



**HAL**  
open science

## Several HSMM training methods to estimate the health of a system

Lestari Handayani, Pascal Vrignat, Frédéric Kratz

► **To cite this version:**

Lestari Handayani, Pascal Vrignat, Frédéric Kratz. Several HSMM training methods to estimate the health of a system. Congrès Lambda Mu 22 “ Les risques au cœur des transitions ” (e-congrès) - 22e Congrès de Maîtrise des Risques et de Sécurité de Fonctionnement, Institut pour la Maîtrise des Risques, Oct 2020, Le Havre (e-congrès), France. hal-03462787

**HAL Id: hal-03462787**

**<https://hal.science/hal-03462787>**

Submitted on 2 Dec 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.



# Several HSMM training methods to estimate the health of a system

## Qualification de HSMM pour estimer l'état de santé d'un système

Lestari Handayani  
PRISME Laboratory, EA 4229,  
INSA-CVL,  
F18020, Bourges, France;  
Informatics Engineering  
UIN Suska Riau  
Pekanbaru, Indonesia  
lestari.handayani@insa-cvl.fr

Pascal Vrignat  
PRISME Laboratory, EA 4229,  
Univ. Orléans, INSA-CVL,  
F45072, Orléans, France

Frédéric Kratz  
PRISME Laboratory, EA 4229,  
INSA-CVL,  
F18020, Bourges, France

**Abstract**— This paper presents a measure of relevance based on HSMM similarity and information criteria. We created a reference model (verified by ACP) and put it in competition with three other models. This work involved several parameters and two training methods for the HSMM.

**Keywords**—Hidden Semi Markov Model, information criteria, PCA, similarity, training method

**Résumé**— Cet article présente une mesure de pertinence basée sur la similarité HSMM et des critères d'information. Nous avons créé un modèle de référence (vérifié par ACP) et l'avons placé en concurrence avec trois autres modèles. Ce travail a impliqué plusieurs paramètres et deux méthodes de formation pour le HSMM.

**Mots clés**—modèle semi-markov caché, critères d'information, ACP, similitude, méthodes de formation

### I. INTRODUCTION

One of the big challenges in the manufacturing industry of modern economic systems is increasingly producing with maximum results, low environmental risk, and worker safety. Current technology has brought increased performance, functionality, and complexity of equipment to achieve automation, condition monitoring, and error diagnosis produced. Another challenge was maintaining and optimizing the quality of services provided by the industrial objects throughout its life cycle. The way is by curative or preventive care [1]. However, the state of the equipment is often difficult to observe directly, only can be expressed from the symptoms of the output [2]. Most automated systems are almost impossible to identify and predict failure conditions at the right time. Many methods and tools [1], [2], [3] can be used to predict the value of the remaining useful life (RUL). Baruah and Chinnam [4] first demonstrated that a standard HMM could be applied in the prognosis field in the machining process. However, this HMM-based method has a weakness to estimate RUL of the durational modeling because HMM use exponential time durations, so the new HSMM method is

derived by adding a temporal component [1], [2]. Thus, the HSMM is applied here to estimate the health of a system that allows arbitrary sojourn time distributions for the hidden process.

Traditional health assessments based on HSMM usually require various machine operating data for different HSMM training models. However, the data needed can hardly be obtained in most situations. In the real case, it only has a few observations. Moreover, it doesn't know the topology, the initial values of the system model, and the maximum duration of each state. Therefore, we propose an HSMM based method that is trained can be used to measure the closeness of a system model to a real case.

In this paper, we establish a reference HSMM that is controlled and suitable for real industrial processes. The sequence of observation symbols is produced by the reference model, using a normal emission distribution and several duration distributions. Our aim is to evaluate 3 different HSMM topologies (Fig. 10) that are used for training observational data from reference models where the output is the new HSMM parameter. The training uses two HSMM training methods [6, 7].

This paper contributes to the selection of the best HSMM model that considers sojourn time in each hidden state. It is used to estimate the health of a real case system. This paper discusses the measurement of the relevance of HSMM parameters based on the maximum likelihood criteria [8, 9]. We propose a similarity measurement HSMM modified using the similarity value [10] to measure the closeness of the model to the reference model. Following the observation generation process, we also used Principal Component Analysis (PCA) to verify the intrinsic characteristics of the HSMM [11]. The hypothesis will be obtained from a HSMM model that is similar to the reference model and has the optimal HSMM parameters.

This paper is organized as follows. In section II, explain HSMM theory and several HSMM training methods. Section III provides a relevance measurement detail of HSMM. We show the evaluation process, the results, and discussion in section IV. Finally, conclusions are formulated in section V.

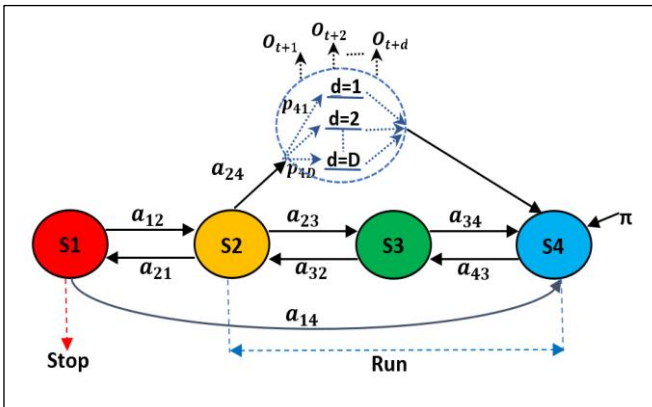
## II. HIDDEN SEMI-MARKOV MODEL

Let  $\{S\}$  be the states of a semi-Markov chain with finite-state space  $\{1, \dots, M\}$ . The state sequence  $Q = \{S_1, \dots, S_T\}$  is denoted by  $S_{1:T}$ , where  $S_t \in S$  is the state at time  $t$ . And the observation sequence  $O = \{O_1, \dots, O_T\}$  by  $O_{1:T}$  where  $O_t \in V$  is the observation at time  $t$  with  $V = \{V_1, \dots, V_k\}$  is the set of observable values. This semi-Markov chain is defined by the following parameters:

- Initial probabilities  $\pi = \{\pi_m\}$  where  $\pi_m = P(S_1 = m)$  with  $\sum_m \pi_m = 1$ ;
- Transition probabilities  $A = \{a_{mn}\}$  where  $a_{mn} = P(S_{t+1} = n | S_t = m)$  with  $\sum_{m \neq n} a_{mn} = 1$  and  $a_{mm} = 0$ ;
- Observation or emission probabilities  $B = \{b_m(v_k)\}$  where  $b_m(v_k) = P(O_t = v_k | q_t = S_m)$ ;
- Duration probabilities  $C = \{p_{md}\}$  where  $p_{md} = P(S_{t+1:t+d} = m | S_{t+1} = j)$  where  $d \in (1, 2, \dots, D)$  and  $D$  is maximum duration in state  $m$ .

