



**HAL**  
open science

# Making use of partially observed states in Markov switching autoregressive models: application to machine health diagnosis

Fatoumata Dama, Christine Sinoquet

## ► To cite this version:

Fatoumata Dama, Christine Sinoquet. Making use of partially observed states in Markov switching autoregressive models: application to machine health diagnosis. CNIA 2021 : Conférence Nationale en Intelligence Artificielle, Jun 2021, Bordeaux (en ligne), France. pp 14-21. hal-03321148

**HAL Id: hal-03321148**

**<https://hal.science/hal-03321148v1>**

Submitted on 17 Aug 2021

**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.

# Making use of partially observed states in Markov switching autoregressive models: application to machine health diagnosis

F. Dama<sup>1</sup>, C. Sinoquet<sup>1</sup>

<sup>1</sup> LS2N / UMR CNRS 6004, Nantes University, France

{fatoumata.dama, christine.sinoquet}@univ-nantes.fr

## Résumé

*Le diagnostic de bon fonctionnement des machines est une tâche fondamentale dédiée à la surveillance des systèmes dans un souci de sécurité, de prévention des accidents et de programmation des opérations de maintenance. Cette tâche est généralement accomplie grâce à des méthodes d'analyse de données (recourant par exemple à des modèles statistiques). Ces méthodes sont appliquées aux séries temporelles décrivant la dynamique du système analysé. Ces séries temporelles sont sujettes à des changements de régimes reflétant les changements d'état du système. Dans cet article, nous proposons de modéliser de telles séries temporelles par un nouveau modèle de changement de régimes Markovien appelé PHMC-LAR (Partially Hidden Markov Chain AutoRegressive model) où le processus des états décrit l'état de santé du système à chaque pas de temps. Ce modèle possède la capacité d'inclure des connaissances partielles sur le processus des états Markovien. La connaissance partielle est traduite par la présence d'états observés à certains pas de temps aléatoires. Les paramètres de notre modèle sont estimés grâce une variante de l'algorithme d'Espérance-Maximisation, que nous avons développée. La procédure d'inférence des états cachés consiste à identifier la séquence d'états la plus probable pour une série temporelle donnée; elle est réalisée au moyen de l'algorithme de Viterbi. Les études expérimentales ont été conduites sur des données décrivant des états de machine de façon réaliste. Les résultats montrent que, pour les données utilisées, l'intégration de connaissances partielles sur le processus des états améliore considérablement les performances de l'inférence.*

## Mots-clés

*Séries temporelles, modèle autorégressif, modèle à changement de régimes, états partiellement observés, chaîne de Markov, inférence, diagnostic de l'état d'une machine.*

## Abstract

*Machine health diagnosis is a fundamental task dedicated to monitor systems' safety in order to prevent incidents, and to program maintenance operations. Such diagnosis is achieved through analyzing the system's features (generally recorded by sensors and depicted by time series), using data analytics methods such as statistical models. These time se-*

*ries are subject to regime switches reflecting changes in the system health conditions. In this paper, we propose to model such time series by a new Markov switching autoregressive model called PHMC-LAR (Partially Hidden Markov Chain AutoRegressive), where the state process depicts system health condition at each time-step. This model has the particularity to include partial knowledge about the Markovian state process. This partial knowledge is depicted by the states observed at some (random) time-steps. The parameters of our model are learnt through a variant of the Expectation-Maximization algorithm, which we developed. The inference procedure, that consists in segmenting a given time series into the most likely sequence of states, is addressed by the Viterbi algorithm. Experimental studies are performed on realistic machine condition data. The results show that, for the used datasets, the incorporation of partial knowledge substantially improves inference performance.*

## Keywords

*Time series analysis, autoregressive model, regime-switching model, Markov chain, inference, machine health diagnosis, CMAPSS datasets*

## 1 Introduction

Time series subject to switches in regime have been widely studied in domains such as econometry, finance or meteorology. The underlying dynamical system of such time series is associated with a state process which specifies the system behavior, in other words, its functioning mode at each time-step. Thus, two dynamics can be highlighted : (i) transitions from one state to another, which drive the global nonlinear dynamics of the system and (ii) local stationary dynamics of the time series that unfold within the regimes. The former dynamics is usually modelled through Markov models (HMMs). Linear autoregressive models are widely-used to capture the latter dynamics [5], hence the name of Markov switching autoregressive models.

Two categories of models are depicted in the literature. The first category considers observed states, and corresponds to a classical Markov mixture model with an autoregressive dynamics. The second category considers hidden states. In this case, it is usual to rely on Hidden Markov Models (HMMs) [12, 2, 8]. Models belonging to the first category are referred to as **observed regime-switching models** (OR-

SARs) [3]. The second category describes **hidden regime-switching models** (HRSARs) [9].

In this paper, we present a novel regime-switching autoregressive model which implements the intermediate case where states are partially observed, *i.e.* they are known at some random time-steps and hidden at other time-steps. This model, referred to as the Partially Hidden Markov Chain Linear AutoRegressive (PHMC-LAR), capitalizes on the observed states while the hidden states are inferred. PHMC-LAR is a unification of ORSAR and HRSAR models when the state process is a Markov Chain. The contributions reported in this paper are two-fold : (i) design of a PHMC-LAR learning algorithm, and (ii) analysis of the ability of PHMC-LAR to infer hidden states on time series generated by NASA' Commercial Modular Aero-Propulsion System Simulation (CMAPSS) model [18].

This paper is organized as follows. The PHMC-LAR model is presented in Section 2. Section 3 describes a learning algorithm that allows to estimate the model's parameters. Section 4 presents a hidden state inference algorithm. Section 5 depicts the application of the PHMC-LAR model to real-world datasets, and focuses on state inference. The last Section concludes our paper.

