HMM Framework, for Industrial Maintenance Activities

Bernard Robles, Manuel Avila, Florent Duculty, Pascal Vrignat, Stéphane Begot, Frédéric Kratz

▶ **To cite this version:**

Bernard Robles, Manuel Avila, Florent Duculty, Pascal Vrignat, Stéphane Begot, et al.. HMM Framework, for Industrial Maintenance Activities. QUALITA2013, Mar 2013, Compiègne, France. hal-00823159

HAL Id: hal-00823159
https://hal.science/hal-00823159

Submitted on 16 May 2013

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.

HMM Framework, for Industrial Maintenance Activities

Bernard Roblès, Manuel Avila, Florent Duculty, Pascal Vrignat, Stephane Bégot, *Orleans University, PRISME Laboratory (EA 4229), IUT de l'Indre, Châteauroux, France* and Frédéric Kratz, *ENSI, PRISME Laboratory (EA 4229), Bourges, France*

***Abstract*—This paper uses the Hidden Markov Model to model an industrial process seen as a discrete event system. Different graphical structures based on Markov automata, called topologies, are proposed. We designed a Synthetic Hidden Markov Model based on a real industrial process. This Synthetic Model is intended to produce industrial maintenance observations (or "symbols"), with a corresponding degradation indicator. These time series events are shown as Markov chains, also called "signatures". The production of symbols is generated by using a Uniform and a Normal distribution. Hence, we implemented these various symbols in proposed topologies using Baum-Welch learning algorithm decoding by Forward Variable and Segmental K-means learning, decoding by Viterbi. Through different measurements on model outputs, these frameworks (a topology with a learning & decoding algorithm and a distribution) are compared to determine the best part of criteria applied to observations. Assessment results show significant differences between the various frameworks studied. After determining the most relevant framework, we developed an industrial application and compared it with the best model framework found. Finally, we propose a model adjustment to fit the industrial maintenance activities studied. Our aim is to produce the best Synthetic Model framework to be used to improve maintenance policy, worker safety and process reliability in the industrial sector.**

***Index Terms*—Hidden Markov Model, model selection, learning algorithms, decoding algorithms, statistical tests, uncertainties, predictive maintenance.**

I. Introduction

HIDDEN Markov Models can be used for modeling complex systems. In this study, we use this kind of model to detect faults in systems. Nowadays, industrial robots operating in a several environment need upstream fault detection in order to prevent any breakdowns. Indeed, it is conceivable for poorly maintained equipment to break down, bringing the entire production line to a halt. The two key concepts in maintenance are: *maintain* and *restore*. The first one refers to a preventive action, while the second refers to a corrective action. Thus, maintenance optimization for reliability determines "optimal" preventive maintenance. The events preceding a problem in maintenance activities are often recurrent. Series of unusual events should inform us about the next failure.

Most published papers aim to optimize performance using given risk and reliability strategies. Our work, [1], has shown that it is possible to model degradation levels of a process. Moreover, we built a *Synthetic Model* based on a *reference model*, which fits real industrial processes ([1]). We also showed in [1], that this *reference model* provided good failure prediction. This *Synthetic Model* produced observations or symbols, commonly used in a CMMS[1] and degradation indicators: S1, S2, S3 and S4 (see figure 1). In this paper, we try to assess the relevancy of these observations based on different HMM topologies, depicted in figure 1. Different kinds of framework[2] are tested: Uniform and Normal distributions, Baum-Welch learning [2], decoding by Forward Variable [3] and Segmental K-means learning [4], decoding by Viterbi. Assessment of the produced observations was made by using criteria usually used in model selection: Shannon Entropy, Maximum Likelihood, Bayesian Information Criterion and several statistical tests. We also evaluated epistemic uncertainties in order to frame the margin error of the model design.

The structure of this paper is as follows: in section II we outline HMM, describe the Synthetic Model design and present methods to assess the relevancy of model parameters. The results are given in section III and discussed in section IV. Finally the findings are compared with an industrial application in section V and some adjustments of the Synthetic Model are proposed.

The purpose of this study is to validate a model framework and show that our design choice meets objectives in the improvement of preventive maintenance and breakdown prediction, in the industrial sector.

II. Methodology

A. Hidden Markov Model

In this study, we chose HMM to describe industrial maintenance events. An HMM consists of a hidden stochastic process modeled by a Markov chain and an observable stochastic process. This kind of model is represented by automaton with hidden states which consists of unobservable variables ([3]). These unobservable variables represent the system status to be modeled. Only output variables are observable. Moreover, this automaton is intended to generate observation sequences from its outputs. (see an example of model topology depicted in figure 1). Indeed, we attempted to assess symbols relevancy.

HMM is characterized by:

- State number;

[1]A Computerized Maintenance Management System is an information database about maintenance operations. This information is intended to make a decision support for maintenance experts.

[2]A framework includes: a topology, a learning & decoding algorithm and a distribution.

- Number of distinct observation symbols per state, observation symbols corresponding to the physical output of the system being modeled;
- Distribution probability of state transitions;
- Distribution probability of observation symbols;
- Initial states distribution.

1) Markov assumption: States prediction is not made any more accurate by adding a priori knowledge of the information, i.e. all useful information for future prediction is contained in the present state of the process (*i.e.* it's a Markov chain of order 1).

$$P(X_{n+1} = j | X_0, X_1, \ldots, X_n = i) = P(X_{n+1} = j | X_n = i). \quad (1)$$

We used this assumption to take into account several orders of a Markov chain.

