An Extension Graphical Duration Models Integrating Conditional Sojourn Time Distributions

J. FOULLIARON*, L. BOUILLAUT*, P. AKNIN**, and A. BARROS***

* IFSTTAR - 14-20 Bd Newton 77420 Champs-sur-Marnes, France
** IRT-SystemX - 8 Avenue de la Vauve, 91120 Palaiseau, France
*** Department of Production and Quality Engineering, 7491, Trondheim, Norway

Abstract: System degradation modelling is a key problem when performing any type of reliability study. It is used to determine the quality of the computed reliability indicators and prognostic estimates. However, the mathematical models that are commonly used in reliability studies (Markov chains, gamma process. etc.) make certain assumptions that can lead to a loss of information regarding the degradation dynamics. Many studies have shown how Dynamic Bayesian Networks (DBNs) can be relevant in representing complex multicomponent systems and in performing reliability studies. In a previous paper [10], Donat et al. introduced a type of degradation model based on DBNs called a graphical duration model (GDM) for discrete-state systems to represent a wide range of duration models. This paper introduces a new type of degradation model based on the GDM approach that integrates the concept of conditional sojourn time distributions (CSTDs) to improve the degradation modelling. It introduces the possibility of considering many degradation dynamics simultaneously. It allows the degradation modelling to be adapted based on newly available observations of a system to account for changes in dynamics over time. A comparative study of the presented methodology and the GDM approach was conducted using simulated data to demonstrate the advantages of this new approach in performing prognostic computations. Only two coexisting dynamics are considered in the experiments for the sake of simplicity.

Keywords: *Reliability study, Degradation modelling, Dynamic Bayesian Networks, Conditional Sojourn-Time Distributions, Prognostic computations*

1.   Introduction

Over the last fifty years, the systems used in most industrial fields have increased in complexity. A failure can lead at best, to poorer performance in some parts of the system, and, at worst, to a complete shutdown that can pose a critical security risk. When a failure is detected, it is necessary to find its cause, obtain new components for replacement and then repair the damaged part. All these steps can cause the system to be unavailable for quite a long time which costs money. Thus, this is necessary both to increase the availability of such systems and to guarantee a high level of security while keeping maintenance costs low. For this reason, optimizing the maintenance of complex systems has become a key concern. To optimize systems maintenance, industrial managers need to perform reliability studies. Such studies rely on computed reliability indicators and prognostic estimators such as the failure rate (FR) and the mean time to failure (MTTF). In particular, there is increasing interest in predictive maintenance strategies that attempt to avoid unexpected failures. These strategies rely on the computation of prognosis indicators such as the failure time (FT) or the remaining useful life (RUL).

---

*Corresponding author's email: jfoulliaron@gmail.com

All these indicators and the accuracy of their estimation depend on the way in which the degradation process of the considered system is modelled. Two major types of degradation models are commonly used in reliability studies. The first is a set of analytical degradation models derived from mechanics [19]. However, such models are generally difficult to validate in practice. Therefore, the class of models known as stochastic models constitutes the most commonly used approach. Many projects have been conducted based on such models [2]. Some stochastic models are directly based on probability distributions, such as the gamma process [24; 14], the Bertholon model [4] and the exponential distribution, whereas others use probabilistic graphical models (PGMs), such as neural networks [26], stochastic Petri nets [30], hidden Markov chains (HMMs) and dynamic Bayesian networks (DBNs). DBNs are a type of PGM that are used in many areas of research, such as speech recognition [31; 20], medical diagnosis [1; 7], industrial system diagnosis [16; 17], and degradation and maintenance modelling [21; 5; 22].

DBNs offer a particularly powerful and intuitive way to represent complex discrete multicomponent systems [25; 15]. Many papers have shown how they can be relevant in performing reliability studies [3; 28; 29]. The simplest way to model the degradation process of a system using a DBN is to use a simple Markov chain. However, this approach implies the strong assumption of a geometric distribution of the sojourn times in each state. To overcome this limitation, a specific type of DBN called "Graphical duration model" (GDM) has been proposed [10]. Such DBNs based on duration models [23], can be used to model the degradation of discrete states systems using any discrete sojourn-time distribution (STD). However, the GDM approach assumes that the sojourn times elapsed in each state are independent of each other. This assumption can lead to a loss of information regarding the degradation process and dynamics. This paper proposes an extension of the GDM approach called Conditional graphical duration model (CGDM) that does not require this assumption. The existence of a dependence between sojourn times give rise to the concept of conditional STDs (CSTDs), which can be used to consider multiple degradation processes and follow them through successive degradation states. This new type of model can describe a system whose overall dynamic is a mixture of many types of degradation dynamics and can dynamically detect changes in these degradation dynamics. Its advantage is that it offers the ability to perform failure prognosis computations with higher accuracy than those obtained using standard degradation models. The proposed methodology applies to periodically observable discrete state systems. Such systems are commonly met for example in the railway or road infrastructure fields.

The first section of this paper presents the problematic of the study and the main assumptions, then, it briefly introduces DBNs and GDMs. The second section presents the proposed model (CGDM), in particular, the concept of CSTD is introduced. Finally, before a discussion of the conclusions and future prospects for related research, a comparative study of the CGDM approach with respect to the original GDM approach is presented. In a final section, a prognostic computation is performed by following the evolution of the FT distribution of a system of interest and computing the risk of error.

2.   Context and Notations

In the paper, to simplify, we will consider a one single component system. The methodology proposed on this paper applies to systems that have a finite discrete state space. The

.

degradation states can be ordered in ascending order according to the level of gravity of the degradation. The last state will be considered as the failure state meaning that it is an absorbent state that requires a corrective maintenance operation to be left.

In this paper, considering a component with n states, the component state evolution will be modelled by a stochastic process denoted $X^{(t)}$, where $X^{(t)} \in E = \{x_1, x_2, \ldots, x_n\}$, where x_1 is the "new" state, and x_n is the "failure" state. Time is modelled as discrete variable. It is assumed that the degradation process of the component is increasing monotone. It means that once in a given state, the system can only go a more degraded state. Thus, the following inequality is respected: $\forall t$, $X^{(t+1)} \geq X^{(t)}$.

Systems with such properties are common in many industrial fields, especially in transportation systems applications such as in the railway field or in the roadway field. For example, according to the UIC (Railways International Union) norms, the state space of a rail is divided into four discrete states denoted $\{OK, X_1, X_2, S)$, where *OK* denotes a rail without any measurable defect, X_1 and X_2 are intermediate degradation states defined by a range in respect of the size of an observed internal rail crack. Finally, the last state denotes a visible rupture of the rail and is considered as the failure state.

In our framework, the degradation of the system is considered from $t=0$. At each new diagnosis, a new estimation of the real state of the system is available. The objective is to compute an estimation of the failure time of the system and update this estimation each time a new diagnosis is available. It is assumed that an experiment feedback database is available with P observation patterns, denoted $\{O_i\}_{i=1\ldots P}$. A degradation pattern is a sequence of periodic observed states of the system until the failure state. Then, $O_i = (o_i)_{oi \in E}$. If a four states system is considered, $E = \{x_1, x_2, x_3, x_4\}$ where x_4 denotes the failure state.

The general framework of the degradation processed considered in this study is given in Figure 1, where $t_{xi \to xj}$ denotes the time when a state transition occurs between states x_i and x_j. Due to the discretisation of the auscultation process, these transitions will be never really observed. The accuracy of the prognosis estimation depends on the accuracy of the degradation process modelling.