## 2 PHMC-LAR model

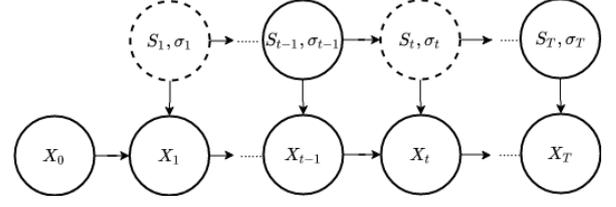
Let  $\{X_t\}_{t \in \mathbb{Z}}$  be a time series subject to regime-switching. Let  $\{S_t\}_{t \in \mathbb{N}^*}$  the state process of  $\{X_t\}$  where  $S_t \in \mathbf{K} = \{1, 2, \dots, K\}$  stands for the state at time-step  $t$  and  $K$  is the number of states. We suppose that  $\{S_t\}$  can be observed at some random time-steps. Let denote by  $\sigma_t \subseteq \mathbf{K}$  the set of possible states at time-step  $t$  with  $\sigma_t = \mathbf{K}$  when  $S_t$  is latent and  $\sigma_t = \{k\}$  when  $S_t$  is observed to be state  $k$ . Thus, the  $\sigma_t$ 's represent precise partial knowledge available about  $\{S_t\}$ .

The following notations will be used hereinafter :

- Symbol ' $:=$ ' stands for the *definition symbol*.
- $A_{t_1}^{t_2}$  denotes the sequence  $A_{t_1}, A_{t_1+1}, \dots, A_{t_2-1}, A_{t_2}$ , where  $t_1$  and  $t_2$  ( $> t_1$ ) are time-steps.  $A_{t_1}^{t_2}$  may be a sequence of any kind (*e.g.*, a sequence of random variables, a sequence of observed values, a sequence of annotations on a time series).
- By convention,  $X_{1-p}^0$  denotes the  $p$  first variables of process  $\{X_t\}$ ; that is why  $\{X_t\}$  is indexed in  $\mathbb{Z}$ .
- Symbols in bold are used to indicate nonscalar variables or vectors of constants.

In this work, the dynamics of the time series  $\{X_t\}$  is captured by a PHMC-LAR model. This model takes into account the partial knowledge about the state process  $\{S_t\}$  and operates in two stages as follows.

- **Modelling of state process.** In the PHMC-LAR, we use available partial knowledge ( $\sigma_t$ 's) to model the state process  $\{S_t\}$  through a *Partially* Hidden Markov Chain (PHMC) [19, 17]. PHMC is an extension of the classical HMM of order 1 [11, 12] in which some states have been observed. PHMC, as classical HMC, is parametrized by a transition



**FIGURE 1** Conditional independence graph of the PHMC-LAR model when the autoregressive order  $p = 1$ . Observed variables (time series and observed states) are displayed in continued line and hidden states are represented in dash line. Note that if the state  $k$  has been observed at time-step  $t - 1$  then  $\sigma_t$ , the set of possible states at this time-step, is reduced to a singleton (*e.g.*,  $\sigma_{t-1} = \{k\}$ )

matrix

$$a_{i,j} = P(S_t = j | S_{t-1} = i), \quad a_{i,j} \in [0, 1], \quad \sum_{j=1}^K a_{i,j} = 1$$

and a stationary law

$$\pi_i = P(S_1 = i), \quad \pi_i \in [0, 1], \quad \sum_{i=1}^K \pi_i = 1.$$

- **Modelling of the dynamics given the state.** Knowing  $S_t$ , the state-value at time-step  $t$ , together with past values of  $X_t$ , we model  $X_t$  relying on a linear autoregressive model (LAR) defined as follows :

$$X_t | X_{t-p}^{t-1}, S_t = k := \mu_k + \sum_{i=1}^p \rho_{i,k} X_{t-i} + v_k \epsilon_t, \quad (1)$$

where  $\{\epsilon_t\}$  are white noises,  $p$  is the number of past values of  $X_t$  to be used in modelling,  $k$  is the state at time-step  $t$ . Within each state  $k$ ,  $\mu_k$  is the intercept,  $(\rho_{1,k}, \dots, \rho_{p,k})$  are the  $p$  autoregressive coefficients and  $v_k$  is the standard deviation. Note that Eq. 1 is not defined for the  $p$  initial values denoted by  $X_{1-p}^0$  which are modelled by some initial law  $g_0(x_{1-p}^0; \psi)$  parametrized by  $\psi$ .

Thus, the PHMC-LAR model is parametrized by  $(\theta, \psi)$  where  $\theta = (\theta^{(S)}, \theta^{(X)})$  with  $\theta^{(S)} = ((\pi_i)_{i=1, \dots, K}, (a_{i,j})_{i,j=1, \dots, K})$ ,  $\theta^{(X)} = (\theta^{(X,k)})_{k=1, \dots, K}$  and  $\theta^{(X,k)} = (\mu_k, \rho_{1,k}, \dots, \rho_{p,k}, v_k)$ . Figure 1 shows the conditional independence graph of the PHMC-LAR model.

## 3 Parameter learning

In this section, an approximation of the maximum likelihood estimate (MLE) of the PHMC-LAR model parameters is presented. We consider a training data set of  $N$  independent time series  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}$ , with  $\mathbf{x}_0^{(1)}, \dots, \mathbf{x}_0^{(N)}$  the corresponding initial values and  $\Sigma^{(1)}, \dots, \Sigma^{(N)}$  (where  $\Sigma^{(i)} = \sigma_{t=1}^{T_i}$ ) the partial knowledge available on the state processes. Notice that the  $\mathbf{x}^{(i)}$ 's have respective lengths  $T_1, \dots, T_N$ , that are not necessarily equal, and that the  $\mathbf{x}_0^{(i)}$  terms have length  $p$  where  $p$  is the autoregressive order.

MLE is the set of parameters that maximizes the likelihood of the training data. It is well-known that in models with latent variables as PHMC-LAR, MLE computation results in an untractable problem. In this case, it

is usual to maximize the **expectation (w.r.t. the latent variables) of the complete data likelihood** noted  $\mathcal{L}^c$ .  $\mathcal{L}^c$  denotes the evidence/likelihood of the training data  $\mathbf{x}_0 = (\mathbf{x}_0^{(1)}, \dots, \mathbf{x}_0^{(N)})$  and  $\mathbf{x} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)})$  when latent/hidden variables are set at  $\mathbf{s} = (\mathbf{s}^{(1)}, \dots, \mathbf{s}^{(N)})$ . The  $\mathcal{L}^c$  writes

$$\begin{aligned} \mathcal{L}^c(\boldsymbol{\theta}, \boldsymbol{\psi}) &= P(\mathbf{X}_0 = \mathbf{x}_0, \mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s}; \boldsymbol{\theta}, \boldsymbol{\psi}) \\ &= P(\mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s} | \mathbf{X}_0 = \mathbf{x}_0; \boldsymbol{\theta}) \times P(\mathbf{X}_0 = \mathbf{x}_0; \boldsymbol{\psi}) \\ &= \mathcal{L}_c^c(\boldsymbol{\theta}) \times \prod_{i=1}^N g_0(\mathbf{x}_0^{(i)}; \boldsymbol{\psi}), \end{aligned} \quad (2)$$

with  $\mathcal{L}_c^c$  the **conditional complete data likelihood** and  $g_0$  the initial law of  $X_t$ .

