

Remaining useful life estimation of critical components with application to bearings

K. Medjaher, D.A. Tobon-Mejia, N. Zerhouni

Abstract

Prognostics activity deals with the estimation of the Remaining Useful Life (RUL) of physical systems based on their current health state and their future operating conditions. RUL estimation can be done by using two main approaches, namely model-based and data-driven approaches. The first approach is based on the utilization of physics of failure models of the degradation, while the second approach is based on the transformation of the data provided by the sensors into models that represent the behavior of the degradation. This paper deals with a data-driven prognostics method, where the RUL of the physical system is assessed depending on its critical component. Once the critical component is identified, and the appropriate sensors installed, the data provided by these sensors are exploited to model the degradation's behavior. For this purpose, Mixture of Gaussians Hidden Markov Models (MoG-HMMs), represented by Dynamic Bayesian Networks (DBNs), are used as a modeling tool. MoG-HMMs allow us to represent the evolution of the component's health condition by hidden states by using temporal or frequency features extracted from the raw signals provided by the sensors. The prognostics process is then done in two phases: a learning phase to generate the behavior model, and an exploitation phase to estimate the current health state and calculate the RUL. Furthermore, the performance of the proposed method is verified by implementing prognostics performance metrics, such as accuracy, precision, and prediction horizon. Finally, the proposed method is applied to real data corresponding to the accelerated life of bearings, and experimental results are discussed.

Index Terms

Condition monitoring, prognostics, remaining useful life, performance metrics, mixture of Gaussians hidden Markov models, condition-based maintenance, dynamic Bayesian networks.

ACRONYMS

RUL	Remaining Useful Life
HMM	Hidden Markov Model
MoG-HMM	Mixture of Gaussians Hidden Markov Model
DBN	Dynamic Bayesian Network
CBM	Condition Based Maintenance
ISO	International Standard Organization
ETTF	Estimated Time To Failure
FMECA	Failure Modes, Effects, and Criticality Analysis
S.P	Signal Processing
CPD	Conditional Probability Distribution
MAPER	Mean Absolute Percentage Error
HP	Horizon of Prognostics
RA	Relative Accuracy
ERPI	Electric Power Research Institute
ER	Extern Ring
IR	Inner Ring
RMS	Root Mean Square
WPD	Wavelet Packet Decomposition

Corresponding author: kamal.medjaher@ens2m.fr (K. Medjaher)

K. Medjaher, D.A. Tobon-Mejia and N. Zerhouni are with FEMTO-ST Institute, AS2M department, UMR CNRS 6174 - UFC / ENSMM / UTBM, 25000 Besançon, France. e-mail: kamal.medjaher@ens2m.fr, diego.tobon@femto-st.fr, noureddine.zerhouni@ens2m.fr

D.A. Tobon-Mejia is with ALSTOM Transport, 7, avenue De Lattre De Tassigny, BP 49, 25290 Ornans, France.

NOTATION

N	Number of states in a Hidden Markov Model
V	Number of distinct observations for each state
A	State transition probability distribution
B	Observation probability distribution
π	Initial state distribution
$S = (X_1, X_2, \dots, X_T)$	Sequence of hidden states of length T
$O = (O_1, O_2, \dots, O_T)$	Sequence of observations of length T
C_{jm}	Mixture coefficient for the m^{th} Gaussian in state j
ξ	Gaussian density with mean vector μ_{jm} and covariance matrix U_{jm} for the m^{th} mixture component in state j
M	Mixture matrix
Mix	Number of Gaussians

I. INTRODUCTION

Maintaining equipment in operational condition is an industrial, economical, and societal requirement. This requirement can be satisfied by implementing appropriate maintenance strategies, which can be curative or preventive. In the first case, the maintenance interventions are done only after a fault has been observed. This action may lead to non desired situations, especially for critical industrial systems (one can cite the petrochemical explosion in 2010 in the Gulf of Mexico). To avoid such situations, a systematic maintenance can be used. However, this type of maintenance can be a source of unnecessary spending due to unnecessary interventions or replacement of components. Thus, for more efficiency, one can implement condition Based Maintenance (CBM) [1]. Indeed, contrary to a systematic maintenance, a CBM is done according to the real equipment's health condition, estimated or measured, potentially through the sensors present on the industrial equipment. In addition, by using appropriate mathematical or modeling tools, the estimated or measured equipment's state can be projected to predict its future health condition, and thus support appropriate decisions.

In the framework of CBM, failure prognostics is considered to be one of the main processes, which permits one to estimate the Remaining Useful Life (RUL) before the failure of a given industrial system [2]. Numerous methods and tools can be used to predict the RUL's value. These methods can be classified into two principal approaches: model-based prognostics (also called physics of failure prognostics), and data-driven prognostics. Model-based prognostics [3] deals with the prediction of the RUL of critical physical components by using mathematical or physical models of the degradation phenomenon (crack by fatigue, wear, corrosion, etc.). The data-driven prognostics [3] aims at transforming the data provided by the sensors into relevant models (which can be parametric or non-parametric) of the degradation's behavior.

This paper deals with a data-driven prognostics method for the estimation of the RUL of critical physical components. The proposed method belongs then to the category of data-driven approaches, and uses mainly Mixture of Gaussians Hidden Markov Models (MoG-HMMs), represented by Dynamic Bayesian Networks (DBNs), to model the degradation, and to estimate the value of the RUL before a potential failure. MoG-HMMs are chosen for their capability to transform the features extracted from the raw signals to relevant behavioral models representing the physical component's degradation.

The proposed failure prognostics method is done in two phases: a learning phase, and an exploitation phase. During the first phase, the raw signals coming from the sensors are processed to extract appropriate, useful features. These features are then used to learn the parameters of the MoG-HMMs. In the second phase, the learned models are exploited in order to assess the current condition of the component, and to predict its future one, leading to an estimation of the component's RUL. In addition to RUL, a confidence value can be calculated. Note that the RUL and the confidence are important information that may be used in the decision process. For example, they can be used to delay the maintenance interventions, or stop a machine before its future maintenance due to early fault.

To assess the performance of the predictions provided by the proposed method, performance prognostics metrics are implemented. Finally, the method is applied on real experimental data related to the accelerated life tests of bearings.

The paper is organized as follows. Section 2 presents the steps of the proposed method, Section 3 is dedicated to its application on real degradation data related to bearings, and Section 4 concludes the paper.

II. DATA-DRIVEN PROGNOSTICS METHOD BASED ON MOG-HMMs

Compared to fault diagnostics, which consist of detecting and isolating the probable cause of the fault [2], failure prognostics aims at anticipating the time of the failure [4]. Thus, the diagnostic is realized after the occurrence of the fault, whereas the prognostics is done before the occurrence of the failure.

Several definitions related to prognostics have been reported in the literature [1], [5], [6]. The International Standard Organization [7] defines failure prognostics as "the estimation of the Time To Failure (ETTF) and the risk of existence or later appearance of one or more failure modes." Note that most of the definitions reported in the literature use the terminology Remaining Useful Life (RUL) instead of ETTF. An illustration of a RUL is given in Fig. 1.

