

Ascertainment-adjusted parameter estimation approach to improve robustness against misspecification of health monitoring methods

Pablo Juesas^a, Emmanuel Ramasso^{*a,b}

^a*Department of Applied Mechanics*

^b*Department of Automatic Control and Micro-Mechatronic Systems
FEMTO-ST institute, UMR CNRS 6174-UBFC/ENSMM/UTBM, 26 chemin de l'Épithaphe,
25000 Besançon, France * Corresponding author: emmanuel.ramasso@femto-st.fr*

Abstract

Condition monitoring aims at ensuring system safety which is a fundamental requirement for industrial applications and that has become an inescapable social demand. This objective is attained by instrumenting the system and developing data analytics methods such as statistical models able to turn data into relevant knowledge. One difficulty is to be able to correctly estimate the parameters of those methods based on time-series data. This paper suggests the use of the Weighted Distribution Theory together with the Expectation-Maximization algorithm to improve parameter estimation in statistical models with latent variables with an application to health monitoring under uncertainty. The improvement of estimates is made possible by incorporating uncertain and possibly noisy prior knowledge on latent variables in a sound manner. The latent variables are exploited to build a degradation model of dynamical system represented as a sequence of discrete states. Examples on Gaussian Mixture Models, Hidden Markov Models (HMM) with discrete and continuous outputs are presented on both simulated data and benchmarks using the turbofan engine datasets. A focus on the application of a discrete HMM to health monitoring under uncertainty allows to emphasize the interest of the proposed approach in presence of different operating conditions and fault modes. It is shown that the proposed model depicts high robustness in presence of noisy and uncertain prior.

Keywords: Time-series, noisy and uncertain labels, weighted distribution theory, health monitoring, CMAPSS datasets

1. Introduction

The statistical representation of multi-dimensional time-series originating from a dynamical system consists in finding a concise and meaningful mathematical model that can be easily interpreted and used to understand the behavior of the system. It is an important problem encountered in a wide range of

real applications such as localization and mapping for mobile robot exploring an unknown environment [44, 43], structural health monitoring under different loading conditions [1, 49, 40], forecasting and prognostics of various systems [53, 47, 19, 24] or human motion analysis [39, 20]. The context of real-world applications generally involves temporal processes subject to uncertainty which is generally managed either by the theory of belief functions, set-membership approaches or probability theory [23, 48, 4, 5].

This paper is focused on the management of uncertainty in statistical models of time-series originating from in-situ monitoring of the health state of industrial equipments. The monitoring is ensured by sensors that continuously record data (observations) based on which the current and future degradation levels of the system has to be inferred for detection (or diagnostics) and prognostics purposes [52].

The literature on data-driven Prognostics and Health Monitoring (PHM) is mainly focused on supervised or unsupervised models [51, 28, 24, 41]: For training, the degradation level (or state) is either known precisely or hidden and, during testing, the current and future levels are inferred from sensor data. In situations where the degradation level can be represented as a sequence of discrete states, statistical models with discrete latent variables have been widely used. Hidden Markov Model (HMM) [34, 45] represents one of those models. HMMs have indeed been widely exploited for PHM [26, 9, 15, 30, 24, 35] where it is assumed that the degradation of the equipment follows a doubly stochastic process: One for the dynamics of the hidden states and one to account for the distribution of the observations.

Generally, the *generative* form of HMM is used: One model is built for each possible degradation level, and, during monitoring, a similarity-based approach is applied to find the likeliest model, that is then used to infer the current state (detection and diagnostics) and the future trends (forecasting and prognostics). We propose an alternative in using HMM for PHM by considering *discriminative* learning where the parameters are estimated with the aim to improve the classification into degradation levels.

Compared to classical generative approaches for HMM-based PHM, and more generally to latent model-based PHM, the proposed approach is based on the idea that sensor data related to different degradation levels have to be put together in a training dataset. Each level is then assigned a *soft label*: It allows one to use uncertain or noisy labels [10] according to the quantity and the quality of the prior knowledge about the degradation level. By this way, some of the latent variables may now represent one degradation level and multiple levels can share common feature subspaces through the use of uncertain labels.

The solution for incorporating soft labels in models with latent variables is based on a modification of the conditional expectation of the log-likelihood in the Expectation-Maximization algorithm by using the Weighted Distribution Theory [33]. An application of this criterion is shown for Gaussian Mixture Models, continuous and discrete HMM. A particular attention is paid to Maximum Likelihood Estimation (MLE) of parameters in HMM with discrete-valued observations (DHMM) in presence of soft labels.

Standard DHMMs have been used in the past in many applications [31, 57, 42, 8, 3, 26] and in particular for noisy speech and character recognition [7, 50, 17, 12]. In the context of PHM, DHMM has also been widely used, for instance in [25] for predictive modeling dedicated to intelligent maintenance in complex semiconductor manufacturing processes, in [29] for incipient fault detection and diagnosis in turbine engines, in [2] for failure isolation for cognitive robots and in [16] for anomaly detection in electronic systems.

While the inference phase is very similar between HMM with continuous (CHMM) and discrete observations (DHMM), the learning phase presents a fundamental difference since the observations in the latter are discrete. The learning phase for huge datasets is thus generally faster for the DHMM by using matrix encoding. Moreover, it has been shown that the DHMM may be better in presence of noise with unknown (non Gaussian) characteristics [50].

The contribution of this paper is two-fold:

- We propose a framework for PHM based on latent variables. It is based on the Expectation-Maximization algorithm and the Weighted Distribution Theory. This work is inspired from a previous work proposed by T. Denoeux [14] based on Dempster-Shafer’s theory of belief functions and plausibility weights. The main difference with the present work is the consideration of almost unrestricted weights (satisfying only positiveness) based on the work of Patil [33].
- This framework is used to improve the performance of DHMM for PHM. An application to turbofan health monitoring is presented where we evaluate the sensitivity of DHMM to vector quantization with respect to the quality and quantity of prior.

The problem of incorporating partial knowledge about latent variables is discussed in Section 2. Section 3 presents the application of this method to DHMM. Section 4 is dedicated to the analysis of the proposed model on several datasets.

2. Incorporating prior knowledge on latent variables in EM

The Expectation-Maximization algorithm (EM) [13] allows to estimate the parameters of a statistical model with latent variables. It has been adapted in [14] in order to take *uncertain* prior information into account. The prior is supposed to be represented by a collection of T Basic Belief Assignments (BBA) denoted as $\mathbf{m} = \{m_1^{\Omega_{\mathbf{y}}}, \dots, m_T^{\Omega_{\mathbf{y}}}\}$ defined on the set of discrete states $\Omega_{\mathbf{y}}$ with $\sum_{A \subseteq \Omega_{\mathbf{y}}} m_t^{\Omega_{\mathbf{y}}}(A) = 1$, and such that $m_t^{\Omega_{\mathbf{y}}}(\emptyset) = 0$ (normalized BBA). The adaptation of EM concerns the auxiliary function (conditional expectation of the log-likelihood) to be maximized at each iteration q :

$$\begin{aligned}
 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(q)}) &= \mathbb{E}_{\boldsymbol{\theta}^{(q)}}[\log L(\boldsymbol{\theta}; \mathbf{z}) | \mathbf{x}, \mathbf{m}] \\
 &= \frac{\sum_{\mathbf{y}} p_l(\mathbf{y}) p(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}^{(q)}) \log L(\boldsymbol{\theta}; \mathbf{z})}{L(\boldsymbol{\theta}; \mathbf{p}l)} \quad (1)
 \end{aligned}$$

where \mathbf{pl} represents the prior knowledge on latent variables \mathbf{y} encoded by contour functions in place of BBA \mathbf{m} and computed as: $pl : \Omega_{\mathbf{y}} \rightarrow [0, 1]$ such that $pl(\omega) = Pl(\{\omega\})$ with $Pl(C) = \sum_{D \cap C \neq \emptyset} m(D)$, $C \subseteq \Omega_{\mathbf{y}}$. The quantity $L(\boldsymbol{\theta}; \mathbf{pl})$ represents the amount of conflict between both the model (parameterized by $\boldsymbol{\theta}$ and evaluated on the observed data \mathbf{x}) and the prior (represented by \mathbf{pl}).

As it can be observed, the uncertain prior play a role of weight on the posterior distribution in the auxiliary function. This can be interpreted as an adjustment of the distribution over hidden states due to, for example, model misspecification or to improve data interpretation.