This paper aims to propose a degradation model that could take into account the existence of several degradation dynamics that could mix together in the degradation process. This model, named *Conditional Graphical Duration Model* is an extension of the graphical duration model presented by Donat in [10]. The next section will introduce the different steps of the degradation modelling by a CGDM.

3.   Mathematical Framework of the Proposed Degradation Model

The graphical duration model and the proposed extension of this model lay on the formalism of Dynamic Bayesian Network. This formalism is briefly explained in this first subsection.

3.1   From Bayesian Networks to Dynamic Bayesian Networks

Probabilistic Graphical Models [15], or Bayesian networks (BNs), are mathematical tools relying on both the probability theory and the graph theory. They allow to qualitatively and quantitatively representing uncertainty. Basically, BN are used to compactly describe the joint distribution of a collection of random variables $V=(X_1, \ldots, X_N)$ taking their values in $\Xi = \{\Xi_1, \ldots, \Xi_N\}$. In this section and the rest of the paper, only discrete finite Bayesian Networks will be considered.
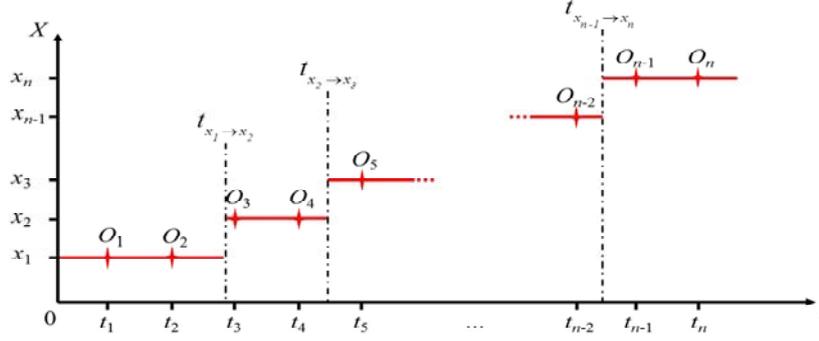
Figure 1: Degradation Process Framework

Formally, a BN denoted by ς is defined as a pair $(\Theta, \{p_n\}_{1 \leq n \leq N})$ where $\Theta = (\xi, \varepsilon)$ is a Directed Acyclic Graph (DAG) and $\{p_n\}_{1 \leq n \leq N}$ a set of Conditional Probability Distributions (CPD) associated with the random variables. These distributions aim to quantify the local stochastic behavior of each variable. The graph nodes and the associated random variables being both represented by $\xi = \{X_1, \ldots, X_N\}$. ε is the set of edges encoding the conditional independence relationships among these variables. Finally, Θ is said to be the qualitative description of the BN. Besides, both the qualitative (i.e. Θ) and quantitative (i.e. $\{p_n\}$) parts can be automatically learnt, if some complete or incomplete data or experts opinions are available [15].

Using BN is also particularly interesting because of the easiness for knowledge propagation through the network. Indeed, various inference algorithms allow computing the marginal distribution of any sub-set of variables. The most classical one relies on the use of a junction tree [18]. Finally, note that such modelling is able to represent dynamic systems (e.g. which contain variables with time dependent distributions) via the DBN solution [23]. A DBN is a convenient extension of the BN formalism to represent discrete sequential systems. In-deed, DBNs are dedicated to model data which are sequentially generated by some complex mechanisms (time-series data, bio-sequences, number of mechanical solicitations before failure…). It is therefore frequently used to model Markov chains. This DBN is described by the probabilities that quantify the transitions from one state of Ξ to another. More precisely, a DBN defines the probability distribution of a collection of random variables $(V^{(t)})_{t \geq 1} = (X_1^{(t)}, \ldots, X_N^{(t)})_{t \geq 1}$ where t is the discrete time index.

It is commonly considered that the stochastic process used in a DBN has the first order Markov property. It means that the states of all random variables at time t only depend on states at the previous time. In other words, in the graph, the dependence arrows from a node on slice t only go to nodes on the same slice or to other nodes on slice $t+1$. It is assumed that the graph is homogeneous (conditional probability of $V^{(t+1)}|V^{(t)}$ does not depend on t).

From these assumptions, it is just necessary to formally define a first order DBN (i.e a 2 slice 2-DBN) like this:
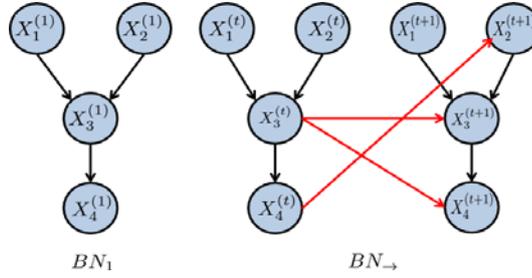
- A static model, BN_1, at initial time:

$$P(X_1^{(1)}, X_2^{(1)}, .., X_N^{(1)}) = \prod_{k=1}^{N} P(X_k^{(1)}|pa(X_k^{(1)}))$$

(1)

- A transition model, $BN_{\rightarrow}$:

$$P(X_1^{(t+1)}, X_2^{(t+1)}, .., X_N^{(t+1)} | X_1^{(t)}, X_2^{(t)}, .., X_N^{(t)}) = \prod_{k=1}^{N} P(X_k^{(t+1)} | pa(X_k^{(t+1)}))$$

(2)

Where $pa(X_k)$ denotes the set of parent nodes of X_k.

Figure 2 illustrates this property, introducing a two slice DBN representing the distribution of 4 random variables denoted ($X_1^{(t)}, X_2^{(t)}, X_3^{(t)}, X_4^{(t)}$).

Figure 2: A Two Slice DBN

In this paper, all considered nodes have discrete state spaces. The disadvantage is that there are many parameters that must be obtained for each node to allow the model to be built. Moreover, one or more nodes have a large state spaces and many parents, the size of the CPDs can be very large, causing the inference computations to be very heavy. The Markov property implies that the sojourn times for each state of each node are geometrically distributed. To overcome this limitation, a particular type of DBN is used, as described in the next section. In the rest of the paper, one node denoted $X^{(t)}$ models the state of the considered one single component system.
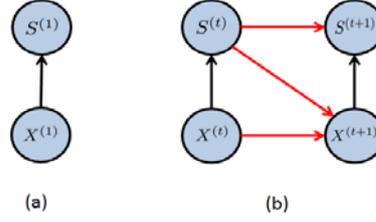
3.2   Graphical Duration Models

This section briefly reviews the concept of graphical duration models (GDM), which is explained in detail in [10]. A GDM is a specific type of DBN in which each considered variable $X^{(t)}$, modelling the state evolution of a component, has its own duration variable denoted by $S^{(t)}$, that controls the sojourn time spent in each state of X. Then, $S^{(t)}$ represents the remaining time, at time t, in state $X^{(t)}$ before the next transition of $X^{(t)}$. If the variable X has n states, the degradation process until failure is thus completely defined by the sojourn time distributions (STD) of the n-1 first states.

Because, a discrete-time approach is considered, $S^{(t)} \in \{1, …, T_S\}$, where T_S is the maximal possible sojourn time among all states of $X^{(t)}$ and $X^{(t)}$ has a discrete state space $X^{(t)} \in \{x_1, …, x_n\}$. The STD in state x_i is a discrete probability function denoted by ψ_{xi}.

A representation of the GDM is shown in Figure 3. The distribution of $X^{(t+1)}$ is conditioned by $X^{(t)}$ and $S^{(t)}$. At each time step, the value of $S^{(t)}$ decreases by 1 unit. While the value of $S^{(t)}$ is different from 1, the state of $X^{(t)}$ is blocked. If $S^{(t)}$ is equal to 1, then $X^{(t)}$ is free to transition to another state, and the new value for $S^{(t+1)}$ is chosen in accordance with the STD associated with the new state of $X^{(t+1)}$.

