Parameter estimation algorithm for bearing prognostics through monovariate generalized Gaussian hidden Markov models

E. Soave, G. D'Elia, G. Dalpiaz, E. Mucchi
University of Ferrara, Department of Engineering,
via Saragat 1, 44122 Ferrara, Italy

Abstract
Nowadays, the industrial scenario is driven by the need of costs and time reduction. In this contest, system failure prediction plays a pivotal role in order to program maintenance operations only in the last stages of the real operating life, avoiding unnecessary machine downtime. In the last decade, Hidden Markov Models have been widely exploited for machinery prognostic purposes. The probabilistic dependency between the measured observations and the real damaging stage of the system has usually been described as a mixture of Gaussian distributions. This paper aims to generalize the probabilistic function as a mixture of generalized Gaussian distributions in order to consider possible distribution variations during the different states. In this direction, this work proposes an algorithm for the estimation of the model parameters exploiting the observations measured on the real system. The prognostic effectiveness of the resulting model has been demonstrated through the analysis of several run-to-failure datasets concerning both rolling element bearings and more complex systems.

1 Introduction

In the last decades, the industrial attention has been even more pointed out on the system reliability, pivotal aspect for the target of cost and time reduction. In this contest, a suitable tool for the reduction of unnecessary maintenance operation is represented by the Condition Based Maintenance (CBM) which allows the definition of a maintenance plan based on the real time condition monitoring of the system. In this field, starting from the 1960s the machine prognostics has been widely studied with the aim of describing the actual degradation level basing on the historical and ongoing damaging trend [1].

Firstly, the prognostic approaches were based on physical models that reproduce the failure propagation with mathematical models that take into account the stress levels and the material properties. In this family of method, Paris and Erdogan can be considered as the pioneer of the failure description [2]. Over the years, their description of the crack grown has been considered as the base for the development of several physical based models applied on the most common mechanical components as described in Ref [3].

Several years later, the difficult in describing the damaging process in complex systems led to the development of another family of prognostic model based on the Artificial Intelligence (AI). In this contest, the most exploited algorithm is represented by the Artificial Neural Network (ANN) which tries to describe the working process of the human brain for analysing huge amount of data from the physical system. In the last years, several AI based prognostic models have been presented for the assessment of the damaging level on mechanical systems and the estimation of the Remaining Useful Life (RUL), i.e. the time left until the final failure. Between them a particular mention has to be given to the work of Gebraeel [4] and Xiao [5]. The general requirement of the AI is the high amount of data needed for the training of the model and this is not even possible to obtain in real industrial application.

A good compromise between the application on complex system and the need of a reasonable number of training data is represented by the statistical based approaches which define the degradation process in form

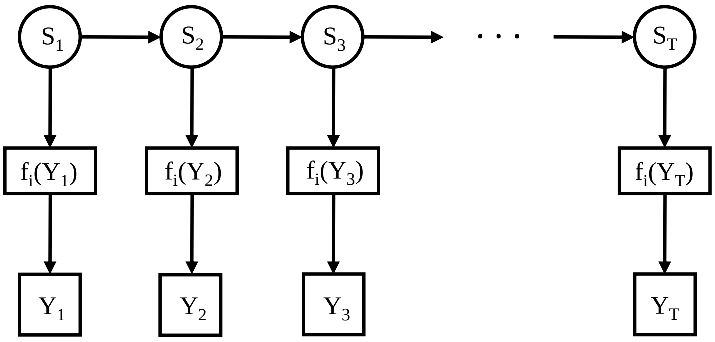


Figure 1: Dependence graph for observations and state sequence inside an HMM.

of a Probability Density Function (PDF) depending on some diagnostics indicators, i.e. the observations, that describe the failure process. In this contest, the Hidden Markov Model (HMM) enables the description of the damaging process as a sequence of finite states where the transition between them follows the principle of the Markov chain [6]. The theory of the HMMs, presented by Bum et al. [7], has been firstly applied for speech recognition [8] but in the last decades it has been taken into account for the prognostics of rotating machines and mechanical systems [9].

Statistically speaking, the conditional distribution that relates the physical observation with the actual health state is the component of a mixture distribution and consequently the PDFs related to each model state should belong to the same distribution family. Unfortunately, the system modifications resulting from the damaging process modify the observation distribution in the last part of the working life. The aforementioned assumption could lead to estimation errors in the last damaging stages affecting the quality and the effectiveness of the maintenance program. This issue may be overcome through the exploitation of a generalized distribution as the mixture component. This assumption allows the consideration of the distribution modification within the model states through different values of the distribution parameters. In this direction, this work proposes a novel iterative algorithm for the estimation of the model parameters for a Generalized Gaussian Distribution (GGD) HMM based on the observation measured directly on the physical system.

