1. INTRODUCTION

One of the challenges in maintenance optimization is to propose the most effective balance between preventive actions which avoid the occurrence of failures, and corrective ones if a failure occurs. The system considered in this paper is subject to continuous monotone random deterioration. It is supposed that the system condition is observable and can be summarised by a scalar ageing variable, see [15]. The system is assumed to be in a failure state when the aging variable is greater than a predetermined threshold $L$ even if it is still functioning. In this situation, its high level of deterioration is unacceptable either for economic reasons (poor products quality, high consumption of raw material, etc.) or for safety reasons (high risk of hazardous breakdowns or service disruption). The aim of this paper is to
propose a Bayesian methodology to deal with parametrical uncertainty of Gamma deterioration process in the framework of maintenance decision-making. An industrial case study is considered.

The remainder of the paper is organized as follows. In section 2, the deterioration model is presented. Section 3 is devoted to data description. The deterioration parameters estimation method is developed in 4. The maintenance decision rule is described in section 6.

2. Deterioration Model

Consider a stochastically deteriorating system described by a scalar ageing variable \((X_t)_{t \geq 0}\) which summarizes the condition of the system at time \(t\). To model this stochastic deterioration a gamma process seems adequate, refer to [3,14]. Note that the gamma process is a positive process with independent increments, hence it is sensible to use this process to describe the deterioration caused by accumulation of wear. Another interest of the gamma process is the existence of an explicit probability distribution function which permits feasible mathematical developments.

As it is assumed that the deterioration process is a gamma process for all \(0 \leq s \leq t\), the increment of \((X_t)_{t \geq 0}\) between \(s\) and \(t\), \(X_t - X_s\), follows a gamma probability distribution function with shape-parameter \(\alpha(t-s)\) and scale parameter \(\beta\). This probability distribution function can be written as follows:

\[
f_{\alpha(t-s),\beta}(x) = \frac{1}{\Gamma(\alpha(t-s))} \cdot \frac{x^{\alpha(t-s)-1}e^{-\frac{x}{\beta}}}{\beta^{\alpha(t-s)}} 1(x \geq 0).
\] (1)

The average deterioration speed rate between \(s\) and \(t\), for \(0 \leq s \leq t\), is \(\alpha(\beta(t-s))\) and its variance is \(\alpha(\beta^2(t-s))\). It is supposed that the parameters \((\alpha,\beta)\) are unknown.

3. Data description

The industrial case that motivates this study is the following. We consider a set of \(M\) identical components from French power plants submitted to intensive effort. The deterioration of a individual component \(k \in \{1, \ldots, M\}\) is characterized by a crack size \(X_{k,t}\) at time \(t\). \(X_{k,t}\) is therefore positive and monotonously increasing through time. When they appear, cracks are not observed until reaching a size \(z\) because of the limitations of observational (ultrasonic) processes. However, such measures can prove the existence of a crack even though its size cannot be correctly determined. The failure (e.g.,
corrective maintenance) is assumed to occur from a size $L > z$. In this paper, we consider only the problem of optimizing the preventive maintenance times given that cracks have been detected, and the problem is simplified by considering only one crack (the major one) per component.

In a more general setting, it is assumed that the considered deteriorating system cannot be continuously monitored. For this reason, the deterioration level can be only detected or observed at preventive inspection times. The available data are denoted $Y_{k,t} = \max(X_{k,t}, z)$, observed at inspection times $t_{1,k}, \ldots, t_{nk,k}$ for the component $k$ (to alleviate notations we remove the second index $k$ for the inspection times). Denote $Z_{k,i} = X_{k,t_i} - X_{k,t_{i-1}}$ the increments of the crack size for the component $k$. They are assumed to be independent and follow the gamma distribution previously described. Let $t_{r_k}$ indicate the last inspection time before the first determination of the true crack size, i.e.

$$t_{r_k+1} = \min_{i \in \{1, \ldots, n_k\}} \{t_i, X_{k,t_i} \geq z\}.$$ 

The likelihood of these data incorporates two censoring terms.