The CPD of node $S^{(t)}$ represents the conditional probability $P(S^{(t)} | S^{(t-1)}, X^{(t)})$. The initial and transition laws of $X^{(t)}$ and $S^{(t)}$ are defined below :

Figure 3: Structure of the Specific DBN Modelling a Graphical Duration Model

- Initial laws: (represented in Figure 3a)

$$P\left(X^{(1)} = x\right) = Q_1(x) \tag{3}$$

$$P\left(S^{(1)} = s \mid X^{(1)} = x\right) = F_1(x,s) = \Psi_x(s) \tag{4}$$

- Transition laws : (represented in figure 3b)

$$P\left(X^{(t+1)} = x \mid X^{(t)} = x', S^{(t)} = s'\right) = Q_\rightarrow(s',x',x)$$
$$= \begin{cases} Q_{sys}(x',x) & \text{if } s' = 1 \\ I(x',x) & \text{if } s' \geq 2 \end{cases} \tag{5}$$

$$P\left(S^{(t+1)} = s \mid X^{(t+1)} = x, S^{(t)} = s'\right) = F_\rightarrow(s',x,s)$$
$$= \begin{cases} F_{sys}(x,s) = \Psi_x(s) & \text{if } s' = 1 \\ T(x',x) & \text{if } s' \geq 2 \end{cases} \tag{6}$$

Where:

- Q_1 is the initial distribution of the system state at time t=0
- Q_{sys} is an $n*n$ matrix that controls the probability for a state to proceed to the next one when a transition is occurring.
- I is the $n*n$ identity matrix.
- Ψ_x is the sojourn time distribution in state x. It is a discrete probability distribution taking its value in the set $\{1, 2, …, T_S\}$
- T is a deterministic matrix of size N_S*N_S, defined such that $T(s', s)$=1 if $s=s'$-1 and 0 otherwise
    - F_1 is the $n*T_S$ matrix that controls the sojourn time drawn at time t=1. The row $F_1(x, :) = \Psi_x$ is the STD for state x.
    - F_{sys} is a transition matrix that contains all STDs in each state. For each degradation state, the probability distributions are learnt with the maximum likelihood method (MLM). Because the approach is completely discrete, the STDs (Ψ_x) are the frequency histograms of the sojourn times in each state. In this paper, the GDM concept is slightly modified. Unlike in the original model described in [10], the context node denoted by Z_t is not considered in this paper and node $S^{(t)}$ does not depend on $X^{(t-1)}$. This modification is made to simplify the representation of the model and the notation. Context nodes can still be used if necessary; doing so will simply increase the complexity of the inference computations.

4.   Extension to Conditional Sojourn Time Distributions

4.1   Introduction and Concept

The graphical duration models make the strong assumption that different sojourn times are independent of each other. An equivalent assumption is also adopted in other types of models that are commonly used in reliability studies.

In the continuous case, many degradation models use stochastic processes belonging to the class of Lévy processes (for example, the gamma process) to model the degradation process. An important assumption is that the increments in the process are stationary and independent. The distribution of the increments in the degradation process between times t_2 and $t_2 + \Delta t$ depends only on Δt. This assumption is equivalent to the assumption of sojourn-time independence in the discrete case. In the case of prognosis computations with new incoming observations from diagnosis devices such as that performed in our framework, however, this assumption could not be correct and could lead to a loss of information about the degradation process.

To explain this statement, let us consider a system with discrete states and a given STD associated with one state. The variance of this distribution arises from a combination of several unknown factors (internal or external) that affect the sojourn time on this state. Without considering the hazards associated with material production, three major scenarios can be considered to explain this variance:

- There are many degradation dynamics that lead to the same observed failure. These dynamics become mixed with each other during the learning step, leading to a large variance in the obtained STD.
- Certain disruptive phenomena can particularly affect the dynamic of certain observations in the database.
- The visible degradation process on the system scale can be a function of several degradation processes of hidden subcomponents of the system.

The main idea is therefore to regard the degradation process as the product of several degradation dynamics. A dynamic is a degradation trend that is linked to a configuration of external or internal factors. However, this assumption implies that the STDs in each state must be dependent on each other. The reason for this is illustrated in figure 4. Consider a periodically observed four-state system with periodic available observations of its real state. The three learned STDs are represented by the blue curves. At the first observed transition from state x_1 to state x_2, a sojourn time in x_1 can be deduced. Suppose that a very short sojourn time has been observed in x_1, much smaller than the mean of the associated STD (case O_1). In this case, the system is likely to follow an "accelerated" degradation dynamic that has no reason to stop affecting the system behaviour and should therefore influence the degradation process until failure.

It can thus be expected that the sojourn times that will elapse in x_2 and x_3 will also be small. In other words, the STD in x_2 given the sojourn time elapsed in x_1 should be different and more centered on small values (green curve) than the global distribution that was initially learned (blue curve). By contrast, if the observed sojourn time in x_1 is greater than the average, then a slower degradation dynamic (case O_2) could be expected and the next sojourn time is similarly likely to be longer. To explain it in other words, the sojourn time in any given state is not just seen as a single statistical quantity but as an indicator, at a given period, describing a global degradation trend (which is called "degradation dynamic") that have effects until failure and particularly on all sojourn times in next states.
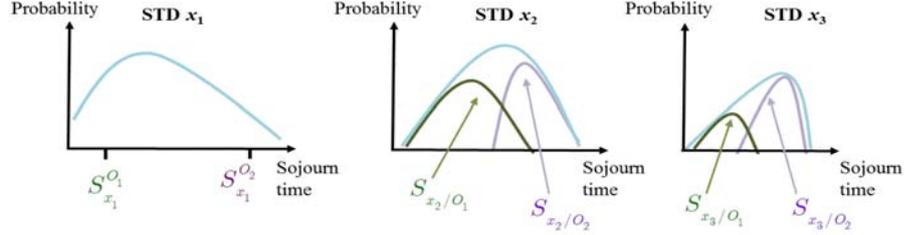
Figure 4: Example of Non-Independent Sojourn Time Distributions

This reasoning leads to reject the independence assumption on sojourn times. In this example illustrated on Figure 6, the existence of two degradation dynamics (fast and slow, respectively in green and purple) has been considered. Consequently, any information regarding the sojourn time in a state leads to an inference regarding the active degradation dynamic. This additional information modifies the considered STDs in subsequent states, which implicitly become dependent and conditioned to a given active degradation dynamic.

The fundamental idea is thus to attempt to detect, as soon as possible, which active degradation dynamic is currently affecting the system and then use this information to predict the next sojourn time using CSTDs given the active dynamic.

The primary expectation is to obtain CSTDs with smaller variances to improve the quality of the computed prognostic estimators. Additionally, we wish to be able to dynamically modify the considered active dynamic, when a particular disruption is detected. Therefore, the active dynamic is re-evaluated each time a new observation becomes available. The estimation of the new active dynamic is computed from the best estimate of sojourn time elapsed in the state in which the system existed during the last observation.

### 4.2 Definition of Condition Graphical Duration Models

In the CGDM, a sojourn time distribution in any state x is not just a discrete distribution as in the GDM but is seen as the mixture of many conditional sojourn time discrete distributions (CSTD). Each component of the mixture is a sojourn time distribution conditionally to a specific active degradation dynamic. A CSTD in a state given the kth dynamic is denoted $\Psi_{x/mk}$. By denoting n_M the number of existing dynamics and ω_k the weight of each component of the mixture, the sojourn time distribution of any state x is given by:

$$\forall x \in E, \Psi_x = \sum_{k=1}^{n_M} w_k \Psi_{x/m_k} \tag{7}$$

An example of sojourn time distribution in the CGDM is represented in figure 5 in the case of considering two degradation dynamics ($n_M = 2$) and equal weights $\omega_1 = \omega_2 = 0.5$. The next subsection details how the STD are learnt for each degradation states.

