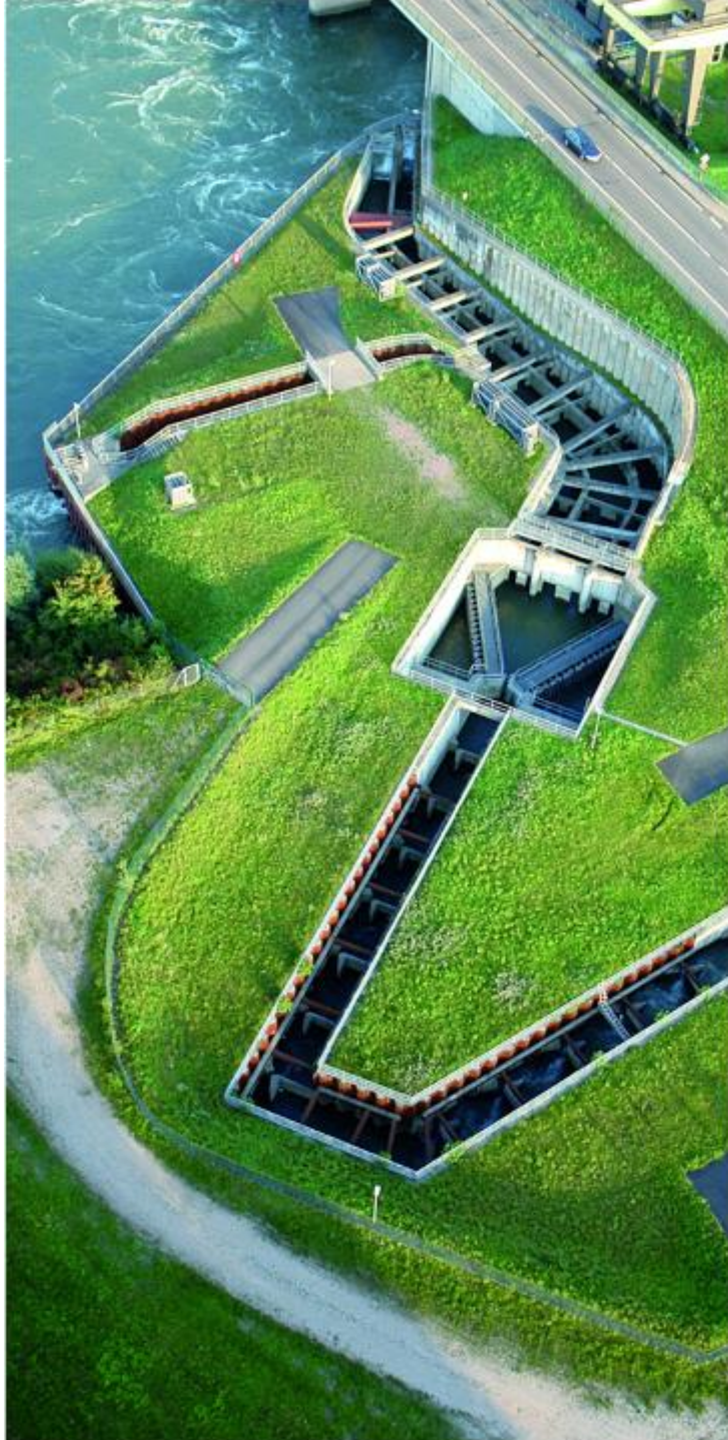
A stochastic process for censored degradation data

