mission completion and efficient production [2]. More precisely, predicting the behavior of critical machinery like bearings [3], cutting tools [4] is a challenging task while handling raw vibration data due to its inherent non-linearity. The degradation of such critical components can also affect the system as a whole and cause major breakdowns or huge wastes [5].

K. Javed, R. Gouriveau, N. Zerhouni & P. Nectoux are with FEMTO-ST Institute (Dept AS2M), UMR CNRS 6174 - UFC / ENSMM / UTBM, 24 rue Alain Savary, Besançon, France. e-mail: firstname.lastname@femto-st.fr

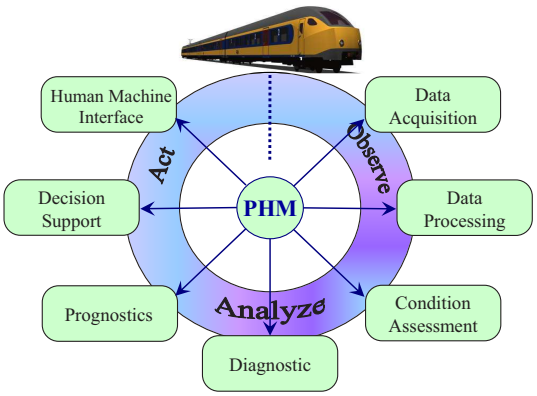


Fig. 1. Prognostics and Health Management cycle adapted from [15]

According to literature, prognostics may face various situations regarding collected information and data from the past, present or future behavior. Among different approaches for prognostics (model-based, data-driven, hybrid [6]–[8]), data-driven approaches are easier to deploy when its hard to understand first principles of a complex machinery to build a diagnostics or prognostics model [9], [10]. They are black-box models that learn system behavior directly from CM data, use that knowledge to infer its current state and to predict future progression of failure to estimate RUL.

Generally, modeling of data-driven prognostics has to go through necessary steps of learning and testing. Firstly, raw data are collected from machinery and pre-processed to extract useful features to learn degradation behavior. Secondly, in the test phase, the learned model is used to predict future behavior and to validate model performance [10]. Most importantly, the quality of extracted features has direct effect on performances of a learned model. Obviously, features that properly reflect failure progression may lead to simple prognostics [11], [12]. In this context, two main problems can be highlighted.

1) Even if most of data-driven approaches are able to cater non-linearity of degrading signals, features with monotonic behavior are likely to lead to better RUL estimates.
2) Some of classical features do not correlate to degradation phenomena and are unable to show variation until failure, which prevents timely RUL estimation [13], [14].

Such situations prevent timely RUL estimates to recommend actions for maintenance and system configuration. Therefore, in this paper mainly, the aspects related to usefulness of features for prognostics are addressed. Developments are focused on improving accuracy of prognostics by proposing a new approach of feature extraction / selection using vibration data.

The main contributions of this paper are as follows.

- Features extraction is performed by using trigonometric functions and cumulative transformation;
- Features selection is performed by assessing their fitness using monotonicity and trendability metrics;
- Generalization of the approach on two real applications.

The paper is organized as follows. In section II, the importance of features extraction / selection is addressed and limits are identified. Section III is dedicated to the proposed approach of features extraction / selection, and prediction modeling to improve accuracy of prognostics. Section IV demonstrates performances of proposed approach on vibration data of cutting tools and bearings. Finally, section V concludes this work.

II. Background - feature extraction / selection

Relevant information related to degrading machinery is often hidden in raw data and should be extracted. Therefore, it is also important to identify features that are sensitive to machine condition and clearly reflect failure progression, to serve need of modeling [11], [16]. Considering importance of such aspects, following topics address recent literature on features extraction / selection and highlight important issues.

A. Features extraction

In literature, a large number of signal processing techniques have been proposed. Prior to any selection among different possibilities, it is required to investigate an appropriate method for a specific application. However, there is vast literature on this topic, which is beyond the scope of our paper. Let us highlight three main categories of features extraction [7], [16].

1) Time domain: Time domain features extraction is directly performed on the sensed waveforms (e.g. acoustic emissions, vibration signal) to identify signatures. Time-domain approach extracts features using statistics like mean, variance, standard deviation, etc. They are suitable for fault detection and applied to stationary signals. Otherwise, extracted features may show sensitivity to variation in data and inherit nonlinearity, which complicates prognostics [16].

2) Frequency domain: Frequency domain techniques are considered more effective for fault diagnostic, because, they have good ability to identify and isolate frequency components. The most widely applied technique in this category is Fast Fourier Transform. Other methods that belong to this category are cepstrum, spectral analysis, higher-order spectra or envelop analysis [7], [16], [17]. The main limitation of such techniques is their inability to deal with non-stationary signals, unfortunately which is the case in degrading machinery.

3) Time-Frequency: Time-frequency techniques aim at investigating signals in both time and frequency domains. They are considered to be powerful to analyze non-stationary signals. Some of popular time-frequency techniques proposed in literature are: Short Time Fourier Transform (STFT) [16], Wavelet Transform (WT) [18], Empirical Mode Decomposition (EMD) [19], etc.

According to literature [20], EMD and WT are the two outstanding examples among signal processing techniques since the last two decades. However, the main weakness of EMD is high sensitivity to noise, and it also runs into the problem of mixing modes [21]. In addition, EMD is also reported to have characteristics like wavelet [22], which encourages to use WT as a substitute in studying the behavior of the time-frequency signals [23]. Moreover, EMD is popular in demodulation applications, whereas WT is commonly used in vibration content characterization [20] and has better applicability [4], especially when vibration data come from rotating machinery like bearings [24], [25] or cutting tools [4], [26]. However, as far as authors know, even application WT cannot guarantee ideal features for prognostics applications, and its performances can vary from case to case.

B. Features selection

Feature dimensionality can be reduced in two ways:

1) By drawing features in a new space with methods like: Principal Component Analysis, Singular Value Decomposition, Self-Organizing Map, or clustering [26];
2) By selecting a feature subset based on essential characteristics like monotonicity and trendability [2], [14] .

Leaving aside conventional approaches for dimensionality reduction, recent works confirm that features selection by essential characteristics in later case are vital to prognostics [2], [11], [13], [14]. In literature, two simple metrics are devised to assess quality of features prior to selection.