4.3 Learning of Conditional Graphical Duration Models

The learning process is divided into two different stages. The first involves the learning of the mixture model for the first degradation state and the second focuses on the other states. The next paragraphs detail the learning process of conditional probability distributions proposed for the specific DBN named Conditional Graphical Duration Models.
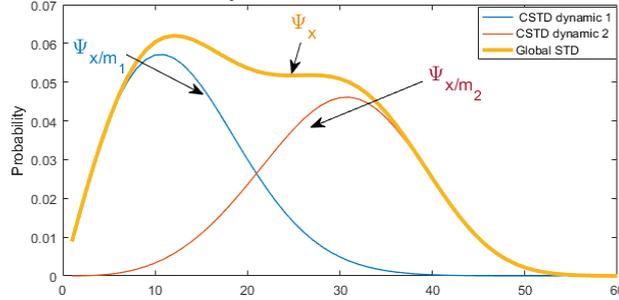
Figure 5: STD and CSTDs in State X Considering Two Degradation Dynamics

Consider an observation database denoted by $\{O_i\}_{i=1..n}$, with n degradation patterns, where O_i denotes the ith degradation pattern. The sojourn time in state x_j associated with the ith observation pattern is denoted S_i^j.

The mixture model for the first degradation state x_1 is learned via an EM algorithm from the set of $\{S_i^1\}_{i \in \{1...n\}}$. The number of mixtures components, denoted by n_M, can be selected sing a BIC or AIC criterion [27]. Thus, the STD in x_1 is:

$$\Psi_1 = \sum_{k=1}^{n_M} w_k \Psi_{1/m_k}$$

(8)

Each mixture component denoted by $\Psi_{1/mk}$ with $k \in \{1, \ldots, n_M\}$ corresponds to an identified degradation dynamic. $\Psi_{1/mk}$ is the CSTD in state x_1 given dynamic m_k. ω_k is the weight of the kth component of the mixture. Then, the observation database is split into n_M subsets $\{K_p\}_{p=1..nM}$. For each observation pattern (O_i), the sojourn time in state x_1, denoted S_i^1, is used to determine the most likely dynamic associated with this pattern (denoted by m_{Oi}) using the a posteriori probabilities deduced from the mixture model generated by the EM algorithm [9].

$$m_{O_i} = \arg \max_k \Psi_{1/m_k}\left(S_1^i\right)$$

(9)

$$K_p = \left\{O_i, m_{O_i} = p\right\}$$

(10)

For the subsequent states, the CSTDs are learned by applying the maximum likelihood method (MLM) [11] to the subsets of observation associated with the relevant dynamics. For example, let $\Psi_{x2/m1}$ be the CSTD in state x_2 given dynamic 1. $\Psi_{x2/m1}$ will be learnt using the MLM applied to the observation subset K_{m1}. At the end of the learning process, each state has n_M CSTDs given each considered dynamic. The global STDs for each state different from x_1 are generated as weighted mixtures of CSTDs using the weights (denoted by ω_k) obtained from the first state mixture as follows:

$$\forall x, \Psi_x = \sum_{k=1}^{n_M} \omega_k \Psi_{x/m_k}$$

(11)

Note that the learning method presented above assumes that once a system has entered one dynamic, it cannot change to another. Another possible approach could be to learn the mixture models for all states using the EM algorithm. It is important to explain why this approach is not used here. If each mixture model for each state is learned independently using the EM algorithm, then the fundamental concept of dependence between the STDs is

not integrated into the learning process. In this case, it is very difficult to connect the obtained mixture models together and, for example, for a given mixture model in state x, to know which component of the mixture is the CSTD associated with a given dynamic m_k. The mixtures that would be obtained for other states would have no reason to have the same number of components as the mixture for state x_1. If a particular number of components were to be imposed to address this issue, then the random shape of the obtained mixtures would make it difficult to follow the effect of a given active degradation dynamic through all states.

4.4   Representation

The integration of the CSTD concept requires the addition of a specific type of node, denoted by $C^{(t)}$, to the standard GDM structure. This node represents the active dynamic considered at time t. Figure 6 illustrates the extension of the GDM approach that includes the CSTD concept. Some text is provided to explain the dependencies that are indicated by the arrows. The initial and transition laws of node $C^{(t)}$ are defined as follows:
   - Initial law:

$$P\left(C^{(1)} = c \mid X^{(1)} = x, S^{(1)} = s\right) = C_1(x,s,c)$$
$$= \begin{cases} 1 \ if \ c = \underset{k}{\arg\max} \ \omega_k \Psi_{x/m_k}(s) \\ 0 \ otherwise \end{cases} \tag{12}$$
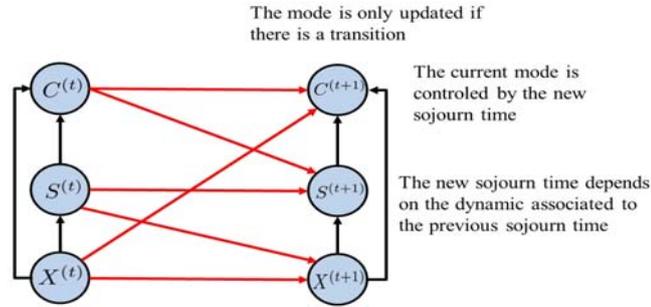
Figure 6: Graphical Structure of a Conditional Graphical Duration Model

C_1 is an $n * T_S * n_M$ deterministic matrix that contains the a posteriori probabilities of the mixture models generated by the EM algorithm.

   - Transition law :

$$P\left(C^{(t+1)} = c \mid C^{(t)} = c',, X^{(t)} = x', X^{(t+1)} = x, S^{(t+1)} = s\right) = C^{\rightarrow}(c',x',x,s,c)$$
$$= \begin{cases} C_{dyn}(x,s,c) = \underset{k}{\arg\max} \ \omega_k \Psi_{x/m_k}(s) & if \ x \neq x' \\ Id(c',c) & if \ s' \geq 2 \end{cases} \tag{13}$$

Each time a state transition occurs, the new most likely active dynamic associated with the newly observed sojourn time is chosen according to the deterministic matrix C_{dyn}, which is built in a similar manner to the matrix C_1 from the a posteriori probabilities of the mixture model in each state. The addition of the node $C^{(t)}$ leads to a modification of the distributions of node $S^{(t)}$ as follows:

$$P\left(S^{(1)} = s \mid X^{(1)} = x\right) = F_1(x,s) = \Psi_x(s) \tag{14}$$

$$P\left(S^{(t+1)} = s \mid S^{(t)} = s', C^{(t)} = c', X^{(t+1)} = x\right) = F^{\rightarrow}(s',c',x,s)$$
$$= \begin{cases} F_{cond}(c',x,s) = \Psi_{x/c'}(s) & \text{if } s' = 1 \\ T(s',s) & \text{if } s' \geq 2 \end{cases} \tag{15}$$

F_{cond} is an $n * T_S * n_M$ matrix, in which the row $F_{cond}(c', x, :)$ is the CSTD in state x given the dynamic c', denoted $\Psi_{x/c'}$. Unlike in the GDM approach, the new sojourn time after a transition is chosen not according to the sojourn time in the new state, but according to the conditional sojourn time in the new state $X^{(t+1)}$ given the active mode $C^{(t)}$, denoted $\Psi_{x/c'}$. Thus, the sojourn time in the new state, $S^{(t+1)}$, is chosen according to the CSTD of the new current state $X^{(t+1)}$ given the last detected active dynamic $\Psi_{X(t+1)/C(t)}$.

