Remaining Useful Life Prediction by Fusing Expert Knowledge and Condition Monitoring Information

Tangfan Xiahou, Zhiguo Zeng, Yu Liu

Remaining Useful Life Prediction by Fusing Expert Knowledge and Condition Monitoring Information

Article *in* IEEE Transactions on Industrial Informatics · May 2020

DOI: 10.1109/TII.2020.2998102

3 authors, including:

Zhiguo Zeng
CentraleSupélec
65 PUBLICATIONS **650** CITATIONS

SEE PROFILE

Remaining Useful Life Prediction by Fusing Expert Knowledge and Condition Monitoring Information

Tangfan Xiahou[1], Zhiguo Zeng[2], and Yu Liu[1,3,*], *Senior Member, IEEE*

1. School of Mechanical and Electrical Engineering,
University of Electronic Science and Technology of China,
Chengdu, Sichuan, 611731, P.R. China
2. Chaire on Risk and Resilience of Complex Systems,
Université Paris Saclay, CentraleSupélec, Laboratoire Genie Industriel,
Gif-sur-Yvette, France
3. Center for System Reliability and Safety,
University of Electronic Science and Technology of China,
Chengdu, Sichuan 611731, P.R. China

[*]Corresponding author at: No. 2006, Xiyuan Avenue, West High-Tech Zone, Chengdu, Sichuan 611731, PR China.
E-mail address: xiahoutf@163.com (T. Xiahou), zhiguo.zeng@centralesupelec.fr (Z. Zeng), yuliu@uestc.edu.cn (Y. Liu).

***Abstract*—**In this paper, we develop a mixture of Gaussians-evidential hidden Markov model (MoG-EHMM) to fuse expert knowledge and condition monitoring information for RUL prediction under the belief function theory framework. The evidential Expectation-Maximization algorithm is implemented in the offline phase to train the MoG-EHMM based on historical data. In the online phase, the trained model is used to recursively update the health state and reliability of a particular individual system. The predicted RUL is, then, represented in the form of its probability mass function. A numerical metric is defined based on the Bhattacharyya distance to measure the RUL prediction accuracy of the developed methods. We applied the developed methods on a simulation experiment and a real-world dataset from a bearing degradation test. The results demonstrate that despite imprecisions in expert knowledge, the performance of RUL prediction can be substantially improved by fusing expert knowledge with condition monitoring information.

***Index Terms*—**Belief function theory, expert knowledge, Mixture of Gaussians-evidential hidden Markov model (MoG-EHMM), remaining useful life (RUL)

I. INTRODUCTION

Prognostics and health management (PHM) has been widely recognized as a useful tool to provide early failure warnings and prevent industrial equipment from unexpected shutdowns. One of the core tasks in PHM

is to accurately predict remaining useful life (RUL) as a large number of engineering decisions rely on the predicted RUL [1]. For example, the results of RUL prediction are used for effective predictive maintenance planning to reduce unplanned maintenance costs and resource wasting. Catastrophic consequences on production and human safety can also be prevented based on the RUL prediction results. RUL prediction has, therefore, attracted tremendous research efforts over the past decade [2],[3].

Existing approaches to RUL prediction can be broadly divided into two categories: physical model-based approaches and data-driven approaches. The physical model-based approaches characterize underlying physical failure processes and predict RUL via analytical models [4]. Often, filtering techniques are used to sequentially update parameters of physical models when new condition monitoring (CM) information is collected [5]. However, as industrial equipment becomes increasingly complicated and sophisticated, the challenge of developing physics-of-failure models emerges and limits the application of physical model-based approaches.

Unlike the physical model-based approaches, data-driven approaches directly learn degradation patterns from CM information. More specifically, we can divide data-driven approaches into direct RUL prediction approaches and statistical model-based approaches [6]. Direct RUL prediction approaches typically resort to artificial intelligence (AI) methods, e.g., neural networks [7], deep learning [8], support vector machines [9], to directly learn a complex mapping from CM information to the RUL. These approaches, with their powerful feature extraction and regression capability, have received tremendous attention in recent years. However, direct RUL prediction approaches require a large number of high-quality data to train the AI-based models. Further, direct RUL prediction approaches have been criticized for being “black boxes” due to their inability to physically explain the data-RUL mapping learned from the data [10].

On the other hand, statistical model-based approaches use rigorous models to learn the degradation evolution regularities from CM information. Typical statistical model-based prognostics including Wiener process-based, Gamma process-based, inverse Gaussian process-based, and hidden Markov models (HMMs)-based approaches. Among these statistical model-based approaches, the HMM-based approach has

widespread applications due to its excellent capability to connect physical degradation processes with observations. Successful applications of the HMM-based approach include bearings [11], turbofan engines [12], and light-emitting diodes (LED) [13]. Therefore, the HMM-based approach is considered in this work for degradation modeling and RUL prediction.

In general, an HMM comprises of two stochastic processes: one is a Markov chain that characterizes the actual degradation of system health states; the other models the observations given the actual degradation states [14]. The HMM allows us to infer the true but unobservable degradation process based on the collected observation data. Traditional HMMs are only able to treat discrete observations. In practice, however, a large number of observation processes are continuous as they come from continuous condition-monitoring signals. To model the relationship between continuous signals and discrete hidden degradation states, the mixtures of Gaussians (MoG) model is frequently combined with HMMs for prognostics. These models are called MoG-HMMs. For example, Tobon-Mejia et al. [15] developed a MoG-HMM-based RUL prediction method for bearings. This method relies on two phases: offline and online. The offline phase trained the MoG-HMM by estimating its parameters. The trained MoG-HMM was exploited in the online phase to assess the current health state of a new system continuously, and to estimate its RUL with the associated confidence. Chen et al. [16] formulated an HMM with auto-correlated observations (HMM-AO) to characterize the degradation of manufacturing systems and developed optimal maintenance policies based on the RUL prediction results. Geramifard et al. [17] introduced a multimodal HMM (m^2HMM) to monitor tool wear. They used three weighting schemes and two switching strategies to combine the continuous wear output from multiple modes.

Nevertheless, most of the existing MoG-HMM-based prognostics use CM information as the only information source for RUL prediction. In reality, apart from CM information, some subjective knowledge on the health state of a system can also be collected from experts. For example, the health state of bearings can be evaluated by experts via direct visual inspections or indirect measurements [23]. In aviation, experts might be able to evaluate the health state of turbofan engines during the breaks between two adjacent missions [12]. Expert knowledge can also provide insight into the health state and could be integrated with the CM

information to achieve more accurate prognostics. However, as pointed out by Si et al. [1] and Lei et al. [10], the effective use of subjective expert knowledge for RUL prediction remains an open challenge. More specifically, the challenges include: (1) to quantify expert knowledge imprecision due to the vagueness of expert judgments and/or the measurement uncertainty; (2) to fuse two different types of information, i.e., expert knowledge and CM information. Existing literature has made some attempts on these challenges. For example, He et al. [19] introduced an exponential model to characterize the degradation of Li-ion batteries, where the model parameters were initialized by combining different imprecise expert knowledge. However, they did not use expert knowledge in the operation phase of Li-ion batteries to support RUL prediction. Ramasso and Denoeux [12] developed a partially-hidden Markov model (PHMM) to estimate model parameters by combining expert knowledge and observations. They found that including expert knowledge drastically improved the performance of parameter estimation. Nevertheless, the PHMM assumed that observations are discrete, and only used one state sequence in the offline training phase. Such a model cannot be straightforwardly implemented on continuous CM information from non-repairable systems. To the best of our knowledge, existing MoG-HMM-based models did not fuse expert knowledge and CM information for RUL prediction.

A mixture of Gaussians-evidential hidden Markov model (MoG-EHMM) is proposed in this paper to fill the aforementioned research gap. The expert knowledge is quantified by the belief function theory which allows modeling knowledge with a range of quality levels (from precise knowledge down to non-informative knowledge) [12],[18]. Compared to the traditional MoG-HMM [15], the hidden states in the developed model are partially, rather than completely hidden, because some expert knowledge on the health states can be elicited during system operations. Under the developed model, the traditional MoG-HMM [15] becomes a special case with non-informative knowledge. Moreover, the Dempster's rule of combination in the belief function theory provides a useful tool to fuse multiple sources of expert knowledge and CM information. The unique contributions of this paper lie in the following three aspects:

1) A MoG-EHMM is developed to fuse expert knowledge and CM information under the belief function

theory framework.

