A stochastic model for competing degradations

S. Mercier & L. Bordes
Laboratoire de Mathématiques et de leurs Applications - Pau (UMR CNRS 5142)
Université de Pau et des Pays de l'Adour - France

E. Remy & E. Dautrême
EDF R&D - Industrial Risk Management Department
Chatou - France.

ABSTRACT: EDF periodically performs in-service inspections of passive components within its electric power plants in order to ensure that their degradation is lower than a critical level and to guarantee the safety and the availability of the installations. These examinations allow to collect successive degradation measurements data (crack sizes) for the components. Unfortunately, these data are incomplete. First, small cracks with size below a specific known threshold may be detected without possible measurement. In that case, the only available information hence is: presence of a crack with size below the threshold. Secondly, one single measurement can be performed by an examination and, in case of several competing cracks on one component, the process cannot measure all cracks sizes but only the largest one (which is not necessarily the same throughout the whole component lifetime) if its size exceeds the previous threshold. However, even if they are not measured, all cracks are detected, so that, in that case, the available information is: number of cracks, size of the largest one. Taking into account this partial information, a specific stochastic model is proposed. In this model, cracks initiate following a Poisson process and propagate according to gamma processes. Parametric estimation procedures are developped, tested on simulated data and then applied to the industrial data. The fitted model is next used to make some prediction over the future degradation propagation and over the residual operation time upon which a critical degradation level is reached.

1 INDUSTRIAL CONTEXT, OBJECTIVE AND DATA

1.1 *Industrial context*

EDF, one of the world's leading electric utilities, periodically performs in-service inspections of the passive components within its electric power plants in order to ensure that their degradation is lower than a critical level and guarantee the safety and the availability of the installations. These examinations allow to monitor the degradation of the components by collecting successive degradation measurements (for instance crack sizes) from which one can predict the degradation propagation and estimate the residual operation time upon which the critical degradation level will be reached.

Unfortunately, the non-destructive testing processes (for instance ultrasonic testing processes) do not always give a perfect image of the degradation. Indeed, because of technical limitations, if a crack size is too small, the measurement process can detect the existence of the crack but is not able to give a value for its size: the crack is said to be non-sizable. Moreover, if several competing cracks have initiated on a component, the testing process can count the number of existing cracks but can only give the size of the largest one, which is not necessarily the same throughout the whole operation period of the component.

Of course, in order to make accurate degradation predictions, it is essential to take into account this partial information coming from the field. That is why the Université de Pau et des Pays de l'Adour and EDF R&D developed a specific stochastic model to tackle this issue.

1.2 *Main objectives*

The aim of the paper is to detail the model and to illustrate its application on a real industrial case study carried out on a passive component from EDF electric power plants. This passive component is subjected to internal crack initiation and propagation. That is why

examinations are regularly performed on this component to monitor the crack size and predict the residual operation time upon which a given regulatory limit, denoted by ℓ, will be reached. Two testing processes with different objectives and performances are used.

1. The first process aims at detecting cracking: thus it gives a binary response "presence" or "absence of cracking", together with the number of initiated cracks.

2. The second one is conceived to measure crack sizes. Of course, the measurement first requires the detection of the crack by the process, which is less sensitive than the first one. A crack is sizeable by this second testing process if its size is larger than a fixed threshold. Due to the use of two different technologies for the measurements, there are two different possible (known) thresholds under which the crack is non sizeable, which are denoted by c_1 and c_2, respectively. Moreover, as previously mentioned, when several cracks have initiated, the second testing process can only give the size of the largest one at the inspection time.

We make the assumptions that the two testing processes give the exact number of initiated cracks when they detect cracking and that the second testing process gives the exact size of the largest crack when sizeable.

1.3 *Available data*

The available data are the following.

- 228 components are studied. They are supposed to be independent and identical.
- The commission date of each component is known.
- Tests are performed at different inspection times with one of the two testing processes, thus providing one of the possible following information about the degradation level:
  - No crack is detected on the component: no cracking has initiated at the inspection time.
  - At least one crack is detected by the first testing process at the inspection time: the number of initiated cracks is known and the size of the largest crack is lower than c_1 (left censoring).
  - At least one crack is detected by the second testing process at the inspection time: the number of initiated cracks is known, but the size of the largest crack is lower than c_2 and is hence non-sizeable (left censoring again).
  - The second testing process detects and measures cracking: the number of initiated cracks is known, as well as the size for the largest crack on the component (which is higher than c_2).

