Skeleton of a Generic Approach for the generation of Health Indicators of Physical Systems

M.C. SEKKAL[1–2-3] & N. BERRACHED[1]
[1] LARESI – laboratory, Electronics department
USTO-MB University
Oran – Algeria.
mc.sekkal@hotmail.fr , laresi.usto.2015@gmail.com

[3]Departement of Electric and Computing Engineering
Dr Yahia FARES University
Medea – Algeria.
sekkal.mohamedchakib@univ-medea.dz

K. MEDJAHER2 & C. VARNIER2
[2]AS2M Department
FEMTO – ST institute, UMR CNRS
Besançon – France
{kamal.medjaher, christophe.varnier,
mohammed.sekkal}@femto-st.fr

***Abstract*— Predictive maintenance of physical systems can only be achieved by monitoring their most critical elements to track their health assessment during operation. The acquired data are processed to extract relevant features, which are used to estimate the state of the system at any time and detect any loss of performance that may occur due to the critical element.**

We propose in this work an architecture for a generic method to supervise this critical element and generate a Health Indicator (HI) for the physical system. The generated HI takes into account the evolution in time of the health status of the physical systems.

The proposed method is based on sensors data that allow us to extract in real time the values of features constituting themselves the HI construction bloc input, through several HI obtaining test.

Block diagram of the approach is made, then checked using benchmark data taken from "NASA data repository prognosis" associated to an element used in different operating conditions.

This approach is classified as data driven method which use sensors data.

Keywords—Healthy Assessment, Health indicator, Physical Systems, Critical Component, Data Driven Methods.

I. INTRODUCTION

Prognostics and Health Management (PHM) is a system engineering discipline focusing on detection, prediction, and management of the health status of complex systems. PHM is all methods that permit the assessment of the reliability of a system under its actual application condition. Among these methods, the health assessment is common to all prognostics applications and can be used to meet several critical goals:

- Advance warning of failure;
- Minimizing unscheduled maintenance and extending maintenance cycles;
- Reducing the life cycle cost of equipment by decreasing inspection costs and downtime;
- Improving qualifications and assisting in the design and logistical support of fielded and future products.

The overall health assessment of complex industrial systems can be achieved by generating a Health Indicator (HI) from sensors data. However, in the literature we don't find any specific definition of Health Indicator (HI), or Degradation Index (DI); authors cite the properties which they are interested in, and then use them in their approaches of prognosis and / or diagnosis. By juxtaposing these properties we arrive at the following definition:

Health Indicator (HI) is a relevant computed index, inferred from the operating parameters and supervision data of a monitored system, reflecting in time its unobserved health state and revealing its hidden degradation level regardless of the application (a particular profile of operation). Figure 1 shows the general form of the HI, where T is total operating time, and t is start time of degradation.

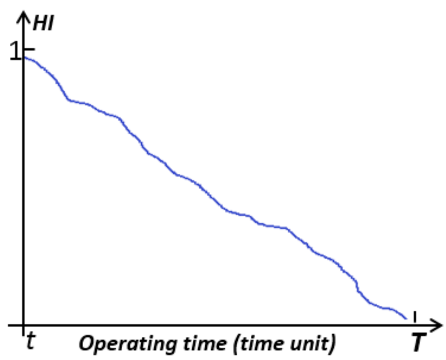


Figure 1. General form of HI.

To solve the problem of a generic approach for HI generation, the first step is to provide an architecture that groups the main stages of this approach. It is the objective of this article.

The study of existing researches clearly shows that until 2013, scientists do not give great importance to the generation of HI in their PHM studies.

The published methods (As part of a prognostic approach generally) concern a single data type, and are even dedicated to specific applications in some cases. In 2014, a work that deals with three types of data was published [1], but the proposed method is not generic.

In this work, generation of the HI is done through a block approach that brings together the main steps encountered in literature.