2) RUL is predicted by integrating expert knowledge and CM information.

3) A numerical metric is defined based on the Bhattacharyya distance to measure the accuracy of the predicted RUL.

The remainder of this paper is organized as follows: Section II provides the necessary background of the belief function theory. The MoG-EHMM is formally defined and used for RUL prediction in Sect. III. Simulation experiments (Sect. IV) and an application on real bearing degradation dataset (Sect. V) are carried out to examine the effectiveness of the MoG-EHMM. Finally, conclusions are drawn in Sect. VI.

II. BACKGROUND

Belief function theory (BFT) was initialized by Dempster [20] and Shafer [21]. In general, let Ω be a set containing all hypotheses/propositions that are presumed to be mutually exclusive. Let Y be a variable taking value in Ω. The uncertain information on Y can be represented by a mass function $m: 2^{\Omega} \to [0,1]$, which satisfies the normalization axiom $\sum_{A\subseteq\Omega} m(A)=1$. The set A with $m(A)>0$ is called a *focal* set. In particular, if a mass function has a single focal set, it is called a *logical* mass. It represents a piece of precise knowledge. The mass function degenerates to a *Bayesian* mass if all focal sets are singletons. A *vacuous* mass $m(\Omega)=1$ corresponds to the non-informative knowledge.

The plausibility function is defined as

$$Pl(B)=\sum_{A\cap B\neq\varnothing} m(A). \tag{1}$$

The plausibility of B equals to the sum of the masses that are not in contradiction with set B, and represents the maximum degrees of support that could be attributed to set B.

The contour function $pl: \Omega \to [0,1]$ restricts the plausibility function to singletons, i.e., $pl(\omega)=Pl(W)$, where $|W|=1$ and $\omega \in \Omega$. The quantity $pl(\omega)$ represents the likelihood of proposition ω.

Evidence from different sources can be fused by the Dempster rule of combination (DRC) "$\oplus$", which is

defined as

$$\left(m_1 \oplus m_2\right)[A] = \begin{cases} \dfrac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1-k} & \text{if } A \subseteq \Omega \\ 0 & \text{if } A = \varnothing \end{cases}, \tag{2}$$

where $k = \left(m_1 \cap m_2\right)[\varnothing] = \sum_{B \cap C = \varnothing} m_1(B) m_2(C)$ is called the *degree of conflict*. If m_1 (or m_2) is a Bayesian mass, then $m_1 \oplus m_2$ is also a Bayesian mass.

III. METHODOLOGY

In this section, we develop a mixture of Gaussians-evidential hidden Markov model (MoG-EHMM) to fuse expert knowledge with CM information. The MoG-EHMM is formally defined as a three-layer model (See details in Sect. III-A). The MoG-EHMM-based RUL prediction comprises of two phases: offline and online. Both phases can elicit Expert knowledge. In the offline phase, training data are collected from a population of similar systems. Then, health indicators (HIs) are extracted from the original training data through feature extraction. Evidential Expectation-Maximization (E^2M) algorithm is implemented to estimate the parameters of MoG-EHMM for model training (See details inSect. III-B). In the online phase, CM information is collected from a new system. Based on the extracted HIs and the online expert knowledge, forward algorithm [14] is exploited recursively for health state inference, system reliability updating, and the RUL prediction. (See details in Sects. III-C, and III-D).

A. Model Formulation

The MoG-EHMM comprises of three-layers: true degradation layer, observation layer, and knowledge layer, as shown in Fig.1. The true degradation layer models the true (but unobservable) degradation process. It is partially hidden because some knowledge of the health state of a system is available from experts. The observation layer represents the HIs extracted from signals, and the knowledge layer quantifies the expert knowledge by the contour functions under the BFT.

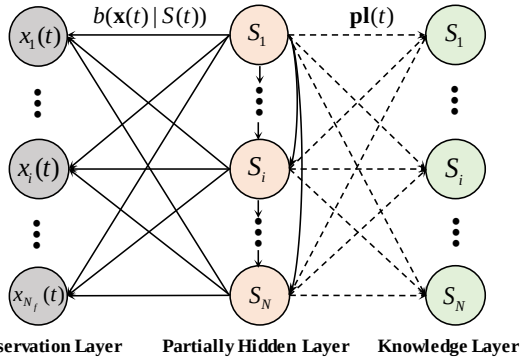


Fig. 1. The proposed MoG-EHMM.

In the true degradation layer, it is assumed that the degradation of a system is multi-state and can be modeled by a discrete- time discrete-state Markov model. Let $S(t)$ denote the system states associated with degradation, where $S(t) \in S_1, S_2, ..., S_N$ and N is the number of system states. The number of states can be determined by expert experience [15]. Let $S_1, S_2, ..., S_N$ be in descending order of performance levels where S_1 is the perfect functioning state and S_N is the complete failure state. The one- step transition probability from state S_i to S_j is denoted as $a_{ij} = \Pr\{S(t+1) = S_j \,|\, S(t) = S_i\}$ ($t = 1, 2, ..., T$, $1 \le i, j \le N$). The corresponding transition probability matrix is denoted by $\mathbf{A} = [a_{ij}]_{N \times N}$, where $\sum_{j=1}^{N} a_{ij} = 1$. Note that maintenance action is not considered in the present study, i.e., $S(t)$ can only transit to worse states. The resulting hidden Markov model is the so-called left-right or Bakis model in HMMs [14]. The initial state probability distribution is represented by $\boldsymbol{\pi} = [\pi_1, \pi_2, ..., \pi_i, ..., \pi_N]$ where $\pi_i = \Pr\{S(0) = S_i\}$ $(1 \le i \le N)$. System reliability $S(t) \in [S_1, S_2, ..., S_i, ..., S_N]$ is defined as the probability that the performance level of the system is not lower than a threshold state S_F ,

$$R_S(t) = \sum\nolimits_{S_i \le S_F} \Pr\{S(t) = S_i\}. \tag{3}$$

In the observation layer, continuous CM information denoted by $\mathbf{c}(t)$, such as vibration signals and acceleration signals, can be collected from sensors. Through feature extraction, HIs can be reconstructed from $\mathbf{c}(t)$. Let $\mathbf{x}(t)$ ($t = 1, 2, ..., T$) denote the extracted HIs, where $\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_{N_f}(t)]$ and N_f is the number of HIs. Examples of the HIs include root mean square, mean value, and kurtosis. A Gaussian

mixture model (GMM) is utilized to characterize the continuous evolution behaviors of the HIs and the uncertainty associated with them. In the GMM, the emission probability $b(\mathbf{x}(t)\,|\,S(t))$ represents the probability of observing the current values of HIs $\mathbf{x}(t)$, given that the current state is $S(t)$, i.e., $b_i(\mathbf{x}(t)\,|\,S(t)) = \Pr\{\mathbf{x}(t)\,|\,S(t)=S_i\} = \mathcal{N}(\mathbf{x}(t)\,|\,\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\mu}_i$ is the mean value of $\mathbf{x}(t)$ under the given hidden state S_i, while the covariance matrix $\boldsymbol{\Sigma}_i$ captures the uncertainty associated with $\mathbf{x}(t)$. Note that the number of Gaussian components is determined in this paper by minimizing the Akaike information criterion (AIC) [12],[15].

In the knowledge layer, expert knowledge on the health state of a system is given in the form of mass function $\mathbf{m}(t) = [m_{S_1}(t), m_{S_2}(t), ..., m_{S_N}(t), ..., m_{\Omega}(t)]$ or contour function $\mathbf{pl}(t) = [pl_{S_1}(t), pl_{S_2}(t), ..., pl_{S_N}(t)]$ ($t = 1, 2, ..., T$) under the BFT. A commonly-used format for experts to express their knowledge in terms of contour function [12]:

$$pl_{S_i}(t) = \begin{cases} 1 & \text{If } S_{Expert}(t) = S_i \\ \rho & \text{Otherwise} \end{cases}, \tag{4}$$

where $S_{Expert}(t)$ is the current health state judged by experts, ρ ($0 \le \rho \le 1$) is the non-specificity coefficient that quantifies the epistemic uncertainty in the expert judgments: $\rho = 0$ indicates precise knowledge elicited by experts, whereas $\rho = 1$ corresponds to the non-informative knowledge. A larger value of ρ means experts are more uncertain in his/her judgments. Note that $S_{Expert}(t)$ is not guaranteed to be coincide with the true health state, denoted by $S_{True}(t)$, as the experts might sometimes provide incorrect judgments.