5.   Application Case Study

To demonstrate the advantage of considering CSTDs, this section presents a comparison of the GDM and the CGDM approaches in a "basic" application case study. The inference computations performed to obtain the reliability indicators will be presented. The purpose of this final section is to show that the prognostic computations are more accurate in many situations when the CGDM approach is used.

5.1   STD Simulations

A single component system with four discrete states is considered, where the state space is denoted by E = {ok; x_1, x_2, F}. The degradation process is assumed to be monotonically increasing: from a given state, the system can proceed only to a more degraded state. State ok consists in the "good as new" state, and F is the "failed" state. The states of the system are assumed to be periodically observable. For this analysis, a discrete observation database was therefore simulated. To show the interest of conditional sojourn time distributions, data was simulated with a mix of two degradation dynamics. The sojourn time distributions for each state were assumed to be a mixture of two Weibull distributions with the following parameters (where β and η are respectively the scale parameter and the shape parameter):
   - state ok : dynamic 1: W(β=15, η= 2) and dynamic 2: W(β=33, η = 6)
   - state x_1: dynamic 1: W(β= 10, η= 6) and dynamic 2: W(β=25, η= 9)
   - state x_2: dynamic 1: W(β= 5, η= 6) and dynamic 2: W(β=15, η= 15)

To test the ability of the CGDM to take into account cases with the existence of many degradation dynamics, the Weibull parameters were chosen so that there exist two visible degradations dynamics. Thus, the components of the mixture were chosen to be enough disjoint. The weights for each dynamic were both chosen to be equal to 0.5 for the following tests. Different weights could have been chosen, but the aim is to eventually see the ability to detect a dynamic change. To increase the probability to observe such behaviour, equal weights have been chosen. The graphical representation of such simulated distributions is given for state ok in Figure 7.

From the STD presented above, a discrete database was simulated. The period between two observations was set to 5 time units. The next section introduces the process of constructing the discrete observation database.
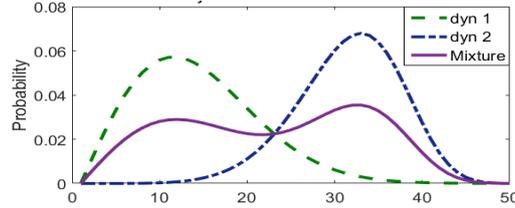
Figure 7: Simulated STD for State ok

5.2   Discrete Learning Database

The procedure considered for constructing a 10000 observation patterns database, illustrated in Figure 8, can be divided into three steps:

- Simulation of the transition times between states deduced from a sojourn time sampling in accordance with the associated Weibull distribution. $t_{x \to y}$ is the notation for the transition time from state x to state y.
- Creation of the observation pattern
- Discretization: separation of the observation pattern into discrete periods Δt
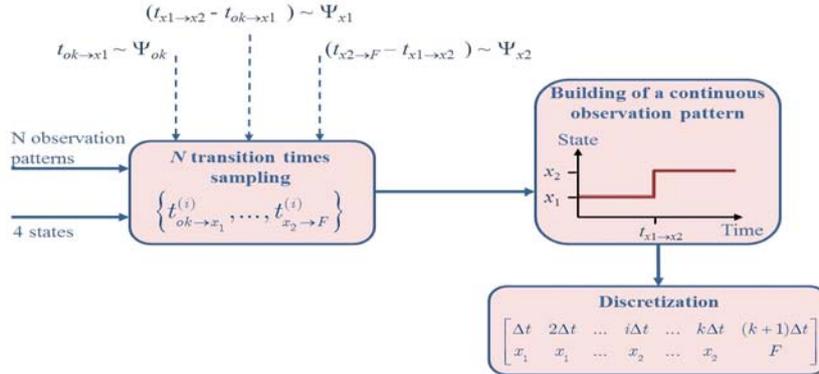
**Figure 8:** Building Process of Observation Patterns

5.3   Learning Process of Conditional Sojourn Time Distributions

For the first state, the parameters of the mixing model were learned using the EM algorithm. Then, the dynamic most likely to be associated with each observation pattern in the database was identified using the a posteriori probabilities. The CSTD for subsequent states were learned by applying the MLM to the subset of observation patterns associated with each dynamic. The fact that the observations are periodic induces a risk of error in the sojourn-time estimations. As an illustration of this problem due to censured data induced by the periodic auscultations, consider the periodic observation pattern shown in Figure 9.

Let Δt be the time between two observations and let S_{x1} be the sojourn time in state x_1. As illustrated, the real sojourn time in x_1 is such that $S_{min} < S_{x1} < S_{max}$. The potential error in the estimation is less than $2\Delta t$. To associate a sojourn time with the presented pattern for state x_1, two methods can be considered:

- The first consists of choosing the sojourn time by means of uniform sampling on the interval $[S_{min}; S_{max}]$

- The second consists of considering the middle of each transition interval and then completing the discrete distribution this obtained through first-order interpolation.
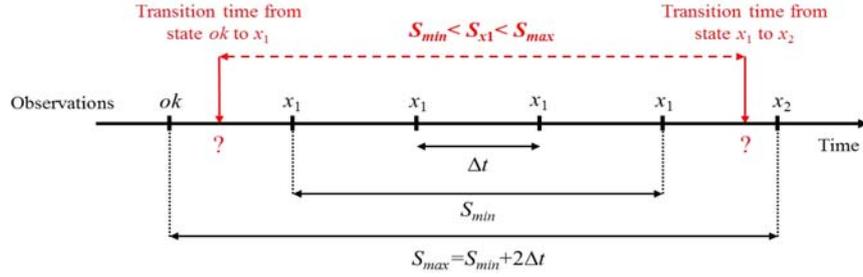
Figure 9: Error in Sojourn Time Estimation due to Periodic Observations
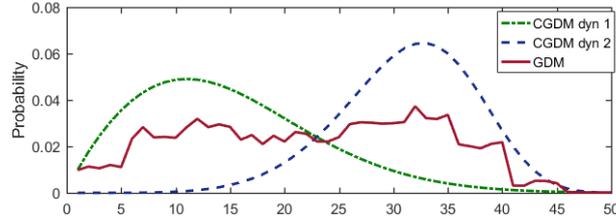
Figure 10: STD and CSTD Learned for States ok

For the analysis presented below, the first method was used. Then, the sojourn-time state space was discrete, and the probability of each possible sojourn time was set equal to its frequency of appearance in the simulated database. Figures 10 illustrates the learning phase of STF in state *ok* using the either the standard GDM approach or the proposed CGDM. Because the original observation database was built using a mixture of two degradation dynamics, as expected, the distributions learned using the CGDM are closer to the original distributions than those learnt using the standard GDM. The dynamics of the system can thus be better modelled using the CGDM.

5.4   Specific DBN for prognosis estimations

The purpose of this section is to introduce a specific DBN used to perform the prognostic computations. The advantage of the proposed methodology is its ability to adapt the considered active degradation dynamic in light of the available observations. Considering the CGDM introduced in the previous section and the classical GDM, two nodes were added to represent the available observation provided by the diagnosis devices. $D^{(t)}$ represents the possible system-state diagnoses. It takes its values from the discrete states space { $\varnothing$,ok,x1,x2,F} where $\varnothing$ is the "disabled" state.