Project AMMSI
Third annual workshop - UTT

E. Dautrême and E. Remy (EDF R&D)
S. Mercier and L. Bordes (University of Pau)

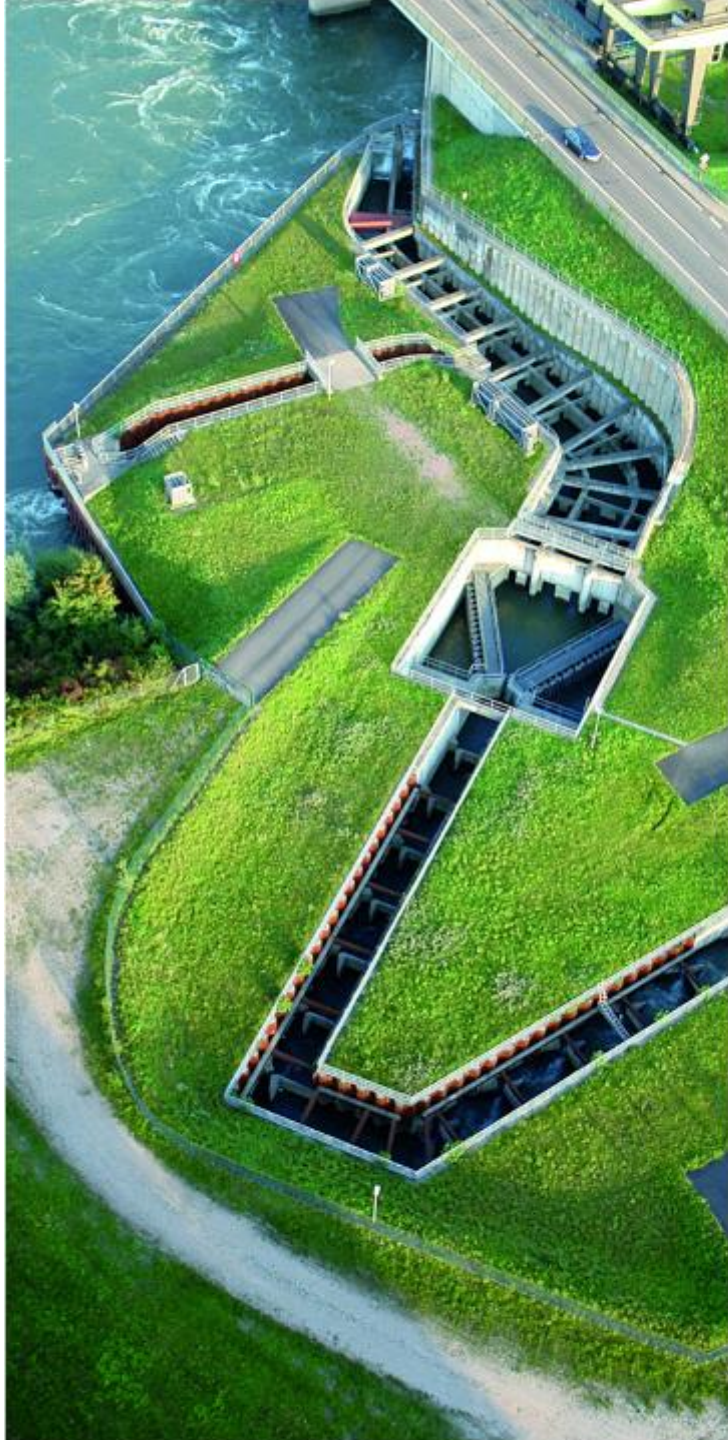
January 29, 2015



OUTLINE

1. INDUSTRIAL CONTEXT, AVAILABLE DATA AND OBJECTIVES
2. STOCHASTIC PROCESS AND STATISTICAL INFERENCE PROCEDURE
3. APPLICATION TO EDF DATA
4. CONCLUSIONS AND PROSPECTS

Industrial context, available data and objectives



INDUSTRIAL CONTEXT AND AVAILABLE DATA

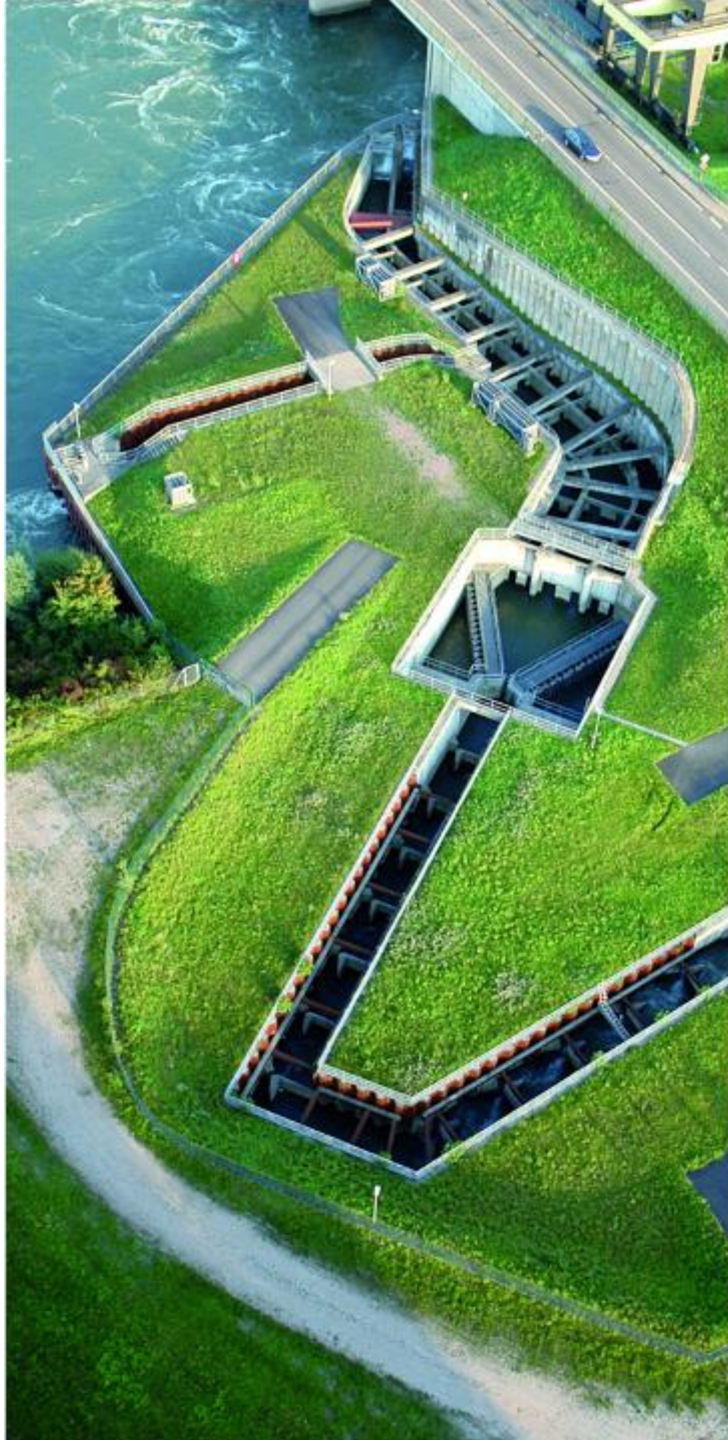
- **Periodic in-service inspections of the components within EDF electric power plants in order to ensure that the flaw sizes remain admissible**
- **No perfect image of the degradations given by the non-destructive testing processes, but only partial and censored information**
- **Two types of information:**
  1. Total number of initiated flaws
  2. Size of the largest initiated flaw
  - Fictitious example for a given component:

Inspection time t_j	Total number of initiated flaws n_j	Lower bound ℓ_j of the size of the largest initiated flaw	Upper bound u_j of the size of the largest initiated flaw	
10	0	0	0	↤ No initiated flaw
90	2	0	20	↤ 2 initiated flaws with sizes lower than 20
300	3	20	25	↤ 3 initiated flaws and size of the largest one between 20 and 25
500	4	37	37	↤ 4 initiated flaws and size of the largest one equal to 37

OBJECTIVES

- **Development of a specific stochastic model:**
  - Driven by this partial censored information coming from the field
  - Allowing to derive useful indicators for the reliability engineer, for instance:
    - Initiation time of the first flaw
    - Propagation of the flaw size over time
    - Hitting time of a given degradation threshold
- **Studying the impact of taking into account the information brought by the total number of initiated flaws**

Stochastic process and statistical inference procedure



NOTATIONS

- **Notations for one component:**
  - N_t: random number of initiated flaws at time $t \geq 0$
  - $T_1 < T_2 < \ldots < T_n < \ldots$: successive random initiation times of the flaws
  - $Z_t^{(i)} = X^{(i)}_{(t-T_i)^+}$: random size of the i-th flaw at time t
  - $Z_t = \begin{cases} \max\limits_{0 \leq i \leq N_t} \left(Z_t^{(i)}\right) = \max\limits_{0 \leq i \leq N_t} \left(X^{(i)}_{(t-T_i)^+}\right) & \text{if } N_t \geq 1 \\ 0 \text{ if } N_t = 0 \end{cases}$: random size of the largest flaw at time t
- **Notations for several components:**
  - N: total number of components
  - $m^{(k)}$: total number of measurements on the k-th component ($1 \leq k \leq N$)
  - $t_j^{(k)}$: j-th inspection time ($1 \leq j \leq m^{(k)}$) of the k-th component
  - $n_j^{(k)}$: observed total number of initiated flaws on the k-th component at inspection time $t_j^{(k)}$
  - $z_j^{(k)}$: measured size of the largest flaw on the k-th component at inspection time $t_j^{(k)}$
  - $\ell_j^{(k)}$ and $u_j^{(k)}$: lower and upper bounds of the size of the largest flaw on the k-th component at inspection time $t_j^{(k)}$ with $\ell_j^{(k)} \leq z_j^{(k)} \leq u_j^{(k)}$
    - If $u_j^{(k)} = 0$: no initiated flaw on the k-th component at inspection time $t_j^{(k)}$
    - If $0 < \ell_j^{(k)} = u_j^{(k)}$: size $z_j^{(k)}$ of the largest flaw on the k-th component at inspection time $t_j^{(k)}$ equal to $\ell_j^{(k)} = u_j^{(k)}$

ASSUMPTIONS

- **Stochastic independence between the:**
  - Components, also considered to be identical
  - Initiated flaws
  - Initiation and propagation phases

- **Non homogeneous Poisson process for initiation:**
  - $\mathbb{P}(N_t = n) = \exp(-\alpha t^\beta) \frac{(\alpha t^\beta)^n}{n!}, t \geq 0, n \geq 0, \alpha > 0, \beta > 0$:
    - Rate function: $\lambda_{\alpha,\beta}(t) = \alpha\beta t^{\beta-1}$
    - Cumulative rate function: $\Lambda_{\alpha,\beta}(t) = \int_0^t \lambda_{\alpha,\beta}(u)\, du = \alpha t^\beta$
    - Expected value: $\mathbb{E}(N_t) = \alpha t^\beta$