- "Monotonicity" characterizes increasing or decreasing trend. It can be measured by absolute difference of "positive" and "negative" derivatives for each feature [2]:

$$M = \left| \frac{no.\ of\ d/dx > 0}{N-1} - \frac{no.\ of\ d/dx < 0}{N-1} \right| \qquad (1)$$

  where "N" represents observations, $M = 1$ is for highly monotonic features and $M = 0$ is vice versa.
- "Trendability" is related to the functional form of a feature and its correlation to time, i.e., how the continuous state of a machinery changes with time [13]:

$$R = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}} \qquad (2)$$

  where, $R \in [-1; 1]$ is the correlation coefficient between feature x and time index y.

C. Problem statement

As raw vibration data exhibit great amount of hidden noise, it can be very challenging to extract features that explicitly reflect failure progression or have essential characteristics like monotonicity and trendability. Most importantly, features are extracted to make inference about health of machinery. They can be seen as time series health indicators that can be nonlinear, noisy or smooth, which may affect performances of a prognostics model. For understanding, consider Fig. 2, where effects of features with different characteristics are presented. Obviously, feature that clearly reflect failure progression (i.e., monotonic and trendable), up to current time t_c may lead to an accurate RUL estimate and with less uncertainty because the predicted degradation is expected to grow in a same fashion

till it reaches failure threshold at time t_{f1}. On the other hand, when a non-monotonic and non-trendable feature is presented to a prognostics model, its performances may impair, or rather, it could be impossible for the model to predict future unknown from current time tc to t_{f2}, leading to large uncertainties that risk decision making. Therefore, it is strictly required to extract features, that not only simplify prognostics modeling, but also lead to accurate RUL estimates. This is the aim of the proposed data-processing approach presented hereafter.

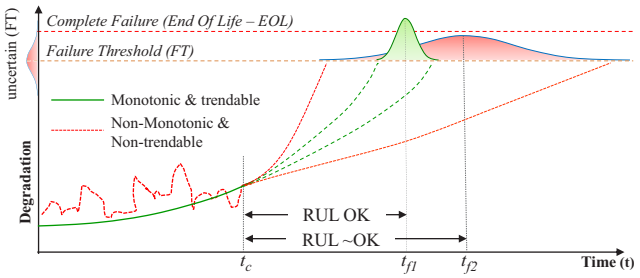


Fig. 2. Effect of features on prognostics results

III. Improving Accuracy of Prognostics

A. Outline: from raw data to accurate predictions

The proposed methodology is demonstrated on vibration data, which are a kind of signal widely used for prognostics, although RUL estimation is difficult to perform [27].
Firstly, time-frequency analysis is performed by applying Discrete Wavelet Transform to raw vibration data. Following that, rather than a classical approach [26], [27], features are extracted / selected (from a decomposed signal) in a new manner: firstly, trigonometric functions are applied to extract features and then smoothed to remove unwanted noisy part. Secondly, extracted features are further transformed into their respective cumulative features by performing a running total and simultaneous scaling, to obtain feature having monotonic behavior and early trend. Note that, cumulative transformation can also be performed on classical features like Root Mean Square (RMS), energy, etc. Thirdly, multivariate data (of extracted features) are analyzed for fitness. A complete representation of proposed data processing procedure is shown in Fig. 3. To show the benefit of our proposition, selected features are used to build a model with a data-driven approach to perform estimation / prediction tasks namely, the Summation Wavelet-Extreme Learning Machine (SW-ELM). Performances with proposed approach are discussed in section IV.

B. Data-Preprocessing to obtain suitable features

The proposed approach is elaborated as follows.

1) Discrete Wavelet Transform and Trigonometric Features: Discrete Wavelet Transform (DWT) is achieved by discretization of Continuous Wavelet Transform, that has the drawback of impracticality with computers. An important implementation of DWT is known as Multi-Resolution Analysis (MRA), which is accomplished by two functions: scaling and wavelet

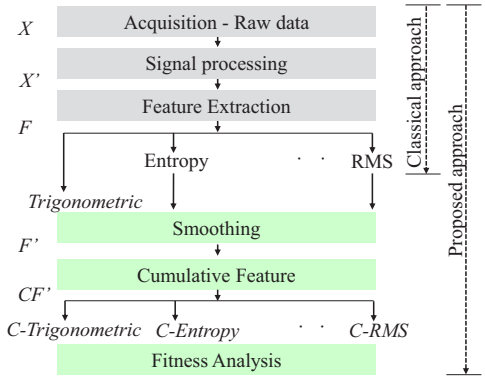


Fig. 3. Proposed approach to obtain monotonic and trendable features

that are associated with a low pass filter (L_{PF}) and a high-pass filter (H_{PF}) respectively. At first level of decomposition, a signal is passed through a L_{PF} that gives approximation coefficients (A), and a H_{PF} that gives detail coefficients (D), followed by a down sampling by 2. The frequency band can be further broken down into lower resolutions by recursive decomposition of approximation part at current level [24], [27], [28]. For applications like vibration signal, MRA has 3 requirements: mother wavelet, decomposition level and features to be extracted.

a) Classical features: the classical approach for feature extraction with DWT is performed at a required level of decomposition (and using approximation coefficients) to extract different features. The extracted features can be entropy, RMS or other statistical features [26], [29]. However, each feature can be sensitive to different faults or the severity of degradation. For example, vibration based features RMS and kurtosis from the degraded bearings show variation only few time before failure (i.e., rate of change increases significantly), which can limit the use of such features for prognostics. Therefore, a new set of features is introduced in the next topic.

b) Proposed trigonometric features: in this case, at a required level of decomposition (of vibration data), features extraction is performed by using a combination of statistics and trigonometric functions. Mainly, the trigonometric functions can be either monotonically increasing or decreasing, e.g. inverse hyperbolic sine, etc. In this context, they can be grouped in two classes.

- Functions that have domain $(-\infty, \infty)$, e.g. inverse hyperbolic sine (asinh), inverse tangent (atan), etc;
- Functions that have different domain value but not $(-\infty, \infty)$, e.g. inverse hyperbolic cosine (acosh), etc.

For the second class, input values outside the domain are transformed to complex outputs which can be further explored. However, the first class appears to be more relevant to a real data, due to domain values from $(-\infty, \infty)$. Therefore, we limit the study to the first class only.
Mainly, benefit of using trigonometric functions is that: they transform the raw input data to different scale thereby, obtained features have better trends and low scale as compared to classical features. To achieve that, a trigonometric function operates on array (X) element-wise ($x_j,\ j = 1, 2, .., n$) to

scale, and the standard deviation (SD) applied to the scaled array for extracting feature value. Different features extracted from vibration data are listed in Table I.

