



HAL
open science

Competing deterioration processes modeling and RUL estimation based on factorial hidden Markov models

Thanh Trung Le, Florent Chatelain, Christophe Bérenguer

► **To cite this version:**

Thanh Trung Le, Florent Chatelain, Christophe Bérenguer. Competing deterioration processes modeling and RUL estimation based on factorial hidden Markov models. MIMAR 2016 - 9th IMA International Conference on Modelling in Industrial Maintenance and Reliability, Institute of Mathematics and its Applications, Jul 2016, Londres, United Kingdom. pp.108-113. hal-01411405

HAL Id: hal-01411405

<https://hal.science/hal-01411405>

Submitted on 7 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Competing deterioration processes modeling and RUL estimation based on Factorial Hidden Markov Models

T.T. LE, F. CHATELAIN and C. BERENGUER

Univ. Grenoble Alpes & CNRS, GIPSA-lab

{thanh-trung.le, florent.chatelain, christophe.berenguer}@gipsa-lab.grenoble-inp.fr

Abstract Deterioration modeling and Remaining Useful Life (RUL) estimation play an important role in implementing a predictive maintenance program since it helps to predict early the failure time caused by an incipient deterioration, and thus to provide sufficient time for maintenance crew to act before the actual failure eventually occurs. One well-identified challenge in deterioration modeling is that equipment does not always deteriorate following only one single mechanism, instead several deterioration modes can co-exist in competition due to the dynamics and variability of operating and environmental conditions. In this paper, we study the phenomenon where different modes co-evolve in parallel within a single component until one of them lead the component to fail. The observations contain hence information of the states of each mode, leading the inference problem being difficult. To deal with this problem, the Factorial Hidden Markov Model (FHMM) and the variational approximation technique are adapted. The associated RUL estimation method is also developed. To evaluate the performance of the proposed approach, a numerical study is given.

1. Introduction

Predictive Maintenance (PM) plays nowadays an important role in maintaining production equipment thanks to its ability to eliminate unnecessary scheduled preventive maintenance operations as well as to avoid unforeseen failures; hence it can significantly reduce the maintenance cost Jardine et al. (2006). Within a PM framework, the RUL estimation is a key enabling task since it could provide sufficient time for maintenance engineers to schedule an intervention or to acquire replacement components before an actual failure. As pointed out in Gorjian et al. (2010), evaluation of the current equipment's health state is an essential task in estimating its RUL in practical applications. To this end, deterioration phenomena of the equipment should be carefully investigated and properly modeled.

In the literature, stochastic models are often studied to model deterioration processes thanks to its capacity of taking into account the stochastic nature of the deterioration phenomena. Depending on the types of its states space, one can divide the stochastic models into two main classes: continuous or discrete. Lévy process Van Noortwijk (2009), general path model Lu & Meeker (1993) are examples of the first class. Regarding the continuous-state processes, we can list here Markovian-based models, such as Hidden Markov Models (HMMs) Rabiner (1989), Hidden semi-Markov models (HsMM) Dong & He (2007). Compared to the first one, the second class models the health states of equipment by a finite number of discrete states which could be easier for maintenance engineers to interpret the obtained results. A detailed review of deterioration models used in reliability analysis can be found in Gorjian et al. (2010).

While modeling the deterioration processes, a well identified difficulty is to deal with the co-existence of multiple modes of deterioration Compare et al. (2015); Le et al. (2016). These different mechanisms, coexisting even within a single component, can lead to different temporal evolution behaviors of deterioration indicators. An example of such phenomenon can be found in a rolling element bearing, an essential component of every rotating machinery in industry. In effect, under operation, point defects such as cracks, pits and spalls appearing on different constituent elements (i.e. outer race, inner race, cage, etc.) can lead to different deterioration dynamics of the bearing Heng (2009). Taking this phenomenon into account is hence crucial both for a comprehensive assessment of the health state of equipment as well as for an accurate RUL estimation. To tackle this problem, the concept of multi-branch models has been proposed recently by the current authors in Le et al. (2016, 2015b). As Markov-based models, the multi-branch methods showed promising results in RUL estimation compared to the ones obtained with the standard "mono-branch" models especially when multiple deterioration modes co-exist. Nevertheless, the works in Le et al. (2016) considered only the modes competition in an incubation period, meaning that once one of them has already initiated, the equipment cannot change

to the others. In Le et al. (2015a), the modes transitions were allowed but the equipment can be only in one of them at a time.