(1) **Left-censoring term.** The likelihood term due to the non-observation of the crack size between $t_0 = 0$ and $t_{r_k}$ for the component $k$ is the following probability (assuming $X_{k,t_0} = 0$, $\forall k \in \{1, \ldots, M\}$)

$$P\left(X_{k,t_1} \leq X_{k,t_2} \leq \cdots \leq X_{k,t_{rk}} < z\right) = P\left(Z_{k,1} \leq Z_{k,1} + Z_{k,2} \leq \cdots \leq \sum_{i=1}^{r_k} Z_{k,i} < z\right) = P\left(\sum_{i=1}^{r_k} Z_{k,i} < z\right)$$

with

$$\sum_{i=1}^{r_k} Z_{k,i} \sim G(\alpha t_{r_k}, 1/\beta)$$

which stands for the distribution of density (1), one has

$$P\left(X_{k,t_1} \leq X_{k,t_2} \leq \cdots \leq X_{k,t_{rk}} < z\right) = \frac{\gamma\left(\alpha t_{r_k}, \frac{z}{\beta}\right)}{\Gamma(\alpha t_{r_k})}$$
where $\gamma(a, b) = \int_0^b x^{a-1} e^{-x} \, dx$ is the lower incomplete gamma function.

(2) **Right-censoring term.** The likelihood term due to the observation of $Z_{k,r_k+1} = X_{k,t_{r_k}+1} - z$, which is a lower bound for the true $Z_{k,r_k+1}$, is

$$P(Z_{k,r_k+1} > x_{k,t_{r_k}+1} - z) = \frac{\Gamma(\alpha t_{r_k+1} - t_{r_k}, (x_{k,t_{r_k}+1} - z)/\beta)}{\Gamma(\alpha t_{r_k+1} - t_{r_k})}$$

where $\Gamma(a, b) = \int_b^\infty x^{a-1} e^{-x} \, dx$ is the upper incomplete gamma function for $x \geq 0$ and $a > 0$.

Finally, the likelihood of data $d_k$ for the $k$th component is

- if there is no exact measure,
  $$\ell_k(\alpha, \beta | d_k) = \frac{\gamma(\alpha t_n_k, \frac{z}{\beta})}{\Gamma(\alpha t_n_k)};$$

- if there is only one exact measure,
  $$\ell_k(\alpha, \beta | d_k) = \frac{\gamma(\alpha t_{n_k-1}, \frac{z}{\beta}) \Gamma(\alpha(t_{n_k} - t_{n_k-1}), (x_{k,t_{n_k}} - z)/\beta)}{\Gamma(\alpha t_{n_k-1}) \Gamma(\alpha(t_{n_k} - t_{n_k-1}))};$$

- if there is more than one exact measure,
  $$\ell_k(\alpha, \beta | d_k) = \frac{\gamma(\alpha t_{r_k}, \frac{z}{\beta}) \Gamma(\alpha(t_{r_k+1} - t_{r_k}), (x_{k,t_{r_k+1}} - z)/\beta)}{\Gamma(\alpha t_{r_k}) \Gamma(\alpha(t_{r_k+1} - t_{r_k})) \prod_{i=r_k+2}^{n_k} \Gamma(\alpha(t_i - t_{i-1}))} \times \beta^{-\alpha(t_{n_k} - t_{r_k+1})} \exp\left\{-\frac{1}{\beta} \sum_{i=r_k+2}^{n_k} z_{k,i} \right\} \left[ \prod_{i=r_k+2}^{n_k} z_{k,i}^{t_i - t_{i-1}} \right]^{\alpha}.$$

Then the likelihood $\ell(\alpha, \beta | d)$ of all observed data $d$ is the product of the $\ell_k(\alpha, \beta | d_k)$ over the $k = 1, \ldots, M$ identical components.