Since  $\mathbf{s}$  is unknown (it can take any value in  $\mathbf{K}^{(\sum_i T_i)}$  where  $\mathbf{K}$  is the set of possible states), the expectation of  $\mathcal{L}^c$  with respect to the *posterior distribution* of  $\mathbf{S}$  is considered. In Eq. 2 the second term does not depend on  $\mathbf{s}$ . Therefore, it can be taken out of the expectation.

$$\begin{aligned} \mathbb{E}_{P(\mathbf{S}=\mathbf{s} | \mathbf{X}, \mathbf{X}_0, \boldsymbol{\Sigma}; \boldsymbol{\theta})}[\mathcal{L}^c(\boldsymbol{\theta}, \boldsymbol{\psi})] &= \mathbb{E}_{P(\mathbf{S}=\mathbf{s} | \mathbf{X}, \mathbf{X}_0, \boldsymbol{\Sigma}; \boldsymbol{\theta})}[\mathcal{L}_c^c(\boldsymbol{\theta})] \\ &\quad \times \prod_{i=1}^N g_0(\mathbf{x}_0^{(i)}; \boldsymbol{\psi}). \end{aligned}$$

By taking the logarithm of the previous expectation we obtain

$$\hat{\boldsymbol{\psi}} = \arg \max_{\boldsymbol{\psi}} \sum_{i=1}^N \ln g_0(\mathbf{x}_0^{(i)}; \boldsymbol{\psi}). \quad (3)$$

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \ln \left( \mathbb{E}_{P(\mathbf{S}=\mathbf{s} | \mathbf{X}, \mathbf{X}_0, \boldsymbol{\Sigma}; \boldsymbol{\theta})}[\mathcal{L}_c^c(\boldsymbol{\theta})] \right), \quad (4)$$

where  $P(\mathbf{S} = \mathbf{s} | \mathbf{X}, \mathbf{X}_0, \boldsymbol{\Sigma}; \boldsymbol{\theta})$  is the *posterior distribution* of  $\mathbf{S}$  and  $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}^{(1)}, \dots, \boldsymbol{\Sigma}^{(N)})$ .

When  $g_0$  is assumed to belong to a family of parametric distributions, Eq. 3 can be easily solved. For instance, supposing  $g_0$  is a multivariate normal distribution  $\mathcal{N}_p(\mathbf{m}, \mathbf{V})$  with mean  $\mathbf{m} \in \mathbb{R}^p$ , variance-covariance matrix  $\mathbf{V} \in \mathbb{R}^p \times \mathbb{R}^p$ , and  $\boldsymbol{\psi} = (\mathbf{m}, \mathbf{V})$ , we can show that

$$\hat{\mathbf{m}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_0^{(i)}, \quad \hat{\mathbf{V}} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_0^{(i)} - \hat{\mathbf{m}})(\mathbf{x}_0^{(i)} - \hat{\mathbf{m}})', \quad (5)$$

where  $'$  stands for matrix transposition.

In contrast, maximization in Eq. 4 is untractable. Therefore  $\boldsymbol{\theta}$  is estimated by an instance of the Expectation-Maximization (EM) algorithm. The main idea behind EM is to maximize a lower bound of  $\ln \left( \mathbb{E}_{P(\mathbf{S}=\mathbf{s} | \mathbf{X}, \mathbf{X}_0, \boldsymbol{\Sigma}; \boldsymbol{\theta})}[\mathcal{L}_c^c(\boldsymbol{\theta})] \right)$ . This lower bound, denoted by  $Q$  and defined in Eq. 6, is obtained by applying Jensen's inequality [6]. EM is an iterative algorithm that alternates between E(xpectation) step and M(aximization) step. The E-step computes  $Q$ . Then  $Q$  is maximized in the M-step. At iteration  $n$ , we obtain

$$\text{E-step.} \quad Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{n-1}) = \mathbb{E}_{P(\mathbf{S}=\mathbf{s} | \mathbf{X}, \mathbf{X}_0, \boldsymbol{\Sigma}; \hat{\boldsymbol{\theta}}_{n-1})}[\ln \mathcal{L}_c^c(\boldsymbol{\theta})], \quad (6)$$

$$\text{M-step.} \quad \hat{\boldsymbol{\theta}}_n = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{n-1}), \quad (7)$$

with  $\hat{\boldsymbol{\theta}}_{n-1}$  the estimated parameters at iteration  $n-1$  and  $P(\mathbf{S} = \mathbf{s} | \mathbf{X}, \mathbf{X}_0, \boldsymbol{\Sigma}; \hat{\boldsymbol{\theta}}_{n-1})$  the associated *posterior distribution* of  $\mathbf{S}$ .

**Step E of EM.** This step consists in computing the expectation  $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{n-1})$ . Following the conditional independence graph of the PHMC-LAR model (Fig. 1), the conditional complete data likelihood  $\mathcal{L}_c^c$  can be written as a product of marginal and conditional probabilities. We recall that  $S_t$  only depends on  $S_{t-1}$ , and that  $X_t$  depends on  $S_t$  and  $X_{t-p}^{t-1}$ , hence the two conditional probabilities exhibited in Eq. 8.

$$\begin{aligned} \mathcal{L}_c^c(\boldsymbol{\theta}) &= \prod_{i=1}^N P(\mathbf{X}^{(i)} = \mathbf{x}^{(i)}, \mathbf{S}^{(i)} = \mathbf{s}^{(i)} | \mathbf{X}_0^{(i)}; \boldsymbol{\theta}) \\ &= \prod_{i=1}^N \left[ P(S_1^{(i)} = s_1^{(i)}; \boldsymbol{\theta}^{(S)}) \times \right. \\ &\quad \prod_{t=2}^{T_i} P(S_t^{(i)} = s_t^{(i)} | S_{t-1}^{(i)} = s_{t-1}^{(i)}; \boldsymbol{\theta}^{(S)}) \times \\ &\quad \left. \prod_{t=1}^{T_i} P(X_t^{(i)} = x_t^{(i)} | [X^{(i)}]_{t-p}^{t-1} = [x^{(i)}]_{t-p}^{t-1}, \right. \\ &\quad \left. S_t^{(i)} = s_t^{(i)}; \boldsymbol{\theta}^{(X, s_t^{(i)})} \right), \end{aligned} \quad (8)$$

with  $\boldsymbol{\theta}^{(X, k)}$  the parameters of LAR process associated with state  $k$ , and  $P(X_t^{(i)} = x_t^{(i)} | [X^{(i)}]_{t-p}^{t-1} = [x^{(i)}]_{t-p}^{t-1}, S_t^{(i)} = k; \boldsymbol{\theta}^{(X, k)})$  the conditional law of  $X_t^{(i)}$  within  $k$ .