The paper is organized as follows: Section 2 provides a brief introduction about the theoretical background of HMMs. Then, the iterative algorithm for the model parameters estimation in the monovariate GGD case is described in Section 3. Section 4 describes an experimental validation carried out on a rolling element bearing run-to-failure test performed on the bearing test rig of the University of Ferrara. Finally, Section 5 provides some final remarks.

2 Theoretical background: hidden Markov models

The basis idea of HMMs is definition of a state sequence S = $\{S_1, S_2, \dots , S_T\}$ starting from a vector of observations, i.e. some diagnostic indicators that describe the damaging process of the system, Y = $\{Y_1, Y_2, \dots , Y_T\}$, where T is the total number of observed time spans. Two consecutive elements of S are related in a probabilistic way according to a first order discrete Markov process:

$$P(S_t|S_1, \dots, S_{t-1}) = P(S_t|S_{t-1}) \tag{1}$$

Given a number of possible damaging state N, the state variables S_t are taken from a finite set $s = \{1, \dots , N\}$ such that $S_t = i, i \in s$, i.e. the HMM is discrete. All the transitions between the state of an HMM are described by a transition matrix $\mathbf{A}(t)$ which entries are represented by the transition probabilities defined in Eq.1, viz:

$$a_{ij}(t) = P(S_{t+1} = j | S_t = i), \qquad i, j = 1, \ldots, N \tag{2}$$

The term hidden refers to the probabilistic relation between the observations and the state sequence, i.e. the state sequence does not correspond to an observable event. This relation, explained in Fig.1, is given by a PDF that makes the state observable mapping S_t into Y_t:

$$f_i(Y_t) = f(Y_t | S_t = i), \qquad i = 1, \ldots, N \tag{3}$$

Due to the assumption of a finite set of S, the marginal distribution of the data is a mixture of N components [10], such as:

$$f(Y_t) = \sum_{i=1}^{N} p_i f_i(Y_t) \tag{4}$$

where p_i are the component proportions.

Finally, the HMM can be completely defined through another parameter, known as the prior (or initial state) probability vector Π, that describes the probability of the system to be in a given state at the first time span, with entries:

$$\pi_i = P[S_1 = i], \qquad i = 1, \ldots, N \tag{5}$$

The previous described model parameters are defined during the training process starting from the physical observations in order to maximize the likelihood function of the model. Once the model parameters have been estimated, i.e. the model have been trained, the optimal state sequence, i.e. the sequence that maximizes the condition probability between state and observations, can be estimated through the Viterbi algorithm [11].

3 Parameter estimation for generalized Gaussian distribution based hidden Markov models

The target of the estimation procedure is the calculation of the model parameters that maximize the likelihood of the observations given a certain state sequence [12]. Before starting with the algorithm description it is mandatory to define some auxiliary variables, i.e. the forward and backward variables. The first type is defined as the combination of different state sequences that leads to the same state S_t [13]:

$$\alpha_1(i) = \pi_i f_i(Y_1), \qquad i = 1, \ldots, N \tag{6a}$$

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} f_j(Y_t), \qquad t = 2, \ldots, T, j = 1, \ldots, N \tag{6b}$$

From the opposite perspective, the backward variables are defined as the probabilities of the observations from $t+1$ to T given the state S_t such as:

$$\beta_T(i) = 1, \qquad i = 1, \ldots, N \tag{7a}$$

$$\beta_t(i) = \sum_{j=1}^{N} \beta_{t+1}(j) a_{ij} f_j(Y_{t+1}), \qquad t = T-1, \ldots, 1, i = 1, \ldots, N \tag{7b}$$

Starting from the aforementioned variables it is possible to define the probability of being in state i at the time t as:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \tag{8}$$

and, analogously, the probability of moving from state i to state j at time t as:

$$\epsilon_t(i,j) = \frac{\alpha_t(i) a_{ij} f_j(Y_{t+1}) \beta_{t+1}(j)}{\sum_{j=1}^{N} \sum_{i=1}^{N} \alpha_t(i) a_{ij} f_j(Y_{t+1}) \beta_{t+1}(j)} \tag{9}$$

From γ and ϵ it is possible to calculate the transition probabilities and the prior probabilities as follows:

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \epsilon_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \qquad i,j = 1,\dots,N \tag{10a}$$

$$\pi_i = \gamma_1(i), \qquad i = 1,\dots,N \tag{10b}$$

For a generic continuous random variable X, the generalized Gaussian PDF is given by:

$$f(x) = \frac{p}{2\Sigma^{\frac{1}{2}} \Gamma\left(\frac{1}{p}\right)} e^{\left[-\left(\frac{|x-\mu|}{\Sigma^{\frac{1}{2}}}\right)^{p}\right]} \tag{11}$$

where μ is the mean value, p is the shape factor, Σ is the scaling factor and Γ represents the gamma function. Eq.11 explains the basis idea of the proposed method: different values of the shape factor p lead to different distributions, e.g. Gaussian ($p = 2$ and $\sigma^2 = \Sigma^2/2$), Laplace ($p = 1$) and uniform distribution ($p \to \infty$).

According to Ref.[12], for a generic ellipsoidal symmetric PDF, i.e. a distribution comprising a quadratic form, given the observation dataset Y = $\{Y_1, Y_2, \dots, Y_T\}$, the mean value and the scale factor that maximize the likelihood function can be calculated as follows:

$$\mu_i = \frac{\sum\limits_{t=1}^{T} \rho_t(i) \beta_t(i) Y_t}{\sum\limits_{t=1}^{T} \rho_t(i) \beta_t(i)}, \qquad i = 1,\dots,N \tag{12a}$$

$$\Sigma_i = \frac{\sum\limits_{t=1}^{T} \rho_t(i) \beta_t(i) (Y_t - \mu_i)^2}{\sum\limits_{t=1}^{T} \alpha_t(i) \beta_t(i)}, \qquad i = 1,\dots,N \tag{12b}$$

where α_t and β_t are the previous defined forward and backward variables and ρ_t is defined as:

$$\rho_t(i) = \sum_{j=1}^{N} \alpha_{t-1}(j) a_{ji} \left[-2 \frac{\partial f_i(x)}{\partial q_i(x)} \bigg|_{x=Y_t} \right], \qquad i = 1,\dots,N \tag{13}$$

where $q_i(x)$ is the quadratic for comprised into the PDF.

According to the basis hypothesis, it is necessary to rewrite Eq.11 to highlight a quadratic form, viz:

$$f(x) = \frac{p}{2\Sigma^{\frac{1}{2}} \Gamma\left(\frac{1}{p}\right)} e^{\left\{-\left[\left(\frac{(x-\mu)^2}{\Sigma}\right)^{\frac{p}{2}}\right]\right.} \tag{14}$$

Starting from Eq.14, the derivative inside the square brackets in Eq.13 becomes:

$$-2 \frac{\partial f_i(x)}{\partial q_i(x)} \bigg|_{x=Y_t} = f_i(x) p_i q_i(Y_t)^{\frac{p_i}{2}-1}, \qquad i = 1,\dots,N \tag{15}$$

Substituting Eq.15 into Eq.13 and remembering the definition of forward variables in Eq.6b, after a simple manipulation ρ_t is given by:

$$\rho_t(i) = \alpha_t(i) p_i q_i(Y_t)^{\frac{p_i}{2}-1}, \qquad i = 1, \ldots, N \tag{16}$$

Finally, substituting Eq.16 into Eq.12a and Eq.12b the expression of mean value and scale factor can be written (taking also into account Eq.8) as:

$$\mu_i = \frac{\sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1} Y_t}{\sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1}}, \qquad i = 1, \ldots, N \tag{17a}$$

$$\Sigma_i = \frac{p_i \sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1} (Y_t - \mu_i)^2}{\sum_{t=1}^{T} \gamma_t(i)}, \qquad i = 1, \ldots, N \tag{17b}$$

Unfortunately, for GGDs, both mean value and scale factor depend on the shape factor and consequently the problem can be solved only with a third equation. As described by Varanasi et al. [14], for a generalized distribution the scale factor and the variance are related as follows:

$$|\Sigma|^{\frac{1}{2}} = \left(\frac{p}{T} \sum_{t=1}^{T} |Y_t - \mu|^p \right)^{\frac{1}{p}} \tag{18}$$

It is possible to express this relation taking into account the Gamma function [15], viz:

$$\Sigma^{\frac{1}{2}} = \left[\sigma^2 \frac{\Gamma\left(\frac{1}{p}\right)}{\Gamma\left(\frac{3}{p}\right)} \right]^{\frac{1}{2}} \tag{19}$$