Expert knowledge can also be elicited in terms of mass function. In fact, as shown in Proposition 1 in the Appendix A in the supplementary file, mass function is equivalent to contour function in terms of eliciting knowledge from the experts. If the mass function is elicited, it can be converted into the contour functions before fusing with CM information. Therefore, we only present the developed methods in terms of contour functions in this work.

B. Parameter Estimation of MoG-EHMM in the Offline Phase

In the offline phase, the training data, denoted as $\mathbf{c}_{Tr}^{(k)}(t)$ ($k=1,2,...,K$, $t=1,2,...,T$) need to be collected from a population of K identical systems. The number of training data, i.e., *K*, should be as large as possible, because the size of the training dataset could directly impact the training performance. Through feature extraction, the HIs $\mathbf{x}_{Tr}^{(k)}(t)$ ($k=1,2,...,K$, $t=1,2,...,T$) can be extracted. Based on the training data $\mathbf{x}_{Tr}^{(k)}(t)$ and its corresponding expert contour function $\mathbf{pl}_{Tr}^{(k)}(t)$ ($k=1,2,...,K$, $t=1,2,...,T$), the parameter of the MoG-EHMM $\hat{\boldsymbol{\theta}}=(\hat{\boldsymbol{\pi}},\hat{\mathbf{A}},\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}})$ can be estimated by maximizing the likelihood of observing $\mathbf{x}_{Tr}^{(k)}(t)$ and $\mathbf{pl}_{Tr}^{(k)}(t)$, ($k=1,2,...,K$, $t=1,2,...,T$)

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{x},\mathbf{pl}\,|\,\boldsymbol{\theta}) = \Pr\{\mathbf{x}_{Tr}^{(1)}(t),...,\mathbf{x}_{Tr}^{(K)}(t),\mathbf{pl}_{Tr}^{(1)}(t),...,\mathbf{pl}_{Tr}^{(K)}(t)\,|\,\boldsymbol{\theta}\} \\ &= \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{x},\mathbf{pl}\,|\,\boldsymbol{\theta}) = b(\mathbf{x}_{Tr}^{(1)}(t),...,\mathbf{x}_{Tr}^{(K)}(t)\,|\,\boldsymbol{\theta}) \oplus (\mathbf{pl}_{Tr}^{(1)}(t),...,\mathbf{pl}_{Tr}^{(K)}(t)) \\ &= \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{x},\mathbf{pl}\,|\,\boldsymbol{\theta}) = \prod_{k=1}^{K}\prod_{t=1}^{T} b(\mathbf{x}_{Tr}^{(k)}(t)\,|\,\boldsymbol{\theta}) \oplus \mathbf{pl}_{Tr}^{(k)}(t)\end{aligned} \tag{5}$$

where $b(\mathbf{x}_{Tr}^{(k)}(t)\,|\,\boldsymbol{\theta})$ is the emission probability of the *k*th training data at time *t* given the parameters $\boldsymbol{\theta}$ of the MoG-EHMM. Note that the result of $b(\mathbf{x}_{Tr}^{(k)}(t)\,|\,\boldsymbol{\theta}) \oplus \mathbf{pl}_{Tr}^{(k)}(t)$ is still a probability measure as $b(\mathbf{x}_{Tr}^{(k)}(t)\,|\,\boldsymbol{\theta})$ can be regarded as a Bayesian mass [12]. Directly resolving (5) is challenging because the true states are partially hidden. The E^{2}M algorithm can be implemented to calculate $\mathcal{L}(\mathbf{x},\mathbf{pl}\,|\,\boldsymbol{\theta})$ iteratively:

***E*-Step**: Compute the expectation of $\mathcal{L}(\mathbf{x},\mathbf{pl}\,|\,\hat{\boldsymbol{\theta}})$ given the current estimates $\hat{\boldsymbol{\theta}}$.

***M*-Step:** Maximize the log-likelihood function obtained in the *E*-Step and calculate a new maximum likelihood estimate for the unknown parameters $\hat{\boldsymbol{\theta}}$.

To implement the E^{2}M algorithm, two auxiliary variables, namely the forward variable $\alpha_{S_j}^{(k)}(t)$ and backward variable $\beta_{S_j}^{(k)}(t)$ are introduced. In this work, $\alpha_{S_j}^{(k)}(t)$ is defined as the probability of observing $\mathbf{x}_{Tr}^{(k)}(1),...,\mathbf{x}_{Tr}^{(k)}(t)$ and $\mathbf{pl}_{Tr}^{(k)}(1),...,\ \mathbf{pl}_{Tr}^{(k)}(t)$ with the current state S_j given the parameter $\boldsymbol{\theta}$:

$$\begin{aligned}\alpha_{S_j}^{(k)}(t) &= \Pr\{\mathbf{x}_{Tr}^{(k)}(t),\mathbf{pl}_{Tr}^{(k)}(t),S(t)=S_j\,|\,\boldsymbol{\theta}\} \\ &= \left(b(\mathbf{x}_{Tr}^{(k)}(t)\,|\,\boldsymbol{\theta}) \oplus \mathbf{pl}_{Tr}^{(k)}(t)\right)[S_j]\end{aligned} \tag{6}$$

for $1 \le j \le N$ ($k = 1, 2, ..., K$, $t = 1, 2, ..., T$). It can be verified that

$$\begin{cases} \alpha_{S_j}^{(k)}(1) = \pi_i \times \left(b(\mathbf{x}_{Tr}^{(k)}(1) \mid \boldsymbol{\theta}) \oplus \mathbf{pl}_{Tr}^{(k)}(1) \right)[S_j] \\ \alpha_{S_j}^{(k)}(t+1) = \left(b(\mathbf{x}_{Tr}^{(k)}(t+1) \mid \boldsymbol{\theta}) \oplus \mathbf{pl}_{Tr}^{(k)}(t+1) \right)[S_j] \times \left[\sum_{i=1}^{N} \alpha_{S_i}^{(k)}(t) a_{ij} \right] \end{cases} \tag{7}$$

The backward variable $\beta_{S_j}^{(k)}(t)$ is defined as the probability of observing $\mathbf{x}_{Tr}^{(k)}(t+1), ..., \mathbf{x}_{Tr}^{(k)}(T)$ and $\mathbf{pl}_{Tr}^{(k)}(t+1), ..., \mathbf{pl}_{Tr}^{(k)}(T)$ ($k = 1, 2, ..., K$, $t = 1, 2, ..., T-1$) given the current state S_j and the parameter $\boldsymbol{\theta}$

$$\begin{aligned} \beta_{S_j}^{(k)}(t) &= \Pr\{\mathbf{x}_{Tr}^{(k)}(t+1), \mathbf{pl}_{Tr}^{(k)}(t+1) \mid S(t) = S_j, \boldsymbol{\theta}\} \\ &= b_j(\mathbf{x}_{Tr}^{(k)}(t+1) \mid \boldsymbol{\theta}) \oplus \mathbf{pl}_{Tr}^{(k)}(t+1) \end{aligned} \tag{8}$$

for $1 \le i, j \le N$ ($k = 1, 2, ..., K$, $t = 1, 2, ..., T-1$). It can also be verified that

$$\begin{cases} \beta_{S_j}^{(k)}(T) = 1 \quad \text{for } 1 \le j \le N \\ \beta_{S_j}^{(k)}(t+1) = \left[\sum_{i=1}^{N} \left(b(\mathbf{x}_{Tr}^{(k)}(t+1) \mid \boldsymbol{\theta}) \oplus \mathbf{pl}_{Tr}^{(k)}(t+1) \right)[S_j] \times a_{ij} \right] \beta_{S_j}^{(k)}(t) \end{cases} \tag{9}$$