- **Homogeneous Gamma process for propagation of one flaw size:**
  - $X_t^{(i)} \sim$ Gamma distribution $\mathcal{G}(at,b), t \geq 0, a > 0, b > 0$:
    - Probability density function: $f_{at,b}(x) = \frac{b^{at}}{\Gamma(at)} x^{at-1} \exp(-bx), x \geq 0$
    - Cumulative distribution function: $\mathbb{F}_{at,b}(x) = \int_0^x f_{at,b}(u)du$
    - Expected value: $\mathbb{E}(X_t^{(i)}) = at/b$
    - Variance: $\mathbb{V}(X_t^{(i)}) = at/b^2$

STATISTICAL INFERENCE PROCEDURE

- **Available data:**

$$\mathbf{n} = \left(n_j^{(i)}\right)_{1 \le j \le m^{(i)}, 1 \le i \le N}, \boldsymbol{\ell} = \left(\ell_j^{(i)}\right)_{1 \le j \le m^{(i)}, 1 \le i \le N}$$

$$\mathbf{u} = \left(u_j^{(i)}\right)_{1 \le j \le m^{(i)}, 1 \le i \le N}, \mathbf{t} = \left(t_j^{(i)}\right)_{1 \le j \le m^{(i)}, 1 \le i \le N}$$

- **Maximum likelihood (ML) method:**
  - Requires the joint distribution of $\{(N_{t_1},Z_{t_1}),\ldots,(N_{t_m},Z_{t_m})\}_{m \ge 1}$ and numerical integration methods
  - Tractable with difficulty
  - Nevertheless, the distribution of $(N_t)_{t \ge 0}$ can be easily written…

- **Two-step statistical inference procedure:**
  - First step: ML estimation of parameters (α,β) using $\mathbf{n}$ and $\mathbf{t}$
  - Second step: maximisation of the composite likelihood function(*), based on the conditional distribution of $(Z_t)_{t \ge 0}$ given $(N_t)_{t \ge 0}$, in order to estimate parameters (a,b) using $\mathbf{n}$, $\mathbf{t}$, $\boldsymbol{\ell}$ and $\mathbf{u}$, and replacing (α,β) by their estimates obtained at the end of the first step
  - Validated on simulated data

(*) D. Cox, N. Reid (2004). A note on pseudolikelihood constructed from marginal densities. Biometrika 91: 729–737
C. Varin, N. Reid, D. Firth (2011). An overview of composite likelihood methods. Statistica Sinica 21: 5–42

STATISTICAL INFERENCE PROCEDURE

- **Conditional distribution of $\mathbf{Z_t}$ knowing $\mathbf{N_t = n \geq 1}$:**
  - Cumulative distribution function:

$$\mathbb{F}_{Z_t|N_t}(z|n; a, b, \alpha, \beta) = \mathbb{P}(Z_t \leq z|N_t = n) = \mathbb{P}\left(\max_{0 \leq i \leq n}\left(X^{(i)}_{t-T_i}\right) \leq z|N_t = n\right)$$
$$= \left(\frac{\int_0^t \mathbb{F}_{ay,b}(z)\lambda_{\alpha,\beta}(t-y)dy}{\Lambda_{\alpha,\beta}(t)}\right)^n$$

  - Probability density function:

$$f_{Z_t|N_t}(z|n; a, b, \alpha, \beta) = \frac{n}{\left(\Lambda_{\alpha,\beta}(t)\right)^n}\left(\int_0^t f_{ay,b}(z)\lambda_{\alpha,\beta}(t-y)dy\right)\left(\int_0^t \mathbb{F}_{ay,b}(z)\lambda_{\alpha,\beta}(t-y)dy\right)^{n-1}$$

STATISTICAL INFERENCE PROCEDURE

- **First step of the statistical inference procedure:**

$$\hat{\beta} = \arg\max_{\beta>0} \mathcal{L}(\beta|\mathbf{n},\mathbf{t})$$

with:

$$\mathcal{L}(\beta|\mathbf{n},\mathbf{t}) = -\left(\sum_{i=1}^{N} n_{m^{(i)}}^{(i)}\right)\log\left(\sum_{i=1}^{N}\left(t_{m^{(i)}}^{(i)}\right)^{\beta}\right) + \sum_{i=1}^{N}\sum_{j=1}^{m^{(i)}-1}\left(n_{j+1}^{(i)} - n_j^{(i)}\right)\log\left(\left(t_{j+1}^{(i)}\right)^{\beta} - \left(t_j^{(i)}\right)^{\beta}\right)$$

then:

$$\hat{\alpha} = \alpha(\hat{\beta}) = \frac{\sum_{i=1}^{N} n_{m^{(i)}}^{(i)}}{\sum_{i=1}^{N}\left(t_{m^{(i)}}^{(i)}\right)^{\hat{\beta}}}$$