When the expectation in Eq. 6 is developed and  $\mathcal{L}_c^c(\boldsymbol{\theta})$  is substituted by its expression (Eq. 8), we show that  $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{n-1})$  only depends on the following probabilities

$$\begin{aligned} \xi_t^{(i)}(k, l) &= P(S_{t-1}^{(i)} = k, S_t^{(i)} = l | [X^{(i)}]_{1-p}^{T_i} = [x^{(i)}]_{1-p}^{T_i}, \\ &\quad \boldsymbol{\Sigma}^{(i)}; \hat{\boldsymbol{\theta}}_{n-1}), \quad \text{for } t = 2, \dots, T_i, \quad 1 \leq k, l \leq K. \end{aligned} \quad (9)$$

$$\begin{aligned} \gamma_t^{(i)}(l) &= P(S_t^{(i)} = l | [X^{(i)}]_{1-p}^{T_i} = [x^{(i)}]_{1-p}^{T_i}, \boldsymbol{\Sigma}^{(i)}; \hat{\boldsymbol{\theta}}_{n-1}), \\ &\quad \text{for } t = 2, \dots, T_i, \quad 1 \leq l \leq K. \end{aligned} \quad (10)$$

The  $\xi_t^{(i)}$  terms will be used to estimate the transition matrix. Intuitively, for  $(k, l)$  fixed,  $\sum_t \xi_t^{(i)}(k, l)$  represents the frequency of occurrence of the two consecutive states  $k, l$ . The  $\gamma_t^{(i)}$  terms are called **smoothed marginal probabilities**; they will be involved in the estimation of parameters  $\boldsymbol{\theta}^{(X)}$ .

The previous probabilities can be computed through an extension of the *forward-backward* algorithm designed to generate the MLE estimates for HMMs [1]. This extension, called *backward-forward-backward*, adds a supplementary

step to the original algorithm in which available partial knowledge about the state process is exploited. Further details about this algorithm can be found in our previous work [4].

**Step M of EM.** At iteration  $n$ , this step consists in maximizing  $Q(\theta, \hat{\theta}_{n-1})$  with respect to parameters  $\theta = (\theta^{(S)}, \theta^{(X)})$ .  $Q(\theta, \hat{\theta}_{n-1})$  can be decomposed as follows

$$Q(\theta, \hat{\theta}_{n-1}) = Q_S(\theta^{(S)}, \hat{\theta}_{n-1}) + Q_X(\theta^{(X)}, \hat{\theta}_{n-1}),$$

where  $Q_S$  (respectively  $Q_X$ ) only depends on parameters  $\theta_S$  (respectively  $\theta_X$ ). Thus, the M-step can be split in two maximization sub-steps defined as

$$\hat{\theta}_n^{(S)} = \arg \max_{\theta^{(S)}} Q_S(\theta^{(S)}, \hat{\theta}_{n-1}), \quad (11)$$

$$\hat{\theta}_n^{(X)} = \arg \max_{\theta^{(X)}} Q_X(\theta^{(X)}, \hat{\theta}_{n-1}). \quad (12)$$

On the one hand, cancelling the first derivative of  $Q_S(\theta^{(S)}, \hat{\theta}_{n-1})$  provides the analytical expressions of  $\hat{\theta}_n^{(S)}$

$$\hat{a}_{k,l}^{(n)} = \frac{\sum_{i=1}^N \sum_{t=2}^{T_i} \xi_t^{(i)}(k, l)}{\sum_{i=1}^N \sum_{t=1}^{T_i} \gamma_t^{(i)}(k)}, \quad \hat{\pi}_l^{(n)} = \frac{\sum_{i=1}^N \gamma_1^{(i)}(l)}{N}, \quad (13)$$

for  $1 \leq k, l \leq K$ , with

$$\begin{aligned} \gamma_1^{(i)}(s) &= P(S_1^{(i)} = s | [X^{(i)}]_{1-p}^{T_i} = [x^{(i)}]_{1-p}^{T_i}, \Sigma^{(i)}; \hat{\theta}_{n-1}) \\ &= \sum_{j=1}^K \xi_2^{(i)}(s, j). \end{aligned}$$

On the other hand, generally, it is difficult to derive the analytical expression of  $\hat{\theta}_n^{(X)}$ . This is the reason why we are compelled to use a numerical optimization method (*e.g.*, the quasi-Newton method) to maximize  $Q_X(\theta^{(X)}, \hat{\theta}_{n-1})$ .

To set the starting point of the EM algorithm, we first run instances of EM using several vectors of  $p$  initial values chosen at random. The MLE parameters that provide the greatest likelihood across these multiple restarts constitute the starting point.