2) Definitions of a discrete HMM: In this section, we describe the variables used in an HMM:

- Let N, be the number of workable hidden states and $\mathbb{S} = \{S_1, S_2, \ldots, S_N\}$, the set of this variable. Let q_t, be the value of this variable at time t;
- The modeled process must match first-order Markov assumption (see §II-A1);
- Let T, be the full number of observation symbols and let $X = \{x_1, x_2, \ldots, x_T\}$, observations sequence of the modeled process;
- Let $A = \{a_{ij}\}$, distribution probability of state transitions where:

$$a_{ij} = P(q_{t+1} = s_j | q_t = s_i) \quad 1 \leqslant i, j \leqslant N, \quad (2)$$

- Let $B = \{b_j(m)\}$, distribution probability of observation symbols in j state, where:

$$b_j(m) = P(X_t = x_m | q_t = s_j) \quad 1 \leqslant j \leqslant N \;\; 1 \leqslant m \leqslant T, \quad (3)$$

  where X_t, is the value of observation variable at time t.
- Let $\pi = \{\pi_i\}$, initial states distribution where:

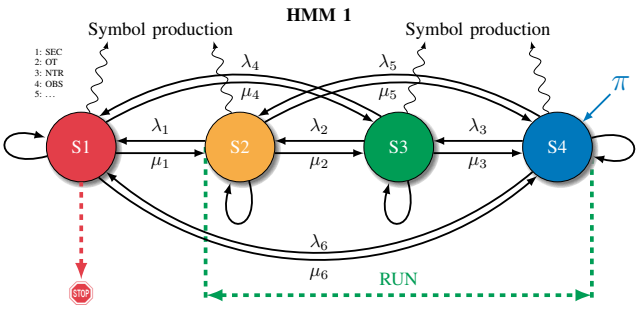
$$\pi = P(q_1 = s_i) \quad 1 \leqslant i \leqslant N, \quad (4)$$

- HMM will be set as: (A, B, π).

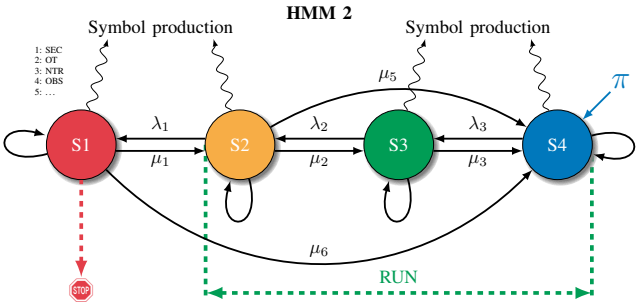
B. HMM topologies

Candidate models are represented by automata with four oriented states. These stochastic automata, depicted in figure 1, represent the degradation level of an industrial process, S4 to S1. {S4, S3, S2} states, when the process is running ("RUN"), and {S1} state, when the process is stopped ("STOP"). Topology 1, depicted in figure 1(a) describes all possible transitions. With topology 2 in figure 1(b) we need to go through all states (S2 and S3) to go from a high level of availability (S4) to a low level of availability (S1). Figure 1(c) depicts the difference between topologies 2 and 3: S1 becomes a first state of a breakdown.

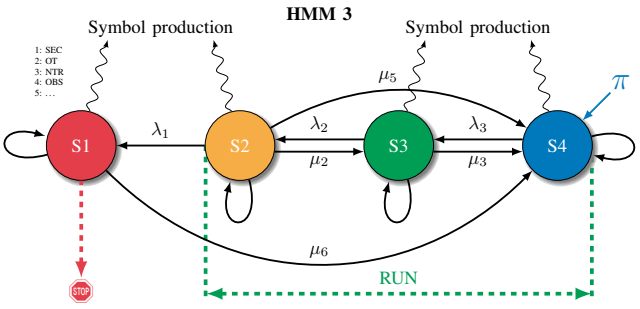
Regular temporal sampling is a requisite to have a Markov process. Both the simulated process and industrial database must have the same temporal sampling. We can therefore use Markov modeling. These 3 topologies are intended to simulate an industrial process as presented in the next paragraph.



(a) Topology 1



(b) Topology 2



(c) Topology 3

Fig. 1. The four-state Hidden Markov Model. Different topologies are represented. All transitions are permitted on 1(a). In 1(b) and 1(c) we removed some transitions, in order to have a more representative topology that was more representative of an industrial situation. S1 to S4 represent the degradation levels. S1: the system is stopped. S4 to S2: progressive degradations of the process. The production of symbols represents the Markov chain given by a maintenance database. λ_i is the failure rate and μ_i is the repair rate. π is the initialization matrix.

C. Simulated industrial Computerized Maintenance Management System

Nowadays, many industrial factories use **preventive maintenance**. Maintenance operators consign their actions and observations in a centralized database. We show an example of such database in table I. For instance, symbols "PM, OT, SP, ..." could characterize maintenance activities carried out on industrial processes. We recall the meaning of selected symbols resulting from observations given in table II. The "SP" symbol corresponds to a stop of production units: process state = "STOP" in table II. It is a critical condition, which our research aims to minimize. Process state = "RUN" when production units are running without failure.