A reference HSMM for the industrial health system is defined as  $\lambda_R = (\pi, A, B, C)$ . We refer to the degradation level of industrial processes [3]. There are a set of states  $S = \{S_1, S_2, S_3, S_4\}$  with topology (Fig. 1). The process runs on  $S_2, S_3$ , and  $S_4$ . While  $S_1$  is used when the process is stopped (failure). We use explicit duration showed in the dotted circle, for example,  $S_2$  to  $S_4$  (where  $m = 2, n = 4$ ). After state  $m$  ends at the  $t^{\text{th}}$  time unit, it transits to state  $m \neq n$  according to the transition probability  $a_{mn}$  and then selects its duration  $d$  according to the duration probability  $p_{md}$ . State  $n$  spends  $d$  time units and produces  $d$  observations  $o_{t+1}, \dots, o_{t+d}$  with emission probability  $b_n(o_{t+1}: o_{t+d})$ . Each state has a variable duration, which is associated with the number of observations produced while in the state.

Fig. 1. HSMM Reference



Moreover, this reference model is designed to generate a sequence of observations. Afterward, we used two algorithm HSMM training for estimated new parameters and decoding through the topology given a specific observation sequence.

### A. Forward-Backward Algorithm

A forward-backward algorithm for an explicit duration hidden Markov model obtained in [6]. The forward variables for HSMM are defined by:

$t = 1, \dots, T :$

$$\alpha_{t|t-1}(m, d) = S_{t-1}(m) p_m(d) + b_m(o_{t-1|t-2}) \alpha_{t-1|t-2}(m, d+1) \quad (1)$$

with the initial value

$$\alpha_{1|0}(m, d) = \pi_m p_m(d) \quad (2)$$

For convenience in the forward recursion, defined variables  $\varepsilon$  and  $\delta$  are respect to the conditional probability of a state ending at  $t$  given  $o_1^t$  and that of a state starting at  $t+1$  given  $o_1^t$ ,

$$\varepsilon_t(m) = \alpha_{t|t+1}(m, 1) b_m^*(o_t) \quad (3)$$

$$\delta_t(m) = \sum_n \varepsilon_t(n) a_{nm} \quad (4)$$

$b_m^*(o_t)$  is the ratio of filtered probability  $\alpha_{t|t}(m, d)$  over the predicted  $\alpha_{t|t-1}(m, d)$  by:

$$b_m^*(o_t) = \frac{\alpha_{t|t}(m, d)}{\alpha_{t|t-1}(m, d)} \quad (5)$$

The backward variables by:

$t = T, \dots, 1 :$

$$\beta_t(m, d) = \begin{cases} \delta_{t+1}^*(m) b_m^*(o_t), & d = 1 \\ \beta_{t+1}^*(m, d-1) b_m^*(o_t), & d > 1 \end{cases} \quad (6)$$

For convenience in the backward recursion, denoted variables  $\varepsilon^*$  and  $\delta^*$  by:

$$\varepsilon_t^* = \sum_n p_m(d) \beta_t(m, d) \quad (7)$$

$$\delta_t^* = \sum_n a_{mn} \varepsilon_t^*(n) \quad (8)$$

The smoothed probability that a transition from state  $m$  to state  $n$  at  $t$  occurs, it is defined by:

$$\theta_t(m, n) = \varepsilon_{t-1}(m) a_{mn} \varepsilon_t^*(n) \quad (9)$$

And the probability that state  $m$  is entered at  $t$  and lasts for  $d$  time units is:

$$\varphi_t(m, d) = \delta_{t-1}(m) p_m(d) \beta_t(m, d) \quad (10)$$

### B. Viterbi Forward-Backward HSMM

We used Viterbi forward backward based on [7]. The forward recursion is given by:

$t = 1, \dots, T - 1 :$

$$\alpha_m(t) = b_m(o_t) \max_{1 \leq u \leq t} \left[ \max \left\{ \prod_{v=1}^u b_m(o_{t-v}) \right\} d_j(u) \max \left\{ p_n(m) \alpha_n(t-u) \right\} \prod_{v=1}^t b_m(o_{t-v}) \right] d_m(t+1) \pi_m \quad (11)$$

The backward recursion is given by:

$$t = T - 1, \dots, 1 :$$

$$\beta_m(t) = \max_{k \neq m} \left[ \max_{\substack{u \\ k \leq u \leq T-t}} [\beta_k(t+u)] \left\{ \prod_{v=1}^u b_k(o_{t+u-v}) \right\} \right. \\ \left. d_k(u) \right] \left[ \prod_{v=1}^{T-t} b_k(o_{T-1-v}) \right] D_k(T-1-t) p_m(k) \quad (12)$$

The smoothed probabilities are:

$$\theta_t(m, n) = \alpha_t(m) a_{mn} \beta_{t+1}(m) \quad (13)$$

and

$$\varphi_t(m, d) = \alpha_{t-1}(m) p_m(d) \beta_t(m, d) \quad (14)$$

$$\text{The last is compute } \gamma_m(t) = \beta_m(t) \alpha_m(t) \quad (15)$$

For numerical stability, we did a log transformation to all the model parameters. As a consequence, all the products are turned into sums in computer implementation.