The probability of being in state S_j at time t given $\mathbf{x}_{Tr}^{(k)}(1), ..., \mathbf{x}_{Tr}^{(k)}(t)$, $\mathbf{pl}_{Tr}^{(k)}(1), ..., \mathbf{pl}_{Tr}^{(k)}(t)$ ($k = 1, 2, ..., K$, $t = 1, 2, ..., T$) and the parameter $\boldsymbol{\theta}$, denoted as $\gamma_{S_j}^{(k)}(t)$, can be calculated by

$$\gamma_{S_j}^{(k)}(t) = \frac{\alpha_{S_j}^{(k)}(t) \beta_{S_j}^{(k)}(t)}{\sum_{j=1}^{N} \alpha_{S_j}^{(k)}(t) \beta_{S_j}^{(k)}(t)} \tag{10}$$

and the probability of the *k*th training data being in state S_i at time *t* while in state S_j at time *t*+1, denoted as $\xi_{S_i, S_j}^{(k)}(t)$, can be computed by

$$\xi_{S_i, S_j}^{(k)}(t) = \frac{\alpha_{S_i}^{(k)}(t) a_{ij} \left(b(\mathbf{x}_{Tr}^{(k)}(t+1) \mid \boldsymbol{\theta}) \oplus \mathbf{pl}_{Tr}^{(k)}(t+1) \right)[S_j] \beta_{S_j}^{(k)}(t+1)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_{S_i}^{(k)}(t) a_{ij} \left(b(\mathbf{x}_{Tr}^{(k)}(t+1) \mid \boldsymbol{\theta}) \oplus \mathbf{pl}_{Tr}^{(k)}(t+1) \right)[S_j] \beta_{S_j}^{(k)}(t+1)} \tag{11}$$

After calculating $\xi_{S_i, S_j}^{(k)}(t)$ and $\gamma_{S_j}^{(k)}(t)$ for all training data, the estimate of the initial state probability $\hat{\pi}_i$ ($1 \le i \le N$) can be calculated by

$$\hat{\pi}_i = \frac{\sum_{k=1}^{K} \gamma_{S_j}^{(k)}(1)}{K} \tag{12}$$

The estimate of the one-step transition probability $a_{ij}\,(1 \le i, j \le N)$ is

$$\hat{a}_{ij} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T} \xi_{S_i,S_j}^{(k)}(t)}{\sum_{k=1}^{K}\sum_{t=1}^{T} \gamma_{S_i}^{(k)}(t)} \tag{13}$$

The estimates of the mean value vector and the covariance matrices of the MoGs can be calculated by

$$\hat{\mu}_i = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T} \gamma_{S_i}^{(k)}(t)\mathbf{x}_{Tr}^{(k)}(t)}{\sum_{k=1}^{K}\sum_{t=1}^{T} \gamma_{S_i}^{(k)}(t)} \tag{14}$$

and

$$\hat{\boldsymbol{\Sigma}}_i = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T} \gamma_{S_i}^{(k)}(t)(\mathbf{x}_{Tr}^{(k)}(t) - \hat{\mu}_i)(\mathbf{x}_{Tr}^{(k)}(t) - \hat{\mu}_i)'}{\sum_{k=1}^{K}\sum_{t=1}^{T} \gamma_{S_i}^{(k)}(t)}, \tag{15}$$

respectively. The MoG-EHMM training procedure is summarized in Algorithm 1. The initial value for $\boldsymbol{\mu}_0$ can be set by the *K*-means clustering algorithm, while $\boldsymbol{\pi}_0, \mathbf{A}_0, \boldsymbol{\Sigma}_0$ can be initialized by assuming non-informative knowledge. Convergence of Algorithm 1 is checked by comparing the relative deviation of the maximum log-likelihood between two adjacent iterations to a pre-specified threshold ε, say ε=10^{-6} as used in this work. Note that the forward and backward variables should be normalized in each step to avoid exponentially converging to zero.

Algorithm 1: Parameter estimation of MoG-EHMM $\hat{\boldsymbol{\theta}}$

Require: Initial values of $\hat{\boldsymbol{\theta}}$, denoted as $\boldsymbol{\theta}_0$;
Training data $\mathbf{x}_{Tr}^{(k)}(1),...,\mathbf{x}_{Tr}^{(k)}(T)$;
Expert knowledge $\mathbf{pl}_{Tr}^{(k)}(1),...,\mathbf{pl}_{Tr}^{(k)}(T)$.

Output: The estimated parameters $\hat{\boldsymbol{\theta}}$ of the MoG-EHMM.

1: Set $\boldsymbol{\theta}^{(q)} = \boldsymbol{\theta}_0$; q=1;
2: **For** k=1 to K **do**
3: **For** t=1 to T **do**
4: Calculate $\alpha_{S_j}^{(k)}(t)$, $\beta_{S_j}^{(k)}(t)$, $\gamma_{S_j}^{(k)}(t)$, $\xi_{S_i,S_j}^{(k)}(t)$ by (7),(9)-(11);
5: Normalize $\alpha_{S_j}^{(k)}(t)$ and $\beta_{S_j}^{(k)}(t)$;
6: **End For**;
7: **End For**;
8: Calculate the parameter $\hat{\boldsymbol{\theta}}$=$\boldsymbol{\theta}^{(q+1)}$ by (12)-(15);
9: **If** $|\mathcal{L}(\mathbf{x},\mathbf{pl}\,|\,\boldsymbol{\theta}^{(q+1)}) - \mathcal{L}(\mathbf{x},\mathbf{pl}\,|\,\boldsymbol{\theta}^{(q)})|/\mathcal{L}(\mathbf{x},\mathbf{pl}\,|\,\boldsymbol{\theta}^{(q)}) \le \varepsilon$, $\hat{\boldsymbol{\theta}}$=$\boldsymbol{\theta}^{(q+1)}$
10: $\hat{\boldsymbol{\theta}}$=$\boldsymbol{\theta}^{(q+1)}$; **Break**;
11: **Else** $\boldsymbol{\theta} = \boldsymbol{\theta}^{(q+1)}$; Go to Step 2;
12: **End If**.

C. Health State Inference and Reliability Updating in the Online Phase

In the online phase, the CM information, denoted as $\mathbf{c}_{CM}(t)\,(t=1,2,...,T)$ is collected from a particular individual system of interest. Similar to the offline phase, once $\mathbf{c}_{CM}(t)$ is collected, the HIs $\mathbf{x}_{CM}(t)$ $(t=1,2,...,T)$ are extracted through feature extraction. Expert knowledge can also be elicited in terms of contour functions for the states, denoted by $\mathbf{pl}_{CM}(t_k)$. The two information sources can be merged to estimate the true health state of the system and update reliability estimation based on the trained MoG-EHMM in the offline phase. Let $\mathbf{p}_{CM}(t_k)=[p_{CM,S_1}(t_k), p_{CM,S_2}(t_k),..., p_{CM,S_i}(t_k),..., p_{CM,S_N}(t_k)]$ represents the posterior state probability distribution of the new system updated by $\mathbf{x}_{CM}(t_k)$ and $\mathbf{pl}_{CM}(t_k)$ up to time t_k, that is

$$p_{CM,S_i}(t_k)=\Pr\{S(t_k)=S_i \mid \mathbf{x}_{CM}(t_k),\mathbf{pl}_{CM}(t_k),\hat{\boldsymbol{\theta}}\}. \tag{16}$$

From Bayesian theorem, $p_{CM,S_i}(t_k)$ can be readily computed based on the forward variable using $\mathbf{x}_{CM}(t_k)$ and $\mathbf{pl}_{CM}(t_k)$, and $\hat{\boldsymbol{\theta}}$ as following

$$p_{CM,S_i}(t_k)=\frac{\Pr\{S(t_k)=S_i,\mathbf{x}_{CM}(t_k),\mathbf{pl}_{CM}(t_k)\mid\hat{\boldsymbol{\theta}}\}}{\Pr\{\mathbf{x}_{CM}(t_k),\mathbf{pl}_{CM}(t_k)\mid\hat{\boldsymbol{\theta}}\}}=\frac{\alpha_{S_i}(t_k)}{\sum_{S_i}\alpha_{S_i}(t_k)} \tag{17}$$

Let $S_{MAP}(t_k)$ be the most likely state at time t_k. It can be determined by maximizing the posterior probability $p_{CM,S_i}(t_k)$ (i=1,2,...,N)

$$S_{MAP}(t_k)=\arg\max_{i=1,2,...,N} p_{CM,S_i}(t_k),\quad \forall t_k=1,2,...,T. \tag{18}$$

Similarly, the system reliability can be updated by the posterior probability distribution $\mathbf{p}_{CM}(t_k)$ and the transition probability matrix $\hat{\mathbf{A}}$ estimated in the offline phase.

$$R_S(t')=\sum_{S_i\le S_F}\mathbf{p}_{CM}(t_k)\times\hat{\mathbf{A}}^{(t')}, \tag{19}$$

where t' is the time elapsed after the running time t_k of the specific new system.