$Act^{(t)} \in \{0,1\}$ is a determinist node that controls whether the diagnosis device is active ($Act^{(t)} = 1$) or not ($Act^{(t)} = 0$), enabling the consideration of both real-time monitored systems and periodically observed systems. The CPDs of $Act^{(t)}$ and $D^{(t)}$ are defined as following :

$$Act^{(t)} = \begin{cases} 1 \text{ if } t = k\Delta t \\ 0 \text{ otherwise} \end{cases} \qquad D^{(t)} = \begin{cases} X^{(t)} \text{ if } Act^{(t)} = 1 \\ \emptyset \text{ if } Act^{(t)} \quad = 0 \end{cases} \tag{16}$$

To compare the testing models to more classical ones, let notice that the subset on nodes ($X^{(t)}$, $Act^{(t)}$, $D^{(t)}$) from the DBN represented in Figure 11 is a simple Hidden Markov Model (HMM). The addition of node $S^{(t)}$ makes possible to be not dependent of the geometrical assumption on sojourn time distribution, the model is also a semi-Markovian hidden model. Then the complete DBN shown in figure 11 including node $C^{(t)}$ includes the ability to take into account the concept of degradation dynamic in the hidden semi-Markovian model. An inference algorithm is then used on this DBN to perform reliability indicators and compare the GDM with the new proposed degradation model CGDM. A scheme of the overall process used for the tests is represented in Figure 12.
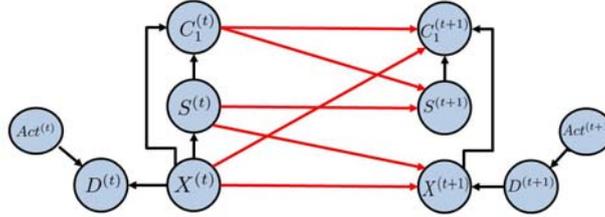
Figure 11: DBN Used for Computing Process
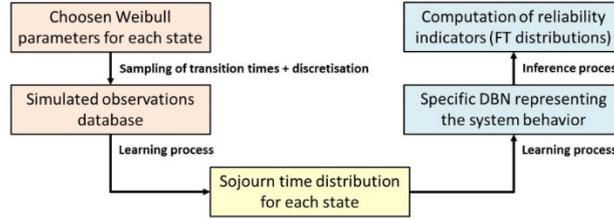
Figure 12: Main Steps of the Overall Process for the Test

The following section presents the testing of this model on three different observation patterns drawn from the simulated database to illustrate the global behaviour of the proposed approach

5.5   Tests and Results

This section details the testing process and discusses the results produced. The model was tested by performing prognostic computations. A pattern from the observations database is considered. $D = \{d_1, d_2, …, d_j\}$ is the set of available periodic observations at at times $\{t_1, t_2, …, t_p\}$. The objective is to compute an estimate of the failure time (FT) and to update this estimate at each observation time t_i. To integrate the observations into the prognosis, the probability for FT computed at time t to be equal to k is expressed with the following conditional probability:

If d_j is the last available diagnosis until time t, $\forall t$, so that $t_1 < t_2 < … < t_j < t$, $\forall$ k $> t_j$

$$P\left(FT^{(t)} = k\right) = P\left(X^{(k)} = F, X^{(k-1)} \neq F \mid D^{(t_1)} = d_1, D^{(t_2)} = d_2, ..., D^{(t_j)} = d_j\right) \quad (17)$$

Because the original sojourn times that were used to build the considered degradation pattern are known, the real failure time (denoted FT_{real}) is also known at each time; in the figures below, the real FT is indicated with dotted lines. The risk of error can also be computed by considering the probability that the FT, computed at time t will have an error less than ε:

$$P\left(Error^{(t)} < \varepsilon\right) = \sum_{k=-\varepsilon}^{\varepsilon} P\left(FT^{(t)} = FT_{real} + k\right)$$

(18)

In this analysis, these inference computations were performed using the variable elimination algorithm [8].

5.5.1 First test

The first considered observation pattern represented in Table 1, corresponds to the case in which the system remains in the first degradation dynamic from its initial state until its failure. It therefore exhibits short sojourn times in each state. The real FT associated with this pattern is 22.

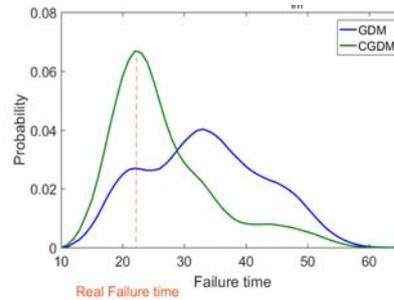**Table 1: First Observation Pattern Considered**

Time	t_1=0	t_2=5	t_3=10	t_4=15	t_5=20	t_6=25
Observed state	ok	ok	x_1	x_1	x_2	F

The FT distribution is computed via inference on the DBN created using the CGDM approach and compared with the results of the same computations performed using the standard GDM at various times (denoted by t_{inf}) to illustrate the behaviour of the model. The inference computation considers all previous available observations. For example, at t_{inf}=10, the inference process will compute the following probability for several values of k, as shown in figure 13b.

$$P\left(FT^{(10)} = k\right) = P\left(X^{(k)} = F, X^{(k-1)} \neq F \mid D^{(0)} = ok, D^{(5)} = ok, D^{(10)} = x_1\right)$$

(19)
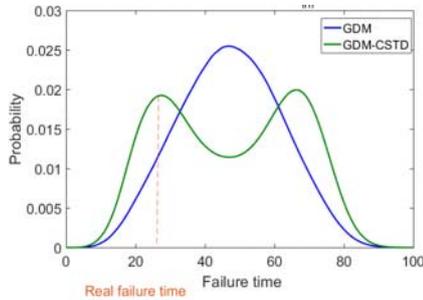
(a) **Failure time distribution at** t_{inf} = 0   (b) **Failure time distribution at** t_{inf} = 10
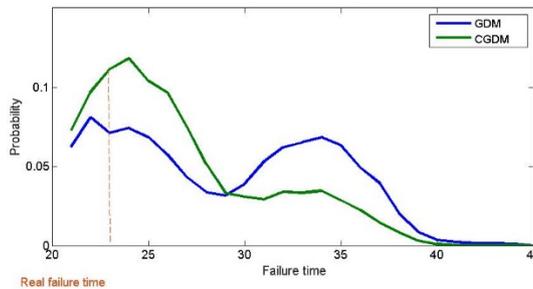
(c) **Failure time distribution at** t_{inf} = 20

Figure 13: Inference Results for Test 1

The results for t_{inf} = 0, t_{inf} = 10 and t_{inf} = 20 are shown in Figure 13. As expected, the FT distribution produced by the CGDM has a smaller variance than that obtained using the

standard GDM. This can be explained by the fact that dynamic 1 is identified earlier as the current active dynamic, and therefore, the convergence to the real failure time is faster with the CGDM.

At time $t_{inf} = 0$, the two dynamics are equivalent in terms of probability. As seen from the existence of two equal peaks centered on $t = 25$ and $t = 70$. At time $t_{inf} = 10$, after the first transition from *ok* to x_1, the CGDM detects that the first dynamic is more likely to be the real one. Therefore, the peak of the CGDM distribution becomes centered around on *FT* = 25 whereas the GDM computes a *FT* distribution with a large variance because of the mixture of both dynamics in the distribution learned by this model. The CGDM can thus produce a better FT estimation. This trend confirms at time $t_{inf} = 20$, the peak given by the CGDM distribution is near to the real FT.