### 4. Estimation

A Bayesian framework is chosen here to carry out the estimation of $\theta = (\alpha, \beta) \in \mathbb{R}_+^2$ and any function of interest $h(\theta)$ (e.g., the level $A$ defined hereafter). Namely the parameter vector $\theta$ is considered random, endowed with a prior distribution of density $\pi(\theta)$, and the estimation task consists in obtaining a detailed description of its posterior distribution $\pi(\theta | d) \propto$
\[ \pi(\theta)\ell(\theta|d), \] having learned from the data through \( \ell(\theta|d) \). The rationale for this choice is twofold. First, the Bayesian approach allows to take into account statistical uncertainties over the estimation of \( b(\theta) \) and in predictive assessments, in a non-asymptotic framework coherent with the rather small sample size of the data. Second, it is by nature adapted to decisional problems in risk management by helping to define conservative estimators. See [9] for details.

To avoid the well-known issues related to the difficult elicitation of a prior distribution [11], an objective Bayesian framework is chosen. Especially defended by [4], a relevant non-informative prior measure \( \pi(\theta) \) is Jeffreys’ prior. It is basically defined as the square root of the determinant of the Fisher matrix, and is therefore linked to the form of the likelihood (see [8] for details). To simplify the elicitation and because the posterior distribution is driven by the data rather than the prior in objective Bayesian framework, the censoring aspects of the likelihood are not taken into account (but discussed further in the text). To be independent on the scale effect imposed by the heterogeneity of time steps along a cracking trajectory and between the components, the prior is defined for a standard gamma distribution, i.e. for the likelihood of the first occurrence of the process \( Z \).

From [16], one has

\[ \pi(\alpha, \beta) \propto \frac{1}{\beta} \sqrt{\alpha \Psi(\alpha) - 1} \]

where \( \Psi(.) \) is the digamma function. It must be noticed that the posterior obtained from this prior and the first likelihood element plays the role of the prior for the second likelihood element, and so on. Besides, the prior is defined only for \( \alpha > \alpha_0 \approx 2.09 \). This is however not restrictive since \( \alpha \) is a scale parameter, which can be upscaled by normalizing the data.

Sampling from the posterior distribution, by using specific numerical techniques as Monte Carlo Markov Chains (MCMC), appears as a relevant way of doing Bayesian estimation since the posterior distribution is not of tractable form. However, it is possible to account for the missing data structure of the problem to implement a Gibbs algorithm [12] taking benefit from conditional conjugation properties. Indeed, if all missing statistics \( S_k = \sum_{i=1}^{r_k} Z_{k,i} \) could be known (available or reconstituted using (2) under the constraint \( S_k \leq z \)) the posterior distribution of \( \beta \) conditional to \( \alpha \)
appears to be inverse gamma:

\[ \beta|\alpha, d \sim IG \left( \alpha \sum_{k=1}^{M} t_{n_k}, \sum_{k=1}^{M} \sum_{i=r_k+1}^{n_k} z_{k,i} \right). \]

Since no such feature can be exhibited for the conditional posterior distribution of \( \alpha \), a Metropolis-Hastings mechanism can be inserted within each Gibbs step to produce an ergodic Markov chain, using (for instance) a random walk defined with a parameter \( \rho \) calibrated from testings. The resulting hybrid algorithm is described hereinafter.