TABLE I
FEATURES EXTRACTED FROM REQUIRED APPROXIMATION LEVEL

Proposed Trigonometric Features	
SD of asinh(X)	$\sigma\left(log\left[x_j+(x_j^2+1)^{1/2}\right]\right)$
SD of atan(X)	$\sigma\left(\frac{i}{2}log\left(\frac{i+x_j}{i-x_j}\right)\right)$
Classical Features	
Entropy (threshold)	$E=\sum_{j=1}^{n}E(x_j)$
Energy	$e=\sum_{j=1}^{n}x_j^2$
Root mean square (RMS)	$RMS=\sqrt{\frac{1}{n}(x_1^2+...+x_n^2)}$
Kurtosis	$\frac{\sum_{j=1}^{n}(x_j-\bar{X})^4}{(n-1)\sigma^4}$
Upper bound	$max(X)+\frac{1}{2}\frac{max(X)-min(X)}{n-1}$

2) Feature smoothing: to reduce variability of extracted features and to filter unwanted noise, smoothing process is performed on each feature F to capture important trends. This step is met by applying a local regression (LOESS) filter with a span value 0.3 (i.e., 30 %). In brief, LOESS is a popular smoothing method based on locally weighted regression function and a second degree polynomial. Given some scattered data, LOESS filter can locally perform a weighted fit with the n-nearest points (see [30] for details).

3) Proposed cumulative features: any machine is bound to degrade as time grows. Thereby, the main idea of this step is obtain health indicators that can clearly reflect failure progression and satisfy requirements mentioned in II-C. In fact, features with non-linear behavior don't clearly indicate state of the machinery under operation, i.e., to distinguish among good, degrading or failure states. Note that, such non-linearity can also represent self-healing. Here we assume that self-healing it is not possible in case of components like bearings, gears, or cutting tools. However, this assumption does not hold for batteries that may undergo self-recovery during non use [2].

A straightforward but an effective strategy is proposed to build health indicators, that aims at transforming an extracted feature into its respective cumulative form, to have monotonicity and trendability characteristics. The feature transformation task is achieved by applying a cumulative function on a given time series, on which a point-wise running total and scaling operations are performed simultaneously to obtain cumulative feature, that can clearly distinguish among different states of degrading machinery:

$$C\acute{F}_v=\frac{\sum_{i=1}^{N}F_v\acute{}(i)}{\left|\left(\sum_{i=1}^{N}F_v\acute{}(i)\right)\right|^{1/2}}, v=1,2....,k \tag{3}$$

where, $\sum_{i=1}^{N}\acute{F}_v(i)$ represents the running total of a "smoothed vth" feature $\acute{F}_v$ up to "N" points, and $C\acute{F}_v$ represents transformed cumulative feature. It should be noted that, the cumulation of a feature can be sensitive to noise, therefore, features smoothing must be performed *a priori* (section III-B2).

4) Features fitness analysis: the choice of a features subset is essential for any prognostics model. Therefore, features selection is performed on the basis of monotonicity and trendability characteristics, i.e., by retaining the subset of cumulative features with higher M (Eq. (1)) and R (Eq. (2)). Those features should lead to accurate RUL estimates.

C. Building a prediction model

1) Data-driven approach: the accuracy of a prognostics model is related to its ability to predict future states of equipment, where prediction is not only dependent on model but observations as well. In the data-driven category of prognostics, a number of techniques are proposed like Bayesian approaches, Artificial Neural Networks (ANNs), etc. Among these methods, ANNs are a special case of adaptive networks, that are widely used among machine learning methods in PHM domain [31]. Although, several learning schemes for ANNs exist [32], relatively a new algorithm was proposed for a Single Layer Feed Forward Networks (SLFN), namely the Extreme learning Machine (ELM) [33]. ELM avoids slow iterative tuning and requires one-pass to train SLFN, by randomly choosing hidden layer parameters (weights and bias) and analytically calculating output layer weights. However, performances of ELM can suffer due to random initialization of parameters and the type of activation functions in hidden layer. Such issues can increase the complexity of model and may lead to ill-condition [34]. To overcome limitations of ELM without compromising its better applicability, an improved variant is presented.

2) The Summation Wavelet-Extreme Learning Machine: basically, SW-ELM combines ANN and wavelet theory and appears to be an effective prediction approach [34]. Like ELM, SW-ELM is a tuning free one-pass algorithm for SLFN, where hidden node parameters are not only independent of training but also each other. SW-ELM differs from ELM, mainly, in structure, activation functions and parameters initialization.

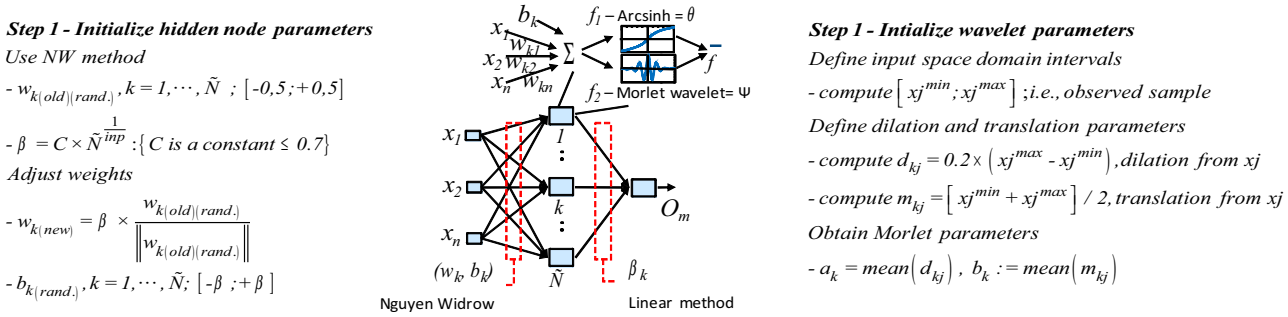
a) Structure: the basic structure of SW-ELM is composed of three fully connected layers of neurons (input, hidden and output layers). As compared to ELM, the main differences of SW-ELM structure are the following (see Fig. 4).