The rest of the article is organized as follows: In the section 2, we will detail the blocks of the proposed approach, specify the metric values calculated thanks to benchmark data associated to element used in different operating conditions. The section 3 is dedicated to the approach tests and results. Then we conclude by giving the next steps of our work.

II. Proposed approach

In this work, the proposed generic approach for the extraction of Health Indicator is based on two essential points:

- A restructuring of all steps necessary to build the HI with a nomenclature of tools for each step.
- A test of obtaining of the HI based on their sufficient conditions which occurs between steps and allows to be sure of the convergence of the approach.

The figure 2 shows the diagram of the proposed approach. Each step details are presented below. However, a step that is not apparent from the state of the art seems necessary in the convergence of this approach: Anomalies detection.

In the end, this approach is subject on some assumptions:

- The physical system monitoring is done by monitoring its most critical components identified previously by system experts;
- No maintenance intervention took place during the data acquisition process;
- The degradation of the monitored components develops gradually over time. i.e. no sudden failure and no regeneration of the physical system.

A. Sensors data

Data acquisition is a process of collecting and storing useful data (information) from targeted physical assets for the purpose of health assessment [2].Condition monitoring data are the measurements related to the health condition/state of the physical asset. Sensors data for condition monitoring are very versatile. It can be vibration data, acoustic data, oil analysis data, temperature, pressure, moisture, humidity, weather or environment data, etc. Various sensors, such as micro-sensors, ultrasonic sensors, acoustic emission sensors, etc., have been designed to collect different types of data. With the rapid development of computer and advanced sensor technologies, data acquisition facilities and have become more powerful and less expensive, making data acquisition for CBM implementation more affordable and feasible [5]. The resulting availability and variety for these data make the condition monitoring sensitive to their types. Hence the need for a generic method for the HI extraction.

B. Preprocessing

After monitoring the system by a set of sensors, the monitored data are pre-processed [6]. In general case, data pre-processing is data cleaning.

This is an important step since data always contains errors. Data cleaning ensures, or at least increases the chance, that clean (error-free) data are used for further analysis and modeling. For condition monitoring data, data errors may be caused by sensor faults. In this case, sensor fault isolation is the right way to go. In general, however, there is no simple way to clean data. Sometimes it requires manual examination of data. Graphical tools would be very helpful to finding and removing data errors [5].

C. Variables selection

One of the simplest method used for the selection of variables is the correlation matrix. It measures the degree of association between two random variables X and Y. Specifically, we speak of linear association. The correlation index is defined as follows [7]:

$$\rho = \frac{Cov\,(X,Y)}{\sqrt{Var(X).Var(Y)}} \quad (1)$$

Where *cov()* designates the covariance and *var()* the variance.

If X, Y are independent, their correlation is zero. It was agreed that if the value of the correlation index is greater than 0.6, there is a strong linear relationship between variables. Otherwise, if the correlation is less than 0.3, X and Y are uncorrelated. Extreme values, $|\rho| = 1$ are only reached if the relationship between X and Y is perfectly linear.

Other measures used for the selection of the variables are the entropy, the mutual information [7] and ensuing mutual uncertainty [1].

For a better representation of degradation phenomenon, the selected variable between two correlated ones is the most monotonous. By definition, “Monotonicity” is the correlation between variable and time [4].

D. Variables reduction

The selected variables are compressed using a method of data combining. The most generic method used is standard principal component analysis (PCA).

The first principle component retains the maximum variance while reducing the dimensionality to one dimension. Therefore, only the first principle component is used to represent the health status evolution with respect to time [1].