Proposition 1. *Eq. 1 is valid for any positive weights \mathbf{w} in place of plausibilities:*

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(q)}) &= \mathbb{E}_{\boldsymbol{\theta}^{(q)}}[\log L(\boldsymbol{\theta}; \mathbf{z}) | \mathbf{x}, \mathbf{w}] \\ &= \frac{\sum_{\mathbf{y}} w(\mathbf{y}) p(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}^{(q)}) \log L(\boldsymbol{\theta}; \mathbf{z})}{\mathbb{E}_{\boldsymbol{\theta}}[w(\mathbf{z})]} \end{aligned} \quad (2)$$

The proof is based on the Weighted Distribution Theory (WDT) which is a general framework for adjusting probabilities. Patil [33] illustrated the WDT with practical examples and the reader may also be interested by the paper of Iyengar and Zhao [21] for a use in maximum likelihood, and by the paper of Nielsen and Signorovitch [32] with an application to single nucleotide polymorphisms discovery.

More specifically, the random variable \mathbf{Z} is supposed to be governed by a probability density function $f(\mathbf{z}; \boldsymbol{\theta})$ (for instance the posterior in Eq. 2). The weight function $w(\mathbf{z})$ allows one to weigh the pdf f when $\mathbf{Z} = \mathbf{z}$ [33]:

$$\begin{aligned} f_w(\mathbf{z}; \boldsymbol{\theta}) &= \frac{f(\mathbf{z}; \boldsymbol{\theta}) w(\mathbf{z})}{\mathbb{E}_{\boldsymbol{\theta}}[w(\mathbf{Z})]} \\ \text{s.c. } w(\mathbf{z}) &\geq 0, 0 < \mathbb{E}_{\boldsymbol{\theta}}[w(\mathbf{Z})] < \infty \end{aligned} \quad (3)$$

The weights are not considered as probabilities and only have to be positive [33]. It is interesting to make a parallel between the denominators of Eq. 1 and 2. Indeed, in the former case, it has been shown that [14]:

$$L(\boldsymbol{\theta}; \mathbf{pl}) = \mathbb{E}_{\boldsymbol{\theta}}[\mathbf{pl}(\mathbf{Z})] \quad (4)$$

which is similar to the denominator of Eq. 3. Taking the conditional expectation of the complete-data log-likelihood given partial knowledge on states encoded as weights $\mathbf{w} = \{w_1 \dots w_T\}$ thus leads to a similar expression as Eq. 2.

3. Learning with soft labels in DHMM

This section shows how to use the proposed model to learn parameters in a Discrete HMM (DHMM). For health monitoring, it follows the plot chart depicted in Figure 1.

[Figure 1 about here.]

3.1. Model and notations

In a DHMM, the observed data take the form of a sequence of discrete symbols $\mathbf{x} = \{x_1 \dots x_T\}$, $x_t \in \{1 \dots V\}$ with V the number of symbols, and where $t = 1 \dots T$ is a discrete time index. The model assumes that the sequence has been generated by a stochastic (first-order) Markov process with hidden states $\mathbf{y} = \{y_1, \dots, y_T\}$, with $y_t \in \{1 \dots K\}$, K is the number of hidden states, and $\Omega_{\mathbf{y}} = \{\omega_1 \dots \omega_K\}$ represents the set of states. The distribution of the output observation x_t at time t depends only on y_t . The complete data is denoted as $\mathbf{z} = (\mathbf{x}, \mathbf{y})$.

A DHMM is described by the following parameters [34]:

- Prior probabilities $\mathbf{\Pi} = \{\pi_1, \dots, \pi_k, \dots, \pi_K\}$, where $\pi_k = P(y_1 = k)$ is the probability of being in state k at $t = 1$;
- Transition probabilities $\mathbf{A} = [a_{kl}]$, where

$$a_{k,l} = P(y_t = l | y_{t-1} = k), \quad (k, l) \in \{1, \dots, K\}^2 \quad (5)$$

is the probability for being in state l at time t given that it was in state k at $t - 1$ with $\sum_l a_{kl} = 1$;

- Observation symbol probabilities $\mathbf{B} = [b_{kv}]$ (emission matrix) where

$$b_{k,v} = P(x_t = v | y_t = k), \quad k \in \{1, \dots, K\}, v \in \{1, \dots, V\} \quad (6)$$

is the probability for being in state k at time t and observing symbol v with $\sum_v b_{k,v} = 1$

The set of parameters in a DHMM is denoted as:

$$\theta = (\mathbf{A}, \mathbf{B}, \mathbf{\Pi}) \quad (7)$$

3.2. Learning procedure based on soft labels

The weighted conditional expectation of the log-likelihood given partial knowledge on states for DHMM (Eq. 2) requires to express the likelihood over hidden and observed variables which, in the DHMM, is given by [34]:

$$L(\theta; \mathbf{z}) = p(y_1; \mathbf{\Pi}) \left(\prod_{t=2}^T p(y_t | y_{t-1}; \mathbf{A}) \right) \prod_{t=1}^T p(x_t | y_t; \mathbf{B}) \quad (8a)$$

$$= \left(\prod_{k=1}^K \pi_k^{y_{1,k}} \right) \left(\prod_{t=2}^T \prod_{k,l} a_{k,l}^{y_{t-1,k} y_{t,l}} \right) \left(\prod_{t=1}^T \prod_{k=1}^K \prod_{v=1}^V b_{k,v}^{y_{t,k}} \right) \quad (8b)$$

This form makes use of the fact that the latent variables are multinomial discrete variables [6], therefore $y_{tk} = 1$ if state k is true at time t . The conditional

expectation of the log-likelihood given partial knowledge on states obtained at iteration q becomes:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(q)}) = \sum_{k=1}^K \gamma_{1,k}^{(q)} \log \pi_k + \sum_{t=2}^T \sum_{k,l} \xi_{t-1,k;t,l}^{(q)} \log a_{k,l} + \sum_{t=1}^T \sum_{k=1}^K \sum_{v=1}^V \gamma_{t,k}^{(q)} \log b_{k,v} \quad (9)$$

where $\gamma_{t,k} = \mathbb{E}_{\boldsymbol{\theta}^{(q)}}[y_{t,k} | \mathbf{x}, \mathbf{w}]$ and $\xi_{t-1,k;t,l} = \mathbb{E}_{\boldsymbol{\theta}^{(q)}}[y_{t-1,k} y_{t,l} | \mathbf{x}, \mathbf{w}]$ represent the mathematical expectations with respect to the posterior distribution over hidden states given by:

$$\xi_{t-1,t;k,l}^{(q)} = \frac{\alpha_{t-1,k}^{(q)} b_{l,v} w_{t,l} a_{k,l}^{(q)} \beta_{t,l}^{(q)}}{L(\boldsymbol{\theta}^{(q)}; \mathbf{w})} \quad (10)$$

$$\gamma_{t,k}^{(q)} = \frac{\alpha_{t,k}^{(q)} \beta_{t,k}^{(q)}}{L(\boldsymbol{\theta}^{(q)}; \mathbf{w})} \quad (11)$$

and

$$L(\boldsymbol{\theta}^{(q)}; \mathbf{w}) = \sum_{k=1}^K \alpha_{T,k} \quad (12)$$

The variables $\alpha_{t,k}^{(q)}$ and $\beta_{t,k}^{(q)}$ can be recursively computed as

$$\alpha_{1k}^{(q)} = \pi_k^{(q)} w_{1,k} b_{k,v} \quad (13a)$$

$$\alpha_{tk}^{(q)} = b_{k,v} w_{t,k} \sum_j \alpha_{t-1,j}^{(q)} a_{j,k}^{(q)} \quad (13b)$$

$$\beta_{Tk}^{(q)} = 1 \quad (13c)$$

$$\beta_{tk}^{(q)} = \sum_l \beta_{t+1,l}^{(q)} b_{l,v} w_{t+1,l} a_{k,l}^{(q)} \quad (13d)$$

for $t = 2 \dots T$. We can observe that the soft labels w do not change the conditional expectation compared to [38] except for the emission probabilities (third term of Eq. 9). Therefore, the reestimation of parameters \mathbf{A} and $\mathbf{\Pi}$ are given by:

$$\pi_k^{(q+1)} = \gamma_{1,k}^{(q)} \quad (14)$$

$$a_{kl}^{(q+1)} = \frac{\sum_{t=2}^T \xi_{t-1,k;t,l}^{(q)}}{\sum_{t=2}^T \sum_{l'=1}^K \xi_{t-1,k;t,l'}^{(q)}} \quad (15)$$

and the MLE of the emission probability model \mathbf{B} is given by applying a similar reasoning as in the unsupervised learning of DHMM [34] adapted as proposed in [38] to take into account the uncertain prior on hidden states leading to

$$b_{k,v}^{(q+1)} = \frac{\sum_{t=1}^T \gamma_{t,k}^{(q)} \mathbb{1}\{x_t = v\}}{\sum_{t=1}^T \gamma_{t,k}^{(q)}} \quad (16)$$

The expression of γ is given by Eq. 11 that makes use of the partial knowledge about hidden states in discrete HMM.