- Each hidden neuron holds a parallel conjunction of two distinct activation functions (f_1 and f_2), where the output from a hidden node is the average value from dual activations ($\bar{f}=(f_1+f_2)/2$).
- For better convergence of the algorithm, an inverse hyperbolic sine (Eq. (4) [35]) and a Morlet wavelet (Eq. (5) [36]) are used.

$$f_1=\theta(X)=log\left[x+(x^2+1)^{1/2}\right] \tag{4}$$

$$f_2=\psi(X)=cos(5x)\,e^{(-0.5x^2)} \tag{5}$$

Let note n and m the numbers of inputs and outputs, N the number of learning data samples (x_i,t_i), where $i\in[1\dots N]$, $x_i=[x_{i1},x_{i2},...,x_{in}]^T\in\Re^n$ and $t_i=[t_{i1},t_{i2},...,t_{im}]^T\in\Re^m$, and $\tilde{N}$ the number of hidden (hid.) nodes, each one with



Step 2 and Step 3 - Adjust linear parameters : hidden to output layer

Compute the hidden layer output matrix H_{avg} *and the output weights* $\beta = H_{avg}^{T}.T$

Fig. 4. SW-ELM structure and learning algorithm

activation functions (f_1 and f_2). For each sample j, the output o_j is mathematically expressed as:

$$\sum_{k=1}^{\tilde{N}} \beta_k \bar{f}\left[(\theta,\psi)\left(w_k.x_j + b_k\right)\right] = o_j \ , j = 1, 2, ..., N \tag{6}$$

where $\bar{f}$ is the average output from two different activation functions θ and ψ. $w_k = [w_{k1}, w_{k2}, ..., w_{kn}]^T \in \Re^n$ is an input weight vector connecting the k^{th} hidden to input layer neurons, $(w_k.x_j)$ is the inner product of weights and inputs, and $b_k \in \Re$ is the bias of k^{th} neuron of hidden layer. Also, $\beta_k = [\beta_{k1}, \beta_{k2}, ..., \beta_{km}]^T \in \Re^m$ is the weight vector to connect the k^{th} neuron of hidden layer and output neurons. To minimize the difference between output o_j and target t_j, $\sum_{j=1}^{\tilde{N}} \|o_j - t_j\| = 0$, there exist β_k, w_k and b_k such that:

$$\sum_{k=1}^{\tilde{N}} \beta_k \bar{f}\left[(\theta,\psi)\left(w_k.x_j + b_k\right)\right] = t_j \ , j = 1, 2, ..., N \tag{7}$$

which can be expressed in matrix form as,

$$H_{avg}\beta = T \tag{8}$$

where H_{avg} is a $\left[N \times \tilde{N}\right]$ matrix expressed as,

$$H_{avg}\left(w_1, \ldots, w_{\tilde{N}}, x_1, \ldots, x_{\tilde{N}}, b_1, \ldots, b_{\tilde{N}}\right) =$$

$$\bar{f}\left(\theta,\psi\right)\begin{bmatrix} (w_1.x_1 + b_1) & \ldots & (w_{\tilde{N}}.x_1 + b_{\tilde{N}}) \\ \vdots & \ldots & \vdots \\ (w_1.x_N + b_1) & \ldots & (w_{\tilde{N}}.x_N + b_{\tilde{N}}) \end{bmatrix} \tag{9}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N}\times m} \qquad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N\times m} \tag{10}$$

Finally, the least square solution of the linear system defined in Eq. (8), with minimum norm of output weights β is:

$$\hat{\beta} = H_{avg}^{\dagger}T = \left(H_{avg}^T H_{avg}\right)^{-1} H_{avg}^T T \tag{11}$$

b) Learning scheme: main learning phase derives from Eq. (8) and (9). However, it is required to properly perform parameters initialization task and to provide a better starting point to algorithm. Two types of parameters have to be considered: those from the wavelets (dilation and translation) adapted by a heuristic procedure [37], and those from the SLFN (weights and bias for input-hidden nodes), initialized by Nguyen Widrow (NW) procedure [38]. Complete learning scheme is synthesized in Fig. 4. Details can be found in [34].

IV. Experiments, Results and Discussion

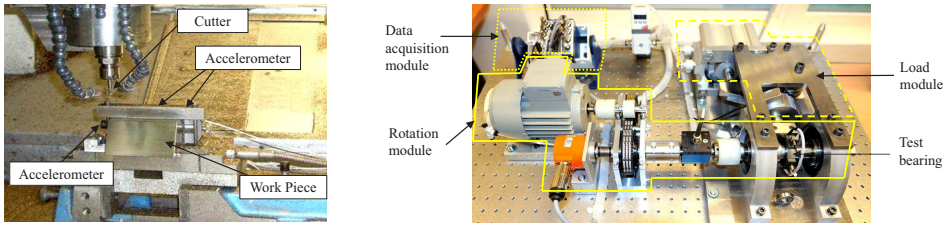
A. PHM challenge datasets

To demonstrate the effectiveness of our contributions, we consider vibration data from two real applications under constant operating conditions: 1) cutting tools from CNC machine [39] (Fig. 5a), and 2) ball bearings from experimental platform PRONOSTIA [25], [40] (Fig. 5b). Key features of both applications are summarized in Table II, and a brief introduction is given below.

- Cutting tools are used for an extremely dynamical cutting process. The *in situ* monitoring during the cutting process can give important information about tool condition, process itself, work piece surface quality and even machine condition [4]. CM systems for the cutting process are normally based on the measurements of vibration, acoustic emission and cutting force. However, vibration measurement benefit from: wide frequency range, less restrictive conditions, and easy to implement [41].
- Bearings are of great importance, because, rotating machinery often includes bearings inspections and replacements, which implies high maintenance costs. However, it is hard to evaluate model performance due to inherent non-linearity in features extracted from raw vibration data [3], [42]. In this context, the platform PRONOSTIA is dedicated to test and validate fault detection, diagnosis and prognostics methods on ball bearings. It allows performing accelerated degradations of bearings by constant and / or variable operating conditions, while gathering CM data (load force, speed, vibration and temperature).