D. RUL Prediction

Given the failure threshold state S_F, if $S_{MAP}(t_k) > S_F$, the RUL of the system is definitely zero. Otherwise, let $\bar{\tau}$ denotes the first passage time to the failure state $\{S_j\}$ where $S_j > S_F$

$$\bar{\tau} = \inf\{\tau : S(\tau) > S_F \mid S_{MAP}(t_k) \le S_F\}. \tag{20}$$

Hence, the RUL of the system, i.e., t', is

$$t' = \bar{\tau} - t_k, \tag{21}$$

and the probability mass function of t' can be calculated by

$$q_{t'=t} = \Pr\{t' = t\} = \Pr\{\bar{\tau} = t_k + t\}, \quad t = 1, 2, \ldots. \tag{22}$$

Based on the total probability law, the failure probability at time $t_k + t$ can be decomposed into

$$\begin{aligned}
&\sum_{S_j > S_F} \Pr\{S(t_k + t) = S_j \mid S(t_k) = S_i\} \\
&= \sum_{S_j > S_F} \Pr\{S(t_k + t) = S_j \mid t' = 1\} \underbrace{\Pr\{t' = 1 \mid S(t_k) = S_i\}}_{q_{t'=1}} \\
&+ \sum_{S_j > S_F} \Pr\{S(t_k + t) = S_j \mid t' = 2\} \underbrace{\Pr\{t' = 2 \mid S(t_k) = S_i\}}_{q_{t'=2}}. \\
&+ \cdots \\
&+ \sum_{S_j > S_F} \underbrace{\Pr\{S(t_k + t) = S_j \mid t' = t\}}_{=1} \underbrace{\Pr\{t' = t \mid S(t_k) = S_i\}}_{q_{t'=t}}
\end{aligned} \tag{23}$$

Hence, the probability mass function of RUL t' can be computed recursively

$$\begin{aligned}
q_{t'=t} = &\sum_{S_j > S_F} \Pr\{S(t_k + t) = S_j \mid S(t_k) = S_i\} \\
&- \sum_{l=1}^{t-1} \sum_{S_j > S_F} \Pr\{S(t_k + t) = S_j \mid t' = l\} \times q_{t'=l}
\end{aligned}, \tag{24}$$

where $\sum_{S_j > S_F} \Pr\{S(t_k + t) = S_j \mid S(t_k) = S_i\} = 1 - R_S(t)$. In general, it is difficult to calculate $\Pr\{S(t_k + t) = S_j \mid t' = l\}$ for $1 \le l \le t - 1$, because the recovery from the failure state to a functioning state is possible for repairable systems. In this paper, as we consider only non-repairable systems, (24) reduces to

$$q_{t'=t} = R_S(t-1) - R_S(t). \tag{25}$$

Detailed derivation of (25) can be found in Appendix B provided as supplementary file.

IV. SIMULATION EXPERIMENTS

In this section, simulation experiments are designed to examine the performance of the MoG-EHMMs. We consider a MoG-HMM with four states and 3-dimensional Gaussian emission probability distribution $b(\mathbf{x}(t)\,|\,S(t)) = \mathcal{N}(\mathbf{x}(t)\,|\,\boldsymbol{\mu}, \boldsymbol{\Sigma}_i)$. The parameters of the MoG-HMM are given as follows

$$\mathbf{A} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0.6354 & 0.3646 & 0 \\ 0 & 0 & 0.7565 & 0.2435 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \boldsymbol{\pi} = [1, 0, 0, 0]$$

$$\boldsymbol{\mu} = \begin{bmatrix} 0.0412 & 0.0916 & 0.0579 \\ 0.1176 & 0.1184 & 0.9168 \\ 0.2002 & 0.2634 & 0.8672 \\ 1 & 1 & 0.8446 \end{bmatrix} \quad \Sigma_1 = \begin{bmatrix} 0.0108 & 0.0018 & 0.0007 \\ 0.0018 & 0.0137 & 0.0014 \\ 0.0007 & 0.0014 & 0.0121 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} 0.0111 & 0.0020 & 0.0012 \\ 0.0020 & 0.0134 & 0.0019 \\ 0.0012 & 0.0019 & 0.0137 \end{bmatrix} \quad \Sigma_3 = \Sigma_4 = \begin{bmatrix} 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}$$

To examine the influence of expert knowledge on the health state inference and reliability updating, two different experiments are designed in both offline and online phases. One hundred sequences of training data were generated based on the true values and used in the offline phase to estimate parameters of the MoG-EHMM, while another sample is generated to serve as online CM information.

A. Reliability Updating of the MoG-EHMM

(1) Performance Under Partial Knowledge: We first examine the scenario where partial knowledge is provided by the experts in the offline phase. As defined in [12], if experts have partial knowledge on the true health state, the contour function takes the following form

$$pl_{S_i}(t) = \begin{cases} 1 & \text{If } S_{Expert}^{(k)}(t) = S_{True}^{(k)}(t) = S_i \\ \rho & \text{Otherwise} \end{cases} \tag{26}$$

where $S_{Expert}^{(k)}(t)$ and $S_{True}^{(k)}(t)$ are the expert-believed and true health state, respectively. Algorithm 1 is implemented to estimate the parameters of the MoG-EHMM by fusing the offline expert knowledge and the

training CM information. Equation (19) is, then, used to dynamically update the system reliability at time t_3 by only using the online CM information, and the results are plotted in Fig. 2(a). As shown in Fig. 2(a), the more precise expert knowledge in the offline phase, i.e., when ρ changes from 1 to 0, the closer that the updated reliability is to the true reliability. This is because the transition probability matrix can be more accurately estimated by integrating expert knowledge in the offline phase. In particular, in the worst case of ρ=1, i.e., the expert has non-informative knowledge, the corresponding reliability is the same as that from only using training data.

To investigate the influence of expert knowledge on system reliability updating in the online phase, we assume non-informative knowledge is given by the experts in the offline phase and using Algorithm 1 to train the MoG-EHMM. Figure 2(b) plots the updated reliability at a particular time t_4 by (19). As shown in Fig. 2(b), even with an imprecision level of $\rho = 0.8$, the updated reliability is more accurate than without integrating online expert knowledge. However, the updated reliability is slightly lower than the true reliability values (See the difference between the blue curve and the true one). This is because the transition probability matrix cannot be accurately estimated without precise expert knowledge in the offline phase (as we assumed non-informative knowledge in the offline phase). Hence, it is concluded that integrating expert knowledge in both the offline and online phase is beneficial to the accuracy of system reliability updating.

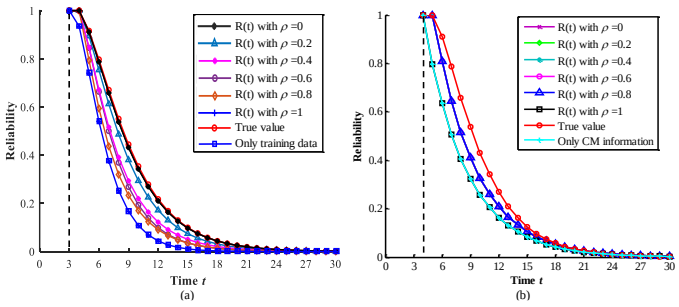


Fig.2. Reliability updating by fusing the expert knowledge. (a) In the offline phase updated at t_3, (b) In the online phase updated at t_4.

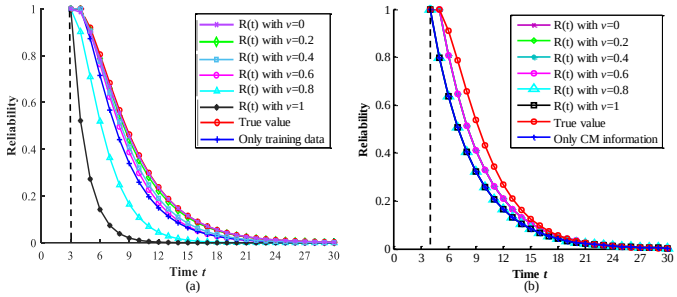


Fig.3. Reliability updating by fusing expert knowledge with noise. (a) In the offline phase updated at t_3, (b) In the online phase updated at t_4.