4. Applications

A first simulated dataset is used to illustrate the behavior of the proposed model (Eq. 2) for Gaussian Mixture Models, CPHMM and DPHMM with respect to different weights. Those three models are available as Matlab codes which allow to reproducing all results of this paper:

<http://www.mathworks.com/matlabcentral/profile/authors/7468430>

The second simulated dataset allows us to study the effect of the soft labels on the both the performance and on the quantization of continuous variables. Finally, complex benchmarking datasets provided by the NASA PCOE [46] are used to illustrate the performance of the proposed model for diagnostics.

In the three applications, the sequences of states estimated (for instance, in order to represent a degradation) with the proposed approach (WDT applied on various models such as DHMM) are compared to a ground truth using the Adjusted Rand Index (ARI) [55]. This performance indicator, widely used for unsupervised classification, allows to evaluating the impact of the quality and quantity of labels. It tends to 1 (resp. 0) if the recognition rate is perfect (resp. bad).

4.1. Toy example dedicated to the application of the methodology on various models from the literature with comparison

The interest of using arbitrary weights in Eq. 2 is illustrated on three different models: partially-Hidden Markov Models with continuous observations (CPHMM) [38], the DHMM presented above and the Gaussian Mixture Model (PGMM) proposed in [10].

A 2-dimensional dataset was generated using the following model made of $K = 3$ clusters:

$$\mathbf{\Pi} = (1/3, 1/3, 1/3)', \mathbf{A} = \begin{pmatrix} 0.6 & 0.1 & 0.3 \\ 0.1 & 0.6 & 0.3 \\ 0.15 & 0.15 & 0.7 \end{pmatrix} \quad (17)$$

$$\boldsymbol{\mu}_1 = (4 \ 0), \boldsymbol{\mu}_2 = (0 \ 4), \boldsymbol{\mu}_3 = (2 \ 2) \quad (18)$$

$$\boldsymbol{\Sigma}_1 = \text{diag}(7 \ 7), \boldsymbol{\Sigma}_2 = \text{diag}(7 \ 7), \boldsymbol{\Sigma}_3 = \text{diag}(2 \ 2) \quad (19)$$

A set of N=300 samples is drawn from this CPHMM and illustrated in Figure 2.

[Figure 2 about here.]

The particularity of this dataset is to be composed of two distinct groups (clusters) plus a third one located on purpose at the middle of both previous ones in order to simulate a fuzzy frontier. The centers of the three clusters are estimated by using the ECM algorithm [27] with K=3 (and with the default parameters proposed in [27]). The interest of using this algorithm here is to get the credal partition, denoted as $m_i^\Omega, \Omega = \{c_1, c_2, c_3\}, i = 1 \dots N$, that gives the membership degree (a BBA) of each data point to each cluster and to all subsets of clusters. The credal partition is assumed to represent a prior knowledge that can be used in the three models.

To be used with Eq. 2, we need to transform the credal partition into a set of weights assigned to singleton clusters. For that, we proceed as proposed in [27] by transforming the partition into three types of weights:

BBA are belief masses on singleton clusters where weights are equal to $w_i(c_k) \equiv m_i^\Omega(c_k)$;

PL are plausibilities on singleton clusters where weights are equal to $w_i(c_k) \equiv pl_i^\Omega(c_k)$;

BETP are pignistic probabilities on singleton clusters [48] where weights are equal to $w_i(c_k) \equiv BetP_i^\Omega(c_k)$.

ECM was run 5 times with random initializations of centers, and the model (position of centers) leading to the highest ARI.

The ARI is used to compare the partition obtained by maximizing the *BBA on singletons* with the real partition obtained by sampling the previous model (the correct cluster is known). By doing so, we explicitly make favourable the choice of BBA as weights. The performance is thus expected to be higher all the more than the plausibility and the pignistic transformations are computed by transferring the belief masses on some subsets towards singletons: Conversely to BBA, the frontier artificially added through the third cluster is expected to collapse with those two transformations.

Learning in CPHMM was performed as in [38] except that the weights took the three forms presented above (BBA, pl and BetP). The states were estimated by maximizing the posterior distribution on states (γ -variable) for both the DPHMM (DHMM with soft labels) and the CPHMM. For the DPHMM, quantization was performed by a Kmeans by minimizing the sum of squared-distance between points and clusters [34].

[Figure 3 about here.]

Figure 3 shows the evolution of the boxplot of ARI over 50 sets of 1000 samples for the CPHMM and for the PGMM learned with plausibility as initially required in E2M (Evidential EM) algorithm [14] using Eq. 1 (E2M+pl), with

pignistic probabilities (E2M+betp) and with the belief masses (E2M+BBA). The performance of ECM is also shown in order to evaluate the performance obtained with the introduction of transitions (which are explicitly used for sampling). The results with noisy labels is also depicted: Noisy labels are obtained by assigning a binary value (0 or 1) for the maximum of BBA, pl or BetP over each cluster and for each data point (for instance $m_i^\Omega(c_k) = 1$ if $k = \operatorname{argmax}_j m_i^\Omega(c_j)$, 0 otherwise).

Results are globally quite low for all models. This is due to the fact that one of the clusters is localized at the very middle of the other clusters with high overlapping. The comparison between the three models has thus to be performed relatively. The absolute performance is indeed not the main purpose of this experiment which is to show that using arbitrary weights can be useful. A deep analysis of the performances is proposed in next sections.

Figure 3 shows that the performance is better when using the belief masses by a factor of almost 2 on average, compared to pignistic probabilities and plausibilities. This is an important result which goes in favor of considering the Weighted Distribution Theory to incorporate the most relevant information in the learning phase when using E2M.

Furthermore, the DPHMM behaves similarly with noisy or uncertain labels on this particular dataset whereas the CPHMM and PGMM with noisy labels provides lower performance compared to uncertain labels. This result was also underlined in [38] for CPHMM. The CPHMM provides the best overall result ($37\% \pm 2\%$) when used with the BBA.

4.2. Impact of soft labels on quantization in DPHMM

The goal of the following experiments is to study the effect of soft labels on both the performance and the quantization in the newly proposed DPHMM. For that, we consider the same continuous-valued data as in [38] generated with a continuous HMM with 3 states in 3 dimensions.

The Kmeans algorithm was applied for vector quantization [34] in order to obtain discrete symbols. To cope with the sensitivity of Kmeans with respect to the initialization, 10 runs were performed and the best partition was selected according to the Davies-Bouldin index [11].

Two different experiments were carried out to study the influence of “label imprecision” and “labeling error” on the performance of the DPHMM. In each case, different numbers of clusters (N) were tested to estimate the impact of the quantization on the performance.

4.2.1. Influence of label imprecision on the performance and on the quantization

We proceed as proposed in [38] to study how the imprecision of knowledge on hidden states affects the performance of the DPHMM. A learning sequence of length T was generated using the model presented in [38]. Imprecision on labels was generated as $pl_{tk} = 1$ if $y_t = k$, or ν otherwise, where ν represents the nonspecificity coefficient ($\nu = 1$ corresponds to a classical DPHMM model with no information available on hidden states, $\nu = 0$ represents a fully supervised learning situation).

To assess the quality of learning, a testing dataset of 1000 observations was generated following the same distribution. The most probable state sequence was computed using the maximum of the γ variable (Eq. 11). It was observed that the Viterbi algorithm may lead to bad results when the transition matrix (Eq. 15) or the emission matrix (Eq. 16) are ill-conditioned (presence of 0). This instability was not observed important in the continuous version of this algorithm [38].

The precision of the predicted state sequences was assessed using the ARI. The whole experiment (data generation, clustering and learning) was repeated 30 times, for different number of clusters and for different length of the learning sequence $T = 100$ and $T = 300$.

Results are shown in Figure 4. We can observe that results degrade from the fully supervised ($\nu = 0$) to the fully unsupervised ($\nu = 1$) case for most of tests. As T increases, the estimation is improved in terms of better ARI and lower variance on results. It allows to decrease the impact of both the quantization phase (ie of N) and of the corruption process (ie of ν).

In addition to the computational problem observed in the Viterbi algorithm mentioned above, we observed that the number of iterations in the learning phase should be low (5 – 10). Otherwise, the convergence leads to very specific emission matrices (Eq. 16) which ultimately lead to overfitting. This problem is mainly due to the size of the training sequence which is probably too small. It can be remarked that this problem was not encountered in the continuous version [38].

