"Modélisation probabiliste et processus stochastique pour le pronostic. Application à un cas d'étude."

Khanh Le Son, Mitra Fouladirad, Anne Barros

Université de Technologie de Troyes
Institut Charles Delaunay - UMR STMR CNRS 6279

Introduction

Objectives

- Using a stochastic approach for prognostic in order to compare with the exciting non-stochastic methods applied on the 2008 Prognostic Health Management data.
- Construction of a degradation indicator from the sensors measurements (2008 Prognostic Health Management (PHM) Challenge data).
- Using a stochastic process to model the deterioration of components (Remaining Useful Life estimation).

Our point of view for prognostic

- Prognostic with probabilistic modeling
  - Data: random variables associated to failure times
  - Many realization of the same law
  - Statistical inference
  - Quantities of interest: probability of failure

  $\Rightarrow$ Independent on time and on actual system state

  $\Rightarrow$ Not so good for prognostic
- Prognostic with stochastic modeling (stochastic process)
  - One scalar indicator (or a vector of reasonable size) that depends on the time and that makes sense with degradation phenomena
  - Many realizations of the same indicator
  - Statistical inference
  - Calculation of probabilistic quantities of interest: time to reach a failure state, law of this time.

  $\Rightarrow$ Dependent on time and on actual system state

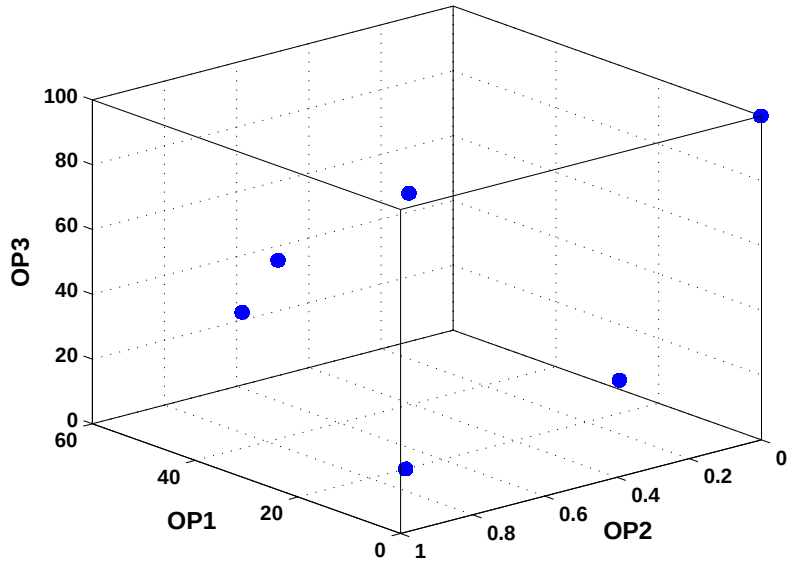
  $\Rightarrow$ Seems to be relevant for prognostic

Experimental data

		3 Operational variables			Measurements of 21 sensors			
Unit	**Cycle**	**OP1**	**OP2**	**OP3**	**SM1**	**SM2**	**...**	**SM21**
1	1							
	...							
	T_1							
...								
218	1							
	...							
	T_{218}							

- Two sub-data set : the training data set and the testing data set.
- The training data set is used to build the prediction model
- The testing data set is used to estimate the RUL for each testing unit.

Experimental data



- Degradation indicator can not be directly deduced from the 21 sensor paths
- All measurements are divided into 6 clusters corresponding to 6 operational modes
- Selection of 7 sensors

Degradation indicator (1)

$\Rightarrow$ STEP 1: identification of a failure space and a failure place for each mode

- Select the measurements of 7 sensors only at the failure time
- Group the failure measurements according to their mode (6 groups)
- For each group:
  - Create a projection space of dimension 2 with PCA (called failure space)
  - Calculate the barycenter of the projected failure measurements in this space to create a failure place.

- Result :
  - 6 plans of PCA $P_1, P_2, ..., P_6$, one for each mode.
  - 6 failure places $L1, L2, ..., L6$, one for each mode.