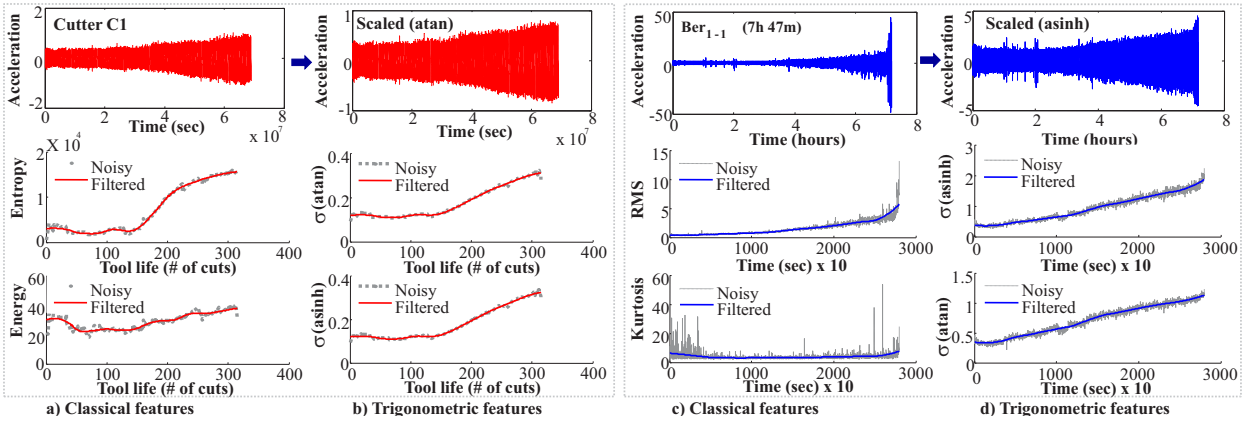
TABLE II
KEY FEATURES OF PHM CHALLENGE DATASETS FROM TWO APPLICATIONS

Validation data	PHM Challenge 2010	PHM Challenge 2012
Experimental platform / component	*Röders* Tech RFM760 / cutters	PRONOSTIA / bearings
Sampling frequency	50kHz	25.6kHz
Failure limit	$165(\times 10^{-3} mm)$ wear level	20g vibration level
Learning set Operating condition / component	$(10400rpm)$ / $C1$ $C4$ $C6$	$(1800rpm$ & $4000N)$ / Ber_{1-1} Ber_{1-2} $(1650rpm$ & $4200N)$ / Ber_{2-1} Ber_{2-2} $(1500rpm$ & $5000N)$ / Ber_{3-1} Ber_{3-2}
Testing set Operating condition / component	$(10400rpm)$ / $C2$ $C3$ $C5$	$(1800rpm$ & $4000N)$ / Ber_{1-3} to Ber_{1-7} $(1650rpm$ & $4200N)$ / Ber_{2-3} to Ber_{2-7} $(1500rpm$ & $5000N)$ / Ber_{3-3}



a) CNC machine, work piece & sensors b) PRONOSTIA bearing testbed

Fig. 5. CNC machine SIMTECH Institute and PRONOSTIA testbed - FEMTO-ST Institute, AS2M department



a) Classical features b) Trigonometric features c) Classical features d) Trigonometric features

Fig. 6. Classical features vs trigonometric features extracted from: a),b) cutting tool (3rd level approximation), and c),d) bearing (4rd level approximation)

B. Feature extraction and selection results

As mentioned in III-B1, decomposition of vibration signal require selection of mother wavelet, and decomposition level. Therefore, as suggested in literature, for cutting tool application Daubichies wavelet of 4th order (db4) [26], and 3rd [43] level of decomposition was used, whereas for bearings db4 and 4th level was considered [27], *prior* to features extraction.

1) Classical features vs. trigonometric features: here, we compare performances of trigonometric features with classical ones, on cutter $C1$ (from CNC machine) and bearing Ber_{1-1} (from PRONOSTIA), see Fig. 6a and Fig. 6c. For both cases, vibration data appear to be noisy with low trendability. Especially for bearing Ber_{1-1} the vibration signal is almost constant until 4th hour, but grows suddenly at the end. Results in Fig. 6a and Fig. 6c show that classical features from both cases ($C1$ and Ber_{1-1}) have low monotonicity / trendability and high noise / scales. Therefore, consider now the proposition of features extraction using a combination of SD and trigonometric functions (Table I). Results in Fig. 6b and Fig. 6d, show that trigonometric features clearly reflect failure progression with high monotonicity and trendability, and also have lower scales as compared to classical features.
Back to the accuracy of prognostics, one can point out that classical features (RMS, Kurtosis, etc.) are not well adapted to catch machine conditions. Also, they can have large scales, which require normalization before feeding a prognostics model. This strengthens the interest of trigonometric features.

2) Classical features vs. cumulative features: it is quite challenging to obtain monotonic and trendable features in cases where raw data do not show clear progression of failure. Cumulative transformation is a straightforward and effective approach to counteract this problem. In brief, following a smoothing task (by LOESS filter), all features (listed in Table I) extracted by classical and proposed approaches were transformed to build respective cumulative features (Eq. (3)). To highlight the improvements, fitness analysis is performed to

compare classical features extraction procedure and proposed approach Fig. 3. Both approaches are thoroughly examined on complete data of cutting tools (from CNC) and bearings (from PRONOSTIA) by assessing trendability (R) and monotonicity (M) characteristics of extracted features. Mean performances by each approach are summarized in Table III and Table IV. According to results from both applications (cutting tools and bearings), one can clearly notice that cumulative features have higher fitness as compared to classical features (listed in Table I). Also, cumulative features based on proposed trigonometric functions (C-σ(asinh) and C-σ(atan) in Table III and Table IV) appear to be the more monotonic and trendable ones. Same conclusion can be drawn qualitatively by considering Fig. 7, that compares the form of features extracted from classical and proposed approaches (having best fitness). Whatever the case is (and thereby, whatever load conditions), cumulative transformation lead to highly monotonic and trendable wear patterns. On the opposite, classical procedure result highly non-linear and complicated patterns that do not clearly reflect machine condition, which impairs performance of prognostics model to very low accuracy (and large uncertainty of RUL estimates, Fig. 2). Note that, classical / cumulative features with higher fitness will only be used for further experiments.