Name	Date	Ope.	Cd	IT	N°	Code
Dupond	11/01/2007	Lubrication	PM	20	1	9
Dupond	11/01/2007	Lubrication	PM	20	2	9
Dupond	12/01/2007	Lubrication	SEC	30	3	5
Dupond	12/01/2007	Lubrication	PM	30	4	5
Dupond	13/01/2007	Padlock	PM	10	5	6
Dupond	13/01/2007	Padlock	NTR	30	6	5
Dupond	13/01/2007	Padlock	NTR	30	7	5
Dupond	16/01/2007	Lubrication	SP	90	8	1
Dupond	19/01/2007	Padlock	OT	10	9	3

TABLE I
EXAMPLE OF RECORDED EVENTS FROM A MAINTENANCE DATABASE.

We conceptualized this kind of maintenance here by using

Process states	
RUN	
STOP	

N°Obs.	Interventions type	
1	SP	(Troubleshooting / Stop Production)
2	SM	(Setting Machine)
3	OT	(Other)
4	OBS	(Observation)
5	PM	(Preventive Maintenance, Production not stopped)
6	SEC	(Security)
7	PUP	(Planified Upgrading)
8	CM	(Cleaning Machine)
9	PMV	(Preventive Maintenance Visit)
10	NTR	(Nothing to report)

TABLE II
SYMBOLIC CODING SYSTEM OF MAINTENANCE INTERVENTIONS.

a **Synthetic Model** presented in the next paragraph (§II-D) to simulate this real industrial environment. We chose "λ_i" (failure rate) and "μ_i" (repair rate) of HMM parameters, to match the maintenance register as closely as possible (table I).

D. Conceptual Synthetic Model

We designed a Synthetic Model with Matlab by using four-state oriented HMM, with the reference **topology 2** presented in figure 1(b).
HMM topology only depends on matrix elements, where $\{a_{ij}\} \neq 0$ (If all matrix elements are different from zero, we have a "total connectivity matrix"). The transition matrix (A) defined in the paragraph II-A2 has been specified in [1] by:

$$A = \begin{pmatrix} 0.500 & 0.250 & 0 & 0.250 \\ 0.100 & 0.070 & 0.500 & 0.330 \\ 0 & 0.005 & 0.495 & 0.500 \\ 0 & 0 & 0.001 & 0.999 \end{pmatrix}. \quad (5)$$

The Synthetic Model built sequences of data (also named "signatures") by using the uniform distribution and the normal distribution.

We used these symbol sequences as a Markov chain (see table III), to model the degradation level of a process (see an example of degradation sequence in figure 2).

We will now describe the framework specifications used for data analysis. We produced 12 sequences distributed among 1000 2-tuple $(Symb_U, State_U)$ for a Uniform

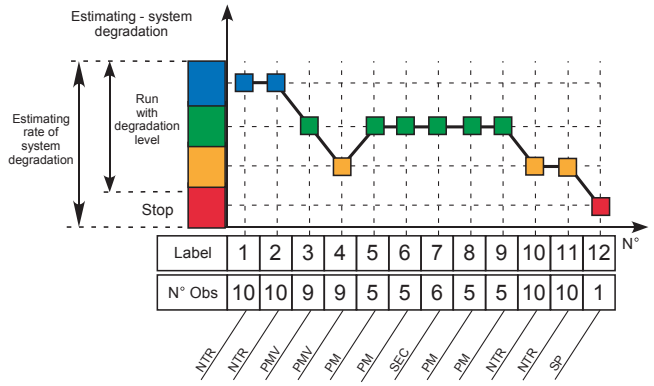


Fig. 2. Example of a degradation sequence of a process. The process starts at level 4. It becomes more and more degraded over time. After a Preventive Maintenance Visit (PMV), you can see that the degradation level increases on level 3. It decreases again until nothing is done (NTR: Nothing To Report) see label 10 and 11. Then, the process stops on label 12 (SP). PM is the symbol for Preventive Maintenance and SEC for a Security report.

PM	PM	SEC	PM	PM	NTR	NTR	SP	...

TABLE III
SEQUENCE OF A MESSAGE FROM MAINTENANCE DATABASE.

distribution and 1000 2-tuple $(Symb_N, State_N)$, for a Normal law. Each sequence ends with a production stop (symbol SP). Hence, we implemented these Symbols and States in the 3 HMM topologies described in figure 1. Thus, implementing the last 2-tuple through the 3 topologies, we obtained new $States$: $(Symb_U, State_Ui)$ and $(Symb_N, State_Ni)$ with $i \in [1, 3]$. Afterwards, we used two different learning and decoding algorithms in order to estimate the new States through the topologies:

- Baum-Welch learning ([2]), decoding by Forward Variable ([3]),
- Segmental K-means learning ([4]), decoding by Viterbi, ([5]).

We then obtained $(Symb_U, State_UiBW)$, $(Symb_U, State_UiSK)$, $(Symb_N, State_NiBW)$ and $(Symb_N, State_NiSK)$. Finally, all theses 2-tuple were compared with $(Symb_U, State_U)$ and $(Symb_N, State_N)$.

E. Evaluation methods

To model industrial processes, different topologies of the Hidden Markov Model have been used, in the aim of finding the best topology, the best learning & decoding algorithm and the best distribution of symbols. We try to evaluate the best HMM topology proposed in [1], by using *Shannon entropy* [6], especially *maximum entropy principle* used in [7]. Calculation is made with states and observations: production of symbols of Synthetic HMM. To emphasize our analysis, we also use some criteria which penalize the likelihood value, in order to overcome over-parameterization models, like *Akaike* (AIC) and *Bayes* (BIC) criteria. Finally, we determine the stochastic nature of our given symbols and use some statistical tests to see if the model fits the reference one. A diagram of the evaluation process is given in figure 3.