Table 2: Probability for the Error to be Less Than ε in Test 1

t_{inf}	ε	GDM	CGDM
0	5	0.07	0.15
	10	0.16	0.28
10	5	0.26	0.57
	10	0.43	0.77
20	5	0.45	0.67
	10	0.67	0.85

This difference can be seen from a computation of the risk of error. The values obtained in this computation are shown in table 2. As expected, the probability that the true state of the system will lie within the confidence interval is always higher with the CGDM than with the original GDM, regardless of the chosen confidence interval. The predictions are thus improved. In this first example, the observation sequence was generated with sojourn times that are close to the average. This is why the peak of the curve produced using the CGDM is centred on the real FT.

5.5.2 Second Test

The next considered observation pattern corresponds to the case in which the system starts in dynamic 2 and remains in this dynamic until its failure. This pattern is represented in table 3. The exact failure time in this case is 67.
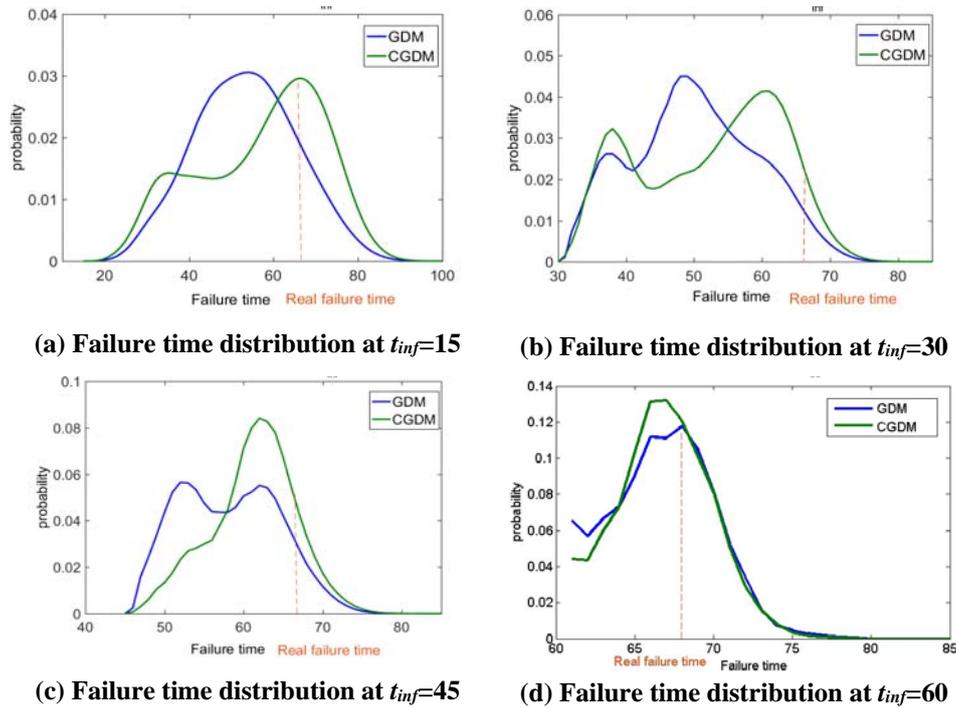
Table 3: Second Observation Pattern Considered

Time	t_1=0	t_2=5	t_3=10	t_4=15	t_5=20	t_6=25	t_7=35	t_8=35
Observed state	*ok*	*ok*	*ok*	*ok*	x_1	x_1	x_1	x_1
	T_9=40	T_{10}=45	T_{11}=50	T_{12}=55	T_{13}=60	T_{14}=65	T_{15}=70	
	x_1	x_1	x_2	x_2	x_2	x_2	F	

The FT distribution was computed at time t_{inf}= 15, 30, 45 and 60. The results are shown in Figure 14. In this example, at every time, the CGDM always provides better estimations than the original GDM.

This advantage can be seen from the error distribution shown in Table 4. The CGDM offers better accuracy in estimating the real FT than does the original GDM. This difference is related to the ability of the CGDM to account for the dynamic distributions. In this

example, sojourn times are longer than in section 5.5.1 because they are generated by the dynamic 2, so the failure occurs latter than in the dynamic 1 case. The CGDM inference algorithm has therefore more time to replace its estimation of the failure time because, as time passes, the variance of CSTD in the considered state decreases. Thus, the FT distribution computed via inference has time to converge to the real FT distribution.

(a) Failure time distribution at t_{inf}=15   **(b) Failure time distribution at t_{inf}=30**

(c) Failure time distribution at t_{inf}=45   **(d) Failure time distribution at t_{inf}=60**

Figure 14: Inference Results for Test 2

Table 4: Probability for the Error to be less than ε in Test 2

t_{inf}	ε	GDM	CGDM
15	5	0.21	0.31
	10	0.40	0.52
30	5	0.14	0.26
	10	0.28	0.46
45	5	0.37	0.57
	10	0.61	0.83
50	5	0.35	0.53
	10	0.58	0.81
60	5	0.90	0.92
	10	0.998	0.999

6.   Conclusions

Beginning from a foundation laid by a former study [10] that proposed a degradation model for a one single component system based on the STD concept using the formalism of DBNs, this paper proposed an extension of the standard GDM that considers CSTDs. This new

study was justified by the fact that failure processes frequently consist of combinations of several dynamics (from various reasons, from the different behaviours of different subsystems component to the existence of multiple degradation processes). The variance of the STDs strongly impacts the estimation of reliability indicators such as the FT or the remaining useful life (RUL). Adopting the CSTD concept might allow the quality of estimation of such reliability indicators to be improved.

The application tests presented in this paper showed that the CGDM models the degradation process more successfully than does the GDM is many cases. It therefore decreases the risk of error in the computations of prognostic indicators such as the FT, as was shown by means of inference computations performed on the DBN. The proposed methodology should thus be better suited for modelling the degradation dynamic of systems that can be subject to many different dynamics and for considering the effect of disruptive phenomena that could impact the degradation speed by changing these dynamics. In that way, a prognosis algorithm, based on the use of CGDM, was proposed in [12]. Encoded through a specific DBN, it allows the computation of the remaining useful life. In this algorithm, the RUL is computed from the current observation considering the estimated active dynamic. The algorithm updates the RUL at each new observation and can detect a dynamic change.

Due to the use of the DBN formalism, one of the main advantage of the proposed degradation model is its ability to be next integrated in a global DBN representing a multi component large scale system. Such DBN make possible to model component with different degradation processes, model the functional dependencies between the components, and to have global vision at the scale of the whole system. An example of such global DBN representing a multicomponent system can be found in [5].

But, the CGDM approach has its own difficulties. It is not always easy to find a meaningful interpretation for each component of a mixture produced by an EM algorithm. The selection of a suitable number of component can also pose a challenge. One potential route for further development could be to improve the classification process in the learning of CSTDs.

Other possibilities could also be investigated, such as considering all existing dynamics simultaneously, weighted by the a posteriori probabilities associated with each component of the mixture models (in contrast to the current approach, which is based on considering a single active dynamic at any given time), or using the node $S^{(t)}$ that controls the remaining sojourn times for each component to represent a multicomponent systems with inter-dependencies between the degradation processes of some of its components.

Finally, further research in another direction is currently underway as part of a new PhD study, addressing the integration of continuous sojourn-time variables into the GDM. The purpose of such improvements is both to improve the accuracy of the degradation process modelling and to solve the potential complexity problems posed, in some cases, by the large ranges of the sojourn-times variables.