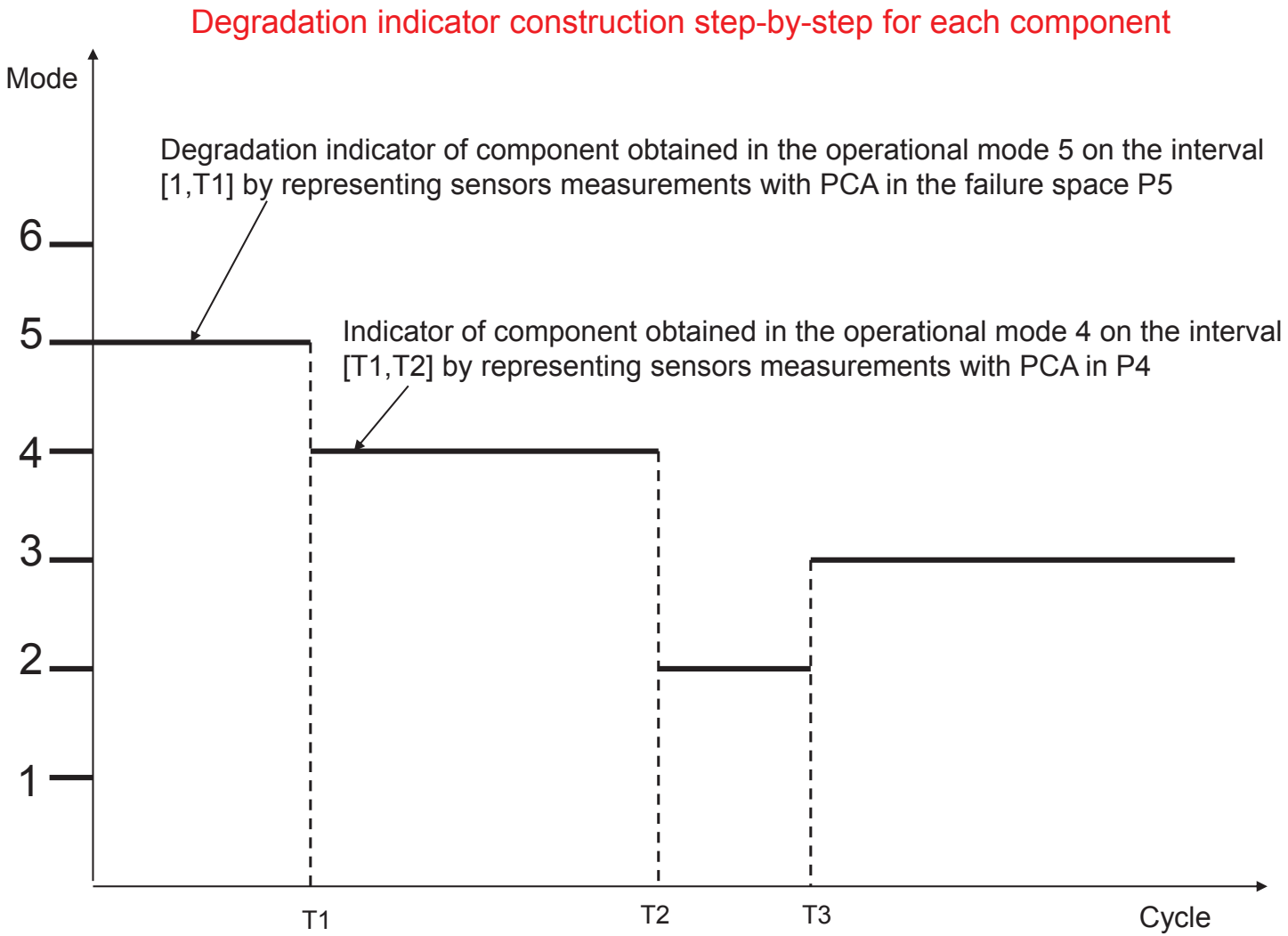
Degradation indicator (2)

$\Rightarrow$ Contribution of principal components for each mode

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
PC1	60.85	72.64	61.45	54.41	58.95	79.65
PC2	38.04	26.75	37.85	44.55	40.07	19.11
PC3	0.66	0.28	0.34	0.56	0.41	0.77

Degradation indicator (3)

⇒ STEP 2: create a degradation indicator at each time for each component



Degradation indicator (4)