2 STOCHASTIC MODEL AND ESTIMATION PROCEDURE

2.1 *Stochastic model*

We here specify the stochastic model for the initiation and propagation of cracks on one single component. We first set N_t to be the random number of cracks already initiated at time t on the component. The points of the counting process $(N_t)_{t\geq 0}$ stand for the initiating times of cracks on the component and are denoted by $T_1, \ldots, T_n, \ldots$, with $0 < T_1 < \cdots < T_n < \cdots$ (almost surely). In case $N_t \geq 1$, we set $Z_t^{(i)}$ to be the (random) size of the i-th crack at time t, with $1 \leq i \leq N_t$. The size of the largest crack at time t hence is:

$$Z_t = \begin{cases} \max\limits_{1\leq i\leq N_t} \left(Z_t^{(i)}\right) \text{ if } N_t \geq 1, \\ 0 \text{ elsewhere.} \end{cases}$$

If $0 \leq t_1 < \cdots < t_m$ are the inspection times of the component (different from one component to another), data for one component are $(n_j, z_j)_{1\leq j\leq m}$ with $n_j = N_{t_j}(\omega)$, $z_j = Z_{t_j}(\omega)$. These data correspond to one observation of $\left(N_{t_j}, Z_{t_j}\right)_{1\leq j\leq m}$.

In order to account for the two possible levels of censoring at time t, we set $C_t \in \{c_1, c_2\}$ to be the (known) censoring level at time t. We also introduce a censoring indicator:

$$D_t = \begin{cases} 0 \text{ if } Z_t \leq C_t, \\ 1 \text{ otherwise,} \end{cases}$$

The data are now of the shape $(n_j, u_j, d_j)_{1\leq j\leq m}$ with $n_j = N_{t_j}(\omega)$, $u_j = U_{t_j}(\omega)$, $d_j = D_{t_j}(\omega)$, and correspond to one observation of $\left(N_{t_j}, U_{t_j}, D_{t_j}\right)_{1\leq j\leq m}$ with

$$U_t = \begin{cases} Z_t & \text{if } Z_t > C_t, \\ C_t & \text{if } Z_t \leq C_t \text{ and } N_t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Assumptions
We assume that the cracking processes are identical for all the components and that they are independent between components. We also assume that once initiated, the processes defining the propagation of all cracks present on one component are independent and identically distributed. For each component, the number of cracks follows a Poisson process $(N_t)_{t\geq 0}$, where N_t is Poisson distributed with parameter $\Lambda(t)$ and distribution denoted by $\mathcal{P}(\Lambda(t))$, see Ross (1996)

for more details on Poisson processes. In the sequel, we either suppose that this process is homogeneous (HP) with $\Lambda(t) = \lambda t$, or non homogeneous (NHP) with $\Lambda(t) = \alpha t^{\beta}$ (power-law process). The rate of this Poisson process is denoted by $\lambda(t)$, with $\lambda(t) = \lambda$ for the HP process and $\lambda(t) = \Lambda'(t) = \alpha\beta t^{\beta-1}$ for the NHP process. When a crack appears, we assume that its size grows according to a Gamma process with parameters $(A(t), b)$, see Abdel-Hameed (1975) or Van Noortwijk (2009) for more details on the use of Gamma processes for modelling deterioration. Here again, we envision the homogeneous case (HG) with $A(t) = at$ and the non homogenous case (NHG) with $A(t) = \eta t^{\gamma}$.

To specifically define the sizes of the various cracks present on a component at time t, we introduce a sequence $\left(X^{(i)}\right)_{i\in\mathbb{N}^*}$ of independent Gamma processes, all with parameters $(A(t), b)$. For each $i \in \mathbb{N}^*$ and each $t \geq 0$, the random variable $X_t^{(i)}$ is Gamma distributed with probability density function (p.d.f.) $f_{A(t),b}$ given by

$$f_{A(t),b}(x) = \frac{b^{A(t)}}{\Gamma(A(t))} x^{A(t)-1} e^{-bx} \mathbf{1}_{\mathbb{R}_+}(x)$$

and cumulative distribution function (c.d.f.) denoted by $F_{A(t),b}$. With these notations, we have $\mathbb{E}\left(X_t^{(i)}\right) = A(t)/b$ and $\text{var}\left(X_t^{(i)}\right) = A(t)/b^2$.

For each $i \in \mathbb{N}^*$, we next set

$$Z_t^{(i)} = X_{(t-T_i)^+}^{(i)}$$

where $(t-T_i)^+ = \max(t-T_i, 0)$. If the i-th crack has already been initiated at time t, we have $(t-T_i)^+ = t - T_i$ and $(t-T_i)^+$ stands for the time elapsed between the initiation time of the i-th crack and time t. Therefore, $Z_t^{(i)} = X_{t-T_i}^{(i)}$ is the size of the i-th crack at time t. On the contrary, if $T_i > t$, we obtain $(t-T_i)^+ = 0$ and $Z_t^{(i)} = X_0^{(i)} = 0$. In each case, $Z_t^{(i)}$ hence is nothing but the size of the i-th crack at time t.