(2) Performance Under Expert Knowledge with Noise: In practice, the experts may sometimes provide misleading state estimations. To capture such phenomenon, it is assumed that as in (26), with a probability P_{Error}, the experts might give a wrong state estimation, i.e., $S_{Expert}^{(k)}(t) \neq S_{True}^{(k)}(t)$. A Beta distribution with mean value v and variance 0.2 is used to quantify the error probability of expert knowledge with respect to the true states i.e., $P_{Error} \sim Beta(v, 0.2)$. Such a model is called expert knowledge with noise [12]. To study the influence of expert knowledge with noise in the offline phase, Algorithm 1 is implemented to estimate the transition probability matrix. The system reliability is, then, updated by (19) at time t_3, the results are shown in Fig. 3(a). As illustrated in Fig. 3(a), an error probability with mean value v=0.6 can still result in a better reliability estimate than without offline expert knowledge. However, when error probability becomes much greater, e.g., $v = 0.8$ or $v = 1$, the reliability estimates are less accurate than when only using CM information. The reason is the estimated transition probability matrix $\hat{\mathbf{A}}$ tends to underestimate the reliability as it overestimates the transition probability to the failure state, which it is caused by the wrong inputs from the expert knowledge with noise.

To investigate the influence of expert knowledge with noise in the online phase, non-informative knowledge is assumed in the offline phase while expert knowledge with different levels of error probability are considered in the online phase. The system reliability is updated at time t_4 by (19), and the results are illustrated in Fig. 3(b). As shown in Fig. 3(b), even with a mean error probability of $v = 0.6$, the reliability estimate is still better than that of without online expert knowledge with noise. Moreover, with the mean error probability of $v = 0.8$ or $v = 1$, the reliability estimate is no worse than when it was without online expert

knowledge with noise.

B. Health State Inference by MoG-EHMM

To assess the quality of health state inference by the MoG-EHMM, two metrics, namely the adjusted Rand index (ARI) and the improvement ratio (IR), are used to quantify the difference between the true state sequence and the predicted state sequence by the MoG-EHMM. The ARI is a well-known partition performance measure which ranges from 0 to 1, where 0 presents two purely random partitions, and 1 for two identical partitions. In this work, the true health states of the generated sequences are viewed as the baseline of the partition, whereas the state sequences inferred by the MoG-EHMM is used to calculate the ARI.

The IR is proposed to measure the improvement of the health state inference by fusing expert knowledge against the result from only using CM information. It is defined as

$$IR = \frac{N_{CM}^{error} - N_{Expert}^{error}}{N_{CM}^{error}} \tag{27}$$

where N_{CM}^{error} is the number of errors in the health state inference made by using only CM information, N_{Expert}^{error} is the number of errors in the health state inference made by fusing the expert knowledge. If $IR > 0$, it means the health state inference by fusing expert knowledge can improve the quality of the health state inference.

The experiment in Sect. IV.A are repeated 50 times (each time, a new data set is generated for training and testing). The results of ARI and IR are shown in Fig. 4(a) and 4(b), respectively. As shown in Fig. 4(a), the quality of health state inference gradually degrades when the imprecision level ρ drops from precise to non-informative knowledge. Moreover, the mean values of ARI with noise are always no better than that of without noise for offline expert knowledge, indicating that the misleading recommendations have negative impacts on the health state inference. As shown in Fig. 4(b), the IR values with offline expert knowledge are almost positive, which indicates that the offline expert knowledge can improve the health state inference as compared to the case of using only CM information. Similarly, the improvement of health state inference can

still be influenced by the misleading recommendations of experts. Figure 5 shows the ARI and IR values after adding expert knowledge in the online phase. As illustrated in Fig. 5, both of the ARI and IR values start from 1, and then gradually degrade as the value of ρ or v gets larger. It is concluded that the MoG-EHMM with online precise knowledge can correctly infer the health state. Similar as the offline phase, the misleading recommendations by experts in the online phase can negatively affect the quality of health state inference.

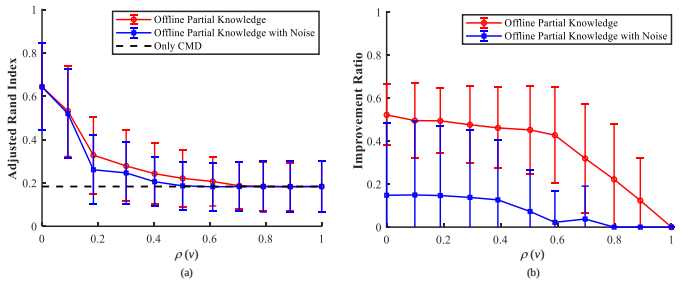


Fig.4. Mean values with one standard deviation of (a) The ARI, (b) The IR, over 50 run times in the offline phase.

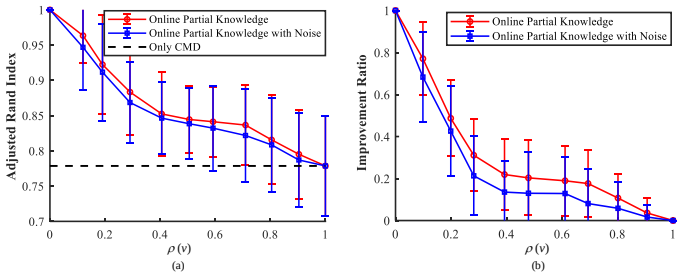


Fig.5. Mean values with one standard deviation of (a) The ARI, (b) The IR, over 50 run times in the online phase.

V. APPLICATION TO NASA BEARING TEST DATA

We further test the performance of the proposed method on a real bearing run-to-failure test dataset provided by NASA prognostics center of excellence [24]. The data acquisition, health indicator construction, health state division and the RUL prediction by the MoG-EHMM are described as follows.

(1) Data Acquisition: The bearing test rig, as shown in Fig. 6, comprises of four test bearings on one shaft. The rotation speed of the shaft was kept constant at 2000rpm. An accelerometer was installed on each bearing housing. Vibration signals were collected every 20 minutes by a National Instruments DAQCard-6062E data acquisition card. The data sampling rate is 20 kHz and the data length is 20480 points [11]. The vibration signals of the four bearings are used to estimate the parameter of the MoG-EHMM in the offline phase.

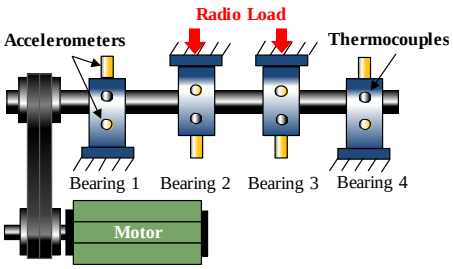


Fig. 6. Bearing test rig [11].

(2) HI Construction: Based on the vibration signals, three features, i.e., the root mean square (RMS), the average power of vibration (APV), and the mean value of vibration (MVV), were extracted via the time-domain feature extraction methods. These three features are selected as they are the most relevant features to the bearing degradation [15],[23]. The extracted health indicators are plotted in Fig. 7.

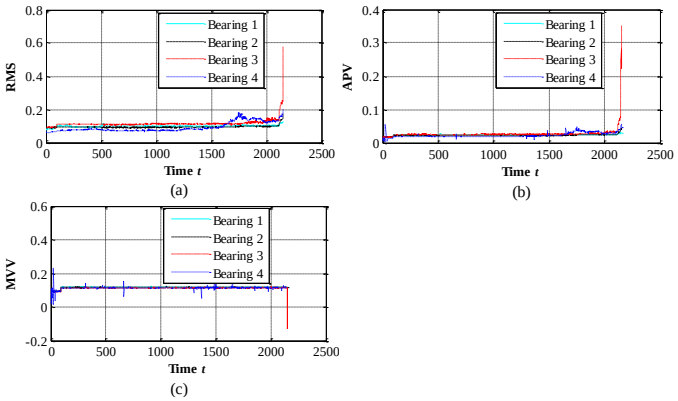


Fig. 7. Extracted features from the four bearings' vibration data. (a) RMS; (b) APV; (c) MVV.

(3) Health State Division: A Gaussian Mixture Model (GMM) is utilized to cluster the degradation data of the three extracted HIs. The number of health states equal to the number of clustering centers, and can be obtained by minimizing the Akaike information criterion (AIC) under different numbers of the components of Gaussians [15]. As shown in Table I, GMM with 4 components of Gaussians has the lowest AIC values. Hence, the degradation of the four bearings can be divided into 4 states, corresponding to the "healthy" state (State 1), and the "moderately worn out" (State 2), the "seriously worn out" (State 3), and the "completely worn out" (State 4). It should be noted that the GMM clustering is a soft partition method which gives the posterior probabilities of all the data belonging to all the clustering centers. In this work, the posterior probabilities are treated as the elicited offline expert knowledge.