E. Features extraction

In literature, a large number of signal processing techniques have been proposed for the features extraction. However, three main categories of features extraction are generally mentioned:

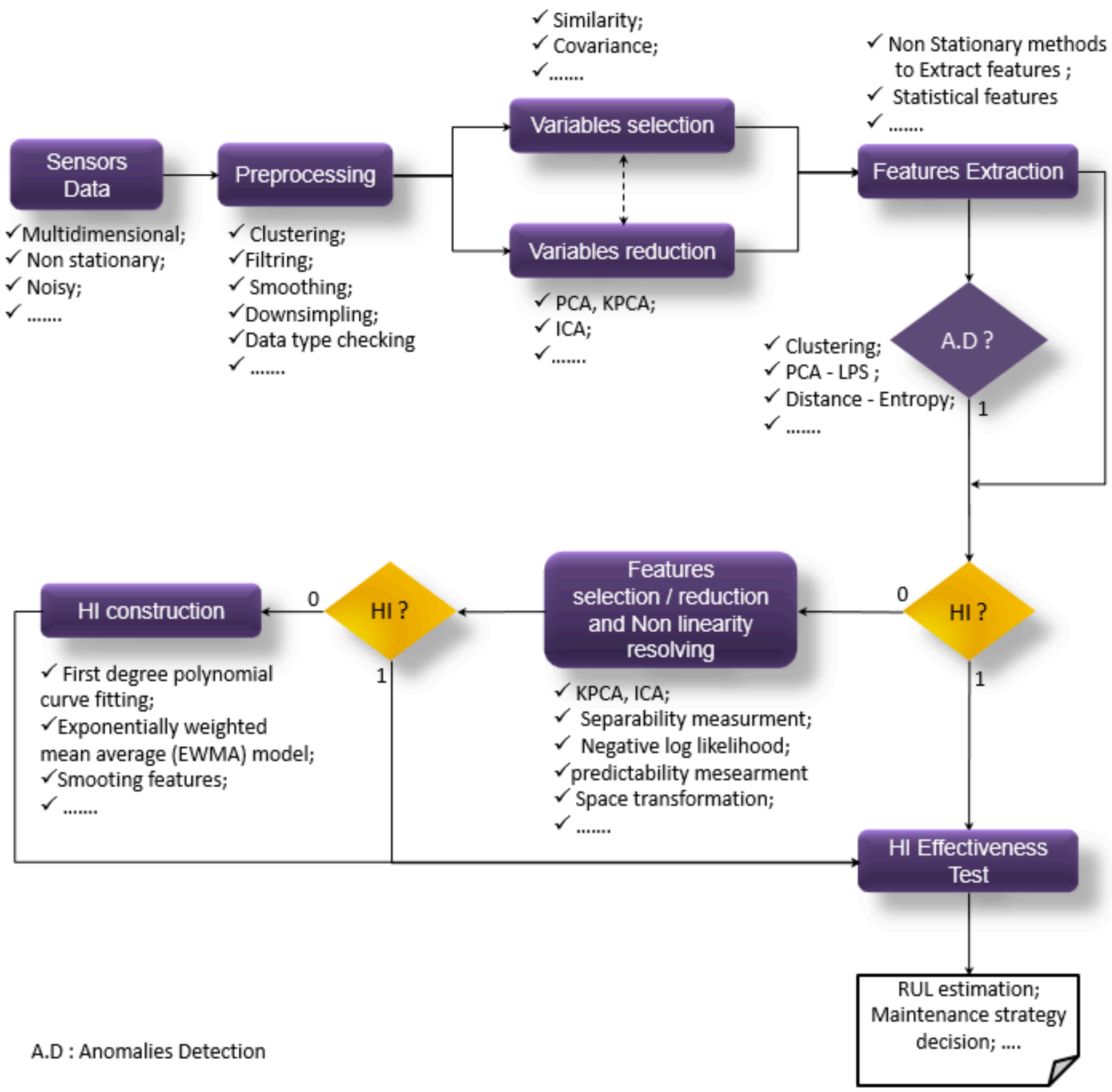


Figure 2. Proposed approach diagram for HI construction.