As only the largest crack can be measured by an examination, without censoring, a measure at time t hence is an observation of

$$Z_t = \max_{i\in\mathbb{N}^*}\left(Z_t^{(i)}\right)$$

$$= \begin{cases} \max\limits_{1\leq i\leq N_t}\left(Z_t^{(i)}\right) = \max\limits_{1\leq i\leq N_t}\left(X_{t-T_i}^{(i)}\right) \text{ if } N_t \geq 1, \\ 0 \text{ otherwise.} \end{cases}$$

2.2 *Estimation principle*

Remember that our data are independent observations of random vectors $\left(N_{t_j}, Z_{t_j}\right)_{1\leq j\leq m}$, possibly censored, where the number of inspections m and the inspection times t_j depend on the component. It is possible to obtain the joint distribution of $\left(N_{t_j}, Z_{t_j}\right)_{1\leq j\leq m}$ by a recursive formula. However this formula is too tricky to be used with the usual maximum likelihood principle. We consequently suggest another method, based on a two-step procedure.

Let us recall that the parameters to be estimated are:

- Parameter(s) of the Poisson process: $\theta_P = \lambda$ (HP) or $\theta_P = (\alpha, \beta)$ (NHP),
- Parameters of the Gamma process: $\theta_G = (a, b)$ (HG) or $\theta_G = (\eta, \gamma, b)$ (NHG).

As a first step, we classically estimate the parameter θ_P of the Poisson process by the usual maximum likelihood method, based on the uncensored observation of $\left(N_{t_j}\right)_{1\leq j\leq m}$and on the joint distribution of $\left(N_{t_j}\right)_{1\leq j\leq m}$ (which is easy to get). The estimator is denoted by $\hat{\theta}_P$.

As a second step, the parameter of the Gamma processes are next estimated, plugging the estimator of the Poisson parameter into some composite likelihood function (see Cox & Reid (2004) and Varin, Reid, & Firth (2011)), based on the conditional distribution of the Z_{t_j}'s given N_{t_j}'s. More specifically, we consider the following composite likelihood:

$$\mathcal{L}\left(\theta_G | (z_j)_{1\leq j\leq m}, (n_j)_{1\leq j\leq m}; \hat{\theta}_P\right)$$

$$= \prod_{\substack{1\leq j\leq m \\ \text{s.t. } n_j\geq 1}} f_{Z_{t_j}|N_{t_j}}\left(z_j | n_j; \theta_G, \hat{\theta}_P\right),$$

where $f_{Z_{t_j}|N_{t_j}}\left(z_j | n_j; \theta_G, \hat{\theta}_P\right)$ is the conditional p.d.f. of Z_{t_j} given that $N_{t_j} = n_j$ with respect to Lebesgue measure for $n_j \geq 1$, and where θ_P is replaced by $\hat{\theta}_P$.

Observations for which $n_j = 0$ do not contain any information on θ_G because $N_{t_j} = 0$ implies that $Z_{t_j} = 0$. That is why only data such that $n_j \geq 1$ are involved in the likelihood function.

In practice, N independent components are observed. Adding exponent (i) to both processes and observations related to the i-th component, the log-composite-likelihood can be written as

$$\mathcal{L}\left(\theta_G | \mathbf{z}, \mathbf{n}, \mathbf{t}; \hat{\theta}_P\right)$$

$$= \sum_{i=1}^{N} \sum_{\substack{1\leq j\leq m^{(i)} \\ \text{s.t. } n_j^{(i)}\geq 1}} \log\left\{ f_{Z_{t_j^{(i)}}|N_{t_j^{(i)}}}\left(z_j^{(i)} | n_j^{(i)}; \theta_G, \hat{\theta}_P\right)\right\}$$

where $\mathbf{z} = \left(z_j^{(i)}\right)_{\substack{1\leq j\leq m^{(i)} \\ 1\leq i\leq N}}$, $\mathbf{n} = \left(n_j^{(i)}\right)_{\substack{1\leq j\leq m^{(i)} \\ 1\leq i\leq N}}$ and $\mathbf{t} = \left(t_j^{(i)}\right)_{\substack{1\leq j\leq m^{(i)} \\ 1\leq i\leq N}}$.

If in addition we take into account censoring, we get:

$$\mathcal{L}\left(\theta_G|\mathbf{u},\mathbf{n},\mathbf{t},\mathbf{d};\hat{\theta}_P\right) \tag{1}$$

$$=\sum_{i=1}^{N}\sum_{\substack{1\le j\le m^{(i)}\\ \text{s.t. } n_j^{(i)}\ge 1}} d_j^{(i)}\log\left\{f_{Z_{t_j^{(i)}}|N_{t_j^{(i)}}}\left(u_j^{(i)}|n_j^{(i)};\theta_G,\hat{\theta}_P\right)\right\}$$

$$+\left(1-d_j^{(i)}\right)\log\left\{F_{Z_{t_j^{(i)}}|N_{t_j^{(i)}}}\left(u_j^{(i)}|n_j^{(i)};\theta_G,\hat{\theta}_P\right)\right\}$$

with $u_j^{(i)}=z_j^{(i)}$ if $d_j^{(i)}=1$ and $u_j^{(i)}=c_j^{(i)}$ otherwise, and $\mathbf{d}=\left(d_j^{(i)}\right)_{\substack{1\le j\le m^{(i)}\\ 1\le i\le N}}$.