- **Second step of the statistical inference procedure:**

$$(\hat{a},\hat{b}) = \arg\max_{(a,b)>0} \mathcal{L}_{\hat{\beta}}(a,b|\mathbf{n},\mathbf{t},\boldsymbol{\ell},\mathbf{u})$$

with:

$$\mathcal{L}_{\beta}(a,b|\mathbf{n},\mathbf{t},\boldsymbol{\ell},\mathbf{u})$$

$$= \sum_{i=1}^{N}\sum_{\substack{1\leq j\leq m^{(i)} \\ \text{such as } 0<\ell_j^{(i)}=u_j^{(i)}}} \log\left(\int_0^{t_j^{(i)}} f_{ay,b}\left(u_j^{(i)}\right)\left(t_j^{(i)}-y\right)^{\beta-1}dy\right) + \sum_{i=1}^{N}\sum_{\substack{1\leq j\leq m^{(i)} \\ \text{such as } 0<\ell_j^{(i)}=u_j^{(i)}}} \left(n_j^{(i)}-1\right)\log\left(\int_0^{t_j^{(i)}} \mathbb{F}_{ay,b}\left(u_j^{(i)}\right)\left(t_j^{(i)}-y\right)^{\beta-1}dy\right)$$

$$+ \sum_{i=1}^{N}\sum_{\substack{1\leq j\leq m^{(i)} \\ \text{such as } n_j^{(i)}\geq 1 \text{ and } 0\leq\ell_j^{(i)}<u_j^{(i)}}} \log\left(\left\{\int_0^{t_j^{(i)}} \mathbb{F}_{ay,b}\left(u_j^{(i)}\right)\left(t_j^{(i)}-y\right)^{\beta-1}dy\right\}^{n_j^{(i)}} - \left\{\int_0^{t_j^{(i)}} \mathbb{F}_{ay,b}\left(\ell_j^{(i)}\right)\left(t_j^{(i)}-y\right)^{\beta-1}dy\right\}^{n_j^{(i)}}\right)$$

- **Confidence intervals on the parameters obtained by non parametric bootstrap**

DERIVED RELIABILITY INDICATORS

- **Initiation time of the first flaw** $U \rightsquigarrow$ **Weibull distribution** $\mathcal{W}(\alpha^{-1/\beta},\beta)$:
  - $\mathbb{P}(U > t) = \mathbb{P}(N_t = 0) = \exp(-\Lambda_{\alpha,\beta}(t)) = \exp(-\alpha t^{\beta})$
- **Mean propagation of the size of one initiated flaw:** $\propto a/b$
- **Quantile of the hitting time of a given degradation threshold:**
  - Denoting $t_\varepsilon^{(t_0,z_0,n_0)}$, $t_0 \geq 0, z_0 <$, $n_0 > 0$, the ε-quantile of the hitting time τ of a given degradation threshold d knowing $Z_{t_0} = z_0$ and $N_{t_0} = n_0$
  - $t_\varepsilon^{(t_0,z_0,n_0)}$ verifies:

$$\mathbb{P}\left(\tau < t_\varepsilon^{(t_0,z_0,n_0)} \middle| Z_{t_0} = z_0, N_{t_0} = n_0\right) = \varepsilon \Leftrightarrow \mathbb{P}\left(Z_{t_\varepsilon^{(t_0,z_0,n_0)}} \leq d \middle| Z_{t_0} = z_0, N_{t_0} = n_0\right) = 1 - \varepsilon$$

  - After some calculations…
    - If $z_0 = 0$, then $n_0 = 0$ and $t_\varepsilon^{(t_0,0,0)}$ is solution of the equation:

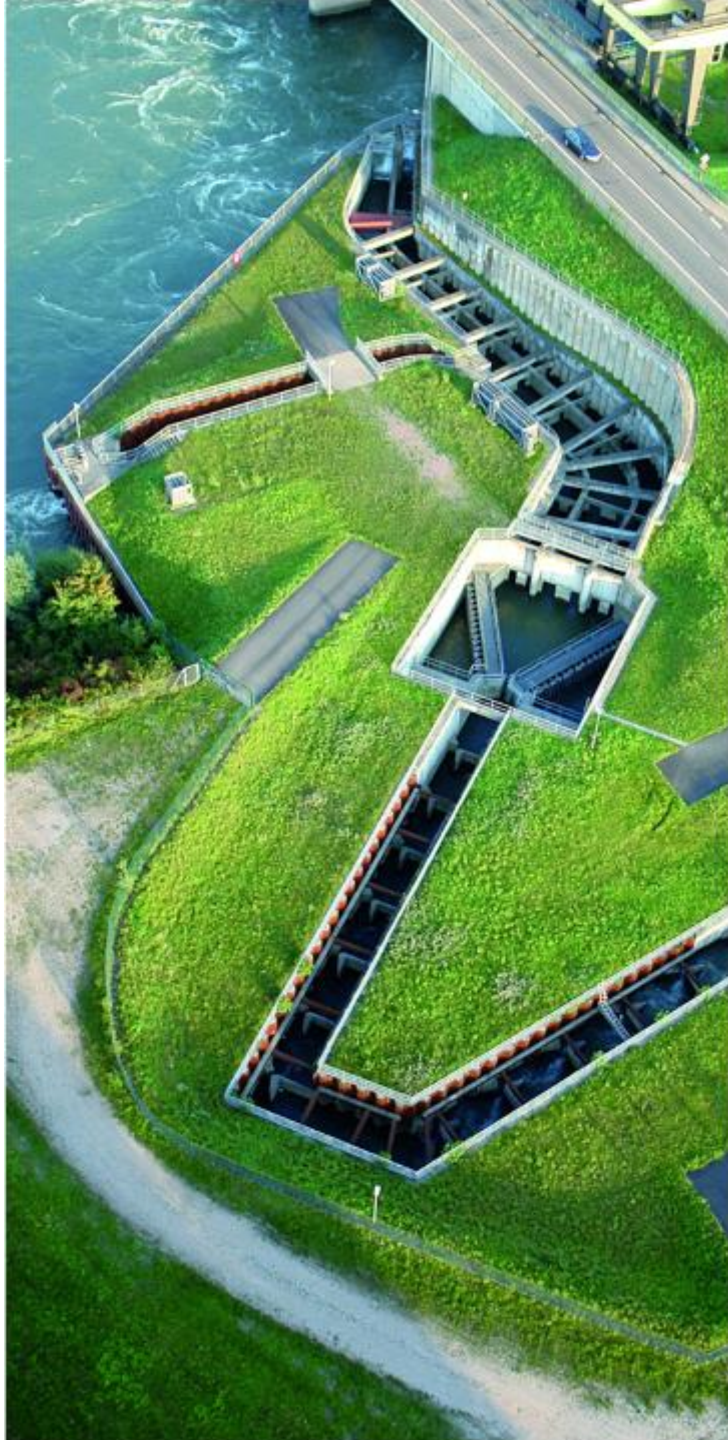
$$\int_0^{t_\varepsilon^{(t_0,0,0)}} \overline{\mathbb{F}}_{ay,b}(d)\left(t_\varepsilon^{(t_0,0,0)} - y\right)^{\beta-1} dy = -\frac{\log(1-\varepsilon)}{\alpha\beta}$$

    - If $z_0 > 0$, then $t_\varepsilon^{(t_0,z_0,n_0)}$ is solution of the equation:

$$\log\left(\mathbb{F}_{a(t-t_0),b}(d - z_0)\right) - \alpha\beta\int_{t_0}^{t} u^{\beta-1}\overline{\mathbb{F}}_{a(t-u),b}(d)du + (n_0-1)\log\left(\int_0^{t_0} u^{\beta-1}\left(\int_0^{z_0}\mathbb{F}_{a(t-t_0),b}(d-x)f_{a(t_0-u),b}(x)dx\right)du\right)$$

$$= \log(1-\varepsilon) + (n_0-1)\log\left(\int_0^{t_0}\mathbb{F}_{a(t_0-y),b}(z_0)u^{\beta-1}dy\right)$$

Application to EDF data



APPLICATION TO EDF DATA

- **Comparison between the results obtained with:**
  - $\mathcal{M}_1$: previously presented model
  - $\mathcal{M}_2$: simplified model in which the information on the total number of initiated flaws is omitted, thus focusing only on the modelling of the size of the largest flaw
    - $Z_t = Y_{(t\text{-}U)^+}$ with U the initiation time $\sim$ Weibull distributed and $(Y_t)_{t \geq 0}$ a non homogeneous Gamma process characterizing the propagation of the size of the largest flaw

- $\mathbf{N} = 228$ **components with a total number of** 623 **inspections**

- **Initiation time of the first flaw:**

Model	Mean [bootstrap 90%-CI]	Standard deviation [bootstrap 90%-CI]
$\mathcal{M}_1$	19.613 [18.396 ; 21.093]	9.11 [8.22 ; 10.315]
$\mathcal{M}_2$	18.994 [18.323 ; 20.392]	17.978 [16.124 ; 22.344]