## 4 Inference or time series segmentation

Let  $\hat{\theta}$  a PHMC-LAR model trained on a partially labelled training dataset. Let  $\mathbf{x} = x_1^T$  be an observed time series and  $\mathbf{x}_0 = x_{1-p}^0$  the associated initial values. As in classical HMMs, the state process of  $\mathbf{x}$  is unknown and inference consists in finding the most likely state sequence  $\mathbf{s}^* = (s_1^*, \dots, s_T^*)$  given  $\mathbf{x}$  and  $\mathbf{x}_0$ . This is equivalent to maximizing, with respect to  $\mathbf{s}$ , the joint probability of  $\mathbf{s} = (s_1, \dots, s_T)$  and  $\mathbf{x}$ , given  $\mathbf{x}_0$

$$\begin{aligned} \mathbf{s}^* &= \arg \max_{\mathbf{s}} P(\mathbf{S} = \mathbf{s}, \mathbf{X} = \mathbf{x} | \mathbf{X}_0 = \mathbf{x}_0; \hat{\theta}) \\ &= \arg \max_{\mathbf{s}} P(\mathbf{S} = \mathbf{s} | \mathbf{X} = \mathbf{x}, \mathbf{X}_0 = \mathbf{x}_0; \hat{\theta}). \end{aligned} \quad (14)$$

Note indeed that  $P(\mathbf{S} = \mathbf{s}, \mathbf{X} = \mathbf{x} | \mathbf{X}_0 = \mathbf{x}_0; \hat{\theta}) = P(\mathbf{S} = \mathbf{s} | \mathbf{X} = \mathbf{x}, \mathbf{X}_0 = \mathbf{x}_0; \hat{\theta}) \times P(\mathbf{X} = \mathbf{x} | \mathbf{X}_0 = \mathbf{x}_0; \hat{\theta})$ , and that the second term in the product does not depend on  $\mathbf{S}$ .

Since state sequence  $\mathbf{s}$  can take  $K^T$  different values where  $K$  is the number of possible states, the "greedy search" method that consists in testing all possible values is extremely costly ( $\mathcal{O}(K^T)$  operations). Alternatively, the **Viterbi algorithm** [7] allows to compute the optimal state sequence in  $\mathcal{O}(TK^2)$  operations.

The Viterbi algorithm operates iteratively, following a dynamic programming algorithm. Let  $\delta_t(l; \hat{\theta})$  the maximal likelihood of subsequence  $(s_1, \dots, s_t = l)$  that ends within state  $l$

$$\begin{aligned} \delta_t(l; \hat{\theta}) &= \max_{s_1, \dots, s_{t-1}} P(X_1^t = x_1^t, S_1^{t-1} = s_1^{t-1}, S_t = l | \\ &\quad \mathbf{X}_0 = \mathbf{x}_0; \hat{\theta}), \quad \text{for } t = 1, 2, \dots, T. \end{aligned} \quad (15)$$

These probabilities can be iteratively computed as follows

$$\begin{aligned} \delta_1(l; \hat{\theta}) &= P(X_1 = x_1 | \mathbf{X}_0 = \mathbf{x}_0, S_1 = l; \theta^{(X,l)}) \times \\ &\quad P(S_1 = l; \hat{\theta}^{(S)}). \end{aligned} \quad (16)$$

$$\begin{aligned} \delta_t(l; \hat{\theta}) &= \max_k \left[ \delta_{t-1}(k; \hat{\theta}) \times P(S_t = l | S_{t-1} = k; \hat{\theta}^{(S)}) \right] \\ &\quad \times P(X_t = x_t | X_{t-p}^{t-1} = x_{t-p}^{t-1}, S_t = l; \theta^{(X,l)}), \end{aligned} \quad (17)$$

for  $t = 2, \dots, T$ .

Therefore, the maximal probability of the complete state sequence is given by  $\max_l \delta_T(l; \hat{\theta})$ . Accordingly, the optimal sequence  $\mathbf{s}^*$  is retrieved by backtracking as follows

$$\mathbf{s}_t^* = \arg \max_l \begin{cases} \delta_T(l; \hat{\theta}) & \text{for } t = T \\ \delta_t(l; \hat{\theta}) \times \hat{a}_{l, s_{t+1}^*} & \text{for } t = T-1, \dots, 1 \end{cases} \quad (18)$$

## 5 Application to machine health diagnosis

In this section, we assess the added value of using partial knowledge about the state process of Hidden Markov Chains. This evaluation is carried out on realistic machine condition data generated by the Commercial Modular Aero-Propulsion System Simulation (CMAPSS) model [18] developed at the NASA Army Research Laboratory. Data description is provided in subsection 5.1. Subsection 5.2 presents the feature extraction procedure used in our experiments. Then the generation of ground truth segmentations is explained in subsection 5.3. Subsection 5.4 describes the experimental setting. Finally, results are presented and discussed in Subsection 5.5.

### 5.1 Data description

The CMAPSS model allows to simulate realistic run-to-failure trajectories of aircraft turbofan engines, under dif-

ferent operational conditions and fault modes. For each simulation, the system begins in healthy state (normal functioning mode) then, at some point, the system health starts degrading and finishes by breaking down (system failure). Besides, each trajectory is described through 3 operational conditions (velocity, altitude and temperature) depicting flight conditions, and 21 time series representing as many system's features.

The NASA dataset repository provides four CMAPSS datasets, with different operational conditions and fault modes (<https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#turbofan>). Each such dataset consists of a training dataset and test dataset. The training datasets are composed of run-to-failure trajectories. In contrast, trajectories within test datasets are stopped before system failure. In this work, we consider the respective training datasets of CMAPSS datasets #1 and #3 (namely *train\_FD001.txt* and *train\_FD003.txt* in the repository). These sets will be further referred to as #1 and #3, to simplify. Both latter sets have a single operational condition, one fault mode for dataset #1, two fault modes for dataset #3 and 100 trajectories each.

## 5.2 Feature extraction : machine health indicator

Our model cannot be directly applied to CMAPSS trajectories which are multivariate time series of dimension 21. To overcome this limitation, we reduce the dimension of our data by aggregating relevant features into a single variable called **health indicator** (HI) [13, 15]. HI is a useful indicator of the system health, computed using the following theoretical model

$$HI_i(t) \equiv 1 - \exp\left(\frac{\log(0.05)}{0.95 T_i} \times t\right), \quad t \in [\sigma_1, \sigma_2], \quad (19)$$

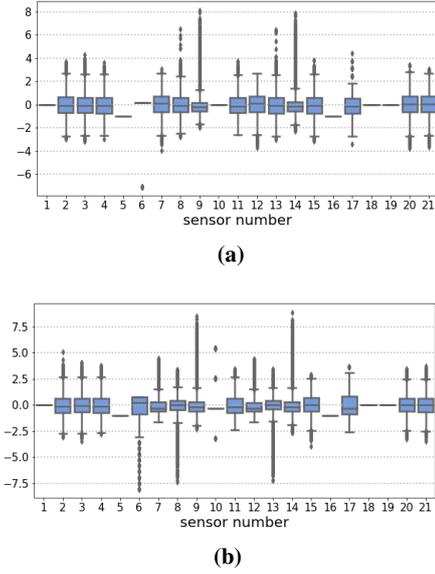
where  $T_i$  is the  $i^{th}$  trajectory length and  $\sigma_1$  and  $\sigma_2$  are respectively set at  $T_i \times 5\%$  and  $T_i \times 95\%$  as proposed in [13]. Note that the theoretical HI roughly decreases from 1 ("healthy") to 0 ("faulty") when  $t$  increases (notice that the term within the exponential is negative).