2.3 *Fitting the Poisson process*

Parameter θ_P of the Poisson process can be estimated from data $(\mathbf{n},\mathbf{t})$ by maximizing the log-likelihood function:

$$\mathcal{L}\left(\theta_P|\mathbf{t},\mathbf{n}\right)=\sum_{i=1}^{N}\log\left[\mathbb{P}\left(\cap_{j=1}^{m^{(i)}}\left\{N_{t_j^{(i)}}=n_j^{(i)}\right\}\right)\right].$$

Using the independent increments of $(N_t)_{t\ge 0}$, we easily get:

$$\log\left[\mathbb{P}\left(\cap_{j=1}^{m}\left\{N_{t_j}=n_j\right\}\right)\right]$$

$$\propto -\Lambda\left(t_m\right)+\sum_{j=0}^{m-1}\left(n_{j+1}-n_j\right)\log\left(\Lambda\left(t_{j+1}\right)-\Lambda\left(t_j\right)\right)$$

for all $0=t_0<t_1<\cdots<t_m$ and $0=n_0\le n_1\le\cdots\le n_m$, where $\propto$ means "is equal to" up to an additive constant independent of the parameter of interest (here θ_P).

The log-likelihood can then be written as

$$\mathcal{L}\left(\theta_P|\mathbf{t},\mathbf{n}\right)\propto -\sum_{i=1}^{N}\Lambda\left(t_{m^{(i)}}^{(i)}\right)+$$

$$\sum_{i=1}^{N}\sum_{j=0}^{m^{(i)}-1}\left(n_{j+1}^{(i)}-n_j^{(i)}\right)\log\left(\Lambda\left(t_{j+1}^{(i)}\right)-\Lambda\left(t_j^{(i)}\right)\right),$$

where we set $t_0^{(i)}=n_0^{(i)}=0$, for all $1\le i\le N$. section 2.3 :

In case of a non homogeneous Poisson process, we have $\Lambda\left(t\right)=\alpha t^{\beta}$ and $\theta_P=(\alpha,\beta)$. Fixing β and solving $\frac{\partial\mathcal{L}}{\partial\alpha}\left(\alpha,\beta|\mathbf{t},\mathbf{n}\right)=0$, leads to

$$\alpha(\beta)=\frac{\sum_{i=1}^{N}n_{m^{(i)}}^{(i)}}{\sum_{i=1}^{N}\left(t_m^{(i)}\right)^{\beta}}. \tag{2}$$

We replace α by $\alpha(\beta)$ in the likelihood function, and then we look for the maximizer $\hat{\beta}$ of

$$\mathcal{L}\left(\beta|\mathbf{t},\mathbf{n}\right)\equiv\mathcal{L}\left(\alpha(\beta),\beta|\mathbf{t},\mathbf{n}\right) \tag{3}$$

$$\propto -\left(\sum_{i=1}^{N}n_{m^{(i)}}^{(i)}\right)\log\left(\sum_{i=1}^{N}\left(t_{m^{(i)}}^{(i)}\right)^{\beta}\right)$$

$$+\sum_{i=1}^{N}\sum_{j=0}^{m^{(i)}-1}\left(n_{j+1}^{(i)}-n_j^{(i)}\right)\log\left(\left(t_{j+1}^{(i)}\right)^{\beta}-\left(t_j^{(i)}\right)^{\beta}\right).$$

Using a numerical optimization method, we first obtain $\hat{\beta}$; then we set $\hat{\alpha}=\alpha(\hat{\beta})$.

The homogeneous case ($\Lambda\left(t\right)=\lambda t$) is a particular case of the NHP case with $\beta=1$ and $\alpha=\lambda$. Thus from (2), we obtain

$$\hat{\lambda}=\frac{\sum_{i=1}^{N}n_{m^{(i)}}^{(i)}}{\sum_{i=1}^{N}t_{m^{(i)}}^{(i)}}.$$

2.4 *Fitting the Gamma process*

The parameter θ_G of the Gamma process is estimated from $(\mathbf{u},\mathbf{n},\mathbf{t},\mathbf{d})$ by maximizing the log-composite-likelihood function given by (1). First we show that for $n\ge 1$ and $z>0$:

$$\mathbb{P}\left(Z_t\le z|N_t=n\right)=\left\{\frac{\int_0^t F_{A(y),b}\left(z\right)\lambda\left(t-y\right)dy}{\Lambda\left(t\right)}\right\}^n \tag{4}$$

and

$$f_{Z_t|N_t}\left(z|n\right)=\frac{n}{\Lambda\left(t\right)^n}\left\{\int_0^t f_{A(y),b}\left(z\right)\lambda\left(t-y\right)dy\right\}$$

$$\times\left\{\int_0^t F_{A(y),b}\left(z\right)\lambda\left(t-y\right)dy\right\}^{n-1}. \tag{5}$$

(Details are not provided here, due to the reduced size of the paper).