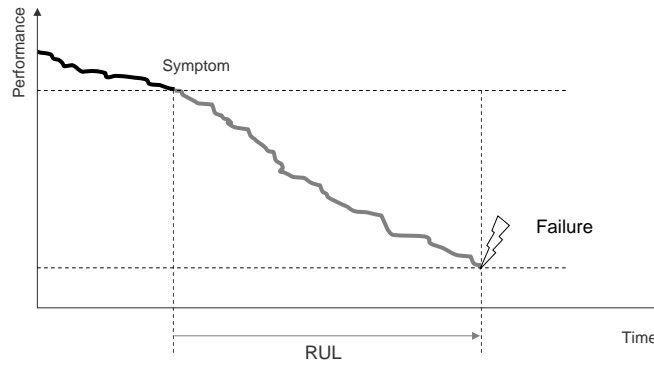


Fig. 1: Illustration of a RUL.

In addition to the absolute value of the RUL, a confidence interval can be associated, which may be useful during the decision process.

Numerous methods and tools regarding failure prognostics have been proposed and reported in the literature [1]–[3], [8]. The reported tools and methods have been subject to several classifications. The first classification (in a pyramidal) form has been proposed in [1] where three main approaches are distinguished: model-based prognostics, data-driven prognostics, and experience-based prognostics. This classification was created according to four criteria: complexity, cost, precision, and applicability. The experience-based approach (also called reliability-based approach, which takes its origins from the reliability domain) uses data gathered from the experience feedback to estimate the parameters of reliability laws. So, it can be put within the data-driven approach. Consequently, the following updated approaches can be considered: model-based (physics of failure), data-driven, and hybrid approaches (Fig. 2). In the model-based (or physics of failure) approach, the physical component and

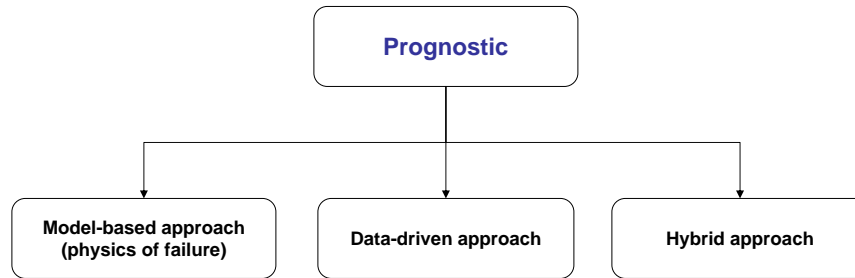


Fig. 2: Main prognostics approaches.

its degradation phenomenon are represented by a set of mathematical laws, which are then used to estimate the RUL [9], [10]. The data-driven approach consists of transforming the monitoring data provided by the sensors installed on the system into reliable behavioral models of the degradations [3], [11]. The modeling tools used in this approach are mainly those used by the artificial intelligence community: temporal prediction series, trend analysis techniques, neuronal networks, neuro-fuzzy systems, hidden Markov models, and dynamic Bayesian networks.

Compared to model-based methods, the data-driven methods offer a trade-off in terms of complexity, cost, precision, and applicability. They are suitable for systems where it is easy to obtain monitoring data and transform them into behavior models of the degradation phenomena. In practice, this condition is the case in several applications, such as bearings, which are the subject in this paper. Indeed, prediction models, such as the L_{10} for bearings or physics of failure laws (Paris-Erdogan) [9] for wear by fatigue, can be used to calculate the remaining useful life; but these laws are valid only for specific components in specific conditions that are difficult to verify in real applications.

In the rest of the paper, MoG-HMMs (represented by DBNs) are used to model the component's degradation, and to estimate its RUL. MoG-HMMs have proved to be a suitable tool in failure diagnostic and prognostics domains [11]. They allow one to model the physical component's degradation by using continuous observations provided by the monitoring sensors. They also permit one to estimate the stay durations in each health state, leading to the prediction of the RUL value. Furthermore, contrary to regression models or neural networks where the structure is not interpretable [12], the states in the MoG-HMM can be interpreted as the health conditions of the component.

A MoG-HMM is primarily an HMM. This latter is defined as a statistical model used to represent stochastic processes, where the states are not directly observed [13]. An HMM is completely defined by the parameters N , V , A , B , and π (more details can be found in [13], and [14]).

For simplicity and clarity of presentation, a compact notation ($\lambda = \pi, A, B$) is used for each HMM. In practice, HMMs are

used to solve the following problems (see [13] for more details).

- Problem 1 (detection): given a model λ , and an observation sequence $O = (O_1, O_2, \dots, O_T)$, compute the probability $P(O|\lambda)$ of the sequence given the model. The solution of this problem is obtained by using the forward-backward algorithm [15].
- Problem 2 (decoding): given an observation sequence $O = (O_1, O_2, \dots, O_T)$, find the hidden state sequence $S = (X_1, X_2, \dots, X_T)$ that has most likely produced the observation sequence. This problem is solved by using the Viterbi algorithm [16].
- Problem 3 (learning): find the model parameters (π, A, B) that better fit the observation sequence O , i.e., that maximize the probability $P(O|\lambda)$. This problem is solved by using the Baum-Welch algorithm [17].

The problem with discrete HMMs is that they cannot be directly applied to model continuous observations. To overcome this difficulty, one can use MoG-HMMs, where the observation probability densities are represented by a mixture of Gaussians in the form

$$b_j(O) = \sum_{m=1}^{Mix} C_{jm} \xi(O, \mu_{jm}, U_{jm}), 1 \leq j \leq N \quad (1)$$

A MoG-HMM is completely defined by the A matrix, the B matrix, and the initial probability π . The observation matrix B is modeled by a Gaussian density with a mean μ , a covariance matrix U , and a mixture matrix M .

To speed up the learning and inference calculations, Dynamic Bayesian Networks (DBNs) and related algorithms are used in this paper to represent the MoG-HMMs (more details on how to model Hidden Markov Models and MoG-HMMs by using DBNs can be found in [18]). An example of a compact representation of a MoG-HMM by a DBN is shown in Fig. 3. In this figure, X_1 is the initial state, X_t represents the hidden states at time t , C_t are the mixture coefficients, and O_t are the observation distributions.

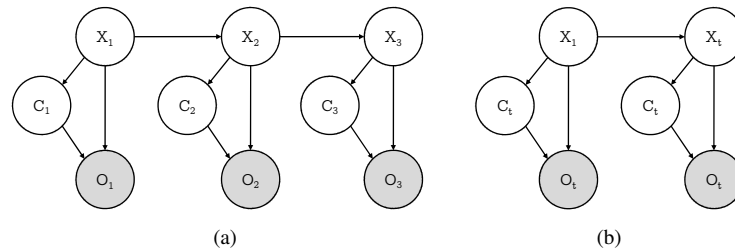


Fig. 3: Representation of a MoG-HMM by a DBN: (a) unrolled model for three time slices, and (b) compact representation.