### C. Estimation of states and parameter HSMM

Given the workload data sequence  $o_{1:T}$ , we can MAP estimation of the hidden states using Eq. (16).

$$\hat{Q}_t = \max(\gamma_m(t)) \quad (16)$$

The re-estimation of the model parameters given below [5]:

$$\hat{a}_{mn} = \frac{\sum_t \theta_t(m, n)}{\sum_n \sum_t \theta_t(m, n)} \quad (17)$$

$$\hat{b}_m(v_k) = \frac{\sum_t \gamma_t(m) I(o_t = v_k)}{\sum_t \gamma_t(m)} \quad (18)$$

$$\hat{p}_m(d) = \frac{\sum_t \varphi_t(m, d)}{\sum_d \sum_t \varphi_t(m, d)} \quad (19)$$

$$\hat{\pi}_m = \frac{\gamma_1(m)}{\sum_m \gamma_1(n)} \quad (20)$$

## III. RELEVANCE MEASUREMENT OF HSMM

We used 3 different HSMM models and processed by two different HSMM training methods. This will be training data observations generated by reference models with several distribution laws. To find the best topology, first measured the similarity of the proposed topology with the reference model. Then analyze the new HSMM parameters and the new states of sequences generated from several training methods. We analyze using maximum likelihood and information criteria (Akaike information criterion (AIC), Bayesian information criterion (BIC), Hannan–Quinn information (HQC)). Finally, we validate the parameters using Principal component analysis (PCA).

### A. Similarity HSMM

To compare the reference model with the new topology that we are proposing we use the HSMM similarity approach to define a distance measure between any pair of HSMM. We propose a similarity measurement HSMM modified using the similarity value [7].

Suppose there are two HSMM  $\lambda^R = (\pi^R, A^R, B^R, C^R)$  and  $\lambda^T = (\pi^T, A^T, B^T, C^T)$ , whose evaluated similarity. One way

to evaluate is to consider the probability of HSMM will generate identical observations.

HSMM similarity measure  $\text{Sim}(\lambda^R || \lambda^T)$  can be computed in the following step:

- 1) Compute stationary distributions  $\mu^R$  and  $\mu^T$ .
- 2) Compute  $D(b_m^R || b_m^T)$  and  $D(p_m^R || p_m^T)$  for every pair of states using Symmetric KLD. [12]
- 3) Evaluate the similarity between states  $S_m^R$  and  $S_m^T$ ,

$$Se(S_m^R || S_m^T) = 1/D(b_m^R || b_m^T) * 1/D(p_m^R || p_m^T) \quad (21)$$

- 4) Estimate the state correspondence matrix  $Qe$  whose elements are a measure of the correspondence between two states

$$Qe(S_m^R, S_m^T) = \frac{\mu_m^R \mu_m^T Se(S_m^R || S_m^T)}{ES(\lambda^R || \lambda^T)} \quad (22)$$

$$ES(\lambda^R || \lambda^T) = \sum_{v_m^R} \sum_{v_m^T} \mu_m^R \mu_m^T Se(S_m^R || S_m^T) \quad (23)$$

- 5) Compute the HSMM similarity measure  $\text{Sim}(\lambda^R || \lambda^T)$

$$\text{Sim}(\lambda^R || \lambda^T) \triangleq \frac{1}{2} \left[ \frac{1}{L^R} \sum_{i=1}^{L^R} H(r_i) + \frac{1}{L^T} \sum_{i=1}^{L^T} H(c_j) \right] \quad (24)$$

where  $r_i$  is the  $i$ th row of  $Qe$ ,  $c_j$  is the  $j$ th column of  $Qe$ , and  $L^R$  and  $L^T$  respectively are the number of rows and columns of  $Qe$ . The function  $H(u)$  is a normalized sparsity measure for the vector  $u$  in [13].

### B. Maximum likelihood

The maximum likelihood (ML) is a method to estimate the parameters of a model and test hypotheses about those parameters. As before, we begin with discrete sample  $X = \{X_1, \dots, X_n\}$  according to probabilities  $P_\theta$ . Besides,  $f(x|\theta)$  will be used to denote the density function for the data when  $\theta$  is the true state of nature. Then, a choice of the estimator  $\hat{\theta}$  as the value for the parameter is maximum.

$$\hat{\theta}(x) = \sum_{i=1}^n \log(P_\theta(x_i)) \quad (25)$$

The most famous maximum likelihood criterion is Akaike's information criterion (AIC) [14]. It is defined by adding a penalty term that expresses the complexity of the model for a negative twofold maximum log-likelihood. The family of log-likelihood-based information criteria (LLBIC) includes BIC [15] and HQC [16], ensure a better estimation by penalizing oversizing models.

The relations between LLBIC and most well-known information criteria are as follows:

$$AIC = -2 \ln V + 2k \quad (26)$$

$$BIC = -2 \ln V + 2k * \ln(n) \quad (27)$$

$$HQC = -2 \ln V + 2k * \ln(\ln(n)) \quad (28)$$

Where  $k$  is the number of free parameters,  $V$  is the likelihood, and  $n$  is the number of data. The best model is the one that has the weakest AIC, the minimum value of BIC and HQC.

### C. Principal Component Analysis (PCA)

PCA is a method included in the family of data analysis and more generally of multivariate statistics. This method consists in transforming variables linked to each other (called "correlated" in statistics) into new variables decorrelated from each other. These new variables are called "main components", or main axes. It allows the expert to reduce the number of variables and make the information less redundant. PCA formulation used to verify the HSMM reference in proof of the emission observation and state estimation. We adopt the PCA theory from the book [11]. It is a way of identifying patterns in data and expressing the data in such a way as to highlight their similarities and differences. The method of PCA as follows:

Step 1. Get some data set (i.e.  $X$  and  $Y$ ),

Step 2. Subtract the mean from each of the data dimensions. The mean subtracted is the average across each dimension. (i.e.  $(X - \bar{X})$  and  $(Y - \bar{Y})$ ),

Step 3. Calculate the covariance matrix,

$$\Sigma = \begin{bmatrix} Var(x) & Cov(x,y) \\ Cov(y,x) & Var(y) \end{bmatrix} \quad (29)$$

Step 4. Calculate the eigenvectors and eigenvalues of the covariance matrix. We are provided with 2-dimensional vectors  $v_1, v_2$ . If we apply a linear transformation  $T$  to our vectors, we will obtain new vectors, called  $b_1, b_2$ .

$$\begin{aligned} T v_1 &= b_1 \\ T v_2 &= b_2 \end{aligned} \quad (30)$$

Those vectors are called eigenvectors, and the scalar which represents the multiple of the eigenvector is called eigenvalue.

$$T v_1 = \lambda_1 b_1 \quad (31)$$

Step 5. Choosing components and forming a feature vector.

$$FeatureVector = (\lambda_1, \lambda_2) \quad (32)$$

Step 6. Deriving the new data set.

$$FinalData = RowFeatureVector \times RowData \quad (33)$$

### IV. EVALUATION PROCESS

We evaluate the best HSMM models by relevance measurement. We present the scenario simulation in Fig. 2. The boot process from creating a reference topology about the model to produce about 10.000 data events (symbol observations or emissions and states). This volume of data is significant to have a good quality of simulation. These symbols (observations) can act as a possible real case study.