TABLE III
COMPARING FEATURES FITNESS (MEAN PERFORMANCES 6 CUTTERS)

Feature	R	M	Cumulative Feature	R	M
σ(asinh)	0.958	0.713	C-σ(asinh)	0.995	1
σ(atan)	0.960	0.709	C-σ(atan)	0.995	1
Entropy	0.949	0.643	C-Entropy	0.991	1
Energy	0.177	0.203	C-Energy	0.981	1
RMS	0.83	0.703	C-RMS	0.996	1
Kurtosis	-0.15	0.418	C-Kurtosis	0.976	1
Upper bound	0.968	0.708	C-Upper bound	0.994	1

TABLE IV
COMPARING FEATURES FITNESS (MEAN PERFORMANCES 17 BEARINGS)

Feature	R	M	Cumulative Feature	R	M
σ(asinh)	0.47	0.28	C-σ(asinh)	0.984	1
σ(atan)	0.47	0.31	C-σ(atan)	0.984	1
Entropy	0.47	0.31	C-Entropy	0.983	1
Energy	0.36	0.29	C-Energy	0.982	1
RMS	0.45	0.28	C-RMS	0.978	0.9
Kurtosis	-0.04	0.15	C-Kurtosis	0.972	1
Upper bound	0.39	0.21	C-Upper bound	0.973	0.9

C. Does monotonic & trendable features improve accuracy?

1) SW-ELM as a suitable model: predicting continuous state of a degrading machinery acts as prerequisite of any prognostics model to estimate RUL. However, for real application, performances of a prognostics model are closely related to critical issues that limit its applicability, like the horizon of prediction, the expected accuracy, the acceptable uncertainty level, human intervention, or convergence of algorithm. Such aspects should be considered for prediction modeling. Therefore, SW-ELM is considered to perform estimation of cutting tool wear, and multi-step ahead prediction (msp) of bearing degradation using iterative approach [10], [31].

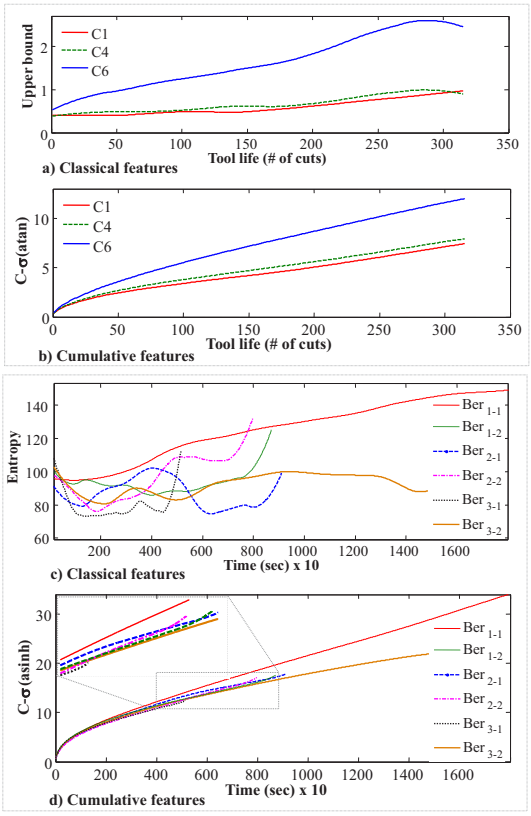


Fig. 7. Classical features vs. cumulative features: a),b) cutters & c),d) bearings

2) SW-ELM learning and testing strategy: for estimation / prediction tasks, simulation settings are given in Table V. For cutting tools application testing set (Table II) is not used due to absence of ground truth (tool wear [39]). Therefore, only learning set is used for experiments using leave-one-out strategy (e.g. learn C1, C4 and test C6). However, for bearings application, learning and testing were performed according to information given in Table II. Assuming that a single model cannot guarantee accuracy of estimation / prediction, therefore, for each application 100 SW-ELM models (with different weights / bias) were learned rapidly, and the best model with minimum learning error was selected for testing. For e.g. in case of bearings, even with the large data from Ber_{1-1} and Ber_{1-2}, the learning time was just 0.0063 sec due to one-pass learning phase (III-C2). Note that, for bearings 3 inputs represent regressors from a particular feature (x_t, x_{t-1}, x_{t-2}), where output represents msp ($x_{t+1\rightarrow t+H}$), H $\in$ msp horizon.

TABLE V
SW-ELM SIMULATION SETTINGS

Parameter	Estimation model	Prediction model
Input / hid. / output node	2 / 4 / 1	3 / 4 / 1
Constant C (see Fig. 4)	0.01	0.001
Inputs (classical)	Entropy, Upper bound	Entropy
Output (classical)	Tool wear	Entropy
Inputs (proposed)	C-σ(asinh), C-σ(atan)	C-σ(atan)
Output (proposed)	Tool wear	C-σ(atan)

3) Results and discussion: As mentioned in previous topic, tool wear estimation is performed for totally unknown cutter, which was not included in learning. For illustration, tool wear estimation results with features from classical and proposed approaches (Table V) are compared for cutter $C1$ in Fig. 8, which shows accurate estimates with cumulative features. All 3 tests on cutting tool application show better performance with proposed features, that are summarized in Table VI. Note that only test on cutter $C6$ showed lower performances by both approaches, which is due to large deviation of unknown test data from the learned cases (see Fig. 7).

As for bearings application, msp were performed from a current time (of test data) to the end of the bearings life (see details in [25], [40]). For illustration, predictions results with features from classical and proposed approaches (Table V) are compared for test bearings Ber_{1-7} which has the longest msp horizon (Fig. 9). One can qualitatively note the accuracy of msp with proposed methodology: the figure depicts very minor difference between the actual and predicted trends, where predictions with classical feature show poor results. This conclusion can be extended for all tested bearings (see Table VII for msp error): predictions with cumulative features are achieved with high accuracy and generalization even for long prediction horizons like "757 steps". Since predictions follow trending behavior of features in an accurate manner, the complete approach (cumulative features with SW-ELM) should lead to low uncertainty of RUL estimates.

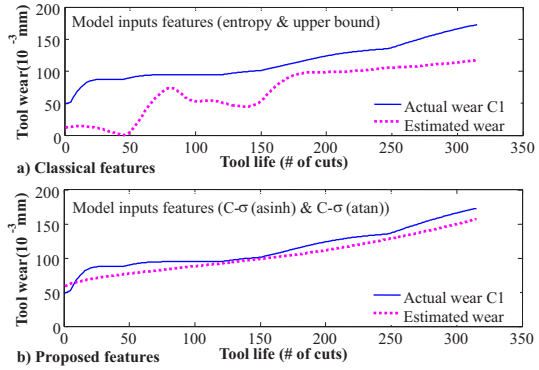


Fig. 8. Examples of estimation (Classical vs. Proposed)

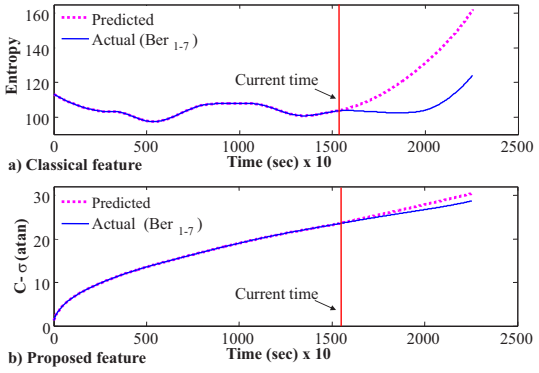


Fig. 9. Examples of multi-step ahead predictions (Classical vs. Proposed)