- **Time domain**: Features are extracted using statistics like mean, variance, standard deviation, kurtosis, etc. They are suitable for fault detection and applied to stationary signals. However, extracted features may show sensitivity to variation in data and inherit nonlinearity, which complicates decisions [8].
- **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 (FFT). Other methods that belong to this category are *Cepstrum*, spectral analysis, higher-order spectra or envelop analysis [15, 12].The main limitation of such techniques is their inability to deal with non-stationary signals, unfortunately which is the case in degrading machinery [4].
- **Time-Frequency:** Time-frequency techniques 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) [8], Wavelet Transform (WT) [9], and Empirical Mode Decomposition (EMD) [1]. EMD and WT are the two outstanding examples among signal processing techniques. The main weakness of EMD is high sensitivity to noise, and it runs into the problem of mixing modes. EMD is also reported to have characteristics like wavelet. Moreover, EMD is popular in demodulation applications, whereas WT is commonly used in vibration content characterization and has better applicability, especially when vibration

data come from rotating machinery. However, even application WT cannot guarantee ideal features for prognostics applications, and its performances can vary from case to case [4].

For a generic approach, the idea is to extract all features in the three domains, then make a selection of those that are most representative of the degradation.

F. Anomalies detection

Anomalies detection can be defined as the process of identifying when a fault/outlier has occurred. [10] reviewed some of the research work which conducted for this task. Note that the PCA is widely used in this sense. In addition, features extracted using statistics (Time domain techniques) and those extracted using frequency domain techniques are suitable for fault detection and more effective for fault diagnostic.

G. Features selection/ reduction

Two way to do are reported to find the feature that best describes the degradation of the system:

- Make a selection and / or reduction of variables on a multivariate dataset. Then apply a features extraction method [1].
- Extract all the features from a variable and then make a selection and / or reduction of features to find the one that best describes the degradation [3].

Our approach hybrids these two ideas to maximize the chance of finding the right feature for a given dataset. The selection of feature is done by measuring correlation with the perfect HI. If the threshold for obtaining HI is not reaches, we Selected feature that maximizes the metric to continue the approach.

H. Features non-linearity resolving

This step is optional and occurs only if all the features are inconclusive. By inconclusive we mean that their metric values are lower than 0.8. This is the average value between the threshold of a strong correlation (0.6) and the maximum 1.

In this case one goes through the resolution of the non-linearity of features which can be done by using non-linear methods for variables reduction and features extraction such as KPCA [3].

If not, ultimately, one can dealing the non -linearity of the selected feature (the one that maximize the value of the metric).

I. HI obtaining test

Knowing the time of onset of degradation τ, we generate the perfect indicator that's equal 1 before τ then linearly decreasing between τ and T, the end of life of the system as shown in the following equation:

$$Perfect\_HI(t) = \begin{cases} 1, & t \leq \tau \\ \frac{-1}{T-\tau}.t + \frac{T}{T-\tau}, & \tau < t \leq T \end{cases} \tag{2}$$

J. HI construction

This step involves estimating the parameters of the transformation that allows to move from selected feature, to the perfect HI.

Our first idea was to use a linear combination, that at time i delivers a number HI (i) such as:

$$\mathrm{HI(i)} = \sum_{k=0}^{H} a_k \,.x(i-k) + \sum_{l=1}^{H} b_l.HI(i-l) \tag{3}$$

H is called the horizon of the digital filter. The simplest filter we can use has the form:

$$\mathrm{HI(i)} = x(i) + b.\mathrm{HI(i-1)} \tag{4}$$

Which means that output value at time i depends on submitted entry and last previous output value, it is a recursive filter, x(i) is the normalized form of the selected feature. HI is then given by:

$$Construct\_HI(i) = \begin{cases} 1, & i \leq \tau \\ x(i) + b.\mathrm{HI(i-1)}, & \tau < i \leq T \end{cases} \tag{5}$$

K. HI effectiveness test

One of methods for evaluate the effectiveness of the proposed approach is by comparing its performance with that of other monitoring indexes [3]. Performances of proposed model, used with different kinds of data can be compared with that of some of the models that may have been used in other studies for developing condition monitoring indexes for their specific kind of data. Each model misses at least one component that is utilized in the proposed model.