#### A. Simulated industrial

We refer to paper [3] to understand these simulations in the industrial field. There are Computerized Maintenance Management System, which maintenance activities carried out on the industrial process. We took Table I. Symbolic coding system of maintenance activities. The symbols resulting from observations. This is grouped into two processes, namely "Run" and "Stop/Failure". Process "Run" held on symbol number 2 until number 10. It happens when production units are running. On the other hand, the symbol process "Stop/Failure" is given by symbol "1(SP)", it is a critical condition.

Fig. 2. Flow chart of HSMM evaluation simulation

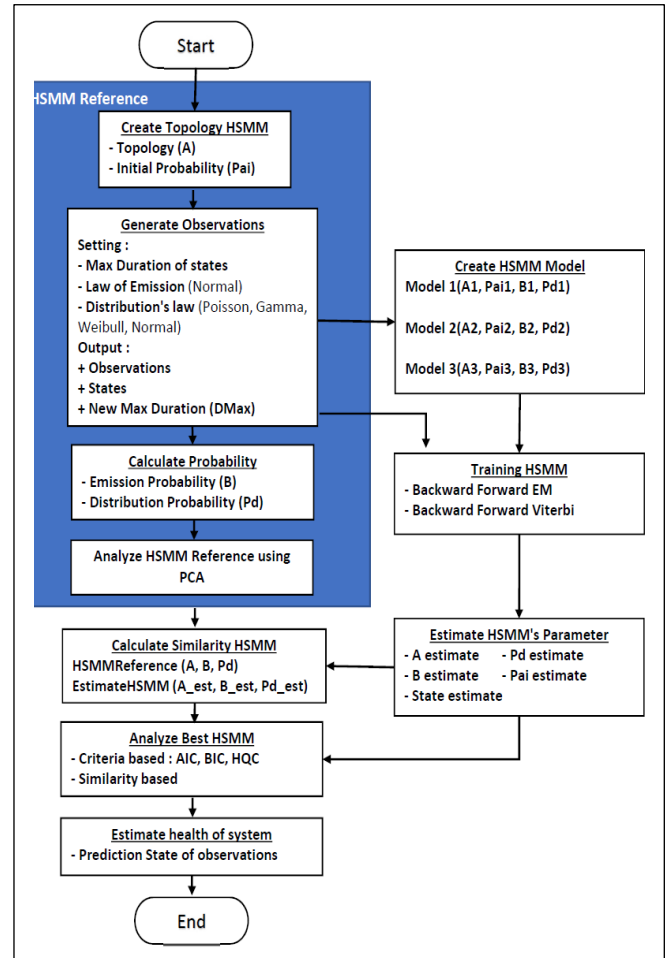


TABLE I. SYMBOLIC CODING SYSTEM OF MAINTENANCE ACTIVITIES

No.	Symbol (measurement data)	Intervention Type	Process status
1	SP	Troubleshooting/Stop production	Stop
2	SM	Setting Machine	Run
3	OT	Other	Run
4	OBS	Observation	Run
5	PM	Preventive maintenance (production not stopped)	Run
6	SEC	Security	Run
7	PUP	Planning upgrading	Run
8	CM	Cleaning machine	Run
9	PMV	Preventive maintenance visit	Run
10	NTR	Nothing to report	Run

#### B. Reference Model

We have set up a reference model that considers sojourn time in each hidden state. We used 4 states because it has good performance in maintenance activities [17]. Fig. 1 describe the topology and matrix transition probability shown in Fig. 3. Each state has a duration before moving to another state. Sojourn time in this state has a maximum duration.

Fig. 3. Matrix transition probability

States	1	2	3	4
1	0	0.3	0	0.7
2	0.2	0	0.5	0.3
3	0	0.6	0	0.4
4	0	0	1	0

### 1) Generate Observations

In generating observations, we have some settings with different and controlled specifications. These controlled specifications concern law of emission and distribution. For emission law, we used normal distribution. Symbols are distributed by modality as shown in Table. II. Value 1 is assigned to state corresponded, otherwise, value 0 is not assigned. For example, State 1 only has symbol “SP”, state 2 could be assigned has symbol “SM”, “OT”, “OBS”, “PM”, or “SEC” etc.

TABLE II. MODALITY OF SYMBOLS

State	Symbol									
	SP	SM	OT	OBS	PM	SEC	PUP	CM	PMV	NTR
1	1	0	0	0	0	0	0	0	0	0
2	0	1	1	1	1	1	0	0	0	0
3	0	0	0	1	1	1	1	1	0	0
4	0	0	0	0	0	1	1	1	1	1

Furthermore, we adjust 4 duration laws that are Poisson, Gamma, Normal, and Weibull distribution. These laws are representative in a real context. So, we have 4 kinds of sequences observation below:

- O<sub>1</sub>: EmissionNormal\_DurationPoisson,
- O<sub>2</sub>: EmissionNormal\_DurationGamma,
- O<sub>3</sub>: EmissionNormal\_DurationNormal,
- O<sub>4</sub>: EmissionNormal\_DurationWeibull.

Also, we provide a maximum duration value of the states (sojourn time) based on the duration distribution used. After obtained the data events visualized in Fig. 4 until Fig. 7, we can calculate the duration maximum of state, probability of emission, and distribution probability for our reference model HSMM.