At this point it should be noticed a pivotal aspect: the left side of Eq.18 can be rewritten according to Eq.17b. At the same time, Eq.19 clearly depicts the proportional relation between variance and scale factor for a given shape factor p. Combining these two consideration, it is clear how the right side of Eq.18 must be weighted in the same way of the left side for proportionality reasons. Consequently, Eq.18 can be rewritten, after some simple manipulations, in the following form:

$$\left[\frac{p_i \sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1} (Y_t - \mu_i)^2}{\sum_{t=1}^{T} \gamma_t(i)} \right]^{\frac{1}{2}} = \left[\frac{p_i^2 \sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1} |Y_t - \mu_i|^{p_i}}{\sum_{t=1}^{T} \gamma_t(i)} \right]^{\frac{1}{p_i}}, \quad i = 1, \ldots, N \tag{20}$$

Finally the shape factor can be found as the zero of the following function:

$$\left[\frac{p_i \sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1} (Y_t - \mu_i)^2}{\sum_{t=1}^{T} \gamma_t(i)} \right]^{\frac{1}{2}} - \left[\frac{p_i^2 \sum_{t=1}^{T} \gamma_t(i) q_i(Y_t)^{\frac{p_i}{2}-1} |Y_t - \mu_i|^{p_i}}{\sum_{t=1}^{T} \gamma_t(i)} \right]^{\frac{1}{p_i}} = 0, \quad i = 1, \ldots, N \tag{21}$$

A first look to Eq.21 brings to the light how the estimated shape factor depends on the value of the mean and the scale factor and consequently the parameter estimation should be based on an iterative algorithm, summarized as follows:

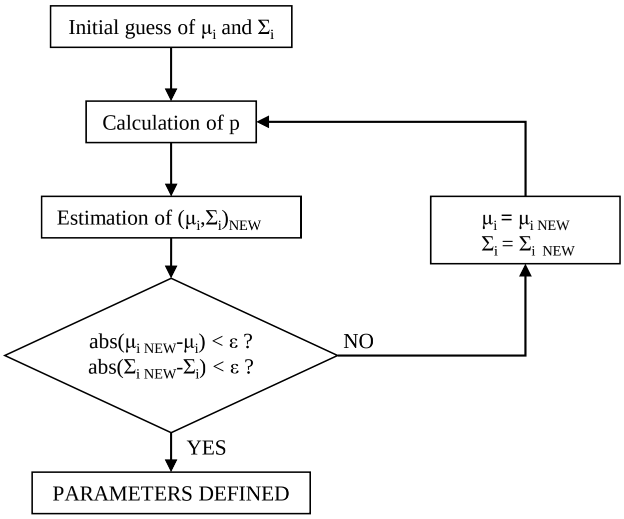


Figure 2: Flow chart of the estimation algorithm for the parameters of a generalized Gaussian HMM.

Step 1: Assume an initial guess for μ_i and σ_i;

Step 2: Calculate the shape factor p_i through Eq.21 by means of a zero finding algorithm;

Step 3: Re-estimate μ_i and σ_i through Eq.12a and Eq.12b, respectively;

Step 4: Repeat Step 2 and Step 3 until convergence.

This iterative parameter estimation algorithm, described with the flow chart in Fig.2, is the base of the proposed HMM, hereafter named Generalized Gaussian Hidden Markov Model (GGHMM).

4 Experimental validation

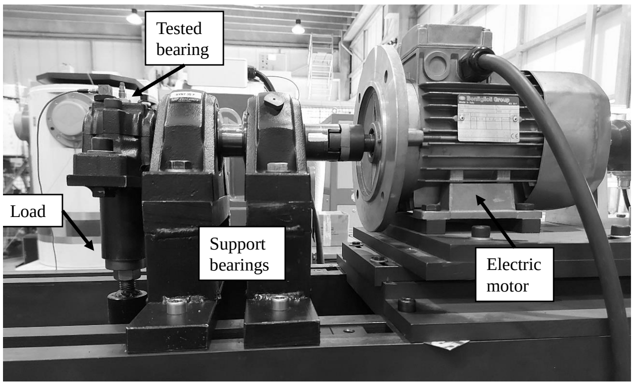
This section provides a comparison between the results obtained with the classic Gaussian based HMM and the proposed GGHMM on a bearing run-to-failure test.