The data-driven method proposed to estimate the RUL of critical physical components is presented next. In this method, it is assumed that the RUL of the machine (or industrial plant) depends on the RUL of its most critical component. It is also considered that an adequate monitoring system, with necessary sensors that capture the evolution of the component's degradation, is available. Thus, the data provided by the sensors can be used to model the component's degradation behavior, and estimate its RUL. The main steps of the proposed prognostics method are given in Fig. 4, and a description of each step is given in the following. The estimation of the RUL is done in two principal phases, as shown in Fig. 5: a learning phase,

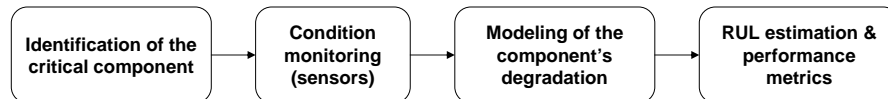


Fig. 4: Main steps of the proposed prognostics method.

and an exploitation phase.

A. Identification of the critical component

The identification of the critical component in an industrial plant or a machine can be done using analysis of the data and information gathered during the exploitation of the machine (experience feedback), and by using Failure Modes, Effects, and Criticality Analysis (FMECA).

The principle of experience feedback consists of improving the knowledge of a system through observation, acquisition, analysis, and information processing related to the real operation of the system in its environment. Then, the knowledge gathered can be used to identify the critical component by providing useful information to formal tools such as the FMECA. For example, in electrical rotating machines, the data gathered during a long time period of exploitation can help with estimating the failure distribution on the main components (stator, rotor, bearings, etc.). Then, the component which presents a high percentage of

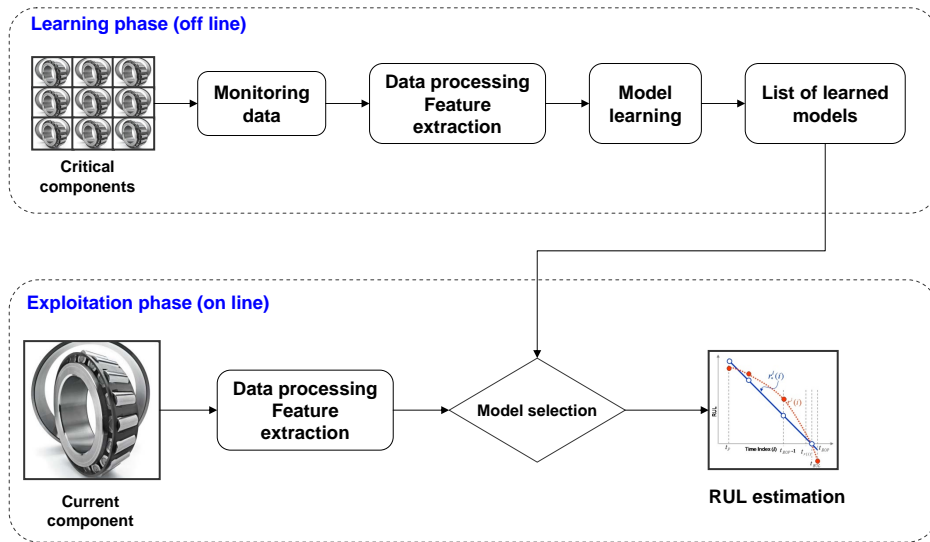


Fig. 5: RUL estimation steps.

failures can be considered as the one to be selected for further analysis.

An example of how to identify the physical critical component in a machine is given in [20].

B. Condition monitoring of the component

Once the critical physical component is identified, and the potential failure known, it is necessary to define a set of appropriate sensors, which will be used to monitor the behavior of the component and its degradation. The type and number of sensors depend mainly on the physical phenomena to monitor. The data provided by the sensors are then processed to extract relevant features and information, which will be used during the modeling step. Depending on the type of the degradation phenomena, and the application domain, several types of features can be extracted. The features can be temporal, frequency, or time-frequency. More details on how to choose the sensors and select the features depending on the application are given in [21].

C. Modeling of the component's degradation

The behavior model of the critical component including its degradation is obtained through a data-driven method, where the features extracted from the raw signals provided by the sensors are used during the learning phase (see Fig. 5) to estimate the parameters of the MoG-HMMs (represented by DBNs) that represent the degradation. Practically, each monitoring signal belonging to a given historical data η (related to a degradation) is transformed into a matrix D_X^η by signal processing (S.P). In this matrix, each column of X cells corresponds to a snapshot of the features at the instant t for the historical η .

$$\text{Raw historical data } \eta \xrightarrow{S.P} D_X^\eta(t) = \begin{pmatrix} D_1(t) & \dots & D_1(T^\eta) \\ \vdots & \ddots & \vdots \\ D_X(t) & \dots & D_X(T^\eta) \end{pmatrix} \forall \begin{matrix} 1 \leq t \leq T^\eta \\ 1 \leq X \leq X \\ 1 \leq \eta \leq H \end{matrix} \quad (2)$$

where

$$T^\eta = \text{total duration of the historical } \eta. \quad (3)$$

These features are then used to estimate the Conditional Probability Distributions (CPDs) (π , A et B) and the temporal parameters (stay duration in each health state) of the DBNs.

The temporal parameters are estimated using the Viterbi algorithm [16]. This algorithm permits us to estimate the sequence of states visited using historical data η during the experiment. The estimation of the sequence is done using the degradation's behavior model (with parameters π , A , and B) learned from the observations (features $O_t = D_X^\eta(t)$). So, by assuming that the stay durations in the states are governed by Gaussian laws, one can estimate the mean stay duration $\mu_d(x_i)$ and its standard deviation $\sigma_d(x_i)$ by using

$$\begin{aligned} \mu_d(x_i) &= \frac{1}{\Omega} \sum_{w=1}^{\Omega} \Delta(x_{iw}) \\ \sigma_d(x_i) &= \sqrt{\frac{1}{\Omega} \sum_{w=1}^{\Omega} [\Delta(x_{iw}) - \mu(\Delta(x_i))]^2} \end{aligned} \quad (4)$$

$\Delta(\cdot)$ stands for the visit duration, i is the state index, w is the visit index, and Ω_{x_i} corresponds to the total number of visits of the analyzed historical data η . A compact representation of each learned DBN is then given by

$$\lambda_{\text{DBN}} = (\pi, A, B, \mu_d(x_i), \sigma_d(x_i), \Omega_{x_{\text{final}}}) \quad (5)$$

Furthermore, the estimated sequence of states is used to identify the final state x_{final} , which corresponds to the failure health state of the component, visited by the historical data.

D. Estimation of the RUL

The estimation of the RUL and the associated confidence value is done during the exploitation phase (see Fig. 5), after having derived the model of the degradation. Indeed, the learned models are used on-line to detect the current health state of the component, and to calculate its RUL. For this purpose, the observations (features) related to the current component are continuously injected to the learned models to select the one that best fits the observed data. The selection process is based on the calculation of the likelihood probability $P[O|\lambda_{\text{DBN}}]$ of the models given the observations. Then, the model that gets the highest probability is chosen to detect the current health state of the component, and to estimate its RUL. The estimation of the RUL and the confidence is done in five steps, which are explained hereafter.