TABLE I
AIC UNDER DIFFERENT NUMBERS OF GAUSSIAN COMPONENTS

No. of Gaussians	1	2	3	4	5	6	7	8
AIC($\times 10^5$)	-1.238	-1.250	-1.251	**-1.253**	-1.250	-1.252	-1.252	-1.252

(4) True Model Parameters and CM Data: In this case study, as the sample size is too small, we do not use cross validation to test the results. Alternatively, we use bootstrap method to generate 100 bootstrap samples from the four bearing data. Algorithm 1 is used to estimate the parameters $\hat{\theta}$ of the MoG-EHMM with these bootstrap samples and offline non-informative knowledge. The MoG-EHMM with $\hat{\boldsymbol{\theta}}$ is treated as the true degradation model and used to generate the CM information in the online phase. The generated CM data, which comprises of three signals $x_1(t)$, $x_2(t)$, and $x_3(t)$, is used in subsequent sections for online reliability updating and RUL prediction.

To study the influence of offline expert knowledge on reliability assessment, both the state partitions (also known as the hard label) and the posterior probabilities (also known as the soft label) of the GMM clustering results are used as offline expert knowledge to train the MoG-EHMM by Algorithm 1, where the data from the four bearings in Fig.7 are used as training CM data. The results of reliability assessment obtained by (3) are presented in Fig. 8. In Fig. 8, the true system reliability is calculated by the MoG-EHMM with the true parameter $\hat{\theta}$. As shown in Fig. 8, the soft label of the training data can result in a better reliability estimate than that of the hard label.

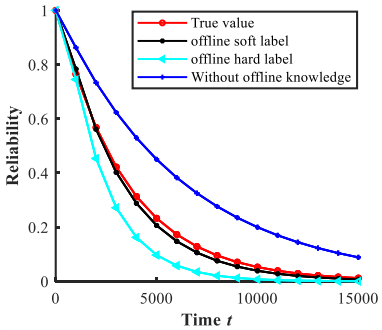


Fig. 8. Reliability estimation of the bearings in the offline phase.

(5) RUL Prediction In The Online Phase: In the online phase, expert knowledge are given at six time

instants, i.e., t=50 (mins), t=81(mins), t=261(mins), t=1205(mins), t=1631(mins), and t=1800(mins). Table II gives the results of the health state inference by the MoG-EHMM with the parameter $\hat{\boldsymbol{\theta}}$. It can be seen that the states at time t=50, 261, 1800 were correctly inferred while the estimations at t=81, 1205, 1631 were inaccurate when using only the CM information. Then, online precise knowledge is given at these time instants. The probability mass distribution (PMF) of the RUL can be calculated via (19) by fusing the precise knowledge, i.e., $pl_{CM,S_{True}(t)}(t)=1$ and $pl_{CM,S_i}(t)=0$ for $S_i \neq S_{True}(t)$. To assess the accuracy of the RUL prediction, a numerical metric based on the Bhattacharyya distance [22] is developed in this work. The smaller the Bhattacharyya distance is, more accurate the RUL prediction is. In this case, the Bhattacharyya distance between the probability mass function of the predicted RUL and the true RUL is presented in Table III. As shown in Table III, all the predicted RULs which uses expert knowledge, no matter the expert knowledge is offline or online, were more accurate than without using expert knowledge. It is noteworthy that when no expert knowledge is used, the model degenerate to the MoG-HMM in [15]. It shows that both the offline knowledge and the online precise knowledge are of benefit to the RUL prediction.

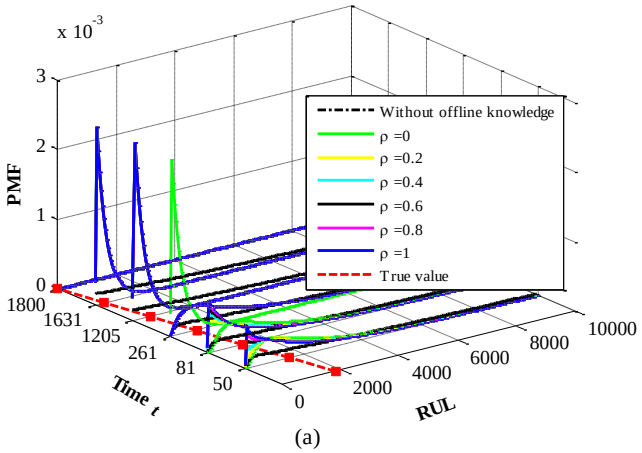
TABLE II
THE RESULTS OF HEALTH STATE INFERENCE

Time t (min)	t=50	t=81	t=261	t=1205	t=1631	t=1800
Only by CM information	1	1	2	2	4	4
True values	1	2	2	3	3	4

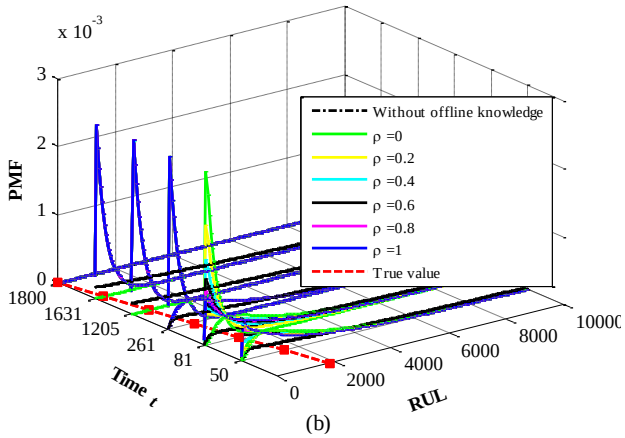
To assess the impact of imprecise expert judgments of the true states of the bearings on the RUL prediction, online partial knowledge at the above six time instants was given in the form of (4). The probability mass distribution (PMF) of the RUL can be calculated via (19) by fusing the expert knowledge from precise knowledge down to non-informative knowledge, i.e., $pl_{CM,S_i}(t)=1$ for all the states. The results are delineated in Fig. 9(a). The Bhattacharyya distance between the probability mass function of the predicted RUL by fusing online partial knowledge and the true RUL is presented in Table IV. As shown in Table IV, the Bhattacharyya distance gradually increases as ρ (indicates the degree of imprecision) increases. However, the RUL prediction with imprecise expert knowledge $0<\rho<1$ is still more accurate than without

using the online expert knowledge ($\rho = 1$). The results indicate that online partial knowledge also contributes to a more accurate RUL prediction.

Furthermore, to assess the impact of incorrect expert judgments of the true states of the bearings on the RUL prediction, the incorrect expert judgments at the six time instants are given as $\mathbf{pl}_{CM}(50) = \rho, 1, \rho, \rho]$, $\mathbf{pl}_{CM}(81) = [\rho, \rho, 1, \rho]$, $\mathbf{pl}_{CM}(261) = [\rho, \rho, 1, \rho]$, $\mathbf{pl}_{CM}(1205) = [\rho, \rho, \rho, 1]$, $\mathbf{pl}_{CM}(1631) = [\rho, \rho, \rho, 1]$, and $\mathbf{pl}_{CM}(1800) = [\rho, \rho, \rho, 1]$ with the imprecision level ρ setting from 0 (corresponding to complete incorrect judgment case) to 1 (corresponding to non-informative knowledge case) with an increment of 0.2. Figure 9(b) illustrates the probability mass functions of the RUL by fusing the incorrect expert judgments. Table V tabulates the Bhattacharyya distances between the probability mass distributions of the RUL by fusing the incorrect expert judgments and their true values. As shown in Table V, incorrect expert judgments of the true states can negatively affect the prediction performance, except for t=81 and $\rho = 0.4$. This is because at time t=81, the inferred state is State 1 if we only use the CM information, while the expert knowledge estimates State 3. Both the two individual information are incorrect since the true state is State 2. Through the DRC for these two incorrect evidences, State 2 has the greatest probability, which coincides the true state at this time instant. Therefore, the incorrect expert judgments of experts under such a scenario can still result in a better RUL prediction.



(a)



(b)

Fig. 9. Probability mass function (PMF) of RUL (a) with online partial knowledge; (b) with online partial knowledge with noise.