Fig. 4. The observations sequence example of O<sub>1</sub>

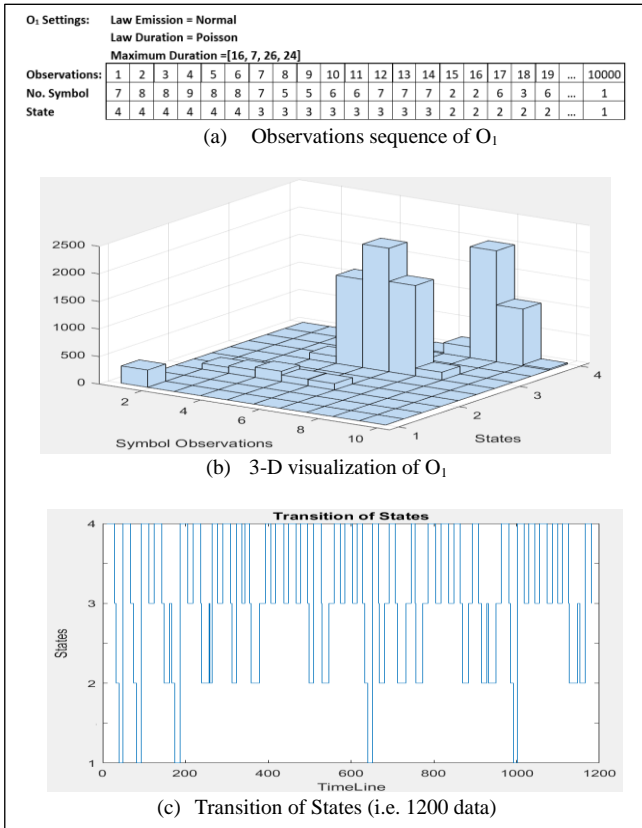


Fig. 5. The observations sequence example of O<sub>2</sub>

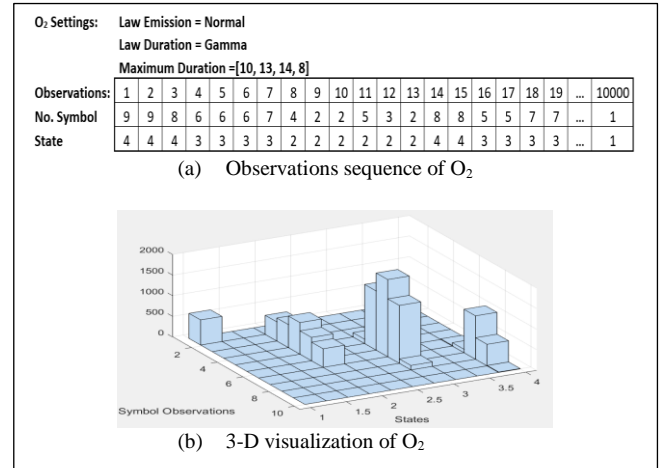


Fig. 6. The observations sequence example of O<sub>3</sub>

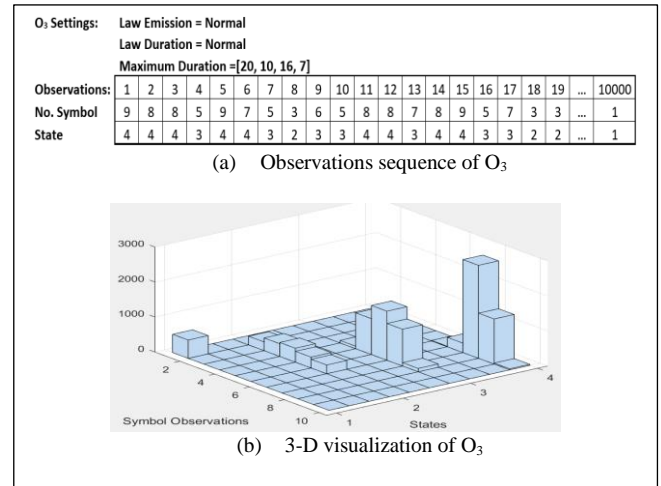
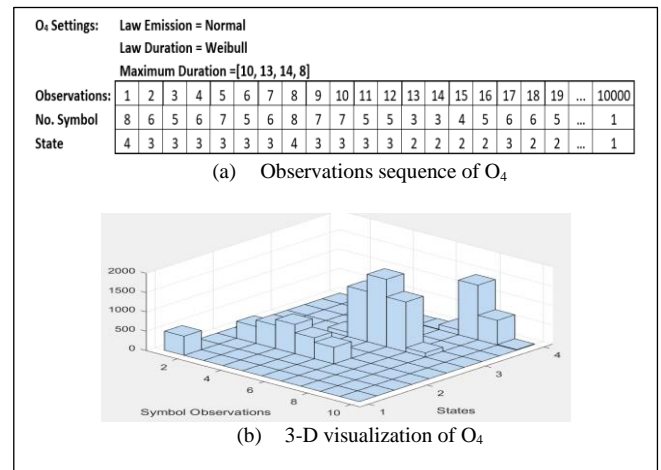


Fig. 7. The observations sequence example of O<sub>4</sub>



### 2) Calculate Dmax

Create a table of sojourn time from observed and states shown in Table. III. For each state, we calculate the maximum value of the sojourn time.

TABLE III. SOJOURN TIME TABLE OF  $O_1$

Timeline	State	Sojourn Time
1-11	4	11
12-28	3	17
29-33	2	5
34-40	1	7
...	...	...
9992-9993	2	1
9993-10000	4	7

3) Calculate Emission and distribution probabilities

We can calculate the probability of emission and distribution from the observations sequence. It will be defined as parameter  $b^R$  and  $p^R$  of the reference model.

4) Analyze HSMM Reference using PCA

We want to analyze the HSMM reference in order to be able to transpose later, that of modeling for a real case. We use PCA to see the correlation between variables and between signatures. However, in our scenario it created sequence observations as signatures which have symbol emissions and states. We did preprocessing before PCA processing.

a) Preprocessing data: make a signature with 3 symbols and it's state. Then, collecting history data of emission symbol ( $E_1: E_{10}$ ) link to the modality of symbol. The Table IV. shown a portion of preprocessing data of  $O_1$  for PCA process.

TABLE IV. PREPROCESSING TABLE OF  $O_1$

Signature	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$	$E_7$	$E_8$	$E_9$	$E_{10}$
'10_8_7'4	0	0	0	0	0	11	172	1520	726	13
'10_9_8'4	0	0	0	0	0	11	173	1521	727	14
'1_1_1'1	351	0	0	0	0	0	0	0	0	0
'1_1_2'1	315	0	0	0	0	0	0	0	0	0
'1_2_2'2	0	120	151	196	141	121	0	0	0	0
'1_6_6'2	0	126	160	204	148	129	0	0	0	0
'2_1_1'1	325	0	0	0	0	0	0	0	0	0
'4_5_6'2	0	124	159	201	144	124	0	0	0	0
'4_5_6'3	0	0	0	127	1619	2201	1673	136	0	0
...	...	...	...	...	...	...	...	...	...	...

b) Processing PCA: Input variables and label of observation. We select the Pearson correlation coefficients. We have 2 results. First, the label is the state and the variables are  $E_1$  to  $E_{10}$ . The objective is to obtain the correlation between emission symbol and state. Second, we choose signature as label, and state as a variable. It aims to view the correlation between state and signature.

c) PCA Result: the result of sequence observations  $O_1$  shown in Fig. 8 and Fig. 9. The correlation of these figures are very good. 82,27% for correlation of states and emission symbol. And 99,98% is correlation of states and signatures. It's mean that the sequence observations of  $O_1$  is strong and we can use it for simulation HSMM.