Setting $\hat{\lambda}(t)=\lambda(t;\hat{\theta}_P)$ and plugging formulae (4) and (5) into (1), we obtain

$$\mathcal{L}\left(\theta_G|\mathbf{u},\mathbf{n},\mathbf{t},\mathbf{d};\hat{\theta}_P\right)\propto \tag{6}$$

$$\sum_{i=1}^{N}\sum_{\substack{1\le j\le m^{(i)}\\ \text{s.t. } n_j^{(i)}\ge 1}}\left[d_j^{(i)}\log\left(\int_0^{t_j^{(i)}}f_{A(y),b}\left(u_j^{(i)}\right)\hat{\lambda}\left(t_j^{(i)}-y\right)dy\right)\right.$$

$$\left.+\left(n_j^{(i)}-d_j^{(i)}\right)\log\left(\int_0^{t_j^{(i)}}F_{A(y),b}\left(u_j^{(i)}\right)\hat{\lambda}\left(t_j^{(i)}-y\right)dy\right)\right].$$

In case of a homogeneous Gamma process, $A(y)$ is replaced by ay in (6). Parameters (a, b) are estimated by $(\hat{a}, \hat{b})$ which maximizes $\mathcal{L}\left(a, b|\mathbf{u}, \mathbf{n}, \mathbf{t}, \mathbf{d}; \hat{\theta}_P\right)$.

In case of a non homogeneous Gamma process, $A(y)$ is replaced by ηy^γ in (6). Parameters (η, γ, b) are estimated by $\left(\hat{\eta}, \hat{\gamma}, \hat{b}\right)$ which maximizes $\mathcal{L}\left(\eta, \gamma, b|\mathbf{u}, \mathbf{n}, \mathbf{t}, \mathbf{d}; \hat{\theta}_P\right)$.

3 MONTE CARLO STUDY

This section is devoted to a simulation study which aims at checking the validity of our estimation method. The next section deals with our application to the real data from EDF. As we will see, EDF data are well fitted by combining the NHP process with the HG process. That is why the present section is devoted to this specific model, which we denote by NHP & HG.

Parameters of the NHP process with cumulative intensity $\Lambda(t) = \alpha t^\beta$ are fixed to $(\alpha, \beta) = (1, 1.5)$. Parameters of the HG process are fixed to $(a, b) = (1, 2)$. Samples are generated according to the NHP & HG model. Two setups are tested:

- we generate samples of size $n = 228$, with identical inspection times as those of the 228 EDF components, up to a multiplicative constant.
- we generate samples of size $n = 4 \times 228 = 912$; the inspection times of the 228 EDF components are used four times each (with the same multiplicative constant).

In order to be as close as possible from the EDF data, the cracks sizes are censored. More specifically, if the largest simulated crack is smaller than 0.2 by an inspection, then its size is put to 0.2 ($u_j^{(i)} = 0.2$) and the censoring indicator is put to 0, indicating a censor ($d_j^{(i)} = 0$).

A point of interest for the industrial study was also to know whether it was worth taking into account censoring. That is why two different estimations are performed. First, we estimate the unknown parameters without taking into account censoring. This means that censoring levels are considered to be observed measurements, or equivalently, that censoring values represent crack sizes at observation times. Of course, this leads to an over-estimation of the cracks sizes. However, this does not infer on the number of cracks at observation times and consequently does not infer on the estimation procedure of the Poisson parameter. In a second step, all parameters are estimated taking into account censoring.

3.1 *Estimation without taking into account censoring*

We use the log-composite-likelihood function (6) with all $d_j^{(i)}$'s equal to 1. Results are provided in Table 1. As for the parameters of the NHP process, we observe that, as expected, they are well estimated, both from a bias and standard deviation point of view. Also, all confidence intervals contain the true values of the parameters.

The results concerning the HG process are less convincing. Indeed, even if empirical means are not too far from their true values, we can see that there is some bias on the estimates of parameters a and b, and also on the mean rates per unit time for both mean (a/b) and variance (a/b^2). As a matter of fact, confidence intervals do not always contain the true values, especially when the sample size is large $(n = 912)$. As will be seen later on, this bias is a simple consequence of the censoring procedure which over-estimates the cracks sizes and which is not taken in consideration here. As a consequence, fitting the model using over-estimated cracks sizes is not convenient. Note however that standard deviations are divided by two when the sample size is multiplied by four, which is an indicator of a normal asymptotic behaviour, that is of a "good" behaviour.