- 1) The first step consists in identifying the model λ_{DBN} , from the list of learned models, that best represents the sequence of current observations (Fig. 5). The best model corresponds to that which gives the highest likelihood probability $P[O|\lambda_{\text{DBN}}]$ calculated by the forward-backward algorithm.
- 2) The second step is the identification of the component's current health state. The Viterbi algorithm is used on the selected model to define the sequence of states that corresponds to the current observed features. The state which is most repeated at the end of the sequence is then retained as the active state.
- 3) The third step corresponds to the utilization of the current health state, the final state, and the transition probability matrix A of the selected DBN to find the critical path, which goes from the current state to the final state (failure). For this procedure, all transition probabilities a_{ij} which are different from zero are used to define the shortest path by considering only one visit per state. Similarly, the longest path is defined as the path which goes through the maximum number of states.
- 4) Finally, in the fourth step, the previously defined paths are used to calculate the RUL. This RUL is obtained by using the stay durations in each state of the selected DBN. In addition, the confidence value is computed by using the standard deviations on the stay durations. The confidence coefficient n (for a confidence interval α) is calculated according to

$$\Phi(n) = \frac{\alpha + 100}{200} \quad (6)$$

where Φ is the distribution function of the centered normal law, and $\alpha \in [0, 100]$. Consequently, three values of RUL are estimated: upper RUL ($\mu + n \cdot \sigma$), mean RUL, and lower RUL ($\mu - n \cdot \sigma$).

$$\text{RUL}_{\text{upper}}(t) = \sum_{i=\text{current state}}^N [v_i \cdot \mu_d(x_i) + n \cdot \sigma_d(x_i)] - \tilde{t}_{ac} \quad (7)$$

$$\text{RUL}_{\text{mean}}(t) = \sum_{i=\text{current state}}^N v_i \cdot \mu_d(x_i) - \tilde{t}_{ac} \quad (8)$$

$$\text{RUL}_{\text{lower}}(t) = \sum_{i=\text{current state}}^N [v_i \cdot \mu_d(x_i) - n \cdot \sigma_d(x_i)] - \tilde{t}_{ac} \quad (9)$$

$$\forall i \in \text{state in the active path} \wedge v_i \in [0, \Omega_{x_i}]$$

In the previous expressions, \tilde{t}_{ac} represents the time spent by the component in the active state. This time is estimated through the equation

$$\tilde{t}_{ac} = \begin{cases} 0, & x_t \neq x_{t-1} \\ \tilde{t}_{ac}(t-1) + \Delta t, & x_t = x_{t-1} \end{cases} \quad (10)$$

The variable ν_i represents the number of possible remaining visits to the state x_i . This variable is initialized to the number of visits to the state x_i , i.e. $\nu_i = \Omega_{x_i}$. Once the process of current state detection has found a change in state ($x_t \neq x_{t-1} \wedge \lambda_{\text{DBN}}(t-1) = \lambda_{\text{DBN}}(t)$) or model ($\lambda_{\text{DBN}}(t) \neq \lambda_{\text{DBN}}(t-1)$), the variable ν_i is reduced by one unit.

E. Prognostics performance metrics

The performance of the proposed prognostics method is assessed through the implementation of prognostics metrics proposed in [8], and [22]. A brief description of the implemented metrics is given hereafter.

- 1) Accuracy [8]: An accuracy value near to zero means that the predictions are not good, while an accuracy near one corresponds to good predictions.

$$Accuracy = \frac{1}{T} \sum_{t=1}^T e^{-\frac{|RUL_{real}(t) - RUL(t)|^\alpha}{RUL_{real}(t)}} \quad (11)$$

In (11) $\alpha \in [1, 2]$, T is the final time of the historical test, $RUL_{real}(t)$ is the real value of the RUL at time t , and $RUL(t)$ is the estimated value of the RUL at time t .

- 2) Precision [22]: this measure quantifies the dispersion of the prediction error around its mean.

$$Precision = \sqrt{\frac{\sum_{t=1}^T (\mathcal{E}(t) - \bar{\mathcal{E}})^2}{T}} \quad (12)$$

where

$$\begin{aligned} \mathcal{E}(t) &= RUL_{real}(t) - RUL(t) \\ \bar{\mathcal{E}} &= \frac{1}{T} \sum_{t=1}^T \mathcal{E} \end{aligned}$$

- 3) Mean Absolute Percentage Error (MAPER): this measure quantifies the mean error in percentage.

$$MAPER = \frac{1}{T} \sum_{t=1}^T \left| \frac{100 \cdot \mathcal{E}(t)}{RUL_{real}(t)} \right| \quad (13)$$

- 4) Prognostics Horizon (HP) [22]: this measure estimates the time at which the prognostics algorithm gives its first prediction within the confidence interval defined by $\alpha_c \in [0, 1]$.

$$HP = T - t \quad \forall RUL(t) \in [RUL_{real}(t) \cdot (1 - \alpha), RUL_{real}(t) \cdot (1 + \alpha)] \quad (14)$$

- 5) Performance measure ($P_{\alpha_c - \lambda_c}$) [22]: it permits one to test if at one or more temporal instants $t = \lambda_c \cdot T \quad \forall \lambda_c \in [0, 1]$ the predicted RUL is within the confidence interval defined by the coefficient α_c .

$$P_{\alpha_c - \lambda_c} = \begin{cases} No & \text{if other} \\ Yes & \text{if } RUL(t) \in [RUL_{real}(t) \cdot (1 - \alpha_c), RUL_{real}(t) \cdot (1 + \alpha_c)] \end{cases} \quad (15)$$

- 6) Relative accuracy measure (RA) [22]: it permits one to assess the accuracy of the estimation of the RUL at different times $t = \lambda_c \cdot T \quad \forall \lambda_c \in [0, 1]$.

$$RA = 1 - \frac{|RUL_{real}(t) - RUL(t)|}{RUL_{real}(t)} \quad (16)$$

III. APPLICATION AND EXPERIMENTAL RESULTS

The method proposed in the previous section is applied hereafter on real experimental data taken from accelerated life tests performed on bearings. First, the choice of bearing as a critical component is motivated, then experimental data and RUL estimation results are given and discussed.

A. Bearing as critical component

During the last decade, several works have been conducted in the field of failure diagnostics and prognostics of critical physical components such as bearings, gears, batteries, etc. Interested readers can find more details in [12], [23]–[25]. The analysis of the experience feedback performed on electrical machines by the Electric Power Research Institute (ERPI), and researchers in the reliability of electrical machines, has shown that the bearings and the stator are the components which present the most failures (table I). In the following application, bearings are considered as the critical components on which the prognostics method is tested.

TABLE I: Failure distribution in asynchronous motors.