[Figure 4 about here.]

4.3. Influence of labeling error on the performance and on the quantization

In the previous experiment information on hidden states was assumed to be always exact, i.e. the true state had always the largest plausibility value. In this part, we proceed as proposed in [10, 38] where at each time step t of a sequence, an error probability q_t was drawn randomly from a beta distribution with mean ρ and standard deviation 0.2. With probability q_t , the state y_t was then replaced by a completely random value \tilde{y}_t (with a uniform distribution over possible states). The plausibilities pl_{tk} were determined as $pl_{tk} = P(y_t = k | \tilde{y}_t) = q_t/K + 1 - q_t$ if $\tilde{y}_t = k$, or q_t/K otherwise.

Uncertain labels generated in this way are more imprecise when the error probability is high. It simulates a more realistic and complex situation compared to the previous section.

Training and test data sets were generated as in the previous section, and results were evaluated in the same way. For each randomly generated data set, the DPHMM was applied with uncertain labels pl_{tk} and noisy labels \tilde{y}_{tk} .

Figures 5 and 6 show the ARI as a function of mean error ρ for uncertain and noisy labels respectively, for different number of clusters and sequence length. As previously, a degradation of the segmentation quality is observed when the mean error probability ρ increases, for both uncertain and noisy labels. Moreover, increasing T from 100 to 300 does improve the quality. It can also be

observed that results are better with noisy labels than with uncertain ones. The explanation is that the observation matrix is better filled with noisy but certain information than with uncertain (fuzzy), the latter tends to smooth the values of the matrix. This is in contrast with results obtained with continuous HMM reported in [38] in which the distributions are better fitted with noisy-fuzzy information than with noisy-certain ones. This is also in agreement with previous experiments.

[Figure 5 about here.]

[Figure 6 about here.]

4.4. Application of DPHMM to turbofan engine diagnostics

4.4.1. Description of the data

The turbofan datasets were generated using the CMAPSS simulation environment that represents an engine model of the 90,000 lb thrust class [18, 46]. A number of editable input parameters was used to specify operational profile, closed-loop controllers, environmental conditions (various altitudes and temperatures). Some efficiency parameters were modified to simulate various degradations in different sections of the engine system. Selected fault injection parameters were varied to simulate continuous degradation trends. Data from various parts of the system were collected to record effects of degradations on 21 sensor measurements and provide time-series exhibiting degradation behaviors in multiple units.

These datasets possess unique characteristics that make them very useful and suitable for developing prognostic and health monitoring algorithms [41]: Multi-dimensional response from a complex non-linear system, high levels of noise, effects of faults and operational conditions, and plenty of units were simulated with high variability.

In the present paper, the learning datasets of the four turbofan datasets were used to illustrate the DPHMM on a complex system. The characteristics of the datasets are described in Table 1. It can be observed that dataset #1 is the simplest one with one operating condition (OC) and one fault mode. Datasets #2 and #4 are the most complex datasets with six OC and one or two fault modes. Dataset #3 presents two fault modes and one OC.

The state-of-the-art results on those datasets are presented in [41]. According to this review, only the publication [38] provided results using soft labels for health monitoring purposes. However, it makes use of particular features for dataset #1 only. In the present work, we apply a more general methodology where a health indicator is extracted as presented in [36] for each training data in each dataset. Then, soft labels are introduced (described below) and the influence of noise and uncertainty on those labels are analyzed.

[Table 1 about here.]

Each dataset is made of a certain number of trajectories with different length and 21 sensor measurements, with a total amount of about 700 training degradation trajectories (sum of values in Table 1). From sensor measurements, a health indicator is built for each trajectory as proposed in [36]. The health indicators (HI) for all trajectories and all datasets are depicted in Figure 7. We can observe that datasets #2 and #4 are made of about 500 trajectories with high variability in terms of noise and length as compared to dataset #1. These four sets of trajectories are used for both training and testing data to illustrate the influence of labeling errors in the DPHMM.

[Figure 7 about here.]

4.4.2. Generation of the ground truth

Those datasets have been used for clustering or classification purposes in many papers as reported in [41]. In order to get a ground truth to both feed the DPHMM and make comparison, a way to automatically generate three states from each trajectory is described in [37] with Matlab codes.

The generation of the state sequences corresponds to a labeling which may be corrupted by errors due, for example, to the noise on the HI (impacted by OC and fault modes), and to the threshold (10% on the absolute error). To evaluate the influence of labeling errors, we proceed as in Section 4.3 (by drawing randomly from a beta distribution and replacing the label by a random value) and considering noisy labels. It was indeed observed that the DPHMM performed badly in presence of uncertain labels for this particular application but better with noisy labels.

The influence of labeling errors was studied for the four datasets using a leave-one-out cross-validation over all training instances for each CMAPSS dataset. Since the HI are continuous, a quantization was performed to generate symbols used in the DPHMM by simply converting the floats into integers as follows:

$$HI^{\text{discrete}}(t) \leftarrow \lfloor HI^{\text{continuous}}(t) \times N \rfloor \quad (20)$$

where N is a parameter playing a similar role as the number of clusters in usual quantization methods [34]. The influence of N is studied in the next tests.

Figure 8, 9, 10 and 11 represent boxplots of the ARI with respect to both the labeling error rate and to the value of N for the four datasets. For dataset #1 (Fig. 8), the best performance is obtained for the supervised case with $N = 5$ yielding 88%, then it decreases monotonically with respect to the labeling error but always remains above the unsupervised case ($\rho = 1$) with 66%. This performance is similar to the state-of-the-art results using soft labels presented in [38]. For dataset #2 (Fig. 9), the performance equals 83% for the supervised case and 63% for the unsupervised case. The performances decreases monotonically with an increasing noise level except for $N = 15$. For dataset #3 (Fig. 10), the best performance is obtained for the supervised case with $N = 5$ yielding 82%. For dataset #4 (Fig. 11), a similar behavior to dataset #2 can be observed. The performance for the supervised case is similar to dataset #2 (84.5%) despite

one additional fault mode. The results for this last dataset shows the impact of the combination of fault modes and operating conditions on the performance.

[Figure 8 about here.]

[Figure 9 about here.]

[Figure 10 about here.]

[Figure 11 about here.]

$N = 5$ yields the best performance for all datasets (over about 700 degradation trajectories). The DPHMM with partial noisy knowledge depicts robustness to the labeling error on the four datasets. Moreover, the performance in the supervised case is generally expectedly better than the unsupervised case. The performance then generally decreases as the level of labeling error increases.

Large values of N leads to a particular results where the performance is almost independent of the noise level. It can be pointed out that the ARI was created for corrected-for-chance evaluation which means that random labelling of states should tend to 0. Therefore, the performance obtained with large N is not due to randomness. Moreover, the performance is better than with $N = 3$. An explanation is given hereafter, remembering that large N implies a high number of symbols. In addition to the noise level on the HI together with the labeling error rate, the estimation of the emission probability model (Eq. 16) is impacted but the performance still remains satisfying (above $N = 3$) since there are always some symbols which allows to represent the sequence. It can also be pointed out that, in presence of more symbols, the model requires more data to estimate the emission matrix. This behavior is confirmed by the interquartile ranges which are important for $N = 15$ for datasets #1 and #3 which are made of 100 trajectories (Tab. 1) whereas the interquartile ranges are lower for datasets #2 and #4 made of about 260 trajectories.

5. Conclusion

The Weighted Distribution Theory (WDT) has been initially proposed to cope with the fact that the probability of a particular observation entering in a data sample gets multiplied by some non-negative weight function w . The data are thus *naturally* distributed according to f_w (Eq. 3) instead of f . Since model specification is of paramount interest in data analysis, such bias should be taken into account.

This paper suggests a means to incorporate noisy and uncertain prior knowledge on latent variables in a sound manner using the Expectation-Maximization algorithm together with the WDT. An application to discrete Hidden Markov Models (DHMM) is developed and illustrated for system's health monitoring. Illustrations are also provided for Gaussian Mixture Models and HMM with continuous outputs.

The interest of such an approach is to cope with model selection for classification with maximum likelihood learning. Instead of considering one model per class, all classes are considered simultaneously and the definition of the classes may be imperfect (noise on labels, uncertainty). The learning phase based on WDT and EM then finds a compromise between the data and the prior.

The first experiment allows to demonstrate the relevance of considering possibly arbitrary weights compared to the algorithm proposed in [14]. The second and third experiments show the DPHMM model behaved better when considering noisy labels than uncertain labels compared to continuous HMM, which seems in agreement with the model. It is also demonstrated that the quantization (transformation of continuous to discrete values) in the DPHMM applied to continuous data does *not* have so much impact on classification results when selected appropriately, using for instance validity index as proposed in this paper. It is practically interesting since DPHMM are less complex.