3.2 *Estimation without taking into account censoring*

The quality of estimation results (see Table 2) is good for both processes (NHP and HG). Bias on the HG process parameters almost disappear. Confidence intervals are well centered on the true parameters values and standard deviations are divided by two as the sample size is multiplied by four. Obviously, it is preferable to take censoring into consideration, even though the impact remains moderate.

4 INDUSTRIAL CASE STUDY

EDF data concern $N = 228$ components. Each component has been inspected several times. The total number of inspection times is 1695. Our approach requires both numerical integration and numerical optimization for the computation and optimization of the log-composite-likelihood function provided by (6). To validate the numerical results, two different computer programs are used, one written with `R` (R Core Team (2013)), the other one with `MATLAB` (MATLAB (2010)). The two programs provide very similar results. In addition, confidence intervals are computed via standard bootstrap method, using 1000 sets of 228 trajectories uniformly drawn (with replacement) from the 228 trajectories of our data set. For each bootstrap sample, estimates are provided for the parameters. Empirical mean, standard-deviation and both 90% and 95% confidence intervals are next derived

for each parameter, based on the 1000 estimation results. The 90% (resp. 95%) confidence interval corresponds to $[q_{0.05}, q_{0.95}]$ (resp. $[q_{0.025}, q_{0.975}]$), where q_α is the α-empirical-quantile of the 1000 estimation results. As often, this method is quite time consuming. Note that, based on the strong evidence provided by the Monte Carlo study for taking into account censoring, all the results of this section are computed under this basis.

4.1 *Crack initiation process*

Estimation results concerning the Poisson process (HP or NHP) are given in Table 3. The results obtained by the bootstrap method are summarized in Table 4.

Table 3: Estimation results for the Poisson process

Model	Estimates	
HP	$\hat{\lambda}$	
	$1.563\,4 \times 10^{-4}$	
NHP	$\hat{\alpha}$	$\hat{\beta}$
	$1.226\,7 \times 10^{-9}$	$2.281\,1$

Table 5 gives the expected annual number of new cracks. For the HP process, it is equal to $\hat{\lambda} \times 365$ whereas for the NHP process it is equal to $\hat{\alpha} \times \left\{(365\ i)^{\hat{\beta}} - (365\ (i-1))^{\hat{\beta}}\right\}$ for the i-th year (data unit is day).

Based on empirical considerations and on the bootstrap confidence interval provided in Table 4 for parameter β, we suggest to retain the non homogenous version of the Poisson process (NHP) to model cracks initiation times.

Table 5: Annual mean number of new cracks.

Model	Annual mean number of new cracks	
HP	5.71×10^{-2}	
NHP	Year 1	8.58×10^{-4}
	Year 2	3.31×10^{-3}
	Year 3	6.35×10^{-3}
	Year 4	9.76×10^{-3}
	Year 5	1.35×10^{-2}
	Year 6	1.74×10^{-2}
	Year 8	2.59×10^{-2}
	Year 10	3.50×10^{-2}
	Year 12	4.47×10^{-2}
	Year 15	6.02×10^{-2}

4.2 *Crack propagation process*

Estimation results are given in Table 6 for the parameters of both HG and NHG processes. The results obtained by the bootstrap method are summarized in Table 7 for the parameters of the Gamma processes for the NHP & HG model.

Table 8 provides the annual mean crack growth. As we can see, in case of a NHG process, the crack growth is strongly non-linear. Based on empirical considerations, we hence suggest to retain the homogeneous version of the Gamma process (combined with an NHP process). However, this point requires further investigation. For example, a bootstrap confidence interval for the unknown parameter γ could be calculated in order to test the null hypothesis $\gamma = 1$ (meaning that the Gamma process is homogeneous). We can use the same bootstrap method as for the NHP & HG model but we have to face the problem of very large calculation time for such an approach.

Table 6: Estimation results on EDF data including censoring

Model	Estimates		
HP & HG	$\hat{a}$	$\hat{b}$	
	2.4354×10^{-4}	7.9669×10^{-2}	
NHP & HG	$\hat{a}$	$\hat{b}$	
	4.8635×10^{-4}	9.8412×10^{-2}	
HP & NHG	$\hat{\eta}$	$\hat{\gamma}$	$\hat{b}$
	0.46800	9.2245×10^{-9}	4.0301×10^{-2}
NHP & NHG	$\hat{\eta}$	$\hat{\gamma}$	$\hat{b}$
	0.35744	2.2407×10^{-10}	3.2708×10^{-2}