In this paper, we are interested in modeling the deterioration phenomenon where several modes coexist in parallel. Specifically, different deterioration modes evolve independently and all contribute to observed default symptoms. It is sufficient that only one of these modes reaches its own failure state can lead to the breakdown of the equipment. The mixed contribution of the modes at the observation level leads to difficulties in diagnosing the equipment's health state as well as in predicting its RUL. To overcome these problems, we propose in this paper a diagnostics/prognostics framework based on the factorial hidden Markov model (FHMM) Ghahramani & Jordan (1997). The remainder of the paper is organized as follows. Section 2 is devoted to review some mathematical background of the FHMM model and explain how to estimate its model parameters via an approximated method. The application of the FHMM model for the diagnostics and prognostics of a bearing is investigated in Section 3. Finally, conclusion follows in Section 4.

2. Factorial Hidden Markov Model

2.1. Model structure

The Factorial Hidden Markov Model (FHMM) is an extension of the HMM one with distributed state representation which is a particular class of probabilistic graphical model Ghahramani & Jordan (1997). The model consists of several branches in parallel in which each one can be considered as a single Markov chain. At any instants, the distributed states all contribute to the observations as shown in Figure 1.

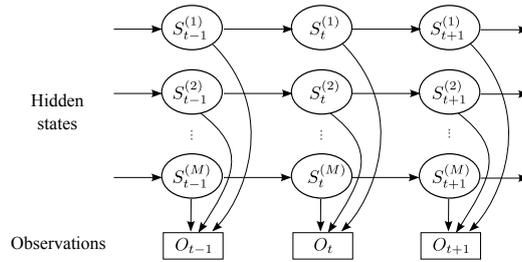


Figure 1. Factorial Hidden Markov Model

Similar to the standard HMM models, states transitions within a branch m of the FHMM model is characterized by an initial state distribution $\pi^{(m)}$ and a state transition matrix $P^{(m)}$ Rabiner (1989) where the superscript is used to denote the corresponding branch in this paper. A combination of the states in all constituent branches at a given time t is called the meta-state, that is $S_t = (S_t^{(1)}, S_t^{(2)}, \dots, S_t^{(M)})$ where M is the number of branches. Furthermore, the states within the branch m can take one of $K^{(m)}$ discrete values, i.e. $S_t^{(m)} \in \{1, 2, \dots, K^{(m)}\}$. For the seek of simplicity, we assume in this paper that $K^{(m)} = K, \forall m$. The generation to the case of different $K^{(m)}$ values is straightforward Ghahramani & Jordan (1997).

In the FHMM model, the observation emitted at a given time depends on the individual states within the branches at that time. Its density can be modeled as a function of the ones of the branch states:

$$p(O_t | S_t) = f \left[P(O_t | S_t^{(1)}), P(O_t | S_t^{(2)}), \dots, P(O_t | S_t^{(M)}) \right] \quad (1)$$

The precise nature of the function f depends on the proportions to which the observations from the individual components are mixed in the current one. The contribution of each component could depend on several factors and is difficult to be determined precisely Reyes-Gomez et al. (2003). For continuous observations, a simple form for this dependence is linear Gaussian Ghahramani & Jordan (1997). That is, the observation O_t is a Gaussian random vector whose mean is a linear function of the component ones:

$$\mu_t = \sum_{m=1}^M w_{\mu}^{(m)} \mu_t^{(m)} \quad (2)$$

where $\mu_t^{(m)}$ represents the D dimensional mean vector of the Gaussian distribution corresponding to the component state $S_t^{(m)}$ and $w_{\mu}^{(m)}$ is a $D \times D$ weighting matrix.