References

[1]. Akutekwe, A., *A hybrid dynamic Bayesian network approach for modelling temporal associations of gene expressions for hypertension diagnosis*, 36th Annual International Conference of the IEEE, pp 804-808, 2014

[2]. Aven, T., Jensen, J., *Stochastic models in reliability*, Stochastic modelling and applied probability, 1999

[3]. Ben-Salem, A., Muller, A.,Weber, P., *Dynamic bayesian networks in system reliability analysis*, 6th IFAC Symposium on Fault Detection, Supervision and Safety of technical processes, Beijing, China. IFAC, pp.481-486, 2006.

[4]. Bertholon, H., Bousquet, N., Celeux, G., *An alternative competing risk model to the Weibull distribution for modelling aging in lifetime data analysis.* Lifetime Data Analysis, 12(4), pp. 481-504, 2006

[5]. Bouillaut, L., Aknin, P., Donat, R., Bondeux, S., *VirMaLab : A generic approach for optimizing maintenance policies for complex systems*, 9th World Congress on Railway Research, 2011

[6]. Bracquemond, C., Gaudoin, O., *A survey on discrete lifetime distribution*, International Journal on Reliability, Quality, and Safety Engineering, 10(1), pp. 69-98, 2003

[7]. Charitos, T., Van der Gaag, L., Visscher, S., Schurink, K., Peter Lucas, P., *A Dynamic Bayesian Network for Diagnosing Ventilator-Associated Pneumonia in ICU Patients*, Expert Systems with Applications 36(2), pp. 1249-1258, 2009

[8]. Dechter, R., *Bucket elimination: A unifying framework for reasoning.* Artificial Intelligence, 113, pp.41-85, 1999

[9]. Dempster, A.P., Laird, N. M., Rubin, D. B., *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal statistical Society, B(39), pp. 1-38, 1977

[10]. Donat, R., Leray, P., Bouillaut, L., Aknin, P., *A dynamic Bayesian network to represent discrete duration models*, Neurocomputing, 73(46), pp. 570-577, 2010

[11]. Fisher, R., *On the mathematical foundations of theoretical statistics.* Transactions of the Royal Society of London. Series A, Containing papers of a Mathematical or Physical Caracter, 222, pp. 309-368, 1922

[12]. Foulliaron, J., Bouillaut, L., Barros, A., Aknin, .P, *A dynamic Bayesian network approach for prognosis computations on discrete states systems*, To appear in Proceedings of the Institution of Mechanical Engineers, Part O: Journal of risks and reliability, 2016

[13]. Freitas, M., de Toledo, M., Colosimo, E., Pires, E., *Using degradation data to assess reliability: a case study on train wheel degradation.* Quality and Reliability Engineering International, 25(5), pp. 607-629, 2009

[14]. Horenbeek, V., Scarf, P.A, Cavalcante, C.A.V., *On the use of predictive information in a joint maintenance and inventory policy*, Safety, Reliability and Risk Analysis: Beyond the Horizon Steenbergen et al., Eds Taylor & Francis, 2014

[15]. Jensen, F.V, *Bayesian Networks and Decision Graphs*, Eds Springer, 2007

[16]. Kao, H., Huangb, C., Li, H., *Supply chain diagnostics with dynamic Bayesian networks*, Computers & Industrial Engineering, 49, pp. 339-347, 2005

[17]. Kawahara, Y., Yairi, T., Machida, K., *Diagnosis method for spacecraft using dynamic bayesian networks*, 8th International Symposium on Artifical Intelligence, Robotics and Automation in Space, Munich, Germany, 2005

[18]. Lauritzen, S.L., Spiegelhalter, D.J., *Local computations with probabilities on graphical structures and their application to expert systems.* Journal of the Royal Statistical Society, 50(2), pp. 157-224, 1988

[19]. Lemaitre, J., Demorat, R., *Engineering Damage mechanics*, Eds Springer, 2005

[20]. Van de Lisdonk, R.H.M., *Automatic speech recognition using dynamic Bayesian networks*, Master Thesis, Delft University of Technology, 2009

[21]. Lorenzoni A., Kempf, M., *Degradation process modelled with dynamic bayesian networks*, 13th IEEE International Conference on Industrial Informatics, Cambridge, United Kingdom, pp. 1694–1699, 2015

[22]. Mc Naught, K.R., Zagorecki, A., *Using Dynamic Bayesian network for Prognostic*

Modelling to inform maintenance Decision Making, IEEE International Conference on Industrial Engineering and Engineering Management, Hong-Kong, China, 2009

[23]. Murphy, K.P, *Dynamic Bayesian network: Representation, inference and Learning*, PhD Thesis, University of California, Berkeley, 2002

[24]. Nguyen, K., Do Van, P., Grall, A., *A predictive maintenance strategy for multi-component systems using importance measure*, Safety, Reliability and Risk Analysis: Beyond the Horizon Steenbergen et al., Eds Taylor & Francis, 2014

[25]. Pearl, J., *Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning*, 7th Conference of the Cognitive Science Society, Irvine, USA, pp. 329-334, 1985

[26]. Rajpal, P.S., Shishodia, K.S., Sekhon, G.S., *An artificial neural network for modeling reliability, availability and maintainability of a repairable system*, Reliability Engineering & System Safety, 91(7), pp. 809-819, 2006

[27]. Schwarz, G., *Estimating the dimension of a model*. Annals of Statistics, 6(2), pp. 461-464, 1978

[28]. Su, A., Fan, M., Li, Z., *Reliability analysis on distribution system based on dynamic bayesian network*, China International Conference on Electricity Distribution, 2012

[29]. Tchangani, A., Noyes, D., *Modeling dynamic reliability using dynamic Bayesian networks*, Journal Européen des Systèmes Automatics, 40(8), pp. 911-935, 2006

[30]. Volovoi, V., *Modeling of System Reliability Using Petri Nets with Aging Tokens*, Reliability Engineering & System Safety, 84, pp. 149-161, 2004

[31]. Zweig, G., *Speech recognition with Dynamic Bayesian networks*, PhD Thesis, University of Berkeley, 1998

Josquin Foulliaron got a master degree in applied mathematics from University Paris 6 in 2011. He also got a PhD in Computer Sciences from the University Paris Est in 2015, about the use and the development of probabilistic graphical models and prognosis algorithms for setting up and optimizing predictive maintenance policies. He is currently in post-doctoral position at IFSTTAR, working on the development of a reliability prediction and maintenance optimization software for the next big Parisian metro line: "Grand Paris Express".

Laurent Bouillaut graduated a French post-doctoral degree allowing supervising PhD students (HDR) in 2014. He received both a PhD in Signal Processing (2000) and his engineering school degree in Computer sciences (1997) from the Université de Technologie de Compiègne. He works as a researcher at IFSTTAR. His research interests include mainly maintenance optimization, probabilistic graphical models, reliability analysis, all applied to public transport systems and infrastructures.

Patrice Aknin. A PhD of Université Paris-Sud Orsay and « Habilitation à diriger des Recherches » of the Ecole Normale Supérieure de Cachan in 2008, the career of Patrice Aknin was one of research at INRETS, later known as IFSTTAR. In 2008 he was appointed Research Director at IFSTTAR. He then joined the Scientific Direction of IFSTTAR as Deputy Scientific Director for doctoral studies and academic partnerships. In 2013, he was appointed Scientific Director of the SNCF Group in the Innovation & Research direction. Since 2016, Patrice Aknin is the Scientific Director of the SystemX technological research institute (IRT).

Anne Barros. Anne Barros is a full Professor in the Department of Mechanical and Industrial Engineering at Norwegian University of Science and Technology (NTNU), Norway. Her research interests focus on the construction of probabilistic decision indicators used for safety assessment and maintenance optimization.