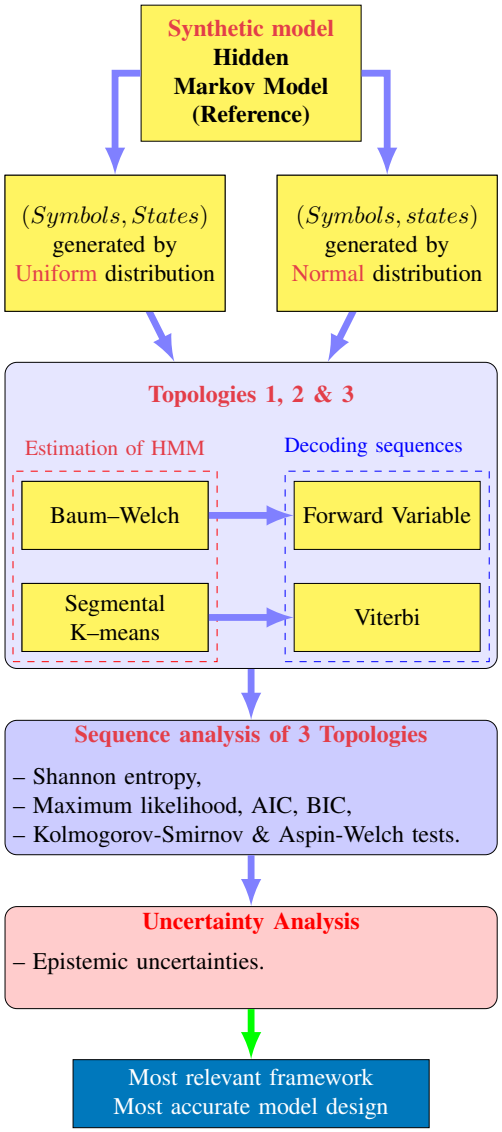


Fig. 3. The matching model method, using a Synthetic Model: We first used the Synthetic Model (also called reference model), to generate stochastic 2-tuple $(Symbols, States)$ by using the Normal and the Uniform distribution. Thus, we injected these signatures into the 3 topologies of Hidden Markov Model. To achieve learning models, we use Baum-Welch learning, decoding by Forward Variable and Segmental K-means learning, decoding by Viterbi. We obtained new sequences of $(Symbols, States)$ that we analyzed. We evaluated the best topology, best learning algorithm and best distribution. To evaluate this "signature", the following widely-used criteria have been applied: Shannon Entropy, Maximum Likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Kolmogorov-Smirnov & Aspin-Welch statistical tests. Finally, epistemic uncertainties make it possible to frame the margin error of the model design. The synthesis of all these criteria, yielded the best relevant framework.

1) Shannon entropy: The first criterion to evaluate relevancy of the "signatures" is the Shannon Entropy. Namely, we evaluated different order Markov chains. Hence, we calculated the Shannon entropy in the one, two, three or four state Markov-chain.

Shannon entropy is a function which calculates the information rate contained in an information source. This source can be a text written in any language, an electrical signal or an unspecified electronic file.

2) Entropy definition: Shannon entropy is defined in [8] as follows:

$$H(S) = -\sum_{i=1}^{n} P_i \log_b P_i, \tag{6}$$

P_i is the average probability to find the i symbol in S.

Our implemented method to evaluate Entropy used the "maximum entropy principles".

3) Maximum likelihood: The second criterion for the evaluation is the maximum likelihood principle: Let P_α, be a statistical model, and X, an observation sequence, the probability of seeing X according to P can be measured by $f(X, \alpha)$ a function which represents the density of X when α appears. Since α is unknown, it seems natural to promote values of α where $f(X, \alpha)$ is maximal: this is the notion of the likelihood of α for observation X.

– Expression of the likelihood V:

$$V(x_1, \ldots, x_n; \alpha) = \prod_{i=1}^{n} f(x_i; \alpha), \tag{7}$$

α is mathematical expectation.

– The maximum likelihood for a discrete sample $P_\alpha(x_i)$ representing the discrete probability where α appears:

$$\log(V(x_1, \ldots, x_n; \alpha)) = \sum_{i=1}^{n} \log(P_\alpha(x_i)). \tag{8}$$

We maximize the logarithm of the likelihood function to compare with the candidate models. According to [9], the principle of maximum likelihood results in over-parameterization of the model to produce good performances. Penalization of the likelihood value can overcome this disadvantage. The most famous penalized log-likelihood criterion is the AIC [10], even if it is not completely satisfactory: it improves the maximum likelihood principle but also leads to an over-parameterization. Other traditional criteria, BIC and HQC, ensure a better estimation by penalizing oversizing models.
In the following paragraphs, we introduce the **Akaike Information Criterion** and **Bayesian Information Criterion**. Both methods of model evaluation penalize the number of estimated parameters.