The third experiment was performed on a complex benchmark representing health indicators originating from a turbofan engine simulator developed at NASA. These tests shown that the proposed model depicted robustness to labelling errors when the level of noise on the initial data was reasonable. A labeling error rate up to 50-60% yields good results. Moreover, the same parameterization of the quantization phase applied to all datasets provided the best results, demonstrating that the model is reasonably robust to quantization on different datasets.

The integration of prior in a sound manner for the modelling of real systems is of key interest, in both data-driven and physics-based approaches. Indeed, the amount of tests required to correctly pave a complex feature space and deduce particular behavior should be sufficiently large to expect a correct identification of parameters. This is for instance the case for identification of failure modes in complex mechanical systems. The integration of prior is all the more important in presence of non stationary systems submitted to external factors (generally hidden and unknown) which creates intermittent faults characterized by sudden changes in system parameters. Making use of prior appears as a means to reduce the number of experiments and include information about external factors with respect to time.

The Bayesian approach is an inescapable way to integrate uncertain prior knowledge on parameters [6]. The main advantage is the possibility to penalize models with too many parameters with the aim to reduce overfitting which can appear in maximum likelihood approaches. As a consequence, this allows to improving model selection which is particularly important for health monitoring [54, 56]. The influence of the prior on results is however not often deeply studied as proposed in this paper.

While the Bayesian approach makes generally use of prior on parameters, the proposed WDT-based approach is focused on the integration of prior on discrete latent variables which appears as a different means to adjust indirectly the distributions on parameters. The proposed approach could be used in Bayesian formulations to give a hand in improving model specifications by driving the estimation using both prior on parameters and on latent variables. We aim at

exploring this path in addition to the application of the WDT approaches to other Markov switching and statistical models with latent variables especially dedicated to nonlinear and non-stationary systems monitoring and prognostics [22].

Acknowledgement

This work has been carried out in the framework of the Laboratory of Excellence ACTION through the program “Investments for the future” managed by the National Agency for Research (references ANR-11-LABX-01-01). The authors are grateful to the Région Franche-Comté and “Bpifrance financement” supporting the SMART COMPOSITES Project in the framework of FRI2. The authors would like to thank the reviewers who provided valuable comments.

References

- [1] Abbiati, G., Ceravolo, R., Surace, C., 2015. Time-dependent estimators for on-line monitoring of full-scale structures under ambient excitation. *Mechanical Systems and Signal Processing* 6061, 166 – 181.
- [2] Altan, D., Sariel-Talay, S., 6-8 March 2014. Hierarchical HMM-based failure isolation for cognitive robots. In: *International Conference on Agents and Artificial Intelligence. ESEO, Angers, Loire Valley, France*, pp. 299–304.
- [3] Baldi, P., Chauvin, Y., Hunkapiller, T., McClure, M., 1994. Hidden Markov models of biological primary sequence information. *Proceedings of the National Academy of Sciences* 91 (3), 1059–1063.
- [4] Beer, M., Ferson, S., 2013. Special issue of mechanical systems and signal processing imprecise probabilities what can they add to engineering analyses?. *Mechanical Systems and Signal Processing* 37 (1-2), 1 – 3.
- [5] Beer, M., Patelli, E., 2015. Editorial: Engineering analysis with vague and imprecise information. *Structural Safety* 52, Part B, 143 –, engineering Analyses with Vague and Imprecise Information.
- [6] Bishop, C., 2006. *Pattern recognition and machine learning*. Springer.
- [7] Bourlard, H., Morgan, N., 1990. A continuous speech recognition system embedding MLP into HMM. In: *Touretzky, D. S. (Ed.), Advances in Neural Information Processing Systems*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 186–193.
- [8] Chen, H.-S., Tsai, W.-J., 2014. A framework for video event classification by modeling temporal context of multimodal features using HMM. *Journal of Visual Communication and Image Representation* 25 (2), 285–295.

- [9] Chinnam, R., Baruah, P., 2003. Autonomous diagnostics and prognostics through competitive learning driven HMM-based clustering. In: International Joint Conference on Neural Networks. Vol. 4. Portland, OR, USA, pp. 2466–2471.
- [10] Côme, E., Oukhellou, L., Denoeux, T., Aknin, P., 2009. Learning from partially supervised data using mixture models and belief functions. *Pattern recognition* 42 (3), 334–348.
- [11] Davies, D., Bouldin, D., 1979. A cluster separation measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-1* (2), 224–227.
- [12] Dehghan, M., Faez, K., Ahmadi, M., Shridhar, M., 2001. Handwritten farsi (arabic) word recognition: a holistic approach using discrete HMM. *Pattern Recognition* 34 (5), 1057–1065.
- [13] Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* 39 (1), 1–38.
- [14] Denoeux, T., 2013. Maximum likelihood estimation from uncertain data in the belief function framework. *Knowledge and Data Engineering, IEEE Transactions on* 25 (1), 119–130.
- [15] Dong, M., He, D., 2007. A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology. *Mechanical Systems and Signal Processing* 21 (5), 2248 – 2266.
- [16] Dorj, E., Chen, C., Pecht, M., 2013. A Bayesian Hidden Markov Model-based approach for anomaly detection in electronic systems. *IEEE Aerospace Conference Proceedings*, 1–10.
- [17] Elms, A., Procter, S., Illingworth, J., 1998. The advantage of using an HMM-based approach for faxed word recognition. *International Journal on Document Analysis and Recognition* 1 (1), 18–36.
- [18] Frederick, D., DeCastro, J., Litt, J., 2007. User’s guide for the commercial modular aero-propulsion system simulation (C-MAPSS). Tech. rep., National Aeronautics and Space Administration (NASA), Glenn Research Center, Cleveland, Ohio 44135, USA.
- [19] Froelich, W., Salmeron, J., 2014. Evolutionary learning of fuzzy grey cognitive maps for the forecasting of multivariate, interval-valued time series. *International Journal of Approximate Reasoning* 55 (6), 1319 – 1335.
- [20] Gong, D., Medioni, G., Zhao, X., 2014. Structured time series analysis for human action segmentation and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 36 (7), 1414–1427.

- [21] Iyengar, S., Zhao, P.-L., 1994. Maximum likelihood estimation for weighted distributions. *Statistics & Probability Letters* 21 (1), 37 – 47.
- [22] Kan, M., Tan, A., Mathew, J., 2015. A review on prognostic techniques for non-stationary and non-linear rotating systems. *Mechanical Systems and Signal Processing* 62, 1–20, cited By 0.
- [23] Klir, G., Wierman, M., 1999. *Uncertainty-Based Information: Elements of Generalized Information Theory*, 2nd Edition. Physica-Verlag.
- [24] Liu, Q., Dong, M., Lv, W., Geng, X., Li, Y., 2015. A novel method using adaptive hidden semi-markov model for multi-sensor monitoring equipment health prognosis. *Mechanical Systems and Signal Processing* 6465, 217 – 232.
- [25] Liu, Y., 2008. Predictive modeling for intelligent maintenance in complex semiconductor manufacturing processes. Ph.D. thesis, University of Michigan.
- [26] Ma, X., Shi, J., 2000. A new method for discrimination between fault and magnetizing inrush current using HMM. *Electric Power Systems Research* 56 (1), 43–49.
- [27] Masson, M.-H., Denoeux, T., 2008. ECM: An evidential version of the fuzzy C-means algorithm. *Pattern Recognition* 41, 1384–1397.
- [28] Mechbal, N., Uribe, J. S., Rbillat, M., 2015. A probabilistic multi-class classifier for structural health monitoring. *Mechanical Systems and Signal Processing* 6061, 106 – 123.
- [29] Menon, S., Uluyol, O., Kim, K., Nwadiogbu, E., June 16-19 2003. Incipient fault detection and diagnosis in turbine engines using hidden Markov models. In: *ASME Turbo Expo 2003, collocated with the Int. Joint Power Generation Conference*. Atlanta, Georgia, USA, pp. 493–500.
- [30] Moghaddass, R., 2013. *Equipment Degradation Diagnostics and Prognostics Under a Multistate Deterioration Process*. Ph.D. thesis, University of Alberta.
- [31] Murphy, K., 2002. *Dynamic Bayesian Networks: Representation, inference and learning*. Ph.D. thesis, UC Berkeley, Computer Science Division.
- [32] Nielsen, R., J., 2003. Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theoretical Population Biology* 63 (3), 245 – 255.
- [33] Patil, G., 2002. *Weighted distributions*. Vol. 4. John Wiley & Sons, Ltd, Chichester, pp. 2369-2377.
- [34] Rabiner, L., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. of the IEEE* 77 (2), 257–286.