Concerning the covariance parameter, there exists two main possibilities in the literature : global or composed covariance approaches Reyes-Gomez et al. (2003). It should be noted that the composed covariance could lead the model learning procedure being more complex than in the global covariance case since it introduces more free parameters to be estimated. For the seek of simplicity, only the global covariance matrix is considered in this paper, meaning that all factorial states have in common a covariance matrix C .

2.2. Parameters estimation

Denote $\Theta = \left\{ \pi^{(m)}, P^{(m)}, w_{\mu}^{(m)}, C \right\}_{m=1:M}$ the parameters vector of the FHMM model. Since the states are hidden, the standard expectation-maximization (EM) algorithm Dempster et al. (1977) appears to be a suitable tool for parameters estimation. However, as pointed out in Ghahramani & Jordan (1997), the expectation (E) step is computationally intractable in this case. Therefore we adapt in this section the variational approximation technique proposed by Ghahramani & Jordan (1997) to overcome this problem.

The basic idea of the variational technique is to approximate the posterior distribution over the hidden variables $P(S_{1:T} | O_{1:T})$ by a tractable distribution $Q(S_{1:T})$ where $S_{1:T}$ and $O_{1:T}$ denote the sequence of meta-states and observations from time 1 to T . More specifically, at any k th iteration of the standard EM algorithm, the Jensen's inequality can be applied in computing the expected complete data log-likelihood as follows:

$$\begin{aligned} \mathcal{Q}(\Theta | \Theta^{k-1}) &= \log \sum_{S_{1:T}} P(S_{1:T}, O_{1:T} | \Theta) = \log \sum_{S_{1:T}} Q(S_{1:T}) \left[\frac{P(S_{1:T}, O_{1:T} | \Theta)}{Q(S_{1:T})} \right] \\ &\geq \sum_{S_{1:T}} Q(S_{1:T}) \log \left[\frac{P(S_{1:T}, O_{1:T} | \Theta)}{Q(S_{1:T})} \right] = \mathcal{F}(Q, \Theta) \end{aligned} \quad (3)$$

This inequality provides a lower bound on the log-likelihood that can be used to obtain an efficient learning algorithm. The difference between \mathcal{Q} and \mathcal{F} is given by the Kullback-Leibler divergence:

$$KL(Q||P) = \sum_{S_{1:T}} Q(S_{1:T}) \log \left[\frac{Q(S_{1:T})}{P(S_{1:T} | O_{1:T}, \Theta^{new})} \right] \quad (4)$$

By minimizing this divergence, the log-likelihood is assured to increase until it reaches a (local) maximum.

It should be noted that the function Q must be chosen to be tractable. Depending on the form of this function, two variational techniques were proposed: completely factorized and structured variational inference Ghahramani & Jordan (1997). Compared to the first one which factors the posterior probability such that all the state variables are statistically independent, the second technique can preserve more the probabilistic structure of the original FHMM model and hence will be used in this paper.

Based on this approximation technique, the parameters vector Θ is computed in the maximization (M) step. More details about the calculation formulas can be found in Ghahramani & Jordan (1997).

2.3. State estimation and RUL prediction

Suppose that we have trained the FHMM model in an offline phase from the historical until-failure data. We move now to an online phase where the deterioration signals of a monitored system are acquired in real time via a condition monitoring system. The objective is to estimate the actual states of all the deterioration modes at a given time and then predict the system's RUL.

The estimation of the deterioration states given the observations in case of the FHMM model is itself a complex task since it can be referred as the source separation problem found in the speech recognition community Reyes-Gomez et al. (2003). Several methods have been developed in the literature for accomplishing this problem. One of the most simple solutions is to transform the FHMM model into a meta-HMM one with K^M "meta-states" and then applied the Viterbi algorithm Rabiner (1989) in a standard way. Although this method leads to more computation time compared to another ones, its performance is still acceptable in several practical applications since only the maximization is retained after each recursive iteration.