	Failure percentage %			
	Bloch and Geitner [26]	O'Donnell [27]	IEEE - ERPI [28]	Albrecht <i>et al.</i> [29]
Bearings	41	45 - 50	45 - 55	41
Stator	37	30 - 40	26 - 36	36
Shaft	10	8 - 12		9
Others	12			14

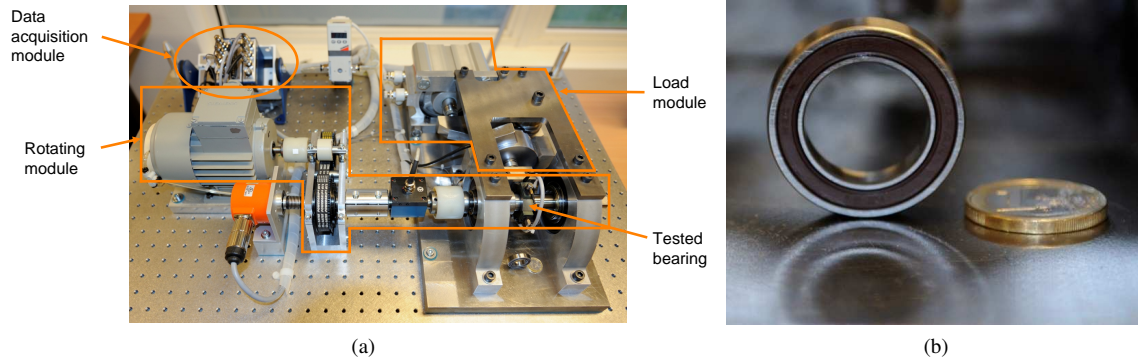


Fig. 6: Experimental platform Pronostia: (a) global overview and (b) type of the tested bearing.

B. Experimental platform: Pronostia

Pronostia is an experimental platform dedicated to test, verify, and validate methods related to bearing health assessment, diagnostics, and prognostics. A general overview of the main parts of the platform and the type of the bearing used for tests are shown in Fig. 6. The main objective of Pronostia is to provide real data related to the degradation of bearings during their useful life. Particularly, in this application, we use the NSK 6804DD bearing, which can operate at a maximum speed of 13000 rpm, and a load limit of 4000 N.

Pronostia is composed of two main parts: a first part related to the speed variation, and a second part dedicated to load profile generation. The speed variation part is composed of a synchronous motor, a shaft, a set of bearings, and a speed controller. The synchronous motor develops a power equal to 1.2 kW, and its operational speed varies between 0 and 6000 rpm. The load profiles part is composed of a hydraulic jack connected to a lever arm used to create different loads to degrade the bearing mounted on the platform. The radial load can be varied between 0 and 10,000 N, and the operating speed of the bearing can be controlled within the interval 0 - 2000 rpm. The force delivered by the pneumatic jack is indirectly applied on the external ring of the bearing through a clamping ring (Fig. 7-(b)). The effort is transmitted by a lever arm in rotation, which applies the load on the clamping ring.

Two high frequency accelerometers (DYTRAN 3035B) are mounted, one horizontally and one vertically, on the housing of

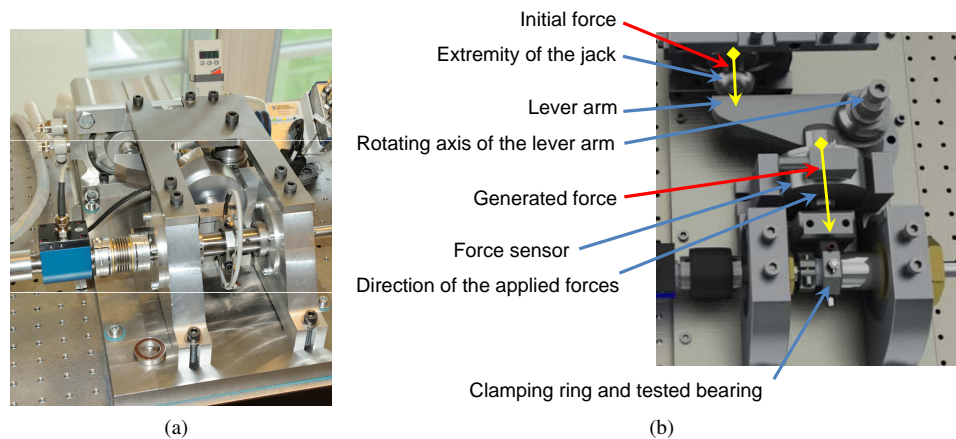


Fig. 7: Multiplication system and load transmission: (a) top view and (b) elements of the transmission chain.

the tested roller bearing to pick up the horizontal and the vertical accelerations. In addition, the monitoring system includes one temperature probe (of type PT100) to record the temperature of the tested bearing. A speed sensor and a torque sensor are also available on the Pronostia platform. The sensors are connected to a data acquisition card (NI DAQCard-9174) to provide the user with monitoring data. The sampling frequency is set to 25600 Hz, and the vibration data provided by the two

accelerometers are collected every 1 second.

C. Experimental data

A total of 12 accelerated life tests (separated in two groups of 6 tests) under two constant operating conditions are realized on the Pronostia platform. In this paper, only the first six tests (first group) are used. The six experiments are realized under specific operating conditions: the constant speed of the shaft is controlled at 1800 rpm, and the radial load at 4000 N (Table II).

For the estimation of the RUL and the associated confidence, the historical data related to tests 1, 3, 4, 5, and 6 are used in the learning phase to estimate the parameters of the DBNs, representing the degradation, while the historical data related to test 2 is used in the exploitation phase to calculate the RUL, and the confidence.

Examples of bearings' failure at the end of experiments are shown in Fig. 8. Two types of features are extracted from

TABLE II: Summary of the accelerated life tests on Pronostia.

Experiment	Speed (rpm)	Load (N)	Duration	Failure mode
1	1800	4000	3H25	IR, ER
2	1800	4000	6H50	ER
3	1800	4000	6H48	ER
4	1800	4000	6H16	ER
5	1800	4000	1H00	ER
6	1800	4000	1H12	ER, IR

ER = Extern ring, IR = Inner ring



Fig. 8: Failure of the bearings of Pronostia: (a) extern ring failure, and (b) inner ring failure.

the raw signal provided by the two accelerometers: temporal features [30] (Root Mean square: RMS, mean value, pick value, crest factor, skewness, etc.), and time-frequency features (energy levels of the signals extracted by using a Wavelet Packet Decomposition: WPD). An example of two temporal features extracted from the horizontal accelerometer is shown in Fig. 9. The drawback of temporal features is that the frequency part is not considered. To overcome this situation, one can use wavelet

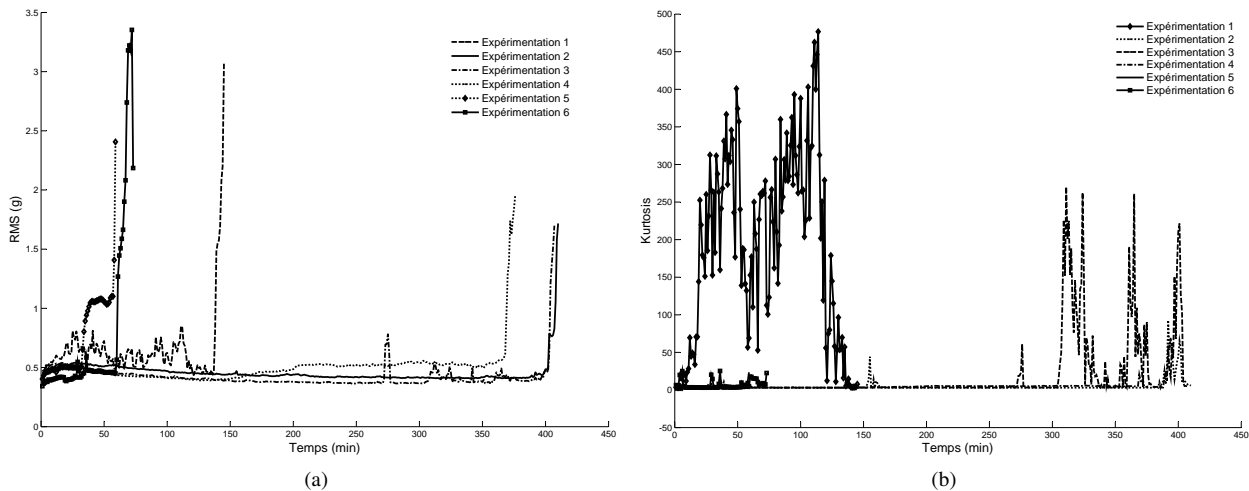


Fig. 9: Features extracted from the radial accelerometer: (a) RMS, and (b) kurtosis.