- [35] Ramasso, E., 2009. Contribution of belief functions to hidden markov models with an application to fault diagnosis. In: IEEE International Workshop on Machine Learning for Signal Processing, MLSP'09. pp. 1–6.
- [36] Ramasso, E., 2014. Investigating computational geometry for failure prognostics. *Int. Journal on Prognostics and Health Management* 5 (5), 1–18.
- [37] Ramasso, E., January 2016. Segmentation of CMAPSS health indicators into discrete states for sequence-based classification and prediction purposes. Tech. Rep. 6839, FEMTO-ST institute.
- [38] Ramasso, E., Denoeux, T., 2014. Making use of partial knowledge about hidden states in HMMs: an approach based on belief functions. *Fuzzy Systems, IEEE Transactions on* 22 (2), 395–405.
- [39] Ramasso, E., Panagiotakis, C., Rombaut, M., Pellerin, D., 2010. Belief scheduler based on model failure detection in the TBM framework. application to human activity recognition. *International Journal of Approximate Reasoning* 51 (7), 846 – 865.
- [40] Ramasso, E., Placet, V., Boubakar, M., (accepted) 2015. Unsupervised consensus clustering of acoustic emission time-series for robust damage sequence estimation in composites. *IEEE Transactions on Instrumentation and Measurement*.
- [41] Ramasso, E., Saxena, A., 2014. Performance benchmarking and analysis of prognostic methods for CMAPSS datasets. *International Journal on Prognostics and Health Management* 5 (2), 1–15.
- [42] Ranjan, R., Mitra, D., 2015. HMM modeling for OFDM-BER performance. *AEU-International Journal of Electronics and Communications* 69 (1), 18–25.
- [43] Reineking, T., 2011. Particle filtering in the Dempster-Shafer theory. *International Journal of Approximate Reasoning* 52 (8), 1124 – 1135.
- [44] Reineking, T., 2014. Belief functions: Theory and algorithms. Ph.D. thesis, Universitat Bremen.
- [45] Renooij, S., 2012. Efficient sensitivity analysis in hidden Markov models. *International Journal of Approximate Reasoning* 53 (9), 1397 – 1414.
- [46] Saxena, A., Goebel, K., Simon, D., Eklund, N., 2008. Damage propagation modeling for aircraft engine run-to-failure simulation. In: *International Conference on Prognostics and Health Management*. IEEE, Denver, CO, USA, pp. 1–9.
- [47] Serir, L., Ramasso, E., Nectoux, P., Zerhouni, N., 2013. E2GKpro: An evidential evolving multi-modeling approach for system behavior prediction with applications. *Mechanical Systems and Signal Processing* 37 (12), 213 – 228.

- [48] Smets, P., Kennes, R., 1994. The transferable belief model. *Artificial Intelligence* 44, 191–243.
- [49] Spiridonakos, M., Fassois, S., 2014. Non-stationary random vibration modelling and analysis via functional series time-dependent ARMA (FS-TARMA) models - a critical survey. *Mechanical Systems and Signal Processing* 47 (1-2), 175 – 224.
- [50] T. Kosaka, M. K., Kohda, M., 2007. Robust Speech recognition and understanding. I-Tech education and publishing, Ch. Discrete-mixture HMMs-based approach for noisy speech recognition, pp. 158–174.
- [51] Tibaduiza, D.-A., Torres-Arredondo, M.-A., Mujica, L., Rodellar, J., Fritzen, C.-P., 2013. A study of two unsupervised data driven statistical methodologies for detecting and classifying damages in structural health monitoring. *Mechanical Systems and Signal Processing* 41 (12), 467 – 484.
- [52] Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A., Wu, B., 2006. *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Wiley.
- [53] Van Den Berg, J., Kaymak, U., Van Den Bergh, W.-M., 2004. Financial markets analysis by using a probabilistic fuzzy modelling approach. *International Journal of Approximate Reasoning* 35 (3), 291 – 305, integration of Methods and Hybrid Systems.
- [54] Vanik, M., Beck, J. L., Au, S., 2000. Bayesian probabilistic approach to structural health monitoring. *Journal of Engineering Mechanics* 126 (7), 738–745.
- [55] Vinh, N., Epps, J., Bailey, J., 2009. Information theoretic measures for clusterings comparison: Is a correction for chance necessary? In: *Proc. of the 26th Annual Int. Conference on Machine Learning*. ACM, New York, NY, USA, pp. 1073–1080.
URL <http://doi.acm.org/10.1145/1553374.1553511>
- [56] Yan, W., Katafygiotis, L. S., 2015. A novel bayesian approach for structural model updating utilizing statistical modal information from multiple setups. *Structural Safety* 52, 260–271.
- [57] Yu, S.-Z., 2010. Hidden semi-Markov models. *Artificial Intelligence* 174 (2), 215–243.

List of Figures

1	Plot chart of the proposed scheme	21
2	Distribution of signal points for the first dataset	22
3	Boxplots of ARI for GMM and HMM (continuous and discrete) learned with E2M based on different weights.	23
4	Medians of the adjusted Rand index as a function of the non-specificity coefficient over 30 repetitions for different number of clusters (3, 5, 15) and with time-series of different length (100, 300).	24
5	Medians of the adjusted Rand index as a function of the labelling error with uncertain labels over 30 repetitions for different number of clusters (3, 5, 15) and with time-series of different length (100, 300).	25
6	Medians of the adjusted Rand index as a function of the labelling error with noisy labels over 30 repetitions for different number of clusters (3, 5, 15) and with time-series of different length (100, 300).	26
7	Evolution of the health indices for all engines in the four datasets.	27
8	Dataset #1: Boxplot of ARI for different numbers of symbols (N, used in quantization).	28
9	Dataset #2, with operating conditions modeling: Boxplot of ARI for different numbers of symbols (N, used in quantization).	29
10	Dataset #3: Boxplot of ARI for different numbers of symbols (N, used in quantization).	30
11	Dataset #4, with operating conditions modeling: Boxplot of ARI for different numbers of symbols (N, used in quantization).	31