This method allows us to estimate the most likely sequence of meta-state given a sequence of observations. The corresponding state sequences of each individual branch of the FHMM model is then obtained by decomposing it into a base-M system. After that, the last state of each decomposed sequence can be considered as the

current health state of the deterioration mode corresponding to that sequence. Given the estimated states, the RUL can be defined as follows:

$$RUL = \min_k RUL^{(k)} \quad (5)$$

where $RUL^{(k)}$ is the remaining time for the mode k to reach the final state from its current state. In other words, $RUL^{(k)}$ is referred as the RUL of a standard HMM model corresponding to the k th branch of the FHMM model. The RUL probability density function (pdf) calculation for a standard HMM model for the discrete time case can be found in Le et al. (2016).

Given the individual RUL density functions, the cumulative distribution function (cdf) of the overall RUL of the FHMM model can be obtained by:

$$CDF_{RUL}(x) = 1 - \mathbb{P}\left(RUL^{(1)} > x, RUL^{(2)} > x, \dots, RUL^{(M)} > x\right) = 1 - \prod_{k=1}^M \left(CDF_{RUL}^{(k)}(x)\right) \quad (6)$$

3. Numerical study

In this section, the diagnostics and prognostics performances of the FHMM model are evaluated through the case of a defected rolling element bearing with multiple point defects. The bearing is chosen because it is an essential component for all rotating machinery found in industry. Firstly, we simulate vibration signals generated by a bearing in case that several point defects coexist within it. The kurtosis features are then extracted and used as observations for training the FHMM model. The RUL estimation performance of the learned FHMM model is compared with the one obtained by a standard HMM model.

3.1. Multiple point defects simulation with severity evolution

McFadden & Smith (1985) proposed a simple mathematical model for simulating the vibration signals produced by a bearing with multiple point defects. By assuming that the bearing operates under constant radial load, the vibration response $v(t)$ can be modeled as the product of repetitious impulses $d(t)$ at the fault frequency with the distribution of load $q(t)$, convolved with an exponential decay $h(t)$ due to damping:

$$v(t) = [d(t) \cdot q(t)] * h(t) + n(t) \quad (7)$$

where the notation $*$ represents the convolution operation and $n(t)$ are the noises added to corrupt the signal.

Figure 2a shows the simulated raw vibration signals of the bearing in case that two point defects co-exist, one is on the inner race and the other on the outer race. The co-existence of these two defects can be verified on the frequency spectrum obtained through an envelope analysis (c.f. Figure 2b).