**Posterior simulation algorithm (step \( j-1 \to j \)).**

1. Sample \( \hat{S}_j \sim G \left( \alpha_{j-1} \sum_{k=1}^{M} t_{n_k}, 1/\beta_{j-1} \right) \cdot \mathbb{1}_{0 \leq S \leq M \cdot z} \).
2. Sample \( \beta_j \sim IG \left( \alpha_{j-1} \sum_{k=1}^{M} t_{n_k}, \hat{S}_j + \sum_{k=1}^{M} \sum_{i=r_k+1}^{n_k} z_{k,i} \right) \).
3. Sample \( \tilde{\alpha}_j \sim \rho(\cdot|\alpha_{j-1}) \equiv N(\alpha_{j-1}, \rho \cdot \alpha_{j-1}^2) \).
4. Sample \( u_j \sim U[0,1] \) and accept \( \alpha_j = \tilde{\alpha}_j \) if \( u_j \leq \eta_j \) where
   \begin{align*}
   \eta_j &= \min \left\{ 1, \frac{\ell(\beta_j, \tilde{\alpha}_j|d) \pi(\beta_j|\tilde{\alpha}_j) \rho(\alpha_{j-1}|\tilde{\alpha}_j)}{\ell(\beta_j, \alpha_{j-1}|d) \pi(\beta_j|\alpha_{j-1}) \rho(\tilde{\alpha}_j|\alpha_{j-1})} \right\} \\
   \text{else accept } \alpha_j &= \alpha_{j-1}.
   \end{align*}

In the applications considered here, the burn-in periods and stationarity of chains were managed using the classical diagnostics provided by [2], implemented within the CODA software [10]. The MCMC chains were decorrelated until reaching a maximal autocorrelation of 5% between the elements of the chains. An illustration of the results produced by the algorithm are provided on Figure 1.

**5. Numerical experiments**

Experiments have been carried out using simulations for various values of \((M, n)\) and some censoring percentages (implied by arbitrarily values of \(z\)). Some of those preliminary results are summarized on Table 1. The values express the relative bias (in percentage) of posterior estimates of \((\alpha, \beta)\), based on a dozen of simulated datasets per situation, using simulation values \( \alpha_0 = 5 \) and \( \beta_0 = 0.1 \). Censoring rate was 20%, due to fixing \( z = 5 \). The results stress the primary importance of getting observations along a trajectory \((n \text{ increasing})\) rather than multiplying the number of independent
Figure 1. Three parallel Gibbs chains approximating the posterior distribution of \((\alpha, \beta)\) for a sample of 30 similar cracks measured at 15 time steps, incurring 20% censoring, sampled using \((\alpha_0, \beta_0) = (5, 0.1)\).

trajectories \((M\) increasing).\n
<table>
<thead>
<tr>
<th>(M)</th>
<th>10</th>
<th>30</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>(-0.24,0.33)</td>
<td>(-0.17,0.14)</td>
<td>(-0.11,0.07)</td>
</tr>
<tr>
<td>25</td>
<td>(-0.23,0.25)</td>
<td>(-0.19,0.08)</td>
<td>(-0.08,0.05)</td>
</tr>
<tr>
<td>50</td>
<td>(-0.21,0.21)</td>
<td>(-0.17,0.07)</td>
<td>(-0.06,0.05)</td>
</tr>
</tbody>
</table>

6. Maintenance policy

In a situation where a sample \(d\) of observed data is not dynamically obtained, suitable maintenance policies are primarily based on the estimation of the time before the probability of failure exceeds a given percentage \(1-\rho\), defined by

\[
t_{1-\rho}(\alpha, \beta) = \max \left\{ t \geq 0, \quad P(X_t > L) = \frac{\gamma(\alpha t, L/\beta)}{\Gamma(\alpha t)} \leq 1 - \rho \right\}.
\]
Considering that a preventive maintenance can be understood as a simple ground visit, there is no special cost associated to the choice of $\rho$ and it is also natural to use $\rho = 50\%$. The overall uncertainty on $t_{1-\rho}(\alpha, \beta)$ is yielded by the posterior distribution $\pi(\alpha, \beta|d)$ and a cost function $\ell\{x, t_{1-\rho}\}$ should be minimized in $x$, over the range of all possible values for $t_{1-\rho}(\alpha, \beta)$, to define a Bayesian estimator of this failure time. Namely, one estimates the failure time by

$\hat{t}_{1-\rho} = \arg \min_x \int \int \ell\{x, t_{1-\rho}(\alpha, \beta)\} \pi(\alpha, \beta|d) \, d\alpha \, d\beta$.