4.1 Experimental setup

The experimental validation regards the analysis of a bearing run-to failure test performed on the bearing test bench at the University of Ferrara. As shown in Fig.3(a), the tested bearing model NSK 1205 ETN9 is cantilever mounted on a shaft supported by two bearings model SKF SYNT 35F and driven by an electric motor. A $3000N$ load is applied on the tested bearing through a lever, regulated by means of a spring system and constantly monitored through a cell load under the bearing housing. The rotating speed has been fixed at $2400rpm$ during the entire test. The vibration signal in radial direction has been measured through a mono-axial piezoelectric accelerometer model PCB 353B18 and continuously acquired by means of a NI CompactRio system with a sampling frequency of $51.2kHz$. The analysis has been carried out on $10s$ length acquisition samples extracted each hour from the dataset. The test has been stopped after 13 days and a deep defect has been found on several rollers as clearly visible in Fig.3(b).

4.2 Results and discussion

Starting from the aforementioned time history, a GGHMM and a classic Gaussian HMM has been built for comparing their results in terms of damaging assessment quality and prediction effectiveness. The time Root



(a) Bearing test bench at the University of Ferrara.

(b) Resulting pitting on the rollers.

Figure 3: Experimental setup and resulting defects at the end of the run-to-failure test.

Table 1: BIC values and distribution intervals of the RMS distribution for the selection of the optimal state number.

Number of states	Distribution intervals	BIC value
1	0-16	1285
2	0-3 , 3-16	131
3	0-3 , 3-10 , 10-16	39
4	0-1, 1-3 , 3-10 , 10-16	45

Mean Square (RMS) shown in Fig.4 has been chosen as the observation vector for describing the degradation trend of the system.

First of all, it is necessary to define the optimal state number N of the model. As previously described, the number of states corresponds to the number of components of the mixture distribution representing the distribution of the physical observations. Fig.4(b) clearly illustrates the problem statement at the base of the proposed model. For low RMS values, i.e. system in healthy conditions, the distribution is Gaussian but moving to the faulty stages, i.e. RMS higher than 4, the data distribution start to move away from the Gaussian form towards a flatterer distribution. The optimal number of state can be found as the number of mixture components that better fits the distribution of the observations. For this purpose, a suitable tool for defining the optimal state number is the Bayesian Information Criterion (BIC). This indicator, proposed by Schwarz [16], estimates the fitting quality through the comparison between the maximized likelihood function and the actual distribution of the data. The BIC increases with the number of model parameters and with the variance error between the fitted and the real distributions, thus the lower the BIC value the better the fitting quality. Tab.1 reports the BIC values calculated for the case of uni-modal Gaussian fitting distribution and Gaussian mixture fitting distribution with two, three and four components. The magnitude order of difference between the uni-modal distribution and the mixture with two components highlights the improvement given by the exploitation of a multi-modal distribution. The best result is given by the model with three components, i.e. a model with three damaging states, representing the healthy stage, the early damaging state and the high failure state.

The availability of a single run-to-failure test led to the need of exploiting the same dataset for both training and validation phases. In this direction, the observation vector has been divided into two different datasets composed by samples alternatively picked from the original observation vector. The selected model is a three state first order left-right model, i.e. a model where the state transition is enabled only for increasing damaging level, due to the monotonically increasing trend of the RMS that describes an irreversible damaging process.

Tab.2 compares the transition probabilities and the prior probabilities obtained under the hypothesis of Gaus-

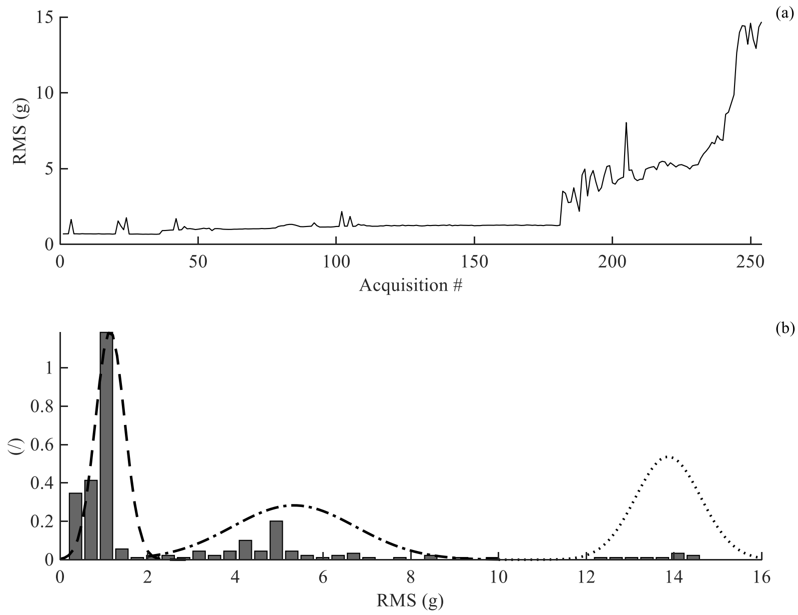


Figure 4: Time history for run-to-failure test: (a) time RMS values from the raw signals, (b) PDF of the RMS with fitted Gaussian distributions.