transforms, such as WPD, which permits one to adjust the size of the temporal window according to the analyzed frequencies. A WPD has two parameters: a scale parameter a for the frequency, and a translation parameter b for the time [31]. The original

signal can then be decomposed into several levels. The decomposition level can be defined using (17).

$$J_f \leq \log_2 \frac{F_s}{3F_d} - 1 \tag{17}$$

where J_f is the decomposition level, F_s is the sampling frequency, and F_d is the maximum frequency among the bearing's frequencies.

Fig. 10 shows two results about the percentage of energy extracted from the last levels of decomposition on the frequency band 1 - 800 Hz.

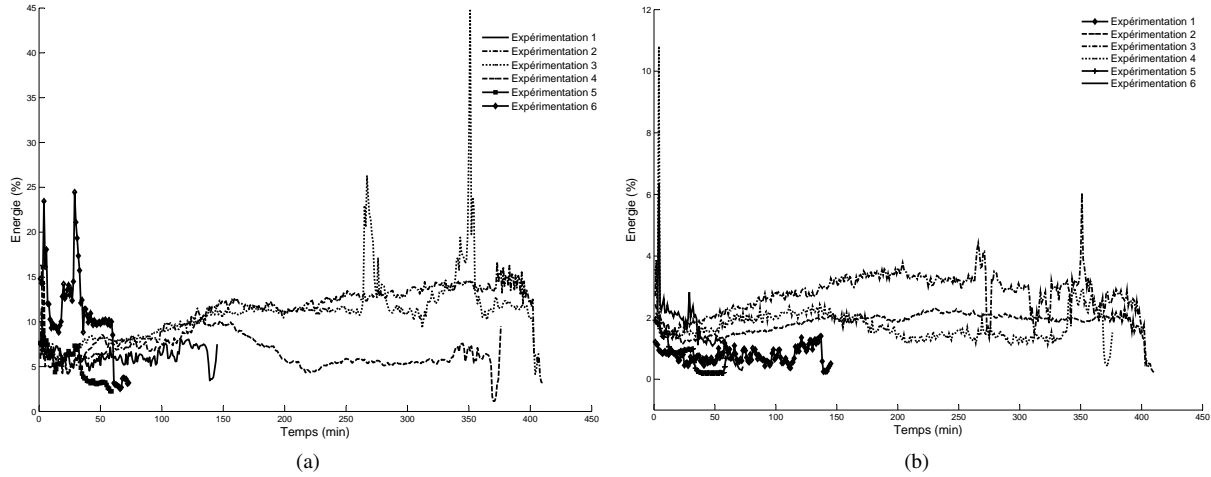


Fig. 10: Percentage of energy of the vertical and horizontal accelerometers after decomposition in wavelet packets for the band 1 - 800 Hz.

D. Degradation modeling

During the learning of the MoG-HMMS models (represented by DBNs) related to the degradations, two parameters are needed: the number of states N , and the number of mixtures of Gaussians M . In this application, the number of states is set to 3 ($N = 3$). This choice of number of states is motivated by the fact the degradation of the bearings can be represented by three levels: the nominal health condition, the initiation and propagation of the defect, and the failure level.

For the definition of the number of mixtures, a prior analysis is done. The results of this analysis are shown in Fig. 11. The number of mixtures has been varied during the model learning related to each experiment. Then, for a given number of mixtures, a likelihood $P[O|\lambda_{DBN}]$ is calculated. Finally, the number of mixtures is taken as small as possible (to avoid model complexity) when the likelihood starts to converge. In the following application, the number of mixtures is set to $M = 4$.

The parameters π , A , and B of the DBNs representing the degradations of the bearings are learned using the Viterbi algorithm,

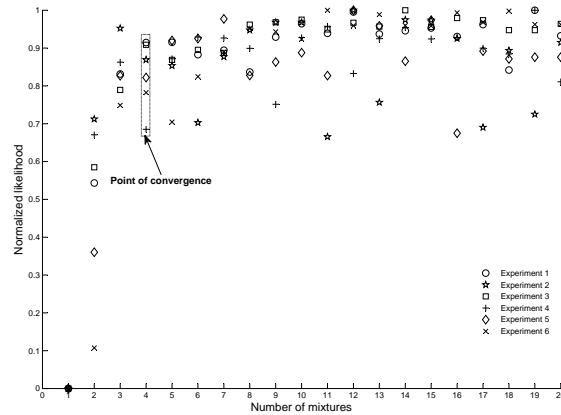


Fig. 11: Sensibility analysis to define the number of mixtures M .

and the extracted features (temporal and time-frequency). An example of a generated Viterbi state sequence is shown in Fig. 12. The obtained parameters of the DBN related to each experiment are given in Tables III, and IV.

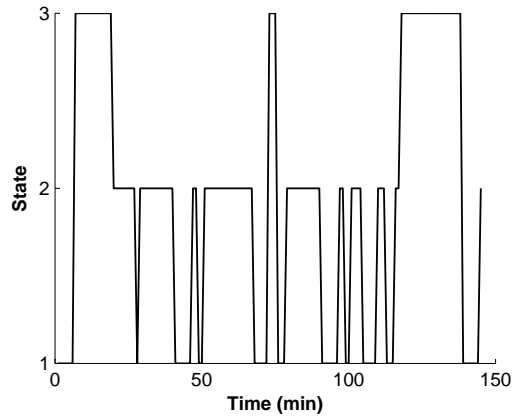


Fig. 12: State sequences obtained by using Viterbi algorithm for experiment 1.

TABLE III: Temporal parameters of the learned model by using temporal features.

State	Parameters (min)	Experiment					
		1	3	4	5	6	
x_1	μ_d	4.1	7.83	9	1	1.37	
	σ_d	1.9	3.3	0	0	0.74	
	Ω	11	6	1	12	8	
x_2	μ_d	6.8	6.6	366	17.5	3.27	
	σ_d	5.6	2	0	9.19	3.84	
	Ω	10	5	1	2	11	
x_3	μ_d	2.66	81.7	1	1	8.66	
	σ_d	9.07	126.3	0	0	5.13	
	Ω	3	4	1	12	3	

E. RUL estimation, and prognostics performance metrics

The models λ_{DBN} obtained during the learning phase are used to estimate the RUL, and to assess the performance of the method through prognostics metrics. In this application, the forgotten coefficient l is set to 10, which means that the recurrent state in the last 10 minutes is chosen as the most probable active state. For the estimation of the confidence limits, ($\text{RUL}_{\text{upper}}$, and $\text{RUL}_{\text{lower}}$), an interval of 95% is defined ($\alpha = 95$).