TABLE III
COMPARATIVE RESULTS OF RUL PREDICTION WITH THE EXISTING APPROACH

Time t	Bhattacharyya Distance			
	Without any knowledge [15]	With only offline knowledge	With only online precise knowledge	With both offline and online knowledge
t=50	0.038831	0.019387	0.029215	**0.017485**
t=81	0.035810	0.017234	0.020863	**0.013333**
t=261	0.038328	0.017540	0.028739	**0.017240**
t=1205	0.523747	0.315424	0.412371	**0.221881**
t=1631	+∞	+∞	0.541708	**0.221881**
t=1800	**0**	**0**	**0**	**0**

TABLE IV
BHATTACHARYYA DISTANCE OF RUL PREDICTION WITH ONLINE PARTIAL KNOWLEDGE

Time t	Bhattacharyya Distance					
	ρ=0	ρ=0.2	ρ=0.4	ρ=0.6	ρ=0.8	ρ=1
t=50	**0.017485**	0.017945	0.018360	0.018736	0.019077	0.019387
t=81	**0.013333**	0.013471	0.013683	0.014049	0.014811	0.017234
t=261	**0.017240**	0.017300	0.017359	0.017419	0.017479	0.017540
t=1205	**0.221881**	0.305304	0.311330	0.313543	0.314705	0.315424
t=1631	**0.221881**	+∞	+∞	+∞	+∞	+∞
t=1800	**0**	**0**	**0**	**0**	**0**	**0**

TABLE V
BHATTACHARYYA DISTANCE OF RUL PREDICTION WITH ONLINE EXPERT KNOWLEDGE WITH NOISE

Time t	Bhattacharyya Distance					
	ρ=0	ρ=0.2	ρ=0.4	ρ=0.6	ρ=0.8	ρ=1
t=50	0.025276	0.022275	0.020944	0.020196	0.019718	**0.019387**
t=81	1.476586	0.019046	**0.016023**	0.017649	0.017453	0.017234
t=261	1.476586	0.018834	0.018015	0.017749	0.017618	**0.017540**
t=1205	+∞	**0.315424**	**0.315424**	**0.315424**	**0.315424**	**0.315424**
t=1631	+∞	+∞	+∞	+∞	+∞	+∞
t=1800	**0**	**0**	**0**	**0**	**0**	**0**

VI. CONCLUSION

In this paper, a mixture of Gaussians-evidential hidden Markov model (MoG-EHMM) was put forth for RUL prediction by fusing expert knowledge and CM information under the belief function theory framework. In the MoG-EHMM, the MoG was used to characterize the behaviors of multi-dimensional CM information, whereas the expert knowledge was elicited through the contour function. The emission probability calculated

by MoG was fused with expert knowledge via the Dempster's rule of combination. Based on the proposed MoG-EHMM, the RUL prediction was divided into two phases. In the offline phase, the parameters of the MoG-EHMM were estimated by the E^2M algorithm, while in the online phase, the estimated parameters, along with online expert knowledge, were used to infer the health states, update the system reliability, and calculate the probability mass distribution of the RUL for a particular individual system of interests. Simulation results and real case study showed that by introducing the expert knowledge, the performances of reliability assessment, health state inference and RUL prediction can be substantially improved.

ACKNOWLEDGMENT

This work was conducted during a short-term visit of Dr. Tangfan Xiahou to CentraleSupélec. He would like to thank the University of Electronic Science and Technology of China and the Chaire on Risk and Resilience of Complex Systems at CentraleSupélec to jointly finance his visit. Dr. Tangfan Xiahou would like to thank Dr. J. Xing from CentraleSupélec for her help in this work. Dr. Tangfan Xiahou and Professor Yu Liu greatly acknowledge grant support from the National Natural Science Foundation of China under contract numbers 71771039 and 71922006.

REFERENCES

[1] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation–a review on the statistical data driven approaches," *Eur. J. Oper. Res.,* vol. 213, no. 1, pp. 1-14, 2011.
[2] Q. Zhai and Z.-S. Ye, "RUL prediction of deteriorating products using an adaptive Wiener process model," *IEEE Trans. Ind. Inform.,* vol. 13, no. 6, pp. 2911-2921, 2017.
[3] B. Yang, R. Liu, and E. Zio, "Remaining useful life prediction based on a double-convolutional neural network architecture," *IEEE Trans. Ind. Electron.,* vol. 66, no. 12, pp. 9521-9530, 2019.
[4] N. Li, Y. Lei, J. Lin, and S. X. Ding, "An improved exponential model for predicting remaining useful life of rolling element bearings," *IEEE Trans. Ind. Electron.,* vol. 62, no. 12, pp. 7762-7773, 2015.
[5] M. Fan, Z. Zeng, E. Zio, R. Kang, and Y. Chen, "A sequential Bayesian approach for remaining useful life prediction of dependent competing failure processes," *IEEE Trans. Reliab.,* vol. 68, no. 1, pp. 317-329, 2018.
[6] F. Cannarile, P. Baraldi, and E. Zio, "An evidential similarity-based regression method for the prediction of equipment remaining useful life in presence of incomplete degradation trajectories," *Fuzzy Sets Syst.,* vol. 367, pp. 36-50, 2019.
[7] Z. Tian, "An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring," *J. Intell. Manuf.,* vol. 23, no. 2, pp. 227-237, 2012.
[8] M. Xia, T. Li, T. Shu, J. Wan, C. W. De Silva, and Z. Wang, "A two-stage approach for the remaining useful life prediction of bearings using deep neural networks," *IEEE Trans. Ind. Inform.,* vol. 15, no. 6, pp. 3703-3711, 2018.
[9] H. T. Pham, B.-S. Yang, and T. T. Nguyen, "Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine," *Mech. Syst. Signal Proc.,* vol. 32, pp. 320-330, 2012.

[10] 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," *Mech. Syst. Signal Proc.,* vol. 104, pp. 799-834, 2018.
[11] A. Soualhi, G. Clerc, H. Razik, and F. Guillet, "Hidden Markov models for the prediction of impending faults," *IEEE Trans. Ind. Electron.,* vol. 63, no. 5, pp. 3271-3281, 2016.
[12] E. Ramasso and T. Denoeux, "Making use of partial knowledge about hidden states in HMMs: an approach based on belief functions," *IEEE Trans. Fuzzy Syst.,* vol. 22, no. 2, pp. 395-405, 2013.
[13] M. S. Hamada, A. Wilson, C. S. Reese, and H. Martz, *Bayesian reliability*. Springer Science & Business Media, 2008.
[14] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE,* vol. 77, no. 2, pp. 257-286, 1989.
[15] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot, "A data-driven failure prognostics method based on mixture of Gaussians hidden Markov models," *IEEE Trans. Reliab.,* vol. 61, no. 2, pp. 491-503, 2012.
[16] Z. Chen, Y. Li, T. Xia, and E. Pan, "Hidden Markov model with auto-correlated observations for remaining useful life prediction and optimal maintenance policy," *Reliab. Eng. Syst. Saf.,* vol. 184, pp. 123-136, 2019.
[17] O. Geramifard, J.-X. Xu, J.-H. Zhou, and X. Li, "Multimodal hidden Markov model-based approach for tool wear monitoring," *IEEE Trans. Ind. Electron.,* vol. 61, no. 6, pp. 2900-2911, 2013.
[18] T. Xiahou, Y. Liu, and T. Jiang, "Extended composite importance measures for multi-state systems with epistemic uncertainty of state assignment," *Mech. Syst. Signal Proc.,* vol. 109, pp. 305-329, 2018.
[19] W. He, N. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method," *J. Power Sources,* vol. 196, no. 23, pp. 10314-10321, 2011.
[20] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," in *Classic Works of the Dempster-Shafer Theory of Belief Functions*: Springer, 2008, pp. 57-72.
[21] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.
[22] E. Choi and C. Lee, "Feature extraction based on the Bhattacharyya distance," *Pattern Recognit.,* vol. 36, no. 8, pp. 1703-1709, 2003.
[23] J. Xing, Z. Zeng, and E. Zio, "A framework for dynamic risk assessment with condition monitoring data and inspection data," *Reliab. Eng. Syst. Saf.,* p. 106552, 2019.
[24] J. Lee, H. Qiu, G. Yu, and J. Lin, "Bearing Data Set, NASA Ames Prognostics Data Repository," ed, 2007.
[25] T. Denoeux, "Maximum likelihood estimation from uncertain data in the belief function framework," *IEEE Trans. Knowl. Data Eng.,* vol. 25, no. 1, pp. 119-130, 2011.