4) AIC and BIC: The best model is the one which has the weakest AIC ([11]) or BIC. BIC was introduced in [12] and penalizes more over-parameterized models.

$$AIC = -2 \ln V + 2k, \tag{9}$$

$$BIC = -2 \ln V + k \ln(n), \tag{10}$$

V is the likelihood, *k is the number of free parameters of Markov Model [13], n is the number of data, $k \ln(n)$ is the penalty term.*

These criteria use the maximum likelihood principle seen in (8). It penalizes models with too many variables, and avoids overfitting models [14].

F. Statistical tests

Another kind of assessment criterion, generally used in model selection, is that of statistical tests. Most statistical tests assume that samples are taken randomly to achieve [15]. That sounds easy but it is actually quite difficult to achieve. The adequacy of the technical design of the Synthetic Model has been improved by using some statical tests.
We used the Kolmogorov-Smirnov test [16] and the Aspin-Welch test [17] These may be used to determine if a set of data comes from a particular probability distribution.

G. Epistemic uncertainties

Finally we tested the model design by calculating Epistemic uncertainties. This uncertainty is explicitly due to the design of the mathematical model. It is related to the human interpretation of the phenomenon which leads to imperfections in the design. We examined epistemic errors on the Synthetic Model and determined design elements with the lowest uncertainty.

For a n measures series of $x_1, x_2, \ldots, x_n$, the uncertainty on the average according to [18] is:

$$\Delta\overline{x} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \overline{x})^2}. \qquad (11)$$

- σ: samples standard deviation.

H. Industrial application

This approach involves understanding the **signatures** of any industrial CMMS by using HMM topologies. Furthermore, our best framework would be able to provide decision support for organizing daily maintenance and would help experts to schedule maintenance activities.

III. Results

- Shannon Entropy: **Topology 2** had significantly greater **Entropy** than the others, with the **Normal distribution** and with the **Baum-Welch learning algorithm decoding by Forward Variable**,
- Likelihood Criteria: The results show a maximum log-likelihood for the **topology 2** and the **Normal distribution**. We have also the same conclusion for the AIC and the BIC criteria,
- The Aspin-Welch and the Kolmogorov-Smirnov adequacy test have the lowest p-value result for the following framework: **Topology 2**, **Normal distribution**, and **Baum-Welch learning, decoding by Forward Variable**,
- The best epistemic uncertainties are given for the **topology 2** with the **Baum-Welch learning algorithm decoding by Forward Variable** and a **Normal distribution**.

The reader can find all results in table IV (a cross indicates a good result).

Evaluation criteria	Topology 1	Topology 2	Topology 3	Learn algo B.W.	Learn algo S.K.	Distribution Nor.	Distribution Uni.
Shannon Entropy 1^{st} order		×		×		×	
Shannon Entropy 2^{nd} order		**No finding**		**No finding**		**No finding**	
Shannon Entropy 3^{rd} order		**No finding**		**No finding**		**No finding**	
Shannon Entropy 4^{th} order		**No finding**		**No finding**		**No finding**	
Maximum likelihood		×		**No finding**		×	
Akaike Information Criterion		×		**No finding**		×	
Bayesian Information Criterion		×		**No finding**		×	
Aspin-welch test		×		×		×	
Kolmogorov-Smirnov test		×		×		×	
Best uncertainty		×		×		×	

TABLE IV
General results for some criteria.

A. Other results

We calculated the Shannon Entropy for the 2^{nd}, 3^{rd} and 4^{th} order Markov chains given by the Synthetic Model. Calculation is made whatever the transition from one state to another. Unfortunately, these different orders did not yield interesting results. Nevertheless, we applied the Entropic Filter ([19]) in the 2^{nd}, 3^{rd} and 4^{th} order Markov chains. We obtained a maximum Entropy only for the 4^{th} order. In particular for all sequences ending with a **SP** symbol (see table V).

Symbols sequences				Entropy
OT	PM	PM	**SP**	0.352
OBS	PMV	PMV	**SP**	0.352
PMV	PMV	OT	**SP**	0.352
Others 4-tuple				$\leqslant$ 0.178

TABLE V
Symbols sequences (4^{th} order) with maximum Entropy, through Entropic Filter.

IV. Discussion

Through various criteria, we validated the best HMM framework on topology, learning & decoding algorithm and distribution:

- **Topology 2** of HMM studied for all criteria,
- **Baum-Welch** learning algorithm decoding by **Forward Variable** excepted for the likelihood criteria,
- **Normal** distribution.

A. Measurement of Shannon Entropy

1) First order measurement: Without a priori knowledge, we evaluated the relevancy of the signatures by measuring the Shannon entropy. We considered this signature as a 1^{st} order Markov Chain. Indeed, it is conceivable that some symbols can disturb the harmony of the signatures we explored. We removed discriminated symbols of zero entropy: Stop Production symbol (SP). We have also removed the most representative symbols, where entropy is maximal: Nothing To Report symbol (NTR). Then, we measured the Shannon entropy on the various topology, learning & decoding algorithm and distribution, to see if it was correlated with the Synthetic Model variables. To analyze the signatures, we made calculations with the Entropic filter and without this filter. This figure shows the most relevant framework according to Shannon. Namely, we verify that the best model (which provides the better estimation

of the degradation level [1]) obtains a good "Entropy" score through the entropic filter. The different distributions tested, give the following results. Indeed, the best framework is the Normal distribution, topology 2 and the Baum-Welch learning, decoding by Forward Variable, where entropy is maximal. According to the two Entropy principles, we expected a better Entropy for the Normal distribution. This is probably due to the biggest value of the Normal distribution. Insofar as the Normal distribution has the greatest differential entropy for a given variance [20].