Concerning the prognostics performance metrics, α_c is set to 0.3. This value permits us to tolerate a deviation of 30% from the real RUL.

For the calculation of $P_{\alpha_c - \lambda_c}$, λ_c is set to 0.5 to verify if at the middle of the historical data length ($T/2$) the predictions are within the confidence interval α_c . Concerning the relative accuracy, it is assessed for three values $\lambda_c = [0.75, 0.5, 0.25]$ corresponding to three time intervals: the beginning, the middle, and the end of the historical data. The results obtained from the implementation of the prognostics performance metrics on temporal and time-frequency features are given in Table V. The prognostics results obtained for the experiments realized on the Pronostia platform are shown in Fig. 13. The results shown in Fig. 13-(a) are obtained from the temporal features, while those obtained from wavelet features are given in Fig. 13-(b).

By observing the results shown in Fig. 13-(a), one can note that there is no change in the confidence limits. In addition, the estimated RUL is a line which is parallel to the real RUL. This result can be explained by the fact that, during the estimation of the RUL, the best model was always that which corresponds to Experiment 4. According to the learning results given in Table III, the historical data of Experiment 4 is represented by the left-right model (only transitions from left states to right states are allowed in the MoG-HMM), which produces null standard deviations σ_d . However, it can be seen that the estimated

TABLE IV: Temporal parameters of the learned model by using time-frequency features.

State	Parameters (min)	Experiment					
		1	3	4	5	6	
x_1	μ_d	63	18.25	29.25	10	8.33	
	σ_d	0	24.54	38.9	9.84	9.29	
	Ω	1	4	4	3	3	
x_2	μ_d	20.5	62.5	50.33	12	10	
	σ_d	3.53	27.58	79.46	11.31	13	
	Ω	2	2	3	2	3	
x_3	μ_d	20.5	69.67	108	1.66	6	
	σ_d	17.68	50.14	0	1.15	2.64	
	Ω	2	3	1	3	3	

RUL is relatively precise, as confirmed by the performance measures given in Table V (accuracy equal to 0.7874). Moreover, the precision is good with a dispersion of 4.91 min. Also, the value of the HP measure is equal to 409 min, which means that 1 min after the launch of the prognostics method, the predictions entered the confidence interval.

The prognostics results obtained by using the wavelet features are shown in Fig. 13-(b). Contrary to the results obtained from

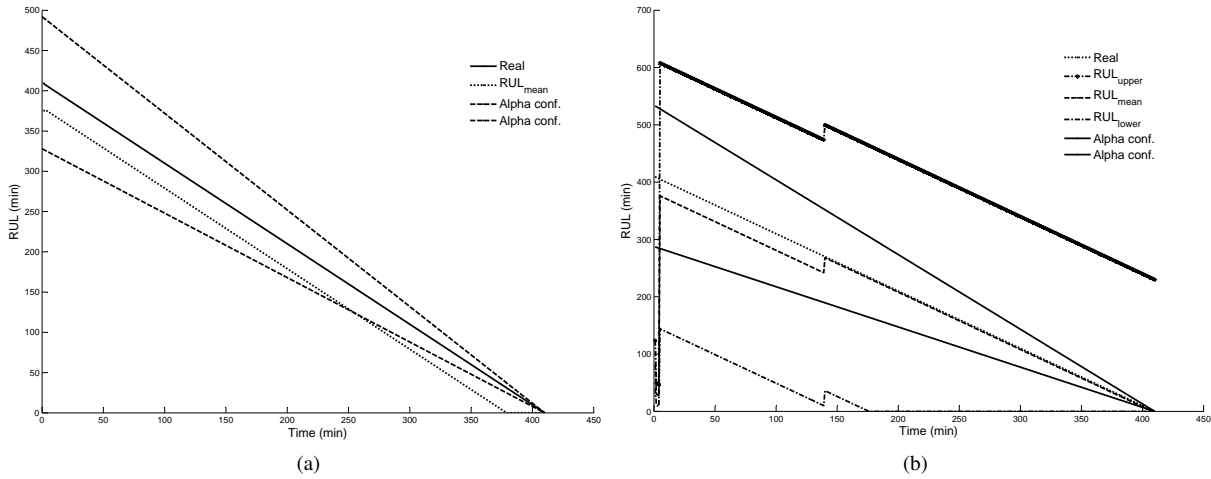


Fig. 13: Estimation of the RUL for experiment 2: (a) temporal features, and (b) time-frequency features.

TABLE V: Prognostics performance metrics.

	Accuracy	Precision	MAPER	HP	$P_{\alpha-\lambda}$	RA
RUL _{Temporal} ^{mean}	0.7874	4.91	26.96	409	No	[0.89, 0.84, 0.69]
RUL _{Wavelets} ^{upper}	0.309	56.79	367.4	False	No	[0.33, -0.12, -1.25]
RUL _{Wavelets} ^{mean}	0.9426	37.18	6.5	405	Yes	[0.9, 0.99, 0.98]
RUL _{Wavelets} ^{lower}	0.402	88.16	91.93	False	No	[0.15, 0, 0]

the temporal features, one can observe a variation in the confidence limits. From Fig. 13-(b), one can see that the real RUL is within the confidence interval during the whole experiment. This result can be explained by the fact that the time-frequency features are more sensitive to the degradation, allowing us to update the models λ_{DBN} , and to calculate the RUL value. This behavior, sensitive to the degradation, has allowed us to create good estimations of the RUL_{mean} . In addition, this estimation is confirmed by a prognostics horizon equal to 405 min, a good value of the measure $P_{\alpha-\lambda_c}$, and relative accuracies near 1.

IV. CONCLUSION

A data-driven prognostics method based on the utilization of data provided by monitoring sensors is presented in this paper. The aim of the method is to estimate the remaining useful life of an industrial plant by focusing on its most critical component. The degradation of this component is then modeled using Mixture of Gaussians Hidden Markov Models (represented by DBNs), which permit us to represent the evolution of the degradation using hidden states. The estimation of the remaining useful life is done in two main phase: a learning phase to derive the degradation's behavior model, and an exploitation phase to estimate the RUL and its associated confidence.

The proposed method is applied to real data related to accelerated life tests on bearings, and the results of the RUL are given and discussed. Two types of features are tested: temporal, and frequency. Furthermore, prognostics metrics are implemented to assess the performance of the prognostics results obtained by the method.

The proposed prognostics method has been applied to the degradation of bearings operating at constant conditions (constant load, and speed). This method can be easily applied to other types of physical components. It can be also extended to variable operating conditions. However, this latter point requires one to adapt the model learning and selection process, and also the computation of the RUL.

REFERENCES

- [1] M. Lebold and M. Thurston, "Open standards for condition-based maintenance and prognostic systems," *Maintenance and Reliability Conference (MARCON)*, 2001.
- [2] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483 – 1510, 2006.
- [3] A. Heng, A. C. Tan, J. Mathew, N. Montgomery, D. Banjevic, and A. K. Jardine, "Intelligent condition-based prediction of machinery reliability," *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1600 – 1614, 2009.