Fig. 9 show that the observations actives with the signature. It has a possibility for emission symbols many state's estimations. For example, the signature '4\_5\_6'2 is emission symbol 5 in class 2. This emission symbol appears

in class 3 by signature '4\_5\_6'3 also. So, it fulfills the requirement of the modality of the symbol in Table II.

Table V describes the PCA results of sequence observations. All values show that the correlation values of these simulations are high. We can use these sequence observations for HSMM evaluation.

Fig. 8. Correlation of states and emission symbols of  $O_1$

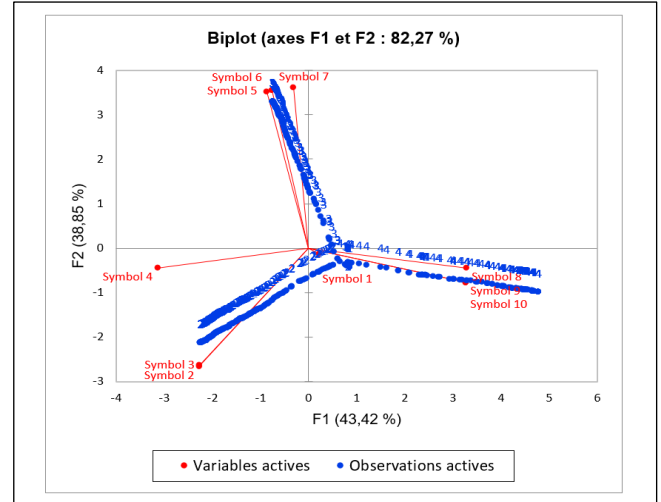


Fig. 9. Correlation of states and signatures of  $O_1$

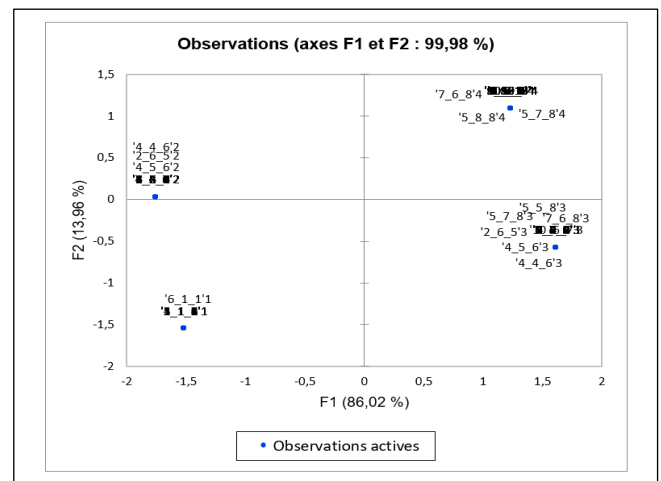


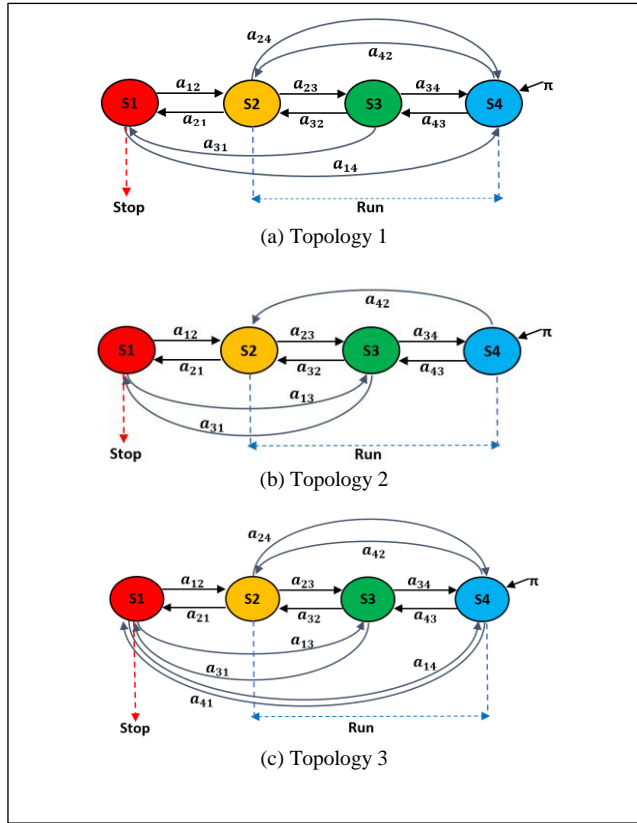
TABLE V. THE PCA RESULTS OF 4 SEQUENCE OBSERVATIONS

Sequence observation	Correlation of state and emission symbol	Correlation of state and signature
$O_1$	82,27 %	99,98 %
$O_2$	84,28 %	93,76 %
$O_3$	81,24 %	91,20 %
$O_4$	83,63 %	99,99 %

C. HSMM Topology

In the next step, we defined 3 different models that we evaluate. These models have different topology notice the maximum duration of the reference model.

Fig. 10. HSMM Topologies



1) *Training HSMM*

We used the forward-backward algorithm and Viterbi forward-backward as a training method. The input data is the observation sequences of the reference model.

2) *Estimate HSMM Parameter*

Re-estimate new parameters of HSMM could obtain by using formula (16) until (20). These values used for evaluation HSMM.