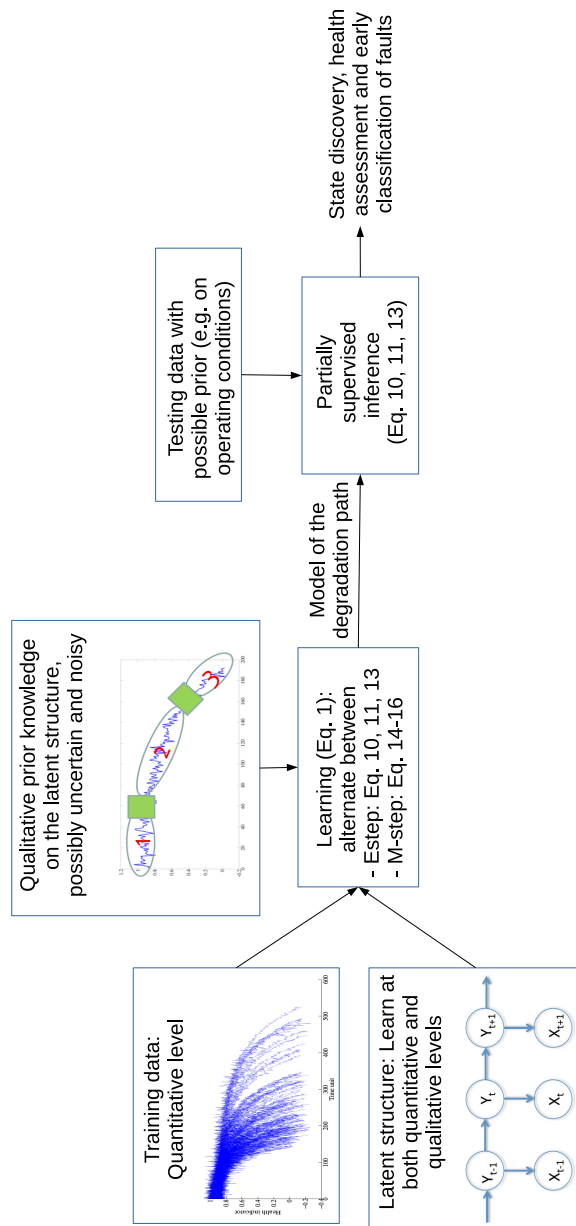


Figure 1: Plot chart of the proposed scheme

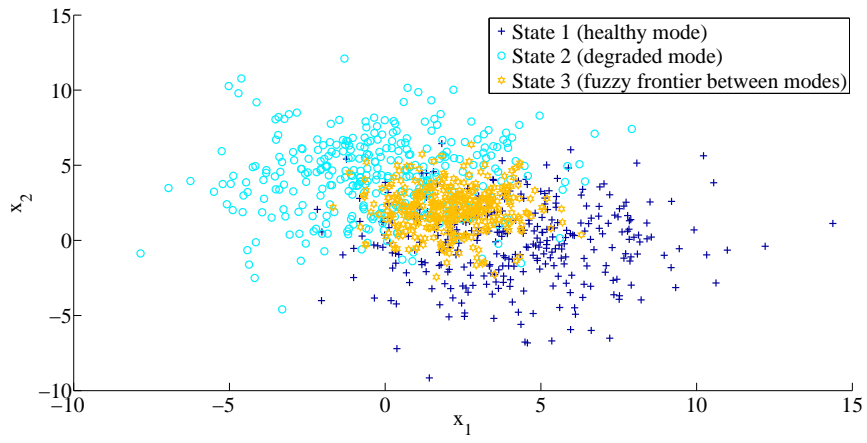
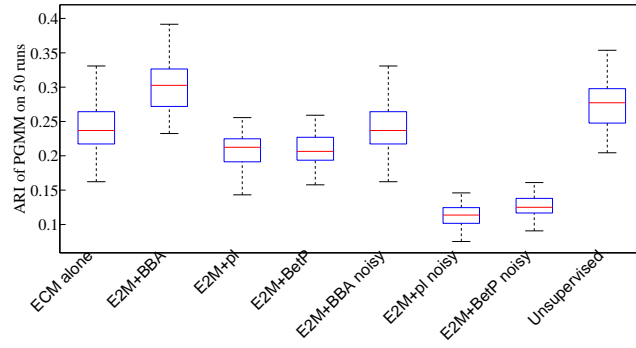
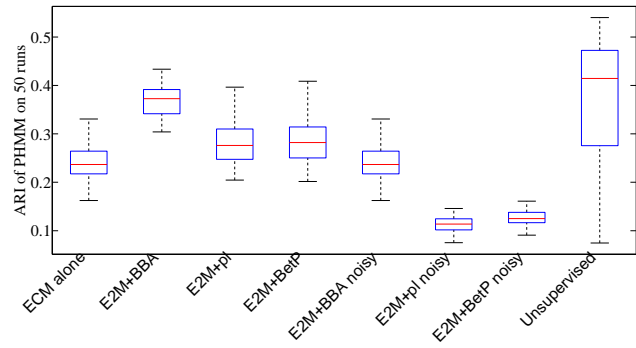


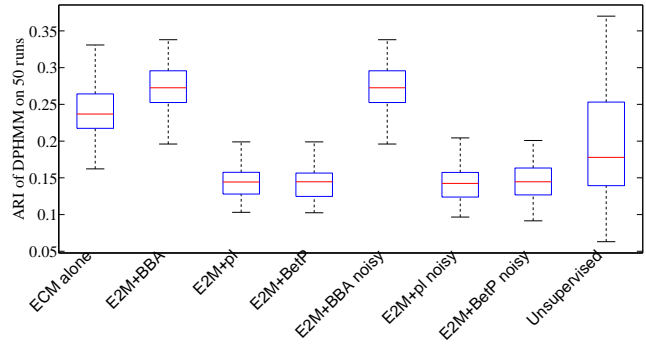
Figure 2: Distribution of signal points for the first dataset



(a) GMM with weights



(b) CPHMM with weights



(c) DPHMM with weights

Figure 3: Boxplots of ARI for GMM and HMM (continuous and discrete) learned with E2M based on different weights.

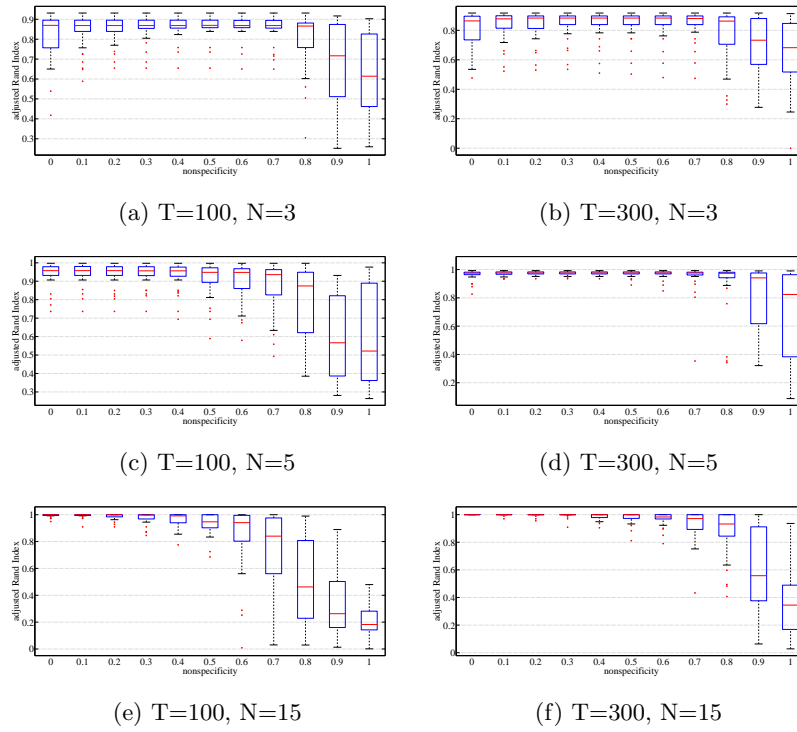


Figure 4: Medians of the adjusted Rand index as a function of the **nonspecificity** coefficient over 30 repetitions for different number of clusters (3, 5, 15) and with time-series of different length (100, 300).

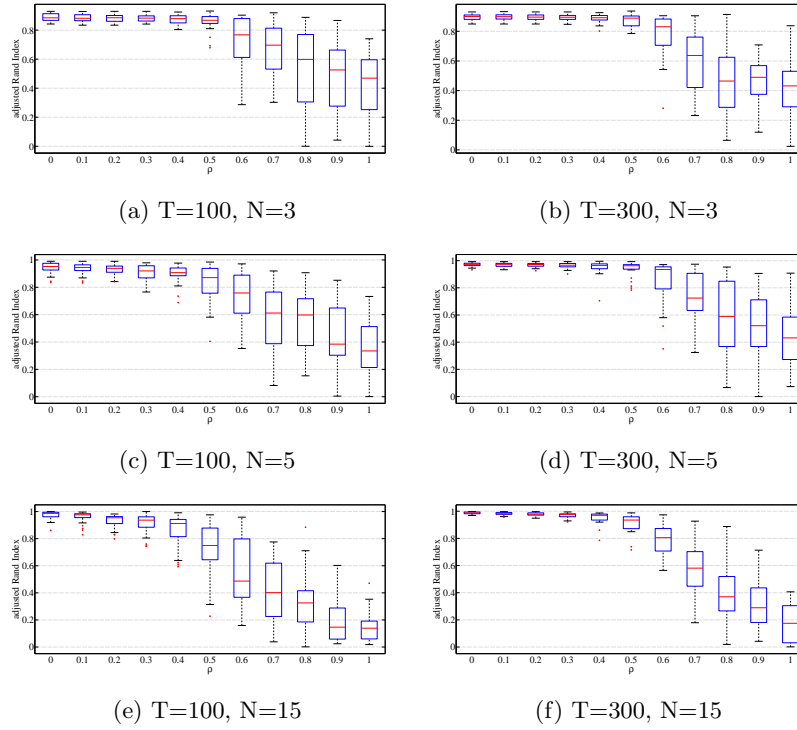


Figure 5: Medians of the adjusted Rand index as a function of the **labelling error with uncertain labels** over 30 repetitions for different number of clusters (3, 5, 15) and with time-series of different length (100, 300).