- [4] V. Venkatasubramanian, "Prognostic and diagnostic monitoring of complex systems for product lifecycle management: Challenges and opportunities," *Computers & Chemical Engineering*, vol. 29, no. 6, pp. 1253 – 1263, 2005.
- [5] W. Q. Wang, M. F. Golnaraghi, and F. Ismail, "Prognosis of machine health condition using neuro-fuzzy systems," *Mechanical Systems and Signal Processing*, vol. 18, no. 4, pp. 813 – 831, 2004.
- [6] A. Muller, M.-C. Suhner, and B. Iung, "Formalisation of a new prognosis model for supporting proactive maintenance implementation on industrial system," *Reliability Engineering & System Safety*, vol. 93, no. 2, pp. 234 – 253, 2008.
- [7] AFNOR, "Condition monitoring and diagnostics of machines - prognostics - part I: General guidelines. NF ISO 13381-1," 2005.
- [8] G. Vachtsevanos, F. L. Lewis, M. Roemer, A. Hess, and B. Wu, *Intelligent fault diagnosis and prognosis for engineering systems*. Wiley, 2006.
- [9] J. Luo, K. R. Pattipati, L. Qiao, and S. Chigusa, "Model-based prognostic techniques applied to a suspension system," *Transactions on Systems, Man, and Cybernetics*, vol. 38, pp. 1156–1168, 2003.
- [10] D. Chelidze and J. Cusumano, "A dynamical systems approach to failure prognosis," *Journal of Vibration and Acoustics*, vol. 126, pp. 2 – 8, 2004.
- [11] M. Dong and D. He, "A segmental hidden semi-markov model (hsmm)-based diagnostics and prognostics framework and methodology," *Mechanical Systems and Signal Processing*, vol. 21, pp. 2248–2266, 2007.
- [12] H. Ocak, K. A. Loparo, and F. M. Discenzo, "Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics," *Journal of sound and vibration*, vol. 302, pp. 951–961, 2007.
- [13] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, vol. 77 (2), 1989, pp. 257–286.
- [14] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot, "A mixture of gaussians hidden markov model for failure diagnostic and prognostic," *IEEE Conference on Automation Science and Engineering, CASE'10*, 2010.
- [15] L. Baum and J. Egon, "An inequality with applications to statistical estimation for probabilistic functions of a markov process and to a model for ecology," *Bulletin of the American Mathematical Society*, vol. 73, pp. 360 – 363, 1967.
- [16] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm," *IEEE Transaction on Information Theory*, vol. 13, pp. 260 – 269, 1967.
- [17] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, pp. 1 – 38, 1977.
- [18] K. P. Murphy, "Dynamic bayesian networks: Representation, inference and learning," Ph.D. dissertation, University of California, 2002.
- [19] G. Carmignani, "An integrated structural framework to cost-based fmeca: The priority-cost fmeca," *Reliability Engineering & System Safety*, vol. 94, no. 4, pp. 861–871, 2009.
- [20] D. A. Tobon-Mejia, K. Medjaher, and N. Zerhouni, "The iso 13381-1 standard's failure prognostics process through an example," *IEEE - Prognostics & System Health Management Conference*, University of Macau, Macau, China, 12-14 January 2010 2010, in Press.
- [21] W. Yan, H. Qiu, and N. Iyer, "Feature extraction for bearing prognostics and health management (PHM)-a survey," Air Force Research Laboratory, Tech. Rep., 2008.
- [22] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, "Metrics for offline evaluation of prognostic performance," *International Journal of Prognostics and Health Management*, 2010.
- [23] Q. Miao, D. Wang, and M. Pecht, "A probabilistic description scheme for rotating machinery health evaluation," *Journal of Mechanical Science and Technology*, vol. 24, no. 12, pp. 2421 – 2430, 2010.
- [24] Y. Pan, J. Chen, and G. Dong, "A hybrid model for bearing performance degradation assessment based on support vector data description and fuzzy c-means," *Journal of Mechanical Engineering Science*, vol. 223, pp. 2687 – 2695, 2009.
- [25] D. Wang, Q. Miao, and R. Kang, "Robust health evaluation of gearbox subject to tooth failure with wavelet decomposition," *Journal of Sound and Vibration*, vol. 324, no. No.3-5, pp. 1141 – 1157, 2009.
- [26] H. P. Bloch and F. K. Geitner, *Machinery failure analysis and troubleshooting*, Elsevier, Ed. Elsevier, 1999.
- [27] P. O'Donnell, "Report of large motor reliability survey of industrial and commercial installations, part I, II & III," *IEEE Transactions on Industry Applications*, vol. 21, pp. 853–872, 1985.
- [28] C. Lanham, "Understanding the tests that are recommended for electric motor predictive maintenance," Baker Instrument Company, Tech. Rep., 2002.
- [29] P. Albrecht, J. Appiarius, and D. Sharma, "Assesment of the reliability of motors in utility applications - updated," *IEEE Transactions on Energy Conversion*, vol. EC-1, pp. 39–46, 1986.
- [30] T. Kurfess, S. Billington, and S. Liang, "Advanced diagnostic and prognostic techniques for rolling element bearings," *Springer Series in Advanced Manufacturing*, pp. 137–165, 2006.
- [31] J. Zarei and J. Poshtan, "Bearing fault detection using wavelet packet transform of induction motor stator current," *Tribology International*, vol. 40, no. 5, pp. 763 – 769, 2007.

Kamal Medjaher is associate Professor at the French high school of mechanics and micro-techniques in Besançon since September 2006, where he teaches control and fault diagnostics and prognostics. After receiving an engineering degree in electronics, he received his MS in control and industrial computing in 2002 at the “Ecole Centrale de Lille,” and his PhD in 2005 in the same field from the University of Lille 1. Since September 2006, Dr. Medjaher leads research work in the field of failure prognostics, and uses artificial intelligence tools, particularly probabilistic graphical models.

Diego Alejandro Tobon Mejia was born in Medellin, Colombia, on February 16, 1985. Diego A. TOBON-MEJIA, was honored in 2002 with the “Excellence scholarship” from the “EEPP de Medellin” (Medellin Public Enterprises) to perform his studies in Colombia. In 2006, he was honored by the French Ministry of Foreign affairs with the “Eiffel Excellence scholarship” to continue his studies in France. He received the B.Sc. and S.M., both in 2008, from the National Engineering School in Metz (France) and the EAFIT University (Colombia). After his master studies, he prepared a PhD thesis at Franche-Comté University, in Besançon (France) sponsored by Alstom Transport where he worked as a research engineer. He is engaged in research on rotating machinery failure prognostics at the FEMTO-ST institute, and Alstom transport.

Noureddine Zerhouni received his engineering degree from National Engineers and Technicians School of Algiers (ENITA) in 1985. After a short period in industry as an engineer, 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 is a Professor at the national high school of mechanics and microtechniques of Besançon. He is now the head of AS2M Department within FEMTO-ST Institute. His main research activities are concerned with intelligent maintenance systems and e-maintenance. Professor Noureddine Zerhouni has been and is involved in various European and National projects on intelligent maintenance systems.