Then for each trajectory, HI is modelled by a linear regression model for which the predictors are system's features and the response variable is the theoretical HI (Eq. 19) :

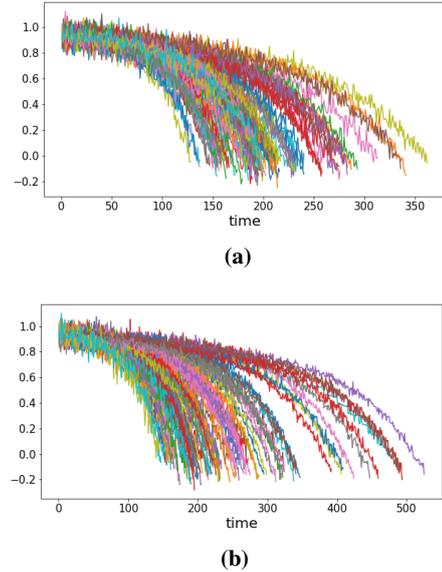
$$HI_i(t) = \eta_0^{(i)} + \sum_{j=1}^q \eta_j^{(i)} y_{t,j}^{(i)} + \delta_t, \quad (20)$$

where  $\mathbf{y}_t^{(i)} = (y_{t,1}^{(i)}, \dots, y_{t,q}^{(i)})$  is the feature vector of the  $i^{th}$  trajectory at time-step  $t$  and  $\boldsymbol{\eta}^{(i)} = (\eta_0^{(i)}, \dots, \eta_q^{(i)})$  are the model parameters which can be estimated by the least square method.  $\delta_t$ 's are independent error terms and  $HI_i(t)$  is defined in Eq. 19.

Once parameters  $\boldsymbol{\eta}^{(i)}$  have been estimated, HI approximations, denoted by  $\hat{HI}_i(\mathbf{y}_t^{(i)}, \hat{\boldsymbol{\eta}}^{(i)})$  are computed following the linear model (Eq. 20). In our experiments, the subset of features  $\{2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 20, 21\}$  that display significant variations over time (as illustrated



**FIGURE 2** Distribution of sensor measurements for the 100 trajectories within (a) training dataset #1, (b) training dataset #3. Data have been standardized in order to show the same scale order across all 21 sensors

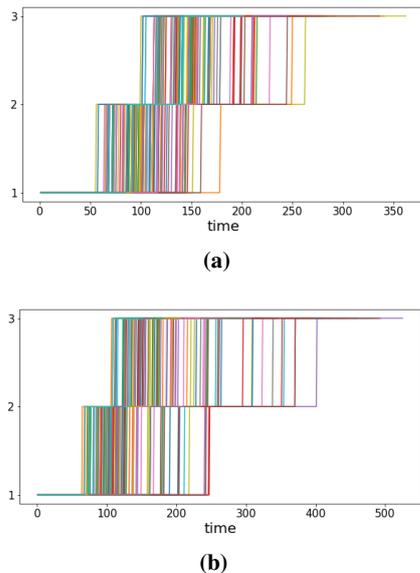


**FIGURE 3** Estimated health indicator for the 100 trajectories within (a) training dataset #1, (b) training dataset #3

in Fig. 2) is considered for both datasets #1 and #3. For dataset #3, feature 6 is added to this latter subset (see Fig. 2b). Figure 3 shows the estimated HI for the 100 trajectories within each of training datasets #1 and #3.

## 5.3 Ground truth segmentation

To note, in CMAPSS trajectories, system degradation levels or health states are not specified. However, we need a "ground truth" segmentation of trajectories to both feed

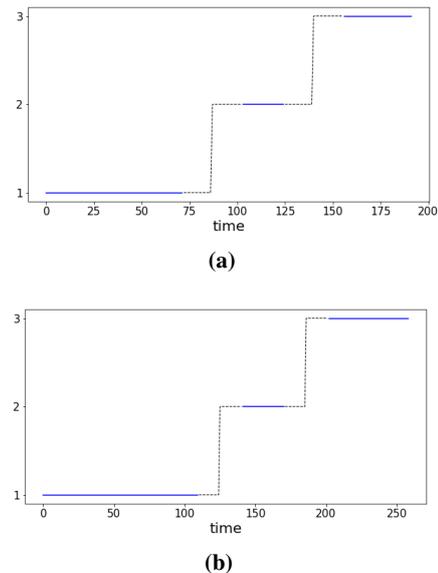


**FIGURE 4** Ground truth segmentation for the 100 trajectories within (a) training dataset #1, (b) training dataset #3. Three health states or degradation levels are considered : 1 for "healthy", 2 for "intermediate" and 3 for "faulty". Each trajectory is assigned a specific color

PHMC-LAR with partial knowledge and validate our models. Three health states or degradation levels are considered : 1 for "healthy", 2 for "intermediate" and 3 for "faulty". For each trajectory, the "ground truth" segmentation is obtained by deterministically splitting the corresponding health indicator time series (see Fig. 3) into 3 regimes ("healthy", "intermediate" and "faulty") where a regime is a succession of time-steps having the same health state. In this work, the automatic segmentation method proposed in [14, 16] and used by [10] has been considered. This method relies on linear regression models. Figure 4 shows the "ground truth" segmentation of the 100 trajectories within training datasets #1 and #3. Note the high variability of regime duration from one trajectory to another. This suggests the difficulty for a model to infer the critical moments where system health degrades.