2) Other order measurements: We calculated this previous criterion for 2^{nd}, 3^{rd} and 4^{th} order Markov Chains. Unfortunately, Shannon Entropy with different orders yielded a lot of small and identical results (not presented in the paper). We only found that Entropy is maximal for the following Markov chains: 2^{nd} order: **(NTR, NTR)**, 3^{rd} order: **(NTR, NTR, NTR)** and 4^{th} order: **(NTR, NTR, NTR, NTR)**. **NTR** is not a representative symbol since it is a no action symbol (Nothing To Report). Then, we applied the Entropic Filter through all symbol sequences. The 2^{nd} and 3^{rd} yielded no interesting results.

On the other hand, with 4^{th} order, we observed in table V that the **SP** symbol only appears in all sequences where Entropy is maximal. We can conclude that these particular signatures of three symbols, **induce a state of a Stop Production**. As we have seen before, this state is the S1 state of the considered automata. These sequences could give us an indication of the order of operations not to do, in order to avoid a critical situation. Indeed, sequences **(OT, PM, PM)**, **(OBS, PMV, PMV)** or **(PMV, PMV, OT)** could lead us to a breakdown that our research aims to avoid. According to the Markov assumption seen in § II-A1, knowledge of the three previous states could tell us about the next state. Hence, with these particular sequences, an expert should decide to undertake preventive maintenance actions, before a halt of the entire production line.

B. Likelihood effects

After implementing these symbols in different learning algorithms, we evaluated the log-likelihood of the frameworks studied: This graph shows that **topology 2** with a **Normal distribution**, gives the best results. Unfortunately, Baum-Welch learning Normal law and Segmental K-means learning Normal law are too close to each other to conclude about the best learning and decoding algorithm. Compared with the other distributions, the Segmental K–means algorithm has a bad distribution of symbols.

AIC and BIC are both similar methods of assessing model fit. Though AIC and BIC are driven and penalize free parameters in an effort to overcome overfitting, they are both maximum likelihood estimates. AIC only differs with log-likelihood by the penalty term (see equation 9). This additional term depends on the number of free parameters. So, there is no interest in presenting AIC results with the log-likelihood ones. On the other hand, BIC is AIC taking into account the number of data. Indeed, the penalty term of BIC depends on the number of data ("$k\ln(n)$" in equation 10). Hence, measurement of BIC is a better way to take into account our 1000 events. The results of BIC highlight the most relevant topology (**topology 2**) (figure 1(b)) and the most relevant distribution (**Normal**). That corroborates the results in [1]. On the other hand, like the log-likelihood, results don't clearly show differences between algorithms then, we can not conclude for the best learning and decoding algorithm.

C. Towards a stochastic generator

We first verified the randomness of the Synthetic Model generator. The results show that all $\text{p-value} \geqslant 0.01$ for all the parameters of all frameworks studied. According to the standard definition of a stochastic process: it is a family of random variables indexed by a parameter ([21]), so we can consider that sequences of the generator are random enough. In this case, the generator of the Synthetic Model gives a real stochastic process.

In this way, we can conclude that the Synthetic Model is not a biased model.

D. Statistical discussion

Next, we applied statistical tests on various frameworks. Aspin-Welch and Kolmogorov-Smirnov test are used to evaluate if two distributions are roughly equal. The results of the Aspin-Welch test (see paragraph II-F) show that only one value achieved the threshold limit of the test. Under this "p-value" limit, the two samples compared are considered to follow the same law. Given a set of observation sequences simulated by the Synthetic Model, we verified that the most relevant topology had the "goodness of fit" i.e. how well model fits the set of observations sequences. It appears now clearly that in a statistical way, **topology 2**, using the **Baum-Welch learning algorithm decoding by Forward Variable**, with a **Normal distribution** is the best one.

To confirm the validity of the last results, we performed the Kolmogorov-Smirnov test. This last test determines if two datasets differ significantly. It has the advantage of making no assumptions about data distribution. This test is less sensitive than the Aspin-Welch test and it is intended for use with samples having unequal variances.

The results give the same conclusions as previous tests for topology, learning & decoding algorithm and distribution of studied symbols.

Two different statistical tests: Aspin-Welch and Kolmogorov-Smirnov showed the same results: the most relevant model is obtained for **topology 2**, using the **Baum-Welch learning algorithm** with **Forward Variable** decoding, with the **Normal distribution** of stochastic symbols.

E. Epistemic uncertainties on the Synthetic Model design

We said that knowledge helps us to conceptualize a phenomenon as different models. Therefore, we can make more or less accurate predictions on the phenomenon. But what can we say about the reliability of these predictions? We can answer this question by given a margin error. But based on what concepts can we calculate this margin?... So many questions,

so many uncertainties. Hence the need to frame this margin of error.

According to [22], including uncertainties in the study, allows:

- Optimization of safety, delimiting qualitatively system failures,
- Conceptual optimization, to improve the system and the model.