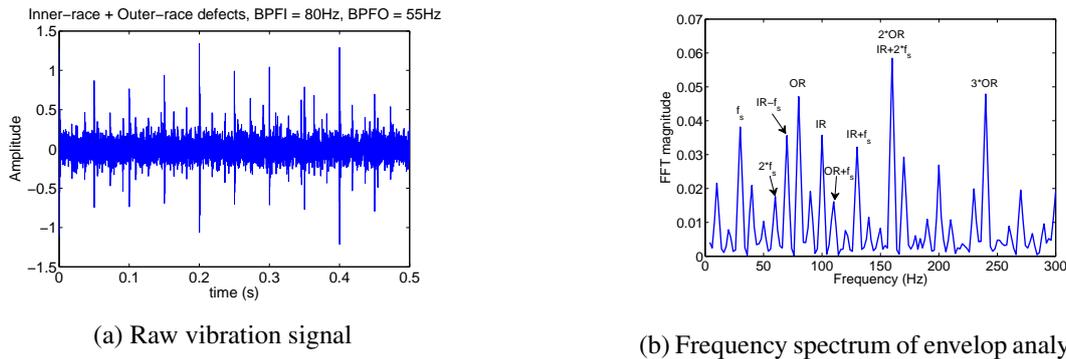


Figure 2. Multiple point defects bearing simulation

This model although represents well the vibration signals of a defected bearing, it does not however take into account the temporal evolution of the defaults. In the literature, vibration data are usually acquired periodically and temporal features such as the kurtosis are extracted and plotted in order to study the evolution trend of fault severity over time for the bearing prognostics purpose. To simulate deterioration processes, the above defective bearing vibration signals are repeatedly generated with progressive impulse amplitudes $d(t)$. Each repetition corresponds to one data acquisition time in practice. In this study, we suppose that the impulse amplitudes $d(t)$

generated by each point defect increase over time following a homogeneous Gamma process. A random delay following a Poisson distribution is also added in order to represent the bias between the appearance times of the two defects. The kurtosis feature is used to represent the overall defect severity of the bearing. Finally, the data generation is stopped once the extracted kurtosis value reaches for the first time a critical threshold $L = 10$. Beyond this boundary, the bearing is considered to be failed.

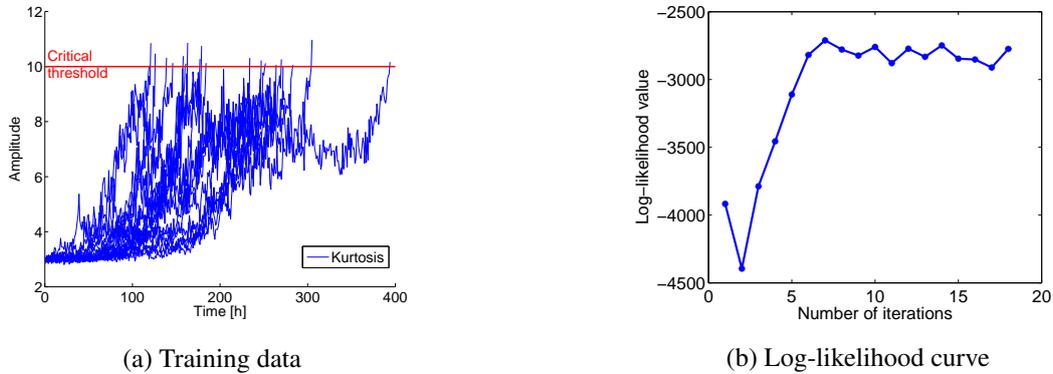


Figure 3. Training data and learned log-likelihood curve

By repeating this procedure several times, different kurtosis evolution curves can be obtained as represented in Figure 3a. These values are considered as the training data for estimating the parameters of the FHMM model. Figure 3b shows the convergence of the log-likelihood curve which demonstrates the effectiveness of the variational approximation technique in training the FHMM model.

3.2. FHMM vs HMM

In studying the FHMM model, an interesting question is to compare it with a “standard” HMM model in term of RUL estimation performance. To this end, we firstly train the two models by the same training data set. The branches of the FHMM model are supposed to have the same number of discrete states as in the HMM one. This number can be determined by using the Bayesian Information Criterion (BIC) as presented in Le et al. (2016).

The next step is to use the two learned models to estimate the RUL of same test data set. The root mean squared errors (RMSE) is used for the comparison purpose. Since the prognostics performance depends on the instant at which the RUL estimation is carried out, we further define a critical level at which the prediction is triggered for all the test cases. In other words, the diagnostic and prognostic tasks for every test cases are started once its corresponding kurtosis feature reaches this level for the first time. Figure 4 shows the RMSE values obtained by the two models as the function of the triggered level.