## 5.4 Experimental setting

In our experiments, trajectories within file *train\_FD001.txt* and *train\_FD003.txt* were split into a training dataset (60 trajectories), a validation and test datasets (20 trajectories each). Gaussian white noises were considered and the initial law  $g_0$  was Gaussian too. The number of states  $K$  was set to 3 and a grid of values  $\{1, \dots, 6\}$  was tested for the autoregressive order  $p$ . For each candidate model, inference performance was evaluated on the validation dataset, relying on the "ground truth" segmentations (see Fig. 4), and using the mean percentage error (MPE) score. Then, the model yielding the highest inference performance was identified. Finally, the global accuracy for health state inference was



**FIGURE 5** Partial knowledge about state process of the 13<sup>th</sup> trajectory within (a) training dataset #1, (b) training dataset #3. Two windows of length 31 centered around the switch from regime 1 to regime 2 and regime 2 to regime 3 respectively are considered. The observations outside these windows are labelled (solid line), whereas dash line denotes hidden states (within the windows)

assessed by computing the confusion matrix for each test dataset.

We remind that we wish to assess the added value of using partial knowledge about the state process of Hidden Markov Chains. Therefore, two modalities are considered : (i) the fully unsupervised case, in which no partial knowledge is included, referred to as Hidden Markov Chain Linear Autoregressive model (HMC-LAR); and (ii) the semi-supervised case denoted by PHMC-LAR. Both HMC-LAR and PHMC-LAR models are fed with health indicator data (see Fig. 3). Note that in (ii), partial knowledge is obtained from the "ground truth segmentations" (see Fig. 4) of the 60 trajectories of previous training datasets as subsequently described. Two windows of length 31 centered around the time-steps at which the system switches from one regime to another are considered; the observations outside these windows are labelled (see Figure 5). Note that these windows represent 22% to 48% of the training trajectories' lengths in dataset #1 against 12% to 39% in dataset #3. The readers' attention is drawn to the fact that the trajectories within the validation and test datasets are kept fully unlabelled.

## 5.5 Results

Table 1 displays the MPE (Mean Percentage Error) values computed on validation dataset, for different values of the autoregressive order. The results show that in the fully unsupervised tuning (HMC-LAR), the highest inference performance (reflected by lowest MPE) is obtained when  $p = 3$  for dataset #1 and  $p = 5$  for dataset #3. However, more

parsimonious models are selected when partial knowledge about state process is included, since for both datasets the best model has an autoregressive order equal to one ( $p = 1$ ). Note that for both datasets, we observe that PHMC-LAR outperforms HMC-LAR with an inference accuracy five times greater.

For the best HMC-LAR models ( $p = 3$  for dataset #1 and  $p = 5$  for dataset #3), confusion matrices resulting from the inference on test trajectories are presented in Table 2. Notice that low inference accuracies are obtained for both datasets : 39% for dataset #1 and 45% for dataset #3. The confusion matrices show that HMC-LAR models globally fail in identifying the three regimes since too many "healthy" states are inferred as "faulty" and reversely.

Considering the best PHMC-LAR models ( $p = 1$  for both datasets), confusion matrices are presented in Table 3. Unlike the fully unsupervised cases (HMC-LAR models), satisfying global accuracies are reached : 88% for dataset #1 and 89% dataset #3. Moreover, no confusion is made between the "healthy" and "faulty" regimes. An analysis of the transition matrices of PHMC-LAR models shows that the system health state never steps back : in other words, transitions "intermediate"  $\rightarrow$  "healthy", "faulty"  $\rightarrow$  "intermediate" and "faulty"  $\rightarrow$  "healthy" have a null probability. Thus, the terms located above the diagonal in the confusion matrices represent anticipations of system health degradation (*i.e.*, the anticipation of the transition from a health condition to a worse one), whereas those located below the diagonal depict delays in the detection of system health degradation. Therefore, almost all inference errors are due to anticipations of system health degradation since the confusion matrices are almost upper triangular. It has to be underlined that in the machine health monitoring literature, health degradation anticipation is a desirable behavior in comparison with models that detect health degradation with some delay. In a sense, this anticipation allows to prevent serious incidents and to program maintenance operations.

| $p$ | CMAPSS dataset #1               |                                  | CMAPSS dataset #3                 |                                  |
|-----|---------------------------------|----------------------------------|-----------------------------------|----------------------------------|
|     | HMC-LAR                         | PHMC-LAR                         | HMC-LAR                           | PHMC-LAR                         |
| 1   | 68 $\pm$ 9 %                    | <b>10 <math>\pm</math> 5.4 %</b> | 75 $\pm$ 7.4 %                    | <b>12 <math>\pm</math> 4.4 %</b> |
| 2   | 69 $\pm$ 4.2 %                  | 12 $\pm$ 8.1 %                   | 73.9 $\pm$ 8.2 %                  | 21 $\pm$ 13 %                    |
| 3   | <b>59 <math>\pm</math> 10 %</b> | 16 $\pm$ 12.5 %                  | 73.7 $\pm$ 8.2 %                  | 25 $\pm$ 13 %                    |
| 4   | 71 $\pm$ 6.8 %                  | 18 $\pm$ 11.2 %                  | 72 $\pm$ 9.5 %                    | 52 $\pm$ 5.1 %                   |
| 5   | 73 $\pm$ 1.8 %                  | 26 $\pm$ 17 %                    | <b>60 <math>\pm</math> 18.5 %</b> | 52 $\pm$ 5.2 %                   |
| 6   | 64 $\pm$ 12 %                   | 55 $\pm$ 4.4 %                   | 73 $\pm$ 5.7 %                    | 35 $\pm$ 12 %                    |

**TABLE 1** Mean percentage inference error (MPE) computed on validation datasets (20 trajectories of different lengths) for both unsupervised case (HMC-LAR) and semi-supervised case (PHMC-LAR).  $p$  is the autoregressive order. The minimum MPE values are displayed in bold

| CMAPSS dataset #1 |              |             |              |            |
|-------------------|--------------|-------------|--------------|------------|
|                   |              | prediction  |              |            |
|                   |              | healthy     | intermediate | faulty     |
| ground truth      | healthy      | <b>1288</b> | 1            | 825        |
|                   | intermediate | 697         | <b>0</b>     | 437        |
|                   | faulty       | 745         | 4            | <b>436</b> |

| CMAPSS dataset #3 |              |             |              |            |
|-------------------|--------------|-------------|--------------|------------|
|                   |              | prediction  |              |            |
|                   |              | healthy     | intermediate | faulty     |
| ground truth      | healthy      | <b>1551</b> | 271          | 625        |
|                   | intermediate | 755         | <b>162</b>   | 242        |
|                   | faulty       | 607         | 183          | <b>445</b> |