TABLE VI
TOOL WEAR ESTIMATION PERFORMANCES FOR 3 CUTTERS

Learned cutters	Test cutter	Classical features R2	Classical features RMSE	Proposed features R2	Proposed features RMSE
C4 C6	C1	-1.85	46.9	*0.84*	*10.9*
C1 C6	C4	0.55	25.7	*0.73*	*19.9*
C1 C4	C6	0.13	43.4	*0.19*	*41.8*

RMSE: Root Mean Square Error - $[0, \infty[$, Perfect score = 0
R2: coefficient of determination - should be closer to 1

TABLE VII
LONG TERM PREDICTION PERFORMANCES FOR 11 TEST BEARINGS

Test Ber	Prediction steps	Entropy (classical) MAPE	Entropy (classical) RMSE	C-σ(atan) (proposed) MAPE	C-σ(atan) (proposed) RMSE
$1-3$	573	2.09%	4.5	*1.9%*	*0.81*
$1-4$	289	32.1%	67.5	*6.4%*	*1.82*
$1-5$	161	*0.38%*	0.5	0.65%	*0.23*
$1-6$	146	0.68%	1.05	*0.66%*	*0.26*
$1-7$	757	16.92%	22.3	*3.0%*	*0.99*
$2-3$	753	1.06%	2.55	*0.19%*	*0.07*
$2-4$	139	3.92%	4.75	*0.09%*	*0.01*
$2-5$	309	3.54%	4.73	*0.55%*	*0.22*
$2-6$	129	5.70%	9.71	*1.26%*	*0.19*
$2-7$	58	5.71%	12.5	*2.60%*	*0.31*
$3-3$	82	2.48%	4.60	*0.78%*	*0.11*

MAPE: Mean Absolute Percent Error - $[0, \infty[$, Perfect score = 0

V. CONCLUSION

Performances of prognostics methods are closely related to the form and trends of extracted features from raw data to serve the need of degradation modeling. However, usefulness of gathered data is highly dependent on the variability of phenomena. Developments of this paper focus on the proposition of enabling features that can lead to simple and accurate prognostics. In the first step, data-preprocessing phase is improved in a new manner by applying trigonometric functions and cumulative transformation to achieve monotonic / trendable features from a decomposed vibration signal (by Discrete Wavelet Transform), which are further selected on the basis of their fitness. To show the benefit of our proposition for prognostics, in the second step, selected features are used to build a model with SW-ELM algorithm to perform estimation / prediction tasks. The proposed approach is generalized on vibration data from two real applications: 1) cutting tools (from CNC machine) and bearings (from platform PRONOSTIA). The results clearly depict effectiveness of the proposition for improving accuracy of prognostics. The future perspective, is to integrate this work with our novel developments of dynamic failure thresholds assignment for prognostics [10]. Another possibility could be, the usage of our proposition on vibration data of gearbox (in wind turbine).

ACKNOWLEDGMENT

This work has been supported by Laboratory of Excellence ACTION dedicated to Smart Systems Integrated into Physical Structures (project ANR-11-LABX-01-01).

REFERENCES

[1] W. Wang and M. Pecht, "Economic analysis of canary-based prognostics and health management," *IEEE Trans. Ind. Electron.*, vol. 58, no. 7, pp. 3077–3089, 2011.

[2] J. Coble and J. W. Hines, "Applying the general path model to estimation of remaining useful life," *International Journal of Prognostics and Health Management*, vol. 2, pp. 1–13, 2011.

[3] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: a neural network approach," *IEEE Trans. Ind. Electron.*, vol. 51, no. 3, pp. 694–700, 2004.

[4] Ž. B. Jakovljević, "Comparative analysis of hilbert huang and discrete wavelet transform in processing of signals obtained from the cutting process: An intermittent turning example," *FME Transactions*, vol. 41, no. 4, pp. 342–348, 2013.

[5] E. Zio, "Prognostics and health management of industrial equipment," in *Diag. and Progn. of Engineering Systems: Methods and Techniques*, S. Kadry, Ed. IGI Global, 2012, pp. 333–356.

[6] A. Heng and S. Zhang, "Rotating machinery prognostic: State of the art, challenges and opportunities," *Mechanical Systems & Signal Processing*, vol. 23, pp. 724–739, 2009.

[7] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems & Signal Processing*, vol. 20, pp. 1483–1510, 2006.

[8] J. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical Systems & Signal Processing*, vol. 25, no. 5, pp. 1803–1836, 2011.

[9] S. Yin, H. Luo, and S. Ding, "Real-time implementation of fault-tolerant control systems with performance optimization," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2402–2411, 2014.

[10] K. Javed, R. Gouriveau, and N. Zerhouni, "Novel failure prognostics approach with dynamic thresholds for machine degradation," in *IEEE IECON 2013-39th Annual Conference on IEEE Ind. Electron. Society, Vienna, Austria*. IEEE, pp. 4402–4407.

[11] F. Camci, K. Medjaher, N. Zerhouni, and P. Nectoux, "Feature evaluation for effective bearing prognostics," *Quality and Reliability Engineering International*, vol. 29, no. 4, pp. 477–486, 2013.

[12] S. Kumar and M. Pecht, "Modeling approaches for prognostics and health management of electronics," *International Journal of Performability Engineering*, vol. 6, no. 5, pp. 467–476, 2010.

[13] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux, "A feature extraction procedure based on trigonometric functions and cumulative descriptors to enhance prognostics modeling," in *IEEE Int. Conf. on Prognostics and Health Management*, Gaithersburg, Maryland, 2013.

[14] L. Liao, "Discovering prognostic features using genetic programming in remaining useful life prediction," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2464–2472, 2014.

[15] M. Lebold and M. Thurston, "Open standards for condition-based maintenance and prognostics systems," in *5th Annual Maintenance and Reliability Conference*, 2001.

[16] W. Yan, H. Qiu, and N. Iyer, "Feature extraction for bearing prognostics and health management (phm)-a survey," in *62nd Meeting of the Society for Machinery*, Virginia Beach, VA, 2008.

[17] J. Yu, "A hybrid feature selection scheme and self-organizing map model for machine health assessment," *Applied Soft Computing*, vol. 11, no. 5, pp. 4041–4054, 2011.