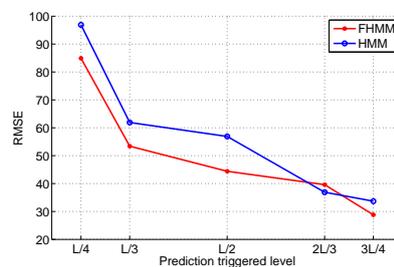


Figure 4. RMSE values at different prediction triggered levels

We can see that the RUL estimation errors for the both models decrease with the increment of the triggered level. This is due to the fact that the latter the prediction is performed, the more information become available and hence leading to the better RUL estimation results. Furthermore, compared to the standard HMM model, the FHMM one shows the better performance in terms of RMSE criterion, especially when one want to trigger the prediction earlier. This can be explained by the fact that the FHMM model helps to take into account

the coexistence of different deterioration modes at early stage of deterioration process. More lately, one of the deterioration modes may already dominate the others, leading to the similar RMSE results between the FHMM and the HMM models.

4. Conclusion

The present paper proposes the use of the factorial Hidden Markov Model (FHMM) in order to deal with the deterioration modeling problem in case that several competing modes co-exist within a single component. The variational approximation technique is adapted to overcome the model parameters learning problems. In terms of diagnostics, this model helps to evaluate the actual deterioration states within every existed modes. The RUL of the component is then estimated as the minimum time for one of these modes to reach its own failure state. The numerical results show that by taking into account the coexistence of several deterioration modes, the FHMM gives better performance in RUL estimation in comparison with the traditional HMM model.

Future research will be focused on the extension of the FHMM model to represent the dependency between the different deterioration modes. The proposed method should also be validated on real world systems data.

References

- M. Compare, et al. (2015). 'Interacting multiple-models, state augmented Particle Filtering for fault diagnostics'. *Probabilistic Engineering Mechanics* **40**:12–24.
- A. P. Dempster, et al. (1977). 'Maximum likelihood from incomplete data via the EM algorithm'. *Journal of the royal statistical society. Series B (methodological)* pp. 1–38.
- M. Dong & D. He (2007). 'A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology'. *Mechanical Systems and Signal Processing* **21**(5):2248–2266.
- Z. Ghahramani & M. I. Jordan (1997). 'Factorial hidden Markov models'. *Machine learning* **29**(2-3):245–273.
- N. Gorjian, et al. (2010). 'A review on degradation models in reliability analysis'. In *Engineering Asset Lifecycle Management*, pp. 369–384. Springer.
- A. S. Y. Heng (2009). *Intelligent prognostics of machinery health utilising suspended condition monitoring data*. Ph.D. thesis, Queensland University of Technology.
- A. K. Jardine, et al. (2006). 'A review on machinery diagnostics and prognostics implementing condition-based maintenance'. *Mechanical systems and signal processing* **20**(7):1483–1510.
- T. T. Le, et al. (2015a). 'Prognosis based on Multi-branch Hidden semi-Markov Models: A case study'. In *Proc. Fault Detection, Supervision and Safety of Technical Processes Conference, Paris, France, September 2015*, pp. 91–96.
- T. T. Le, et al. (2015b). 'Jump Markov Linear Systems for deterioration modeling and Remaining Useful Life estimation'. In *Proc. European Safety and Reliability Conference, Zurich, Switzerland, September 2015*, pp. 1287–1295.
- T. T. Le, et al. (2016). 'Multi-branch hidden Markov models for remaining useful life estimation of systems under multiple deterioration modes'. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*.
- C. J. Lu & W. O. Meeker (1993). 'Using degradation measures to estimate a time-to-failure distribution'. *Technometrics* **35**(2):161–174.
- P. McFadden & J. Smith (1985). 'The vibration produced by multiple point defects in a rolling element bearing'. *Journal of sound and vibration* **98**(2):263–273.
- L. R. Rabiner (1989). 'A tutorial on hidden Markov models and selected applications in speech recognition'. *Proceedings of the IEEE* **77**(2):257–286.
- M. J. Reyes-Gomez, et al. (2003). 'Multi-channel source separation by factorial HMMs'. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, vol. 1, pp. I–664. IEEE.
- J. Van Noortwijk (2009). 'A survey of the application of gamma processes in maintenance'. *Reliability Engineering & System Safety* **94**(1):2–21.