III. TESTS AND FIRST RESULTS

First tests were conducted using Li-Ion batteries dataset taken from "NASA data repository prognosis" [11].Table 1 shows the variables that are available for 3 different operational profiles at room temperature.

Table 1. Available variables for Li-Ion batteries dataset.

	Charge \ Discharge	Impedance
1	Voltage measured (Volts)	Sense current (Amps)
2	Current measured (Amps)	Battery current (Amps)
3	Temperature measured (C°)	Current ratio
4	Current load (Amps)	Battery impedance (Ohms)
5	Voltage load (Amps)	Rectified impedance (Ohms)
6	Time (Secs)	/

The learning phase gives the following results for those data:

A signal is considered as HI if its correlation index with the perfect HI is at least equal to **0.9734**. Figure 3 shows a perfect HI and a real HI constructed manually to be able to measure the threshold of the metric.

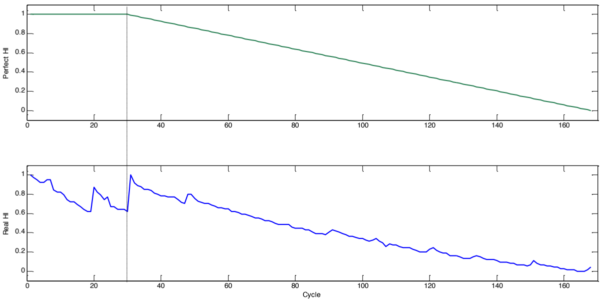


Figure 3: Perfect and real Health Indicators.

The value of the parameter b of the linear filter for the HI construction is 0.1964. Figure 4 shows the different steps of HI construction, while learning the value of the parameter b of the filter.

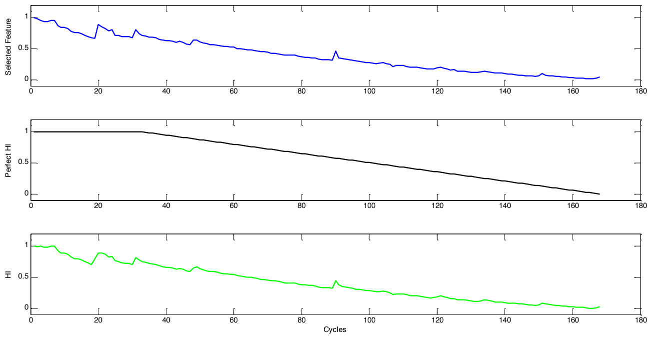


Figure 4: Different steps of HI construction by linear combination.

The program was executed with profiles charge and discharge variables separately.

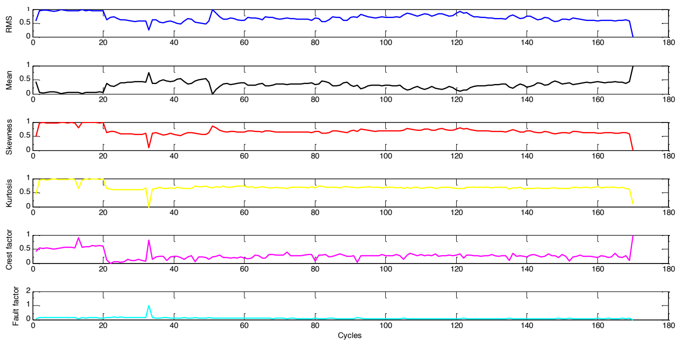
The online program works as follows:

- The HI keeps the value 1 until the detection of the start of degradation.
- Once the start time of degradation τ detected, the perfect HI is generated using τ and the average lifetime of the critical component. The start time of deterioration marked a peak on temporal statistical descriptor: Default factor of the principal component (Figure 6). It is this first method of anomalies detection that is used in this work.
- The real-time measurement of HI at each iteration is as follows:
  - The least correlated variables are selected by removing less monotonous ones. Figure 5 shows the variables correlation matrix for the two profiles. The last line of the correlation matrix represent monotonicity, because the last variable in the data structure is time.
  - One variable representing the degradation is made from the selected data using PCA, then time domain features are extracted.