- N_k number of units in mode k
- $\bar{P}_k = (\bar{a}, \bar{b})$ the barycenter of the failure space P_k, $k = 1, ..., 6$
- $P_i^k = (a_i, b_i)$, $i = 1, ..., N_k$ is the i^{th} failure place in the projection space P_k
- $P_{ij}^k = (a_{ij}, b_{ij})$ is the measure of the 7 selected sensors at time j for component i in the projection space P_k

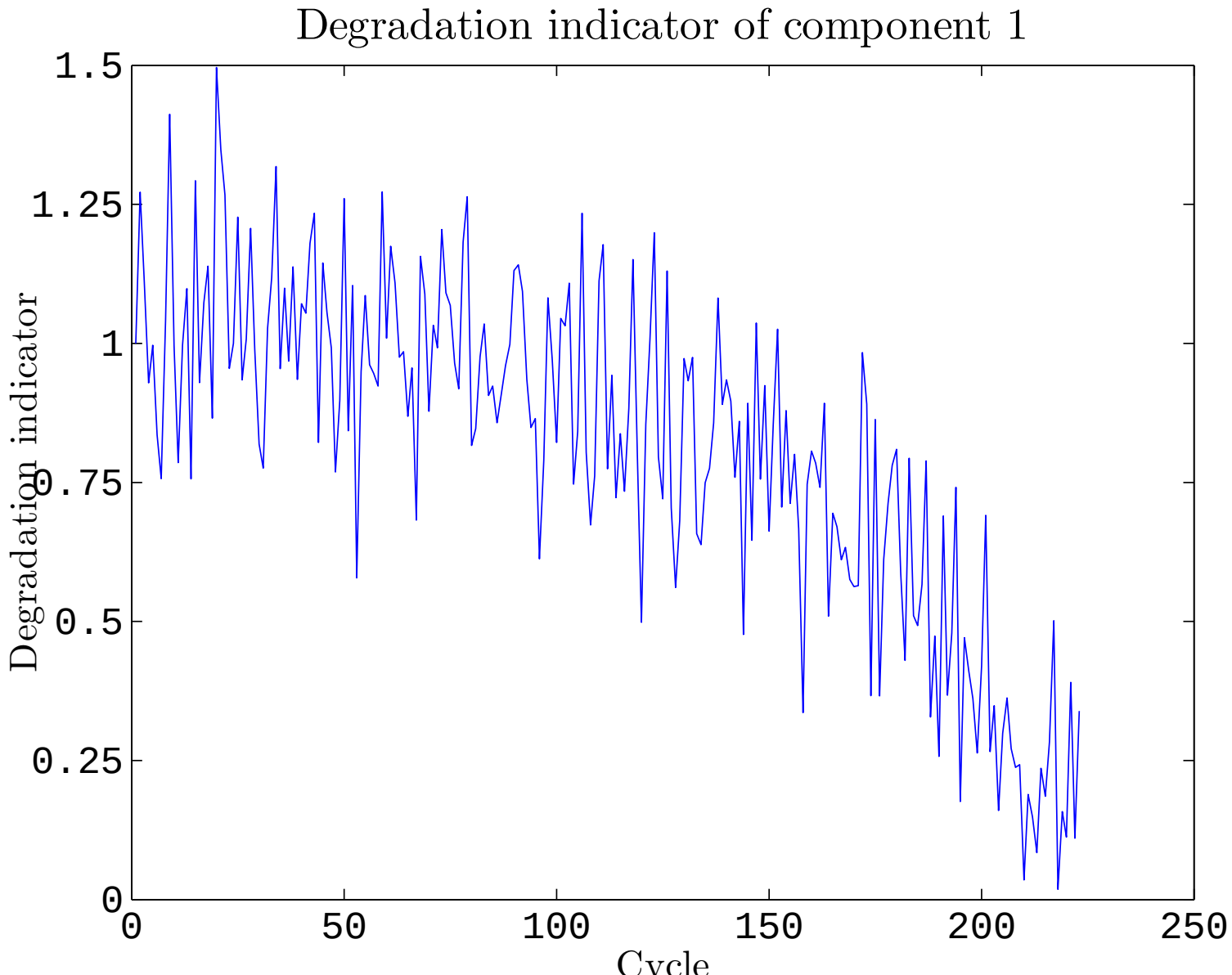
The dispersion of the failure places in mode k at time j (noted $k(j)$) is defined by:

$$Disper_{k(j)} = \sqrt{\frac{1}{N_{k(j)} - 1} \sum_{i=1}^{N_{k(j)}} ((a_i - \bar{a})^2 + (b_i - \bar{b})^2)}$$

$$D_{ij}^{k(j)} = \frac{\sqrt{(a_{ij}^{k(j)} - \bar{a})^2 + (b_{ij}^{k(j)} - \bar{b})^2}}{Disper_{k(j)}}$$