Table 2: Model parameters: initial state and transition probabilities.

	Gaussian			Generalized Gaussian		
A	0.98	0.02	0	0.99	0.01	0
	0	0.93	0.07	0	0.97	0.03
	0	0	1	0	0	1
Π	1	0	0	1	0	0

Table 3: Estimated parameters of the conditional PDFs.

State	Gaussian		Generalized Gaussian		
#	μ	σ	μ	p	Σ
1	1.09	0.23	1.12	2.01	0.35
2	4.78	1.31	5.06	1.93	2.14
3	12.1	2.45	13.8	2.61	1.25

sian distribution and GGD. It should be noticed that the different distribution does not affect the probabilities being them strictly connected to the trend of observations and not on the data distribution. On the contrary, the distribution parameters (Tab.3) and the estimated PDFs (Fig.5) explains the difference between the exploitation of different basis distributions. Taking into account the first state, the healthy conditions of the system reflects on a Gaussian data distributions. This aspect is confirmed by the distribution parameters (the mean value is the same and the shape factor is around 2 as for the Gaussian distribution) and the distribution form can be considered as the same in both cases. Moving to the early damage stage, i.e. the second state, the difference between the results starts to be considerable due to a slight departure of the data distribution from the Gaussian condition (GGHMM estimate a flatter PDF with an higher mean value). The last state, i.e. end of the working life, demonstrates the aim of the proposed method. The actual data distribution is far from a Gaussian PDF and is more similar to a uniform distribution. As a consequence, the Gaussian based

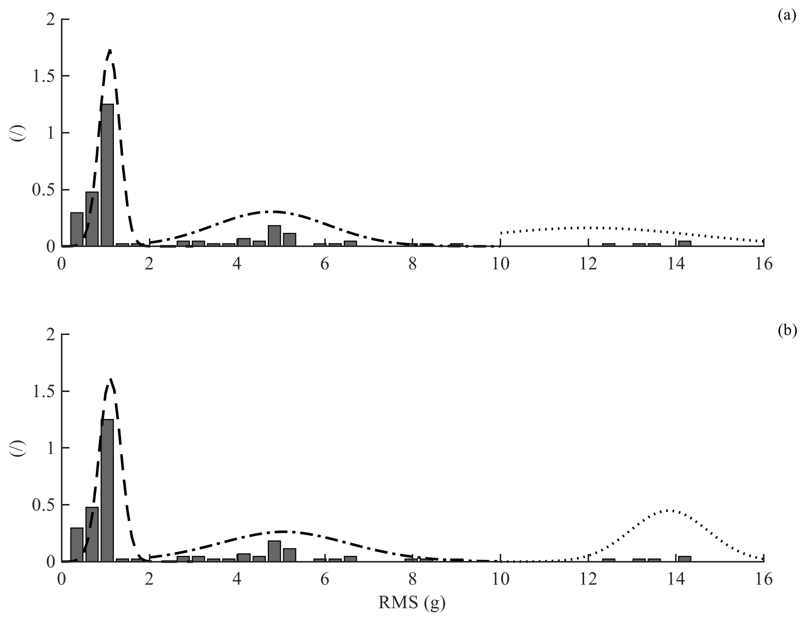


Figure 5: Estimated PDFs after the training process for: (a) Gaussian distribution, (b) generalized Gaussian distribution.

Table 4: Dimensionality and maximized log-likelihood for both estimated mixture PDFs.

Basis distribution	Max Log-likelihood	Dimensionality
Gaussian	-26.78	6
Generalized Gaussian	-22.59	9

analysis is no longer able to identify the correct mean value due to the necessity of exploit a distribution as flatter as possible. On the contrary, the GGHMM is able to estimate the exact mean value of the distribution and the estimated shape factor ($p = 2.61$) describes the departure from the ideal Gaussian form towards to the uniform distribution ($p \to \infty$).