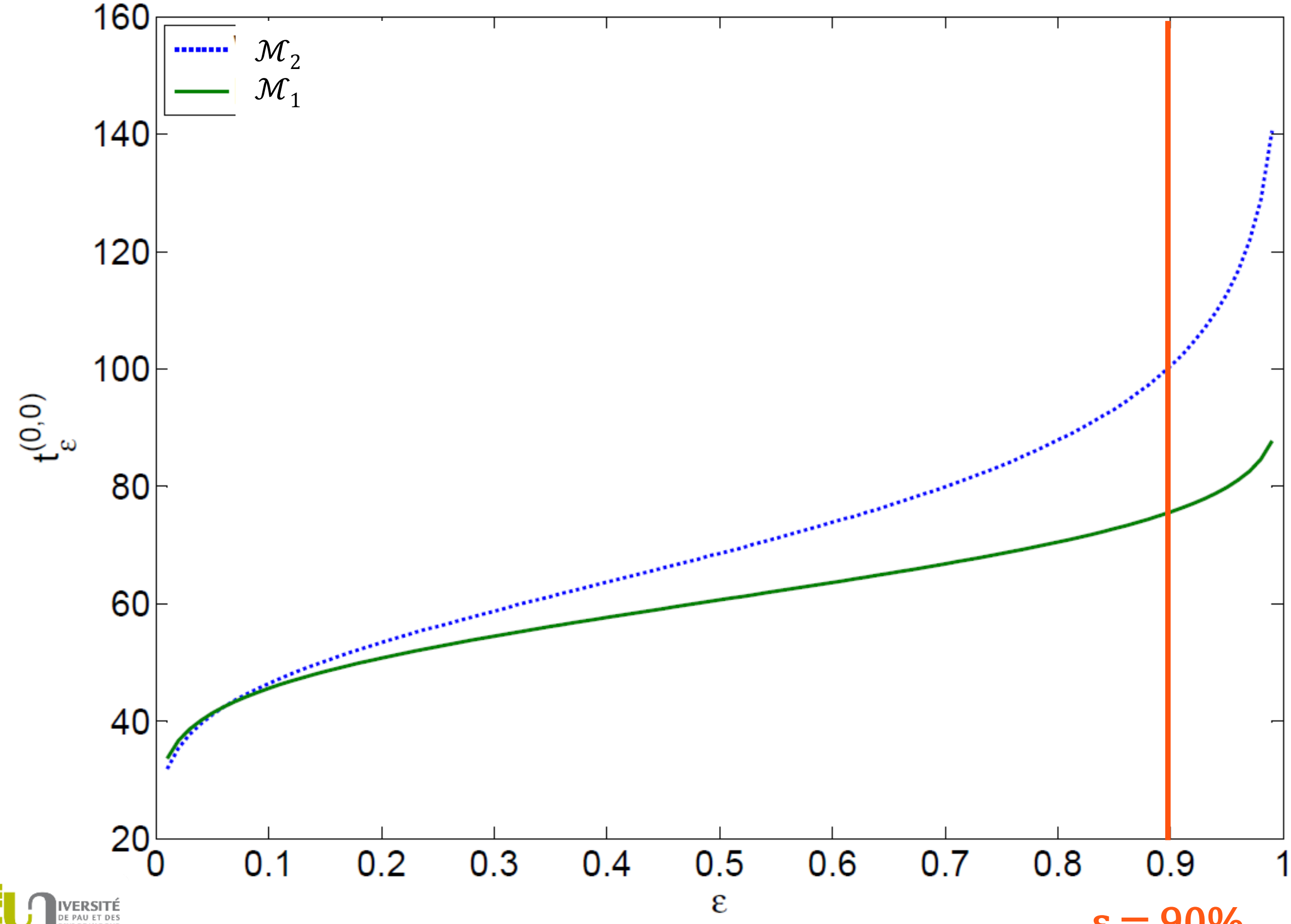
APPLICATION TO EDF DATA

- **Mean propagation over time:**
  - $\mathcal{M}_1$: 1.999 per unit of time for one initiated flaw (bootstrap 90%-CI = [1.762 ; 2.263])
  - $\mathcal{M}_2$: for the size of the largest flaw

Time	Mean propagation [bootstrap 90%-CI]
1	1.235 [0.385 ; 1.421]
2	1.397 [0.694 ; 1.502]
3	1.465 [0.884 ; 1.561]
4	1.511 [1.038 ; 1.6]
5	1.546 [1.16 ; 1.646]
6	1.575 [1.255 ; 1.696]
7	1.599 [1.329 ; 1.747]
8	1.62 [1.377 ; 1.808]
9	1.639 [1.413 ; 1.869]
10	1.656 [1.45 ; 1.934]

APPLICATION TO EDF DATA

- **Quantiles of the hitting time of a given degradation threshold:**
  - From $z_0 = 0$ at $t_0 = 0$ (new component):