[18] C. K. Chui, *An introduction to wavelets*. Academic press, 1992, vol. 1.

[19] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[20] B. Chen, Z. Zhang, C. Sun, B. Li, Y. Zi, and Z. He, "Fault feature extraction of gearbox by using overcomplete rational dilation discrete wavelet transform on signals measured from vibration sensors," *Mechanical Systems and Signal Processing*, vol. 33, pp. 275–298, 2012.

[21] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1, pp. 108–126, 2013.

[22] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 112–114, 2004.

[23] S. Olhede and A. Walden, "The hilbert spectrum via wavelet projections," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 460, no. 2044, pp. 955–975, 2004.

[24] A. Bellini, F. Filippetti, C. Tassoni, and G.-A. Capolino, "Advances in diagnostic techniques for induction machines," *IEEE Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4109–4126, 2008.

[25] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Morello, N. Zerhouni, and C. Varnier, "Pronostia: An experimental platform for bearings accelerated life test," in *IEEE Int. Conf. on Prognostics and Health Management*, Denver, CO, USA, 2012.

[26] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, "Health assessment and life prediction of cutting tools based on support vector regression," *Journal of Intelligent Manufacturing*, pp. 1–11, 2013.

[27] H. Bendjama, S. Bouhouche, and M. S. Boucherit, "Application of wavelet transform for fault diagnosis in rotating machinery," *International Journal of Machine Learning and Computing*, vol. 2, no. 1, pp. 82–87, 2012.

[28] S. G. MALLAT, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, 1989.

[29] D. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot, "A data-driven failure prognostics method based on mixture of gaussians hidden markov models," *IEEE Trans. Rel.*, vol. 61, no. 2, pp. 491–503, 2012.

[30] J. Iwaniec, W. Lisowski, and T. Uhl, "Nonparametric approach to improvement of quality of modal parameters estimation," *Journal of theoretical and applied mechanics*, vol. 43, no. 2, pp. 327–344, 2005.

[31] R. Gouriveau and N. Zerhouni, "Connexionist-systems-based long term prediction approaches for prognostics," *IEEE Trans. Rel.*, vol. 61, no. 4, pp. 909–920, 2012.

[32] M. R. Meireles, P. E. Almeida, and M. G. Simões, "A comprehensive review for industrial applicability of artificial neural networks," *IEEE Trans. Ind. Electron.*, vol. 50, no. 3, pp. 585–601, 2003.

[33] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Int. Joint Conf. on Neural Networks*, Budapest, Hungary, 2004.

[34] K. Javed, R. Gouriveau, and N. Zerhouni, "SW-ELM: A summation wavelet extreme learning machine algorithm with *a priori* parameter initialization," *Neurocomputing*, vol. 123, pp. 299–307, 2014.

[35] M.-B. Li, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "Fully complex extreme learning machine," *Neurocomputing*, vol. 68, pp. 306–314, 2005.

[36] A. Pourtaghi and M. Lotfollahi-Yaghin, "Wavenet ability assessment in comparison to ann for predicting the maximum surface settlement caused by tunneling," *Tunneling and Underground Space Technology*, vol. 28, pp. 257–271, 2012.

[37] Y. Oussar and G. Dreyfus, "Initialization by selection for wavelet network training," *Neurocomputing*, vol. 34, no. 1-4, pp. 131–143, 2000.

[38] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," in *Int. Joint Conf. on Neural Networks IJCNN*, San Diego, CA, USA, 1990.

[39] PHM 2010 data challenge. [Online]. Available: https://www.phmsociety.org/competition/phm/10

[40] IEEE PHM 2012 data challenge. [Online]. Available: http://www.femto-st.fr/f/d/IEEEPHM2012-Challenge-Details.pdf

[41] F. Ding and Z. He, "Cutting tool wear monitoring for reliability analysis using proportional hazards model," *The International Journal of Advanced Manufacturing Technology*, vol. 57, no. 5-8, pp. 565–574, 2011.

[42] B. Zhang, C. Sconyers, C. Byington, R. Patrick, M. Orchard, and G. Vachtsevanos, "A probabilistic fault detection approach: application to bearing fault detection," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 2011–2018, 2011.

[43] N. Fang, P. S. Pai, and N. Edwards, "Tool-edge wear and wavelet packet transform analysis in high-speed machining of inconel 718." *Journal of Mechanical Engineering*, vol. 58, no. 3, pp. 191–202, 2012.

Kamran Javed did his BS. (2006) and MS. (2009) in Computer Engineering from COMSATS University (Abbottabad) Pakistan. In 2014 he received his Ph.D. Degree in Automatic Control from FEMTO-ST Institute, Besançon, France. Currently he is working at FEMTO-ST Institute and FCLAB Belfort (France), as Postdoc researcher in PHM group. His research activities are concerned with condition monitoring and prognostics of aging Fuel Cells.

Rafael Gouriveau received his engineering degree from National Engineering School of Tarbes (ENIT) in 1999. He then got his MS (2000), and his Ph.D. in Industrial Systems in 2003, both from the Toulouse National Polytechnic Institute (INPT). In September 2005, he joined the National Engineering Institute in Mechanics and Microtechnologies of Besançon (ENSMM) as Associate Professor. As for investigation, he is with the AS2M department of FEMTO-ST Institute. His research interests are in the development of industrial prognostics systems and the investigation of reliability modeling using possibility theory. He is also the scientific coordinator of PHM research axes from the FCLAB Research Federation (CNRS).

Noureddine Zerhouni received his engineering degree from National Engineers and Technicians School of Algiers (ENITA) in 1985. He received his Ph.D. Degree in Automatic Control from the Grenoble National Polytechnic Institute in 1991. In September 1991, he joined the National Engineering School of Belfort (ENIB) as Associate Professor. Since September 1999, Noureddine Zerhouni has been a Professor at the national high school of mechanics and microtechniques of Besançon. As for investigation, he is with AS2M department of FEMTO-ST Institute. His main research activities are concerned with intelligent maintenance systems, and e-maintenance.

Patrick Nectoux received his engineering degree in Electronics from Conservatoire National des Arts et Métiers (CNAM) in 2008. He is design engineer with CNRS and currently working at FEMTO-ST Institute AS2M department (Besançon), France. He is in charge of testbed platform development projects in the field of health assessment, diagnostic and prognostic. His current fields of interest are instrumentation, embedded systems, electronic and mechatronic engineering.