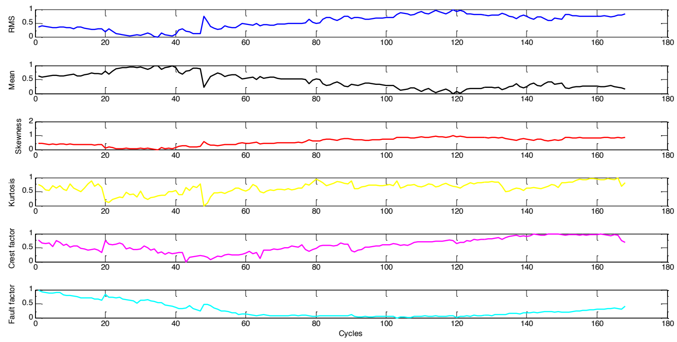
```
M_Discharge =

    1.0000
   -0.0917    1.0000
   -0.9146    0.3546    1.0000
   -0.0345   -0.3508   -0.0917    1.0000
    0.3841   -0.9178   -0.6017    0.2900    1.0000
   -0.8175    0.4169    0.9155   -0.2256   -0.6283


M1_Charge =

    1.0000
   -0.6910    1.0000
   -0.1310    0.6631    1.0000
   -0.6909    1.0000    0.6629    1.0000
   -0.1971    0.4303    0.3923    0.4297    1.0000
    0.6469   -0.8839   -0.6961   -0.8836   -0.5624
```

Figure 5. Correlation matrix for the variables of charge and discharge profiles.

  - The most correlated feature to the perfect HI are selected as Best Feature at current time (iteration). Figure 6 shows the time domain features extracted.



Charging cycles



Discharging cycles

Figure 6: Time domain features extracted from first Principal Component of selected variables.

  - If the correlation index is greater than the metric, the best feature becomes the Current HI. Otherwise, it is used to build the current HI. Figure 7 shows the feature that maximize the metric. It is inversely proportioned to the perfect HI (metric = -0.8578). In this case, *Best feature* = 1 –*Selected feature*.

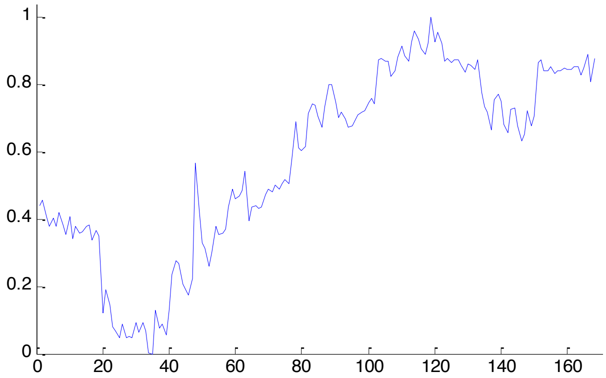
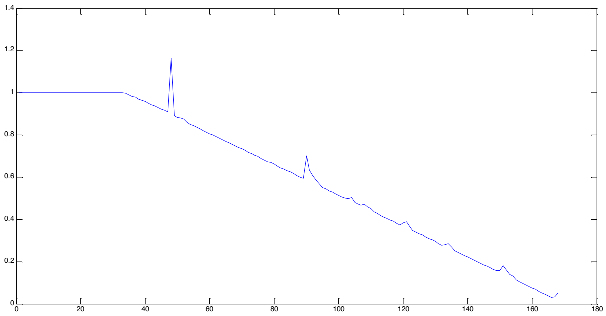
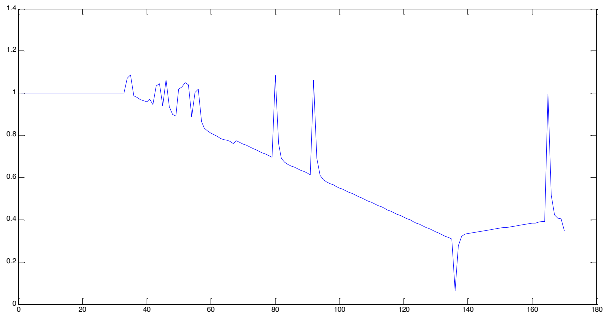