**TABLE 2** Confusion matrices of test datasets (20 trajectories of different lengths) for the best HMC-LAR model. For dataset #1 :  $p = 3$  with a global accuracy equal to 39%. For dataset #3 :  $p = 5$  with a global accuracy equal to 45%

| CMAPSS dataset #1 |              |             |              |             |
|-------------------|--------------|-------------|--------------|-------------|
|                   |              | prediction  |              |             |
|                   |              | healthy     | intermediate | faulty      |
| ground truth      | healthy      | <b>1964</b> | 190          | 0           |
|                   | intermediate | 9           | <b>815</b>   | 310         |
|                   | faulty       | 0           | 0            | <b>1185</b> |

| CMAPSS dataset #3 |              |             |              |             |
|-------------------|--------------|-------------|--------------|-------------|
|                   |              | prediction  |              |             |
|                   |              | healthy     | intermediate | faulty      |
| ground truth      | healthy      | <b>2234</b> | 293          | 0           |
|                   | intermediate | 0           | <b>914</b>   | 245         |
|                   | faulty       | 0           | 7            | <b>1228</b> |

**TABLE 3** Confusion matrices of test datasets (20 trajectories) for the best PHMC-LAR model :  $p = 1$  for both datasets, with a global accuracy equal to 88% for dataset #1 and 89% for dataset #3

## 6 Conclusion

In this work, we have presented a new Markov switching autoregressive model which incorporates partial knowledge about the state process. This partial knowledge is represented by the states observed at some random time-steps. This model, referred to as PHMC-LAR (for Partially Hidden Markov Chain Linear AutoRegressive) model, is a generalization of the observed regime-switching models (ORSARs) and the hidden regime-switching models (HRSARs).

In the evaluation, the inference performance of PHMC-LAR model has been compared to that of the fully unsupervised HMC-LAR (Hidden Markov Chain Linear AutoRegressive) model. To this end, realistic machine condition

data available in NASA's CMAPSS datasets has been used. Three health states have been considered ("healthy", "intermediate" and "faulty") and "ground truth segmentation" has been derived. For both models, six values of the autoregressive order  $p$  (from 1 to 6) have been tested. The results show that parsimonious models (reflected by small values of  $p$ ) are selected in PHMC-LAR model. Moreover, PHMC-LAR substantially outperforms the fully unsupervised HMC-LAR model. A further analysis of the confusion matrices computed on test datasets shows that PHMC-LAR is more able to anticipate the degradation of system health than HMC-LAR. Such anticipation capability is a desirable behavior in the literature of machine health diagnosis.

In future work, the multivariate extension of PHMC-LAR will be considered. Such extension will allow to directly model the relevant features of CMAPSS trajectories without any dimension reduction. On the other hand, following a case-based reasoning approach, PHMC-LAR can be adapted to failure prognostic task, which consists in predicting the remaining useful life, that is the number of time-steps remaining before system failure.

## Acknowledgements

The software development and the realization of the experiments were performed at the CCIPL (Centre de Calcul Intensif des Pays de la Loire, Nantes, France).

## References

- [1] L.E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The annals of mathematical statistics*, 41(1) :164–171, 1970.
- [2] J. Berg, T. Reckordt, C. Richter, and G. Reinhart. Action recognition in assembly for human-robot-cooperation using Hidden Markov Models. *Procedia CIRP*, 76 :205–210, 2018.
- [3] J. Bessac, P. Ailliot, J. Cattiaux, and V. Monbet. Comparison of hidden and observed regime-switching autoregressive models for (u, v)-components of wind fields in the Northeast Atlantic. *Advances in Statistical Climatology, Meteorology and Oceanography*, 2(1) :1–16, 2016.
- [4] F. Dama and C. Sinoquet. Partially Hidden Markov Chain Linear Autoregressive model : inference and forecasting. *arXiv preprint arXiv :2102.12584*, 2021.
- [5] A.B. Degtyarev and I. Gankevich. Evaluation of hydrodynamic pressures for autoregressive model of irregular waves. In *Contemporary Ideas on Ship Stability*, pages 37–47. 2019.
- [6] S.S. Dragomir. Some refinements of Jensen's inequality. *Journal of mathematical analysis and applications*, 168(2) :518–522, 1992.
- [7] G.D. Forney. The Viterbi algorithm. *Proceedings of the IEEE*, 61(3) :268–278, 1973.
- [8] K. Ghasvarian Jahromi, D. Gharavian, and H. Mahdiani. A novel method for day-ahead solar power prediction based on hidden Markov model and cosine similarity. *Soft Computing*, 24(7) :4991–5004, 2020.
- [9] J.D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, pages 357–384, 1989.
- [10] P. Juesas and E. Ramasso. Ascertainment-adjusted parameter estimation approach to improve robustness against misspecification of health monitoring methods. *Mechanical Systems and Signal Processing*, 81 :387–401, 2016.
- [11] S. Morwal, N. Jahan, and D. Chopra. Named entity recognition using hidden Markov model (HMM). *International Journal on Natural Language Computing*, 1(4) :15–23, 2012.
- [12] R. Mouhcine, A. Mustapha, and M. Zouhir. Recognition of cursive Arabic handwritten text using embedded training based on HMMs. *Journal of Electrical Systems and Information Technology*, 5(2) :245–251, 2018.
- [13] E. Ramasso. Investigating computational geometry for failure prognostics. *International Journal of prognostics and health management*, 5(1) :005, 2014.
- [14] E. Ramasso. Segmentation of CMAPSS health indicators into discrete states for sequence-based classification and prediction purposes. Technical report, 6839, FEMTO-ST institute, 2016.
- [15] E. Ramasso. RULCLIPPER algorithm and CMAPSS health indicators. [MATLAB Central File Exchange](#), 2021.
- [16] E. Ramasso. Segmentation of CMAPSS trajectories into states. [MATLAB Central File Exchange](#), 2021.
- [17] E. Ramasso and T. Denoeux. Making use of partial knowledge about hidden states in HMMs : an approach based on belief functions. *IEEE Transactions on Fuzzy Systems*, 22(2) :395–405, 2013.
- [18] A. Saxena, K. Goebel, D. Simon, and N. Eklund. Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 international conference on prognostics and health management*, pages 1–9. IEEE, 2008.
- [19] T. Scheffer and S. Wrobel. Active learning of partially Hidden Markov Models. In *In Proceedings of the ECML/PKDD Workshop on Instance Selection*. Cite-seer, 2001.