$\varepsilon = 90\%$

AMMSI

APPLICATION TO EDF DATA

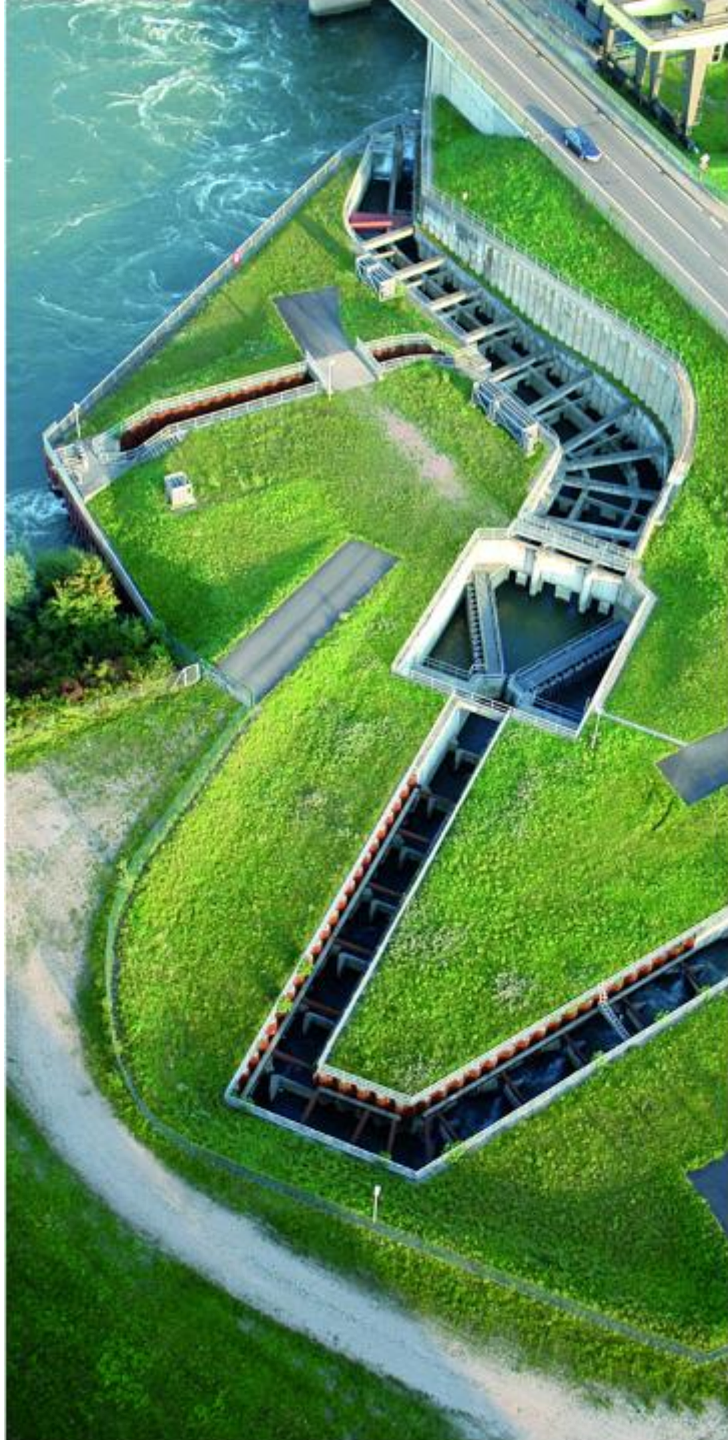
- **Quantiles of the hitting time of a given degradation threshold:**
  - From (t_0, z_0, n_0) with $\varepsilon = 90\%$:

t_0	z_0	n_0	$t_{0.9}^{(t_0,z_0,n_0)}$ with $\mathcal{M}_1$	$t_{0.9}^{(t_0,z_0,n_0)}$ with $\mathcal{M}_2$
25	20	2	41.48	54.959
25	20	4	36.66	54.959
25	20	6	33.58	54.959
25	20	8	31.102	54.959
25	20	10	29.683	54.959
25	30	2	38.362	48.542
25	30	4	34.023	48.542
25	30	6	31.311	48.542
25	30	8	29.258	48.542
25	30	10	27.655	48.542
25	40	2	34.494	42.101
25	40	4	30.974	42.101
25	40	6	28.716	42.101
25	40	8	26.859	42.101
25	40	10	25.395	42.101
25	50	2	29.981	35.532
25	50	4	27.47	35.532
25	50	6	25.56	35.532
25	50	8	24.076	35.532
25	50	10	22.781	35.532

APPLICATION TO EDF DATA

- **Comments:**
  - Remain cautious about the interpretation of the two models:
    - $\mathcal{M}_1$: characterizes the initiation and propagation of one flaw
    - $\mathcal{M}_2$: characterizes the initiation and propagation of the largest flaw
  - Similar mean initiation time between $\mathcal{M}_1$ and $\mathcal{M}_2$, but higher scattering for $\mathcal{M}_2$
  - Mean propagation of the size of one flaw for $\mathcal{M}_1$ higher than the mean propagation of the size of the largest flaw for $\mathcal{M}_2$
  - Results obtained by the simplified model $\mathcal{M}_2$ more "optimistic" than the ones obtained by $\mathcal{M}_1$
  - Omitting the (available) information brought by the total number of initiated flaws with $\mathcal{M}_2$ may lead to non-conservative forecasts!
  - Some numerical instabilities observed (more on $\mathcal{M}_2$ than on $\mathcal{M}_1$)
  - And if a non homogeneous Gamma process had been taken for propagation in $\mathcal{M}_1$?
    - Test carried out…
    - … but no statistical evidence of the relevance of considering this more general process

Conclusions and prospects



CONCLUSIONS

- **Development of a specific stochastic model driven by the partial and censored information coming from the field**
- **Useful indicators for reliability engineers**
- **Importance of:**
  - Taking into account all the available information
  - Reporting it to the operators who collect the data from the field
- **A perfect example of a fruitful partnership between academic and industrial worlds**

PROSPECTS

- **Testing other stochastic process families?**
- **Model selection criteria?**
- **Taking into account:**
  - Measurement error on the flaw sizes?
  - Probability of (non) detection of (small) initiated flaws?

THANK YOU FOR YOUR ATTENTION!