D. *Evaluation HSMM*

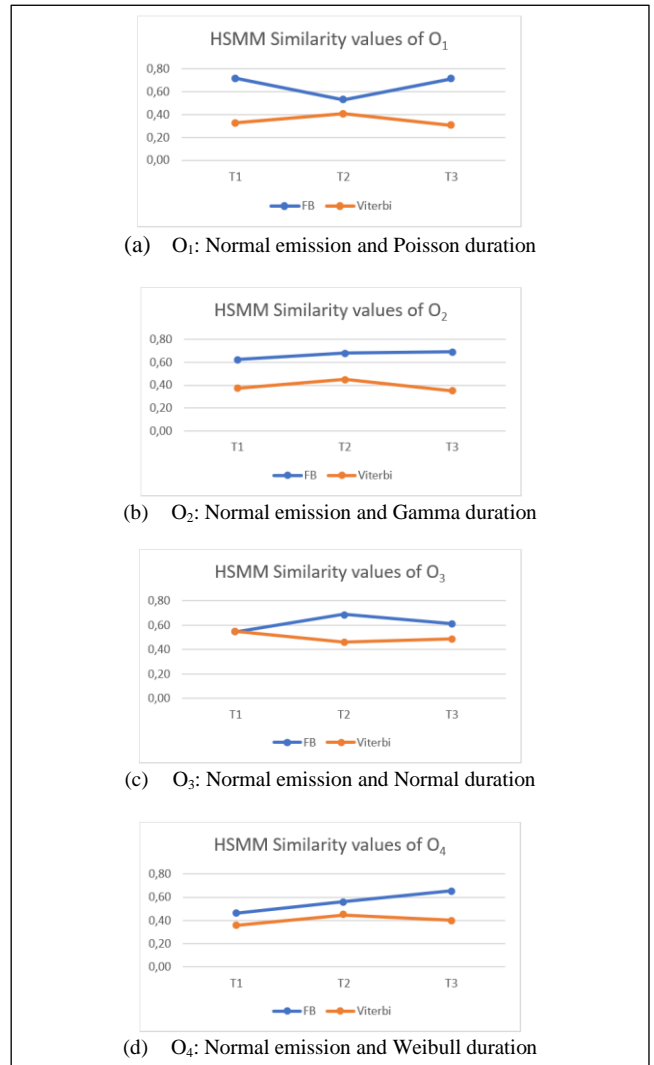
We tried to evaluate the Topology and Reference model by calculating similarity HSMM and maximum likelihood criterion.

1) *Calculate Similarity HSMM*

We obtain 4 types of sequence observations that some settings describe in Fig. 4 until Fig. 7 ( $O_1, O_2, O_3, O_4$ ). Next, the model is trained by the forward-backward algorithm (BF) and the Viterbi forward-backward algorithm (Viterbi). The results of the HSMM similarity shown in Fig. 11.

From Fig. 11, we show that the similarity value of the BF algorithm is often higher than the Viterbi algorithm. The topology who has the highest similarity value is topology 1 for  $O_1$ , topology 3 for  $O_2$ , topology 2 for  $O_3$ , and topology 3 for  $O_4$ .

Fig. 11. Similarity value of HSMM



(a)  $O_1$ : Normal emission and Poisson duration

(b)  $O_2$ : Normal emission and Gamma duration

(c)  $O_3$ : Normal emission and Normal duration

(d)  $O_4$ : Normal emission and Weibull duration

2) *Calculate Criterion HSMM*

During the HSMM training process, it is possible to calculate the maximum likelihood. Then, we can calculate some criterions HSMM (AIC, BIC, HQC). Fig. 12 presents the results of the criterion values from the FB training for  $O_1, O_2, O_3$ , and  $O_4$ . These values are in the 10000s for the FB algorithm and  $1E9$  for the Viterbi algorithm.

From Fig. 12, it appears that the values of AIC, BIC, and HQC have the same tendency. The best topology is that which has the highest value. Based on these criteria, we chose topology 1 is the best for  $O_1$ , topology 3 for  $O_3$ , topology 2 for  $O_3$ , and topology 3 for  $O_4$ .

3) *Estimate Health of System*

In this chapter, we detail the elements of competition between the different models. First, we select the best topology that has the highest similarity value and minimum criterion. The results of the evaluation simulation are shown in Table VI, and Table VII (i.e. sequence observations  $O_1$ , we choose topology 1 which used FB training method).



Fig. 12. Criterion value of HSMM using forward- backward Algorithm

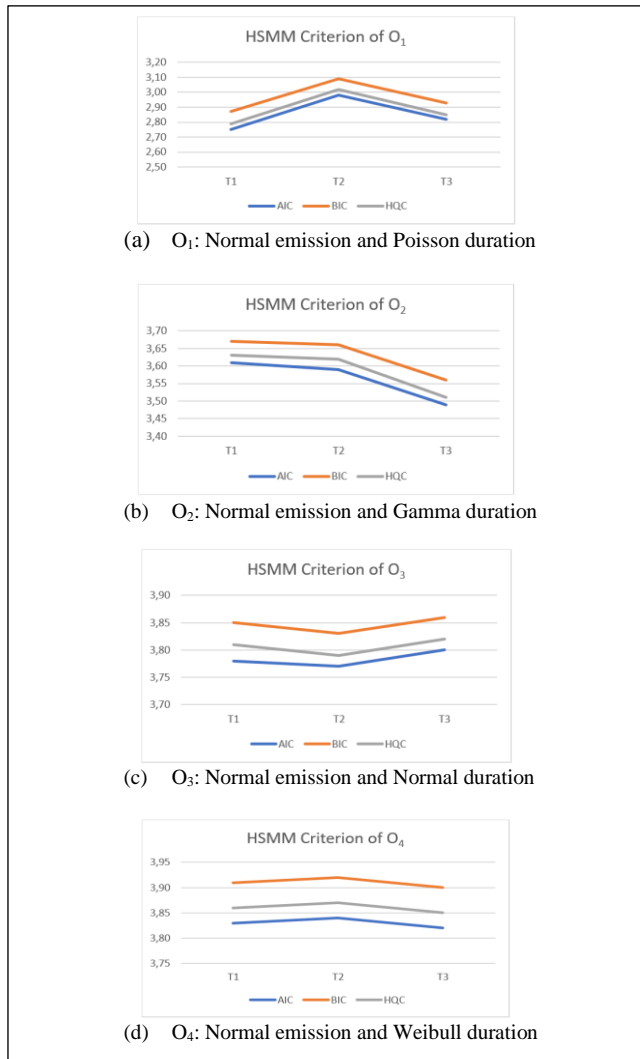


TABLE VI. HSMM EVALUATION OF O<sub>1</sub>, AND O<sub>2</sub>

Evaluate	Similarity	AIC	BIC	HQC
<b>O<sub>1</sub>: Normal emission and Poisson duration</b>				
FB_T1	0,72	2,75	2,87	2,79
FB_T2	0,53	2,98	3,09	3,02
FB_T3	0,71	2,82	2,93	2,85
Viterbi_T1	0,33	1,2110	1,2110	1,2110
Viterbi_T2	0,41	1,2044	1,2044	1,2044
Viterbi_T3	0,31	1,2230	1,2230	1,2230
<b>O<sub>2</sub>: Normal emission and Gamma duration</b>				
FB_T1	0,63	3,61	3,67	3,63
FB_T2	0,68	3,59	3,66	3,62
FB_T3	0,69	3,49	3,56	3,51
Viterbi_T1	0,37	1,4556	1,4556	1,4556
Viterbi_T2	0,45	1,4505	1,4505	1,4505
Viterbi_T3	0,35	1,4812	1,4812	1,4812