Table 8: Expected annual growth of a crack from EDF data

Model	Annual mean of crack growth	
HP & HG	1.1158	
NHP & HG	1.8038	
HP & NHG	Year 1	11.61
	Year 2	7.43×10^{-8}
NHP & NHG	Year 1	10.93
	Year 2	1.70×10^{-9}

4.3 *Additional indicators*

Apart from the previously mentioned indicators (annual mean number of new cracks, annual mean crack growth), other indicators of interest for the applications are linked to the hitting time τ_ℓ of the critical degradation level $\ell > 0$, defined by $\tau_\ell = \inf\{t > 0; Z_t \geq \ell\}$. As an example, we are here interested in the quantile t_α of the reaching time τ_ℓ (where $\alpha \in (0,1)$), which is such that:

$$\mathbb{P}\left(\tau_\ell < t_\alpha\right) = \alpha,$$

or equivalently such that:

$$\mathbb{P}\left(Z_{t_\alpha} > \ell\right) = \alpha.$$

For the NHP & NHG model (the most general), this equation may be written as:

$$\int_0^{t_\alpha} \bar{F}_{A(y),b}\left(\ell\right) \lambda\left(t_\alpha - y\right) dy = -\ln\left(1-\alpha\right),$$

where $\bar{F}_{A(y),b}\left(\ell\right) = 1 - F_{A(y),b}\left(\ell\right)$.

For the NHP & HG model, we have to solve numerically the following equation:

$$\int_0^{t_\alpha} \bar{F}_{ay,b}\left(\ell\right) \left(t_\alpha - y\right)^{\beta-1} dy = -\frac{\ln\left(1-\alpha\right)}{\alpha\beta}. \qquad (7)$$

Table 1: Monte Carlo estimations for the NHP-HG model based on K simulated censored samples of size n, without considering censoring.

	n	K	a	b	a/b	a/b^2	α	β
value			1	2	0.5	0.25	1	1.5
	228	1000	1.125	2.199	0.511	0.234	0.999	1.500
mean	912	108	1.115	2.184	0.511	0.234	1.000	1.500
	228	1000	0.110	0.182	0.011	0.017	0.046	0.019
stand. err.	912	108	0.056	0.091	0.006	0.008	0.023	0.010
90% confidence	228	1000	[0.953,1.313]	[1.921,2.522]	[0.493,0.529]	[0.206,0.260]	[1.170,1.346]	[1.469,1.530]
interval	912	108	[1.032,1.219]	[2.041,2.352]	[0.500,0.521]	[0.211,0.247]	[0.963,1.038]	[1.484,1.516]
95% confidence	228	1000	[0.928,1.358]	[1.862,2.588]	[0.490,0.532]	[0.201,0.267]	[1.156,1.365]	[1.462,1.536]
interval	912	108	[1.074,1.230]	[2.118,2.367]	[0.506,0.521]	[0.229,0.247]	[0.985,1.040]	[1.494,1.521]

Table 2: Monte Carlo estimations for the NHP-HG model based on K simulated censored samples of size n considering censoring.

	n	K	a	b	a/b	a/b^2	α	β
value			1	2	0.5	0.25	1	1.5
	228	825	1.000	2.040	0.503	0.248	0.998	1.501
mean	912	231	1.007	2.011	0.501	0.249	1.000	1.500
	228	825	0.100	0.168	0.011	0.017	0.043	0.018
stand. err.	912	231	0.049	0.083	0.006	0.009	0.022	0.009
90% confidence	228	825	[0.878,1.205]	[1.787,2.347]	[0.484,0.521]	[0.219,0.276]	[0.929,1.071]	[1.471,1.532]
interval	912	231	[0.922,1.084]	[1.874,2.141]	[0.491,0.509]	[0.234,0.265]	[0.962,1.035]	[1.485,1.515]
95% confidence	228	825	[0.852,1.237]	[1.746,2.397]	[0.481,0.524]	[0.212,0.281]	[0.912,1.086]	[1.466,1.537]
interval	912	231	[0.910,1.093]	[1.846,2.156]	[0.489,0.511]	[0.233,0.267]	[0.953,1.042]	[1.480,1.518]

Table 4: Results based on 1000 bootstrapped samples of size 228 for the parameters of the Poisson process.

	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$
mean	1.568×10^{-4}	2.121×10^{-9}	2.289
stand. err.	1.687×10^{-5}	2.775×10^{-9}	0.130
90% conf. int.	$[1.309 \times 10^{-4}, 1.845 \times 10^{-4}]$	$[1.604 \times 10^{-10}, 7.026 \times 10^{-9}]$	$[2.081, 2.511]$
95% conf. int.	$[1.275 \times 10^{-4}, 1.914 \times 10^{-4}]$	$[1.176 \times 10^{-10}, 1.012 \times 10^{-8}]$	$[2.041, 2.550]$

Table 7: Results based on 1000 bootstrapped samples of size 228 for the parameters of the Gamma process, case PNH & GH.