Sensitivity analysis on various frameworks gave the epistemic uncertainties on different HMM studied. Epistemic uncertainties highlight that **topology 2**, **Baum-Welch learning algorithm** with **Forward Variable decoding** and a **Normal distribution** gives the lowest error rate. Obviously, when probability distributions have a finite standard deviation, the uncertainty of [18] converges inexorably to zero. Otherwise, we can see that the other topologies have higher uncertainties. Hence, we can say that our conceptual Synthetic Model has a better design with the following framework: **topology 2, Baum-Welch learning** and a **Normal distribution**. Epistemic errors due to human conceptualization are negligible with this framework. Unfortunately, we failed to establish any ranking between these various criteria. The results emphasized our choice for topology 2 by limiting error on the relevance of the symbols and therefore, in the model design.

V. Confrontation of the Synthetic Model with an Industrial Application

A. Context

The industrial case consists in studying a continuous process of bread production. When any failure occurs on the subsystems of the line that involves stopping the entire production line (a lot of bakery products are wasted). Therefore, preventive maintenance has been scheduled to prevent such cases. The factory operates all over the year, without any interruption. Teams are organized in shifts. Every maintenance operation or problem in the process must be recorded in the CMMS database. According to the internal maintenance policy, one day sampling is chosen. This regular temporal sampling makes it possible to be in a Markov process situation. We made comparisons with data for a period of 2 years. We had 611 records from their database (see table I).

B. Confrontation with Industrial data

To confront the Synthetic Model with the real industrial case, we compared data from this industrial CMMS, with data given by the Synthetic Model $Symb_U$ and $Symb_N$. The two tested distributions are the Normal and the Uniform one. To compare objectively with the most relevant symbols, we used the entropic filter. The densities of the compared models are given in table VI. Comparisons are made by using the Kolmogorov-Smirnov test as the adequacy between models and the correlation coefficient. Hence, measurements of the Synthetic Model adequacy reach up **93.75%** and **0.7554** for the correlation coefficient with a **Normal distribution**. We only found **5.62%** of adequacy for the **Uniform distribution**. By observing the different shapes in figure 4, it appears now clearly that the Normal distribution has almost the same shape as the real industrial case. Whereas the Synthetic Model with the Uniform distribution does not follow the industrial shape studied.

C. Synthetic Model adjustment

Density comparison with the industrial case in figure 4 leads us to make some adjustments. The Synthetic Model should be improved, to have a better correlation coefficient. Several improvements are suggested to fit the real industrial study, by exchanging some symbols:

- SEC $\rightarrow$ PMV (the right arrow means a symbol exchange),
- PM $\rightarrow$ SEC,
- PMV $\rightarrow$ PM.

Green dotted arrows in figure 4, show the proposed improvements. Hence, we obtain the red dotted shape which is now almost the same as the factory symbols densities.

D. Model adjustment results

Despite having the same results of the adequacy test (**93.75%**), the correlation coefficient increased up to **0.9611**. Indeed, our proposed changes made the Synthetic Model more efficient.

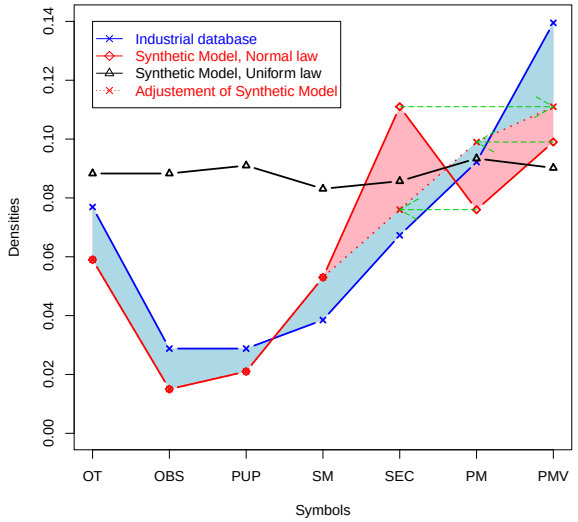


Fig. 4. Densities comparison, between the Synthetic Model and a real maintenance database from a bread production factory. You can see that the Synthetic Model with the Normal distribution, has almost the same shape as the real industrial data. Whereas the Synthetic Model with the Uniform distribution does not follow the industrial shape. We made an adjustment of the Synthetic Model (in red dotted), to have a better correlation coefficient. Green dotted arrows are the proposed changes for the new model. SEC $\rightarrow$ PMV, PM $\rightarrow$ SEC and PMV $\rightarrow$ PM. The blue area is the difference between the industrial database and the adjusted Synthetic model. The (blue + red) area is the difference without the adjustment.

VI. Conclusion

To model industrial processes, we built a Synthetic Model and studied relevancy of the proposed frameworks. Different topologies of HMM, different distributions and different

Symbols	Symbols Densities			
	Industrial Data	Model Normal law	Model Normal law **Adjusted**	Model Uniform law
OT	0.0769	0.0590	0.0590	0.0883
OBS	0.0288	0.0150	0.0150	0.0883
PUP	0.0288	0.0210	0.0210	0.0910
SM	0.0385	0.0530	0.0530	0.0831
SEC	0.0673	0.0760	0.1110	0.0857
PM	0.0922	0.0990	0.0760	0.0934
PMV	0.1395	0.1110	0.0990	0.0902
Correlation coefficient: Industrial Data VS Others Models		0.7554	**0.9611**	0.3750
Kolmogorov-Smirnov: Industrial Data VS Others Models		**93.75%**	**93.75%**	5.62%

TABLE VI
COMPARISON BETWEEN SYNTHETIC MODEL AND MAINTENANCE DATABASE FROM A BREAD PRODUCTION FACTORY (YEARS 2005–2006).