This qualitative comparison can be formalized into a more quantitative analysis by evaluating the fitting quality of the two mixture distributions. For this purpose, the BIC value can not be taken into account due to the different number of free parameters of the Gaussian distribution and the GGD. In order to overcome this limitation, the Likelihood Ratio (LR) test represents a suitable solution for the quantitative evaluation of the fitting quality of two or more distributions. The LR, defined by Giudici et al. [17] as twice the difference between the miximized log-likelihood functions of the two distributions, is exploited as the test statistics and it has been demonstrated that for large number of samples it follows a χ^2 distribution with degrees of freedom corrisponding to the difference between the free parameters of the compared distributions. Starting from the maximized log-likelihood function summarized in Tab.4, the LR is 8.38. Taking into account the 95% quartile of the $\chi^2(3)$ (7.81) the H_1 hypothesis is confirmed and consequently the GGD represents the best fit of the actual RMS distribution.

The trained models have been applied on the aforementioned validation dataset with the aim of estimating the state sequence, final target of each HMMs. Fig.6 reports the state sequences estimated starting from the trained HMMs based on the Gaussian distribution and the GGD. Firstly, comparing Fig.6 and Fig.4(a) it is possible to note how both state sequences correctly reproduce the RMS trend, i.e. correctly describe the fault evolution. On the other hand, the comparison between Fig.6(a) and Fig.6(b) highlights two important differences related to pivotal aspects in the prognostic field. The first one regards the presence of sparse points in the state sequence where the model is not able to detect the actual damaging state. Those points are

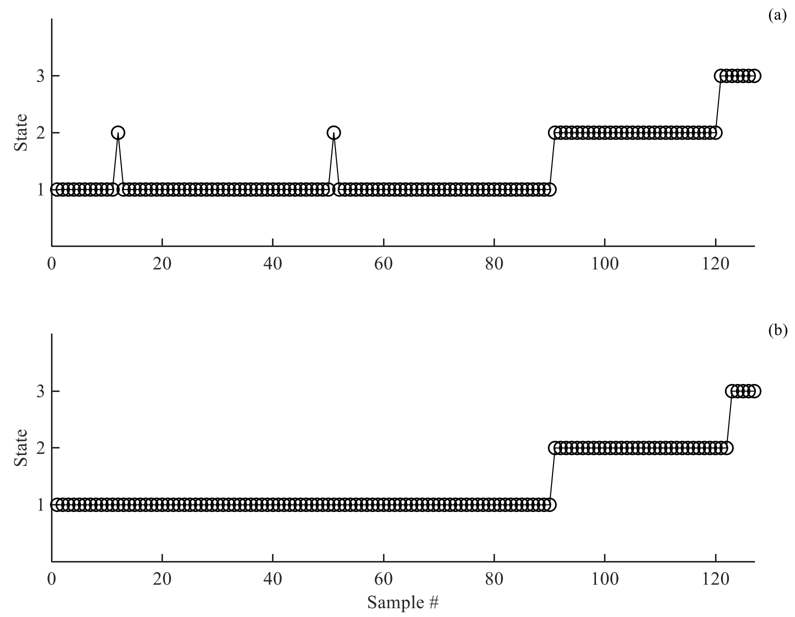


Figure 6: Estimated state sequences on the validation dataset starting from: (a) Gaussian model, (b) generalized Gaussian model.

related to isolated peaks in the RMS trend (see Fig.4(a)) and this common issue is a direct consequence of uncertainties on the physical system, the measuring chain and the signal processing needed for the features extraction. In Fig.6(a) it is possible to note how those RMS peaks reflect on a state transition for the Gaussian based model although the healthy conditions. Regarding this aspect, the GGHMM seems to present ah higher robustness estimating the health state even in presence of measuring uncertainties consequently avoiding unnecessary alarms on possible incipient faults.

The second aspect to be investigated is the state transitions, pivotal for the estimation of the RUL and the planning of the maintenance operations. Fig.6 shows how the transition between state 1 and state 2 is detected by the two models at the same time span. This results is easily explainable taking into account that this transition regards the first appearance of the fault and consequently it mainly depends on the diagnostic capability of the indicator considered as physical observation instead of on the basis model distribution. Moving to the transition between state 2 and state 3, i.e. between early damage level and final failure, the results obtained with the two basis distributions are significantly different. The lower mean value of the estimated Gaussian distribution for the state 3 (Fig.5(a)) reflects on a earlier transition to the last damaging stage that does not describe the actual damaging process described by the RMS trend. On the contrary, the better fitting quality reached by the GGD leads to a correct estimation of the state transitions, pivotal task for diagnostic purposes. The correct estimation of the transition to the final state is fundamental from the predictive maintenance point of view: an early estimation of the state transition reflects on a early estimation of the final failure and consequently leads to the planning of an unnecessary machine downtime due to the maintenance operation.