TABLE VII. HSMM EVALUATION OF O<sub>3</sub>, AND O<sub>4</sub>

<b>O<sub>3</sub>: Normal emission and Normal duration</b>				
FB_T1	0,54	3,78	3,85	3,81
FB_T2	0,69	3,77	3,83	3,79
FB_T3	0,61	3,80	3,86	3,82
Viterbi_T1	0,55	1,8374	1,8374	1,8374
Viterbi_T2	0,46	1,8339	1,8339	1,8339
Viterbi_T3	0,49	1,8285	1,8285	1,8285
<b>O<sub>4</sub>: Normal emission and Weibull duration</b>				
FB_T1	0,46	3,83	3,91	3,86
FB_T2	0,56	3,84	3,92	3,87
FB_T3	0,66	3,82	3,90	3,85
Viterbi_T1	0,36	1,5860	1,5860	1,5860
Viterbi_T2	0,45	1,6393	1,6393	1,6393
Viterbi_T3	0,40	1,5855	1,5855	1,5855

Using this topology, we can predict the state of sequence observations. Fig. 13 shown that state row is the health estimation of the system.

Fig. 13. Prediction of sequence observations

NO. Observation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	...	1000
No. Symbol	7	8	8	9	8	8	7	5	5	6	6	7	7	7	2	2	6	3	6	...	1
State	4	4	4	4	4	4	3	3	3	3	3	3	3	3	2	2	2	2	3	...	1

## V. CONCLUSION

In this study, a method for measuring HSMM relevance is proposed. For this, we have created a benchmark HSMM. A sequence observation generator has been specifically developed for this model. We have successfully used the PCA method to obtain possible correlations in the HSMM. After testing randomly the other models from our reference model, we have applied all relevant measurements of HSMM. We have successfully applied this method for 3 different models in some settings of sequence observations and several HSMM training methods. The results of the method, proven that it could be used to choose the best topology. The highest similarity value and the minimum criterion provide the same topology. It also gives a conclusion about the HSMM training method. The Forward-backward algorithm is top on the graphic of HSMM similarity (it's happened for all observation sequences types). This work is a first step to qualify an HSMM model. It will have to be consolidated later with real data.

## ACKNOWLEDGMENT

We thank the Directorate of Islamic Higher Education at the Ministry of Religion, Republic of Indonesia, for doctoral scholarships.

## REFERENCES

- [1] Diego Tobon-Mejia, K.M., Nouredine Zerhouni, Gérard Tripot, "Hidden Markov models for failure diagnostic and prognostic," Prognostics and System Health Management Conference, 2011.
- [2] Huo Lin, F.S., Lv Chuan, Wang Zili, "A novel methodology based on hidden semi-Markov model for equipment health assessment," Jve International Ltd. Vibroengineering Procedia, 2014. 4: p. 271-276.

- [3] Roblès, B., Avila, M., Duculty, F., Vrignat, P., Bégot, S., & Kratz, F., "Hidden Markov model framework for industrial maintenance activities," Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 2014, 228(3), 230-242.
- [4] Baruah, P. and R.B. Chinnam, "HMMs for diagnostics and prognostics in machining processes," International Journal of Production Research, 2005. 43(6): p. 1275-1293.
- [5] Yu, S.Z., Hidden Semi-Markov Models Theory, Algorithms and Applications, Elsevier, 2016.
- [6] Shun-Zheng Yu and H. Kobayashi, "Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden Markov model," IEEE Transactions on Signal Processing, vol. 54, no. 5, pp. 1947-1951, May 2006..
- [7] Guédon, Y. "Exploring the state sequence space for hidden Markov and semi-Markov chains," Computational Statistics & Data Analysis, 2007, 51, 2379 – 2409.
- [8] Suda, D., Spiteri, L. "Analysis and Comparison of Bitcoin and S and P 500 Market Features Using HMMs and HSMMs," Information, 2019, 10, 322.
- [9] Chen, J. & Hawkes, Alan & Scalas, Enrico & Trinh, Mailan. "Performance of information criteria used for model selection of Hawkes process models of financial data," SSRN Electronic Journal. 10.2139/ssrn.3040511, 2017.
- [10] Sahraeian, S. and B.-J. Yoon, "A Novel Low-Complexity HMM Similarity Measure," Signal Processing Letters, IEEE, 2011. 18: p. 87-90.
- [11] Jolliffe I.T., Principal Component Analysis, Second Edition. Springer, New York, 2012.
- [12] M. Basseville, "Distance measures for signal processing and pattern recognition," Signal Processing, vol. 18, no. 4, pp. 349 – 369, 1989.
- [13] N. Hurley and S. Rickard, "Comparing measures of sparsity," Information Theory, IEEE Transactions on, vol. 55, no. 10, pp. 4723 – 4741, oct. 2009.
- [14] Akaike, H, "A new look at the statistical model identification. IEEE Trans," Automatic Control, AC-19, 716–723, 1974.
- [15] Schwarz, G, "Estimating the dimension of a model," Ann. Statist., 6, 461–464, 1978.
- [16] Hannan, E. J. & Quinn, B. G, "The determination of the order of an autoregression," J. Roy. Statist. Soc. Ser. B, 41, 190–195, 1979.
- [17] Vrignat, P, Avila, M, Duculty, F and Kratz, F, "Modélisation des dysfonctionnements d'un système dans le cadre d'activités de maintenance", 16<sup>ème</sup> Congrès de Maîtrise des Risques et de Sécurité de Fonctionnement, Communication 4A-1, Avignon, 2008.