	$\hat{a}$	$\hat{b}$	$\hat{a}/\hat{b}$	$\hat{a}/\hat{b}^2$
mean	5.8887×10^{-4}	0.11472	4.9527×10^{-3}	4.8829×10^{-2}
stand. err.	1.0881×10^{-3}	0.16965	5.1462×10^{-4}	1.1652×10^{-2}
90% conf. int.	$[2.8948, 9.8237] \times 10^{-4}$	$[6.6464 \times 10^{-2}, 0.17594]$	$[4.1185, 5.7648] \times 10^{-3}$	$[3.0292, 6.8574] \times 10^{-2}$
95% conf. int.	$[2.5470, 11.644] \times 10^{-4}$	$[6.0923 \times 10^{-2}, 0.20922]$	$[3.9469, 5.9619] \times 10^{-3}$	$[2.6929, 7.2000] \times 10^{-2}$

Table 9: Estimated α-quantiles t_α of the hitting time τ_ℓ

α	0.25	0.50	0.75	0.90	0.95
t_α	420.48	443.08	480.20	525.54	619.19

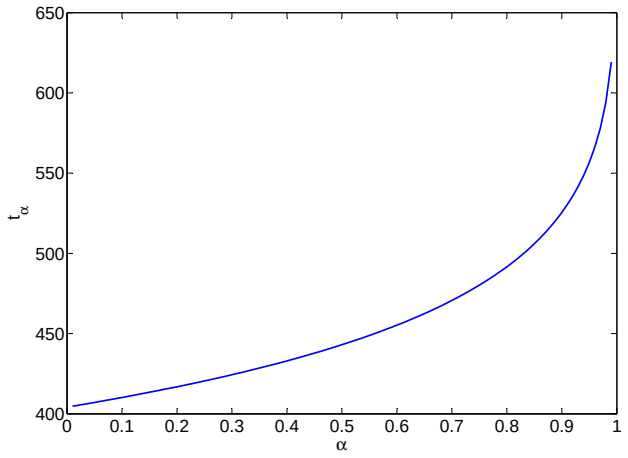

Figure 1: Quantile function $\alpha \mapsto t_\alpha$ of the hitting time τ_ℓ

For a fixed level ℓ, estimation results are obtained by substituting (α, β, a, b) by $(\hat{\alpha}, \hat{\beta}, \hat{a}, \hat{b})$ in (7) and then solving equation (7) with respect to t_α. Results are provided in Table 9 and Figure 1.

5 CONCLUDING REMARKS

We proposed a stochastic model for the bivariate process $(N_t, Z_t)_{t\geq 0}$ describing initiation and growth of cracks on a passive component from EDF electric power plants. This model associates a Poisson process together with a Gamma process, where both processes may be homogeneous or not. This leads to four different models, which have been fitted to EDF data. The finally retained model is based on a bootstrap confidence interval for the shape parameter of the non homogeneous Poisson process and on empirical considerations for the Gamma process. It is a combination of a NHP process for apparition of cracks and a HG process for describing cracks growth. This point however requires further investigation. Indeed, selecting one model is a challenging statistical objective, that requires to use some information criteria like AIC, BIC, etc. that have to be adapted to our estimation method, which is based on composite likelihood method instead of classical likelihood approach (see Gao & Song (2010) for some recent results concerning BIC selection model criteria for composite likelihood approach).

The Monte Carlo simulation study shows that the sample size of the EDF data is sufficient to guarantee the quality of estimation results, except maybe for the parameter α of the NHP process which has a rather large estimated standard deviation. This study also shows that taking into account censoring is necessary to avoid some bias on estimates of the Gamma process parameters. This bias disappears when an estimation method taking into account censoring is used.

Another challenging statistical issue would be to integrate measurement errors in our model. There are two levels of possible errors of measurement. The first level deals with the number of cracks, where typically, too small cracks may remain unrevealed by the testing processes. The second level is linked to the sizing process, which may be spoiled by measurement errors.

Acknowledgement This work was supported by the French National Research Agency (ANR), AMMSI project, ref. ANR-2011-BS01-021.

REFERENCES

Abdel-Hameed, M. (1975). A gamma wear process. *IEEE Transactions on Reliability 24*(**2**), 152–153.

Cox, D. & N. Reid (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika 91*, 729–737.

Gao, X. & P. X.-K. Song (2010). Composite likelihood bayesian information criteria for model selection in high-dimensional data. *Journal of the American Statistical Association 105(492)*, 1531–1540.

MATLAB (2010). *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ross, S. M. (1996). *Stochastic processes*. Wiley, New York.

Van Noortwijk, J. (2009). Bivariate Gamma wear processes for track geometry modelling, with application to intervention scheduling. *Struct. Infrastruct. E 94*(**1**), 2–21.

Varin, C., N. Reid, & D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica 21*, 5–42.