This experimental application of the proposed method demonstrates the improvement given by the exploitation of a generalized distribution for building an HMM for the prognostics of rotating machines. The main aspects of improvement are represented by the higher robustness of the proposed method with respect to possible outliers in the observation vector and the better fitting quality given by the exploitation of a distribution that allows the modifications of the PDF form within the model states.

5 Final remarks

In the last decades, the HMMs have been widely exploited for prognostic purposes in the rotating machinery field. This paper aims to propose a novel HMM that exploits the GGD as the conditional PDF in order to allows the distribution modifications within the different damaging states. A novel iterative algorithm for the estimation of the model parameters under the hypothesis of GGD starting from the observation from the physical system has been proposed. The improvement given by the proposed method with respect to the classic Gaussian based HMM has been demonstrated through the analysis of a bearing run-to-failure test performed on the bearing test bench at the University of Ferrara. The main effect of the exploitation of a GGD can be seen in terms of fitting quality of the estimated PDF with respect to the real observation distribution. The better estimation, specially in the last damaging stage where the distribution moves away from the Gaussian form, leads to a more accurate identification of the state transition enabling a more effective planning of the maintenance operations. At the same time, the proposed method proves an higher robustness with respect to the possible presence of outliers in the distribution vector related to uncertainties in the physical system, the measuring chain and the data processing.

The proposed method regards the case of monovariate mixture distributions, i.e. with a single observation vector, useful for the analysis of single comonents, e.g. bearings and gears. The extension of the multivariate case, i.e. with several observation vectors for the analysis of complex systems, is now under study and development.

References

[1] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, 2018.

[2] P. Paris and F. Erdogan, "A critical analysis of crack propagation laws," *Journal of Fluids Engineering*, vol. 85, pp. 528–533, 1963.

[3] A. Cubillo, S. Perinpanayagam, and M. Miguez, "A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery," *Advances in Mechanical Engineering*, vol. 8.8, pp. 1–21, 2016.

[4] N. Gebraeel and M. Lawley, "A neural network degradation model for computing and updating residual life distributions," *IEEE Transactions on Automation Science and Engineering*, vol. 5.1, pp. 154–163, 2008.

[5] L. Xiao, X. Chen, X. Zhang, and M. Liu, "A novel approach for bearing remaining useful life estimation under neither failure nor suspension histories condition," *Journal of Intelligent Manufacturing*, vol. 28.8, pp. 1–22, 2015.

[6] J. Kharoufeh and S. Cox, "Stochastic models for degradation-based reliability," *IIE Transactions*, vol. 37.6, pp. 533–542, 2005.

[7] L. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *The Annals of Mathematical Statistics*, vol. 37.6, pp. 1554–1563, 1966.

[8] F. Jelinek, L. Bahl, and R. Mercer, "Design of a linguistic statistical decoder for the recognition of continuous speech," *IEEE Transactions on Information Theory*, vol. 21.3, pp. 250–256, 1975.

[9] C. Bunks, D. McCarthy, and T. Al-Ani, "Condition-based maintenance of machines using hidden Markov models," *Mechanical Systems and Signal Processing*, vol. 14.4, pp. 597–612, 2000.

[10] C. Bishop, *Pattern recognition and machine learning*, Springer, Ed., 2006.

[11] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13.2, pp. 260–269, 1967.

[12] L. Liporace, "Maximum likelihood estimation for multivariate observations of Markov sources," *IEEE Transactions on Information Theory*, vol. 28.5, pp. 729–734, 1982.

[13] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257–286, 1989.

[14] M. Varanasi and B. Aazhang, "Parametric generalized Gaussian density estimation," *Journal of the Acoustical Society of America*, vol. 86.4, pp. 1404–1415, 1989.

[15] R. Krupinski and J. Purczynski, "Approximated fast estimator for the shape parameter of generalized Gaussian distribution," *Signal Processing*, vol. 86.2, pp. 205–211, 2006.

[16] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6.2, pp. 461–464, 1978.

[17] P. Giudici, T. Ryden, and P. Vandekerkhove, "Likelihood-ratio tests for hidden Markov models," *Biometrics*, vol. 56.3, pp. 742–747, 2000.