As in such cases it is as much costly to overestimate and underestimate the true failure time, a reasonable (and classical) choice is to define $\ell\{\ldots\}$ as the quadratic loss function. Hence a good estimator is the posterior mean

$\hat{t}_{1-\rho} = \int \int t_{1-\rho}(\alpha, \beta) \pi(\alpha, \beta|d) \, d\alpha \, d\beta$

which can be easily computed by Monte Carlo means. Note however that for each couple $(\alpha, \beta)$ sampled from the posterior distribution, an optimization program must be conducted to get a sample $t_{1-\rho}(\alpha, \beta)$.

Now we consider a situation where the data $d$ can be observed dynamically, and the cost of replacement depends on time. The preventive replacement time should be carefully chosen in order to be able to replace the system before the failure. If the replacement occurs too early then the maintenance policy is conservative. In the opposite, if the replacement takes place too late the system can be failed. The preventive replacement time depends on the parameters of the deterioration process. The decision about a possible replacement has to be taken according to the current state of the system and also with respect to the available information on the deterioration rate (deterioration parameters). In a Bayesian framework, this choice can be traduced by a more accurate selection of the loss function $\ell\{\ldots\}$.

To be clearer, denote $C_1(t_0 + t, t_0)$ the corrective (additional) cost due to replacing the component after the (unknown) date of “normative” failure $t_0$ (at time $t_0 + t$). This cost possibly involves penalties due to the non-respect of safety measures. Similarly, denote $C_2(t_0 + t, t_0)$ the preventive (additional) cost due to replacing the component before $t_0$ (at time $t_0 + t$ with $t \leq 0$). Especially when $t \simeq t_0$, this cost involves a possible large loss
of income. Therefore a reasonable estimator of the failure time is
\[ \hat{t}_{1-\rho} = \arg \min_x \int \int \left\{ C_1(t_{1-\rho}(\alpha, \beta) + x, t_{1-\rho}(\alpha, \beta)) \mathbb{1}_{\{x \geq 0\}} + 
C_2(t_{1-\rho}(\alpha, \beta) + x, t_{1-\rho}(\alpha, \beta)) \mathbb{1}_{\{x < 0\}} \right\} \pi(\alpha, \beta | d) \, d\alpha \, d\beta. \]
It is worth noting that in the simplest case where the two costs can be linearly approximated, i.e. \( C_1(t_0 + t, t_0) \approx c_1|t - t_0| \) and \( C_2(t_0 + t, t_0) \approx c_2|t - t_0| \) where \((c_1, c_2)\) are constant values, then \( \hat{t}_{1-\rho} \) is the \( \frac{c_1}{c_1 + c_2} \)-order percentile of the posterior distribution of \( t_{1-\rho}(\alpha, \beta) \), which can be similarly estimated by Monte Carlo means.

7. Discussion
This article resumes an objective Bayesian methodology of calibrating a Gamma degradation process and computing failure times by accounting of all parametrical uncertainty, in classical situations where a large part of the data is censored due to the limits of the experimental process. Some points should deserve a deeper study, the interest of which having not appeared essential within the scope of the present paper. The elicitation of a Jeffrey’s prior that accounts for the censoring phenomena affecting the likelihood and the Fisher matrix appears of interest, as [5] showed on simple examples that such priors can improve the properties of the posterior distribution in terms of frequentist coverage when dealing with small samples. This approach was not carried out here since it requires laborious technical developments using incomplete polygamma functionals, for a result which is suspected to present low differences in terms of maintenance management.

8. Acknowledgements
This study was supported by the French Ministry of Research in the context of the AMMSI ANR project (http://www-ljk.imag.fr/AMMSI/).

References