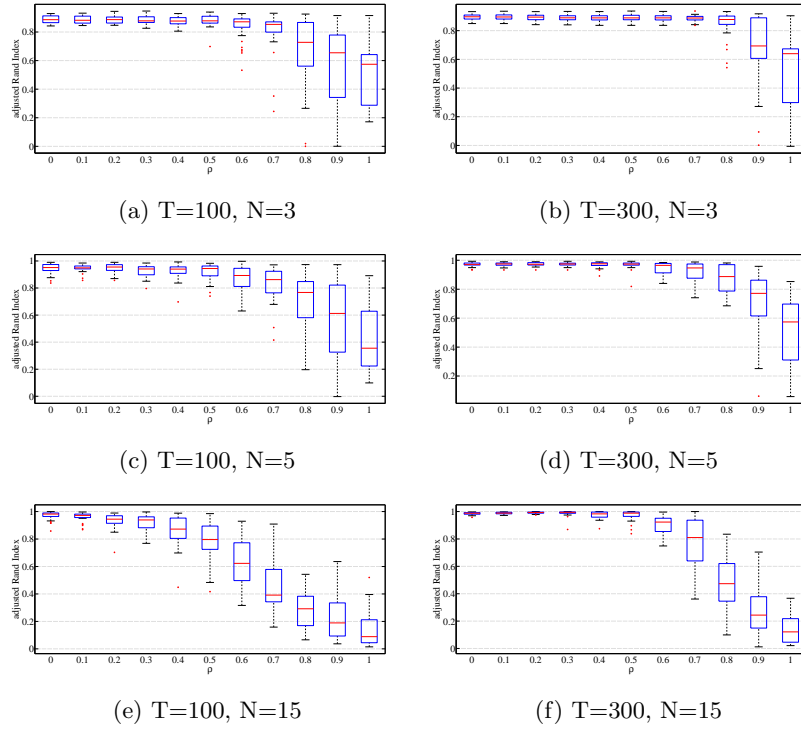


Figure 6: Medians of the adjusted Rand index as a function of the **labelling error with noisy labels** over 30 repetitions for different number of clusters (3, 5, 15) and with time-series of different length (100, 300).

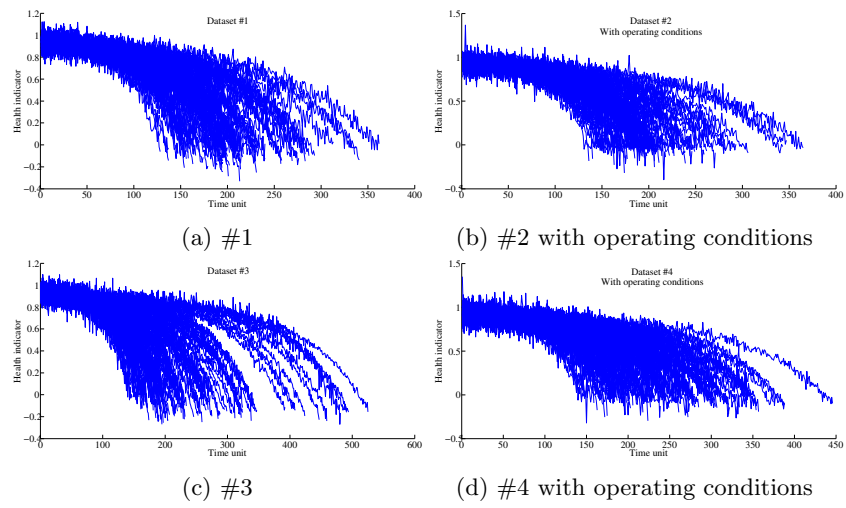


Figure 7: Evolution of the health indices for all engines in the four datasets.

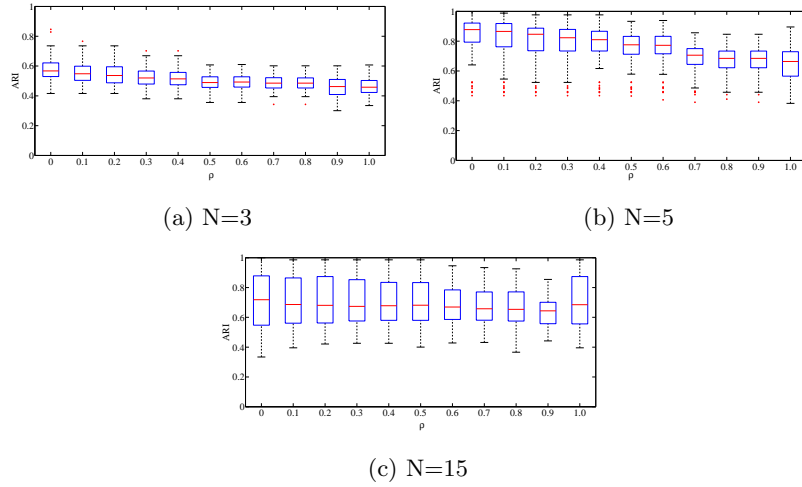


Figure 8: Dataset #1: Boxplot of ARI for different numbers of symbols (N , used in quantization).

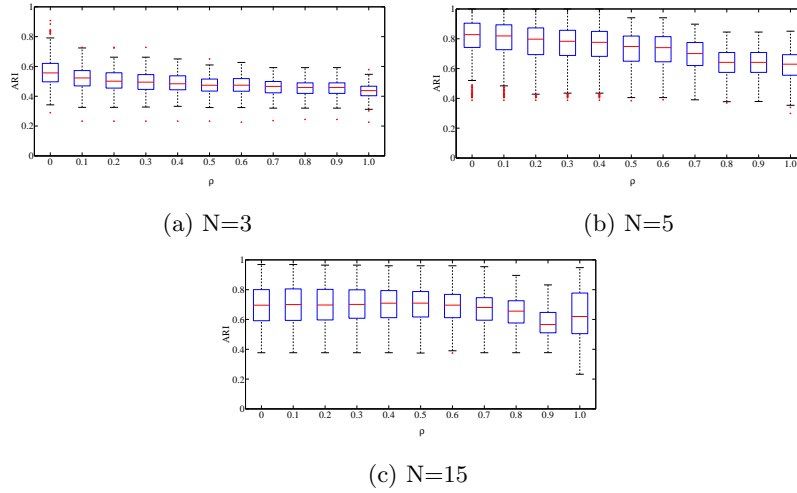


Figure 9: Dataset #2, with operating conditions modeling: Boxplot of ARI for different numbers of symbols (N , used in quantization).

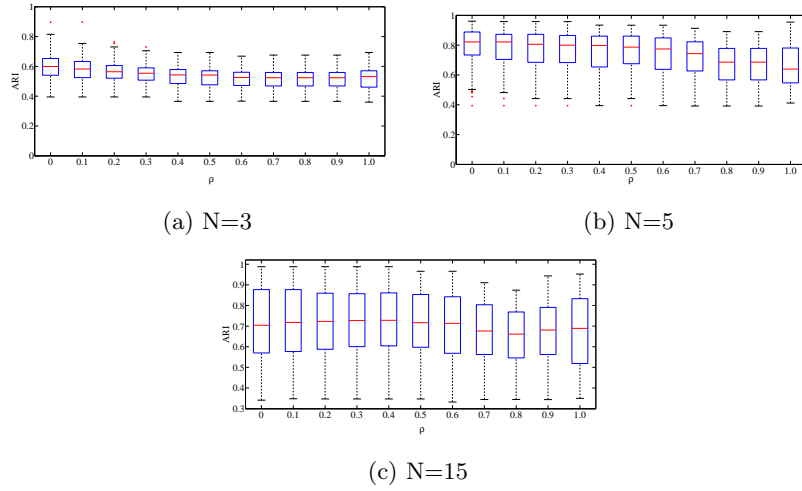
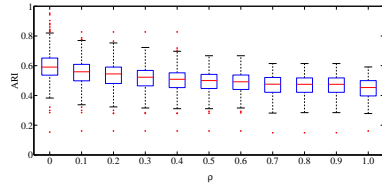
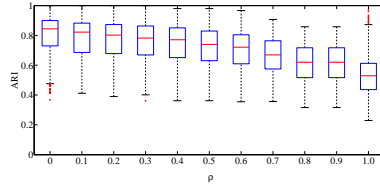


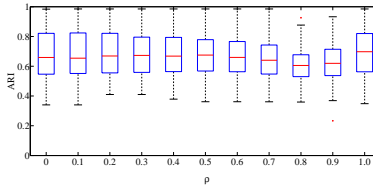
Figure 10: Dataset #3: Boxplot of ARI for different numbers of symbols (N , used in quantization).



(a) $N=3$



(b) $N=5$



(c) $N=15$

Figure 11: Dataset #4, with operating conditions modeling: Boxplot of ARI for different numbers of symbols (N , used in quantization).

List of Tables

1	Description of the turbofan degradation datasets available from NASA repository.	33
---	--	----

Datasets		#Fault Modes	#Conditions	#Train Units	#Test Units
Turbofan	#1	1	1	100	100
data from	#2	1	6	260	259
NASA	#3	2	1	100	100
repository	#4	2	6	249	248

Table 1: Description of the turbofan degradation datasets available from NASA repository.