learning & decoding algorithms were used, with a view to finding the best framework. We have provided several different methods to help maintenance experts to select the best way to optimize their daily maintenance policies. We also presented an approach to assessing the sensitivity of the Synthetic Models. Good relevancy and good error rates for the following framework: **topology 2, Baum-Welch algorithm & decoding by Forward Variable and a Normal distribution**. Moreover, with the 4^{th} order Markov chain, we showed that we could provide potential critical sequences to the maintenance experts, before a breakdown occurs. After a confrontation with an industrial case, we proposed a model adjustment to improve the Synthetic Model. This makes it possible to apply these results as part of preventive maintenance applications. After having chosen the best framework, we should be able to improve their preventive maintenance policy by providing the maintenance experts with indicators of future potential failures. Indeed, when HMM output indicates an **orange level** (S2), the expert can decide to undertake preventive maintenance before a breakdown occurs. In our work on industrial breakdown prediction, determining the best framework of HMM is expected to significantly reduce the failure rate in production. Minimizing the failure rate will reduce dangerous human intervention in maintenance, especially in an unsafe working environment. Decreasing machine failures will furthermore increase processes reliability. These criteria could be used in relevance assessment for HMM modeling. Thus we could propose a best model among a candidate set.

ACKNOWLEDGMENT

We would like to thank our industrial partners: continuous food process (HARRYS), aluminium foundry (Montupet & F2R) and glass factory (Arc International).

REFERENCES

[1] P. Vrignat, M. Avila, F. Duculty, and F. Kratz, "Use of HMM for evaluation of maintenance activities," *IJAIS, International Journal of Adaptive and Innovative Systems, Vol. 1, Nos. 3/4*, pp. 216–232, 2010.

[2] 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*, vol. 41, no. 1, pp. pp. 164–171, 1970.

[3] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceeding of the IEEE, 77(2) SIAM interdisciplinary journal*, pp. 257–286, 1989.

[4] B. H. Juang and L. R. Rabiner, "The segmental k-means algorithm for estimating parameters of hidden markov models," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 38, no. 9, Sep. 1990.

[5] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.

[6] D. Hocker, L. Xiaohu, and S. S. Iyengar, "Shannon entropy based time-dependent deterministic sampling for efficient on-the-fly quantum dynamics and electronic structure," *J. Chem. Theory Comput.*, pp. 256–268, 2011.

[7] V. Chandrasekaran, J. K. Johnson, and A. S. Willsky, "Maximum entropy relaxation for graphical model selection given inconsistent statistics," *Laboratory for Information and Decision Systems, Massachusetts Institute of Technology Cambridge, MA 02139*, 2007.

[8] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York, NY, USA: Wiley-Interscience, 1991.

[9] C. Olivier, F. Jouzel, A. El Matouat, and P. Courtellemont, "Prediction with vague prior knowledge," *Communications in Statistics 25- Theory and Methods*, pp. 601–608, 1996.

[10] J. Shang and J. E. Cavanaugh, "Bootstrap variants of the akaike information criterion for mixed model selection," *Comput. Stat. Data Anal.*, vol. 52, pp. 2004–2021, 2008.

[11] F. Forbes and N. Peyrard, "Hidden markov random field model selection criteria based on mean field-like approximations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1089–1101, Sep. 2003.

[12] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, pp. 461–464, 1978.

[13] M. Avila, "Optimisation de modèles markoviens pour la reconnaissance de l'écrit," Ph.D. dissertation, Université de Rouen, 1996.

[14] A.-K. K. Seghouane and S.-I. Amari, "The AIC criterion and symmetrizing the Kullback-Leibler divergence." *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 18, no. 1, pp. 97–106, Jan. 2007.

[15] J. Steinebach, E. L. Lehmann, and J. P. Romano, "Testing statistical hypotheses," *Metrika*, vol. 64, pp. 255–256, 2006.

[16] Z. Drezner, O. Turel, and D. Zerom, "A modified kolmogorov-smirnov test for normality," *Communications in Statistics - Simulation and Computation*, vol. 39, pp. 693–704, 2010.

[17] L. Held, K. Rufibach, and F. Balabdaoui, "A score regression approach to assess calibration of continuous probabilistic predictions," *Biometrics*, vol. 66, no. 4, pp. 1295–1305, 2010.

[18] L. Pibouleau, *Assimiler et utiliser les statistiques*. Ellipses Marketing, technosup, 2010.

[19] B. Roblès, M. Avila, F. Duculty, P. Vrignat, S. Begot, and F. Kratz, "Methods to choose the best Hidden Markov Model topology for improving maintenance policy," in *9th International Conference of Modeling, Optimization and Simulation - MOSIM'12*, Bordeaux, France, Jun. 2012, p. 1.

[20] M. Payaró and D. P. Palomar, "Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector gaussian channels," *IEEE Trans. Inf. Theor.*, vol. 55, no. 8, pp. 3613–3628, Aug. 2009.

[21] Y. Zheng and R. Xu, "A composite stochastic process model for software reliability," in *Proceedings of the 2008 International Conference on Computer Science and Software Engineering - Volume 02*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 658–661.

[22] B. Iooss and M. Ribatet, "Global sensitivity analysis of computer models with functional inputs," *Reliability Engineering & System Safety / Reliability Engineering and System Safety*, vol. 94, no. 7, pp. 1194–1204, 07 2009.