Figure 7: Selected feature that maximizes the metric to continue the approach.

- Thus, the real time value of HI is the ith value of the ith HI generated at iteration i. Figure 8 shows the real time value of HI for the charge and discharge profiles.



Discharging cycle



Charging cycle

Figure 8: Real time value of HI.

The measurement of correlation between the real and the perfect HI gives a result of r = 0.9973, for discharge profile and r=0.9231, for charge profile. This is because the charge cycles variables did not provide good features in the time domain. This will result in a final approach in a passage to the frequency domain or the time-frequency domain.

IV. CONCLUSION

The estimation of the system health status is provided overall by a health indicator extracted from sensors data placed on the critical components of this system.

Diversity of the critical elements for complex industrial systems and diversity of sensors and ensuing data, fact that there is no global approach for the HI generation as defined in this article, although steps are common to some specific approaches found in the literature.

In this work, we first propose a general definition for the HI. Then gathered in a flowchart the main steps of its generation. For each block of the approach, one mathematic tool was used (usually the simplest in its category). We next tested our approach on dataset of an element that we know the shape of degradation, to be able to estimate the metric of the HI obtaining test that we have inserted between the steps of the approach.

The only prior information which we depend is the average lifetime of the critical element. That is a given generally provided by the manufacturer.

The next step is the addition of various mathematical tools to each block for the execution of their tasks. The selection of the better tool will be automatically and intelligently in order to adapt our method of HI generation to different types of data.

REFERENCES

[1] Ahmed Mosallam, Kamal Medjaher, Noureddine Zerhouni, "Time Series Trending for Condition Assessment and Prognostics", Journal of manufacturing technology management, April 2014.

[2] Aiwina Heng, Sheng Zhang, Andy C.C. Tan, Joseph Mathew, "Review, rotating machinery prognostics: State of the art, challenges and opportunities", Mechanical Systems and Signal Processing 23 (2009) 724–739.

[3] Aye, S.A., Heyns, P.S. and Thiart, C.J.H., "Health diagnostics of slow rotating bearings based on Gaussian process regression", In South African Conference on Computational and Applied Mechanics Proceedings (SACAM 2014), Somerset West, Cape Town, South Africa, 14–16 January, 2014.

[4] Kamran Javed, Rafael Gouriveau, Noureddine Zerhouni, Daniel Hissel, "Improving Accuracy of Long-term Prognostics of PEMFC stack to Estimate Remaining Useful Life", in processing of IEEE International Conference Industrial Technology (ICIT), 17-19 March 2015.

[5] 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.

[6] D.A. Tobon-Mejia, K. Medjaher, N. Zerhouni, "CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks", Mechanical Systems and Signal Processing 28 (2012) 167–182.

[7] Isabelle Guyon, André Elisseeff, "An Introduction to Variable and Feature Selection", Journal of Machine Learning Research 3 (2003) 1157-1182.

[8] 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.

[9] C. K. Chui, « An introduction to wavelets ». Academic press, 1992, vol. 1.

[10] Ahmed Mosallam, Kamal Medjaher, Noureddine Zerhouni, "Time Series Trending for Condition Assessment and Prognostics", Journal of manufacturing technology management, April 2014.

[11] B. Saha and K. Goebel (2007). « Battery Data Set », NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/), NASA Ames Research Center, Moffett Field, CA, online in April 2016.