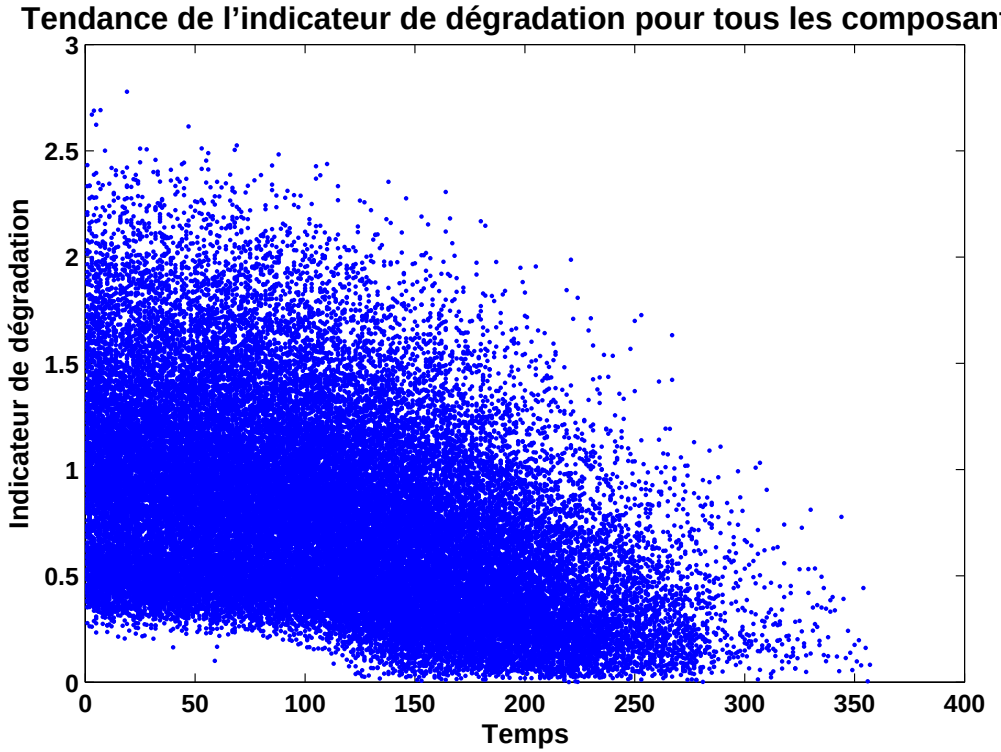
Degradation indicator (5)

$\Rightarrow$ One component



Degradation indicator (6)

$\Rightarrow$ All the components



- Non-linear and decreasing
- At the beginning cycles, the indicator value has an important dispersion and at the end cycles, the failure times, it tends to zero

First degradation model - Definition

$\Rightarrow D_{i,j}^{k(j)}$ = degradation indicator of unit i at cycle j.

- One process X_j for all $D_{i,j}^{k(j)}$ i.e. one unique process:
  - for all the modes,
  - for all the units.
- The increments $\triangle d_{i,j} = D_{i,j+1} - D_{i,j}$, $i = 1, ..., 218$ follow a gaussian distribution $\mathcal{N}(\eta((j+1)^\alpha - (j)^\alpha), \sigma^2((j+1)^\alpha - (j)^\alpha))$
- This law depends only on time j.

$\Rightarrow$ A path $D_{i,j}^{k(j)}$; $j = 2, ..., n_i$ is one realization of this Wiener process and we have 218 realizations of the same stochastic process. $\Rightarrow$ Law for the

first value $D_{i,1}$: $D_{i,1}$; $i = 1, ..., 218$ are the realizations of the same Weibull law with 3 parameters.

First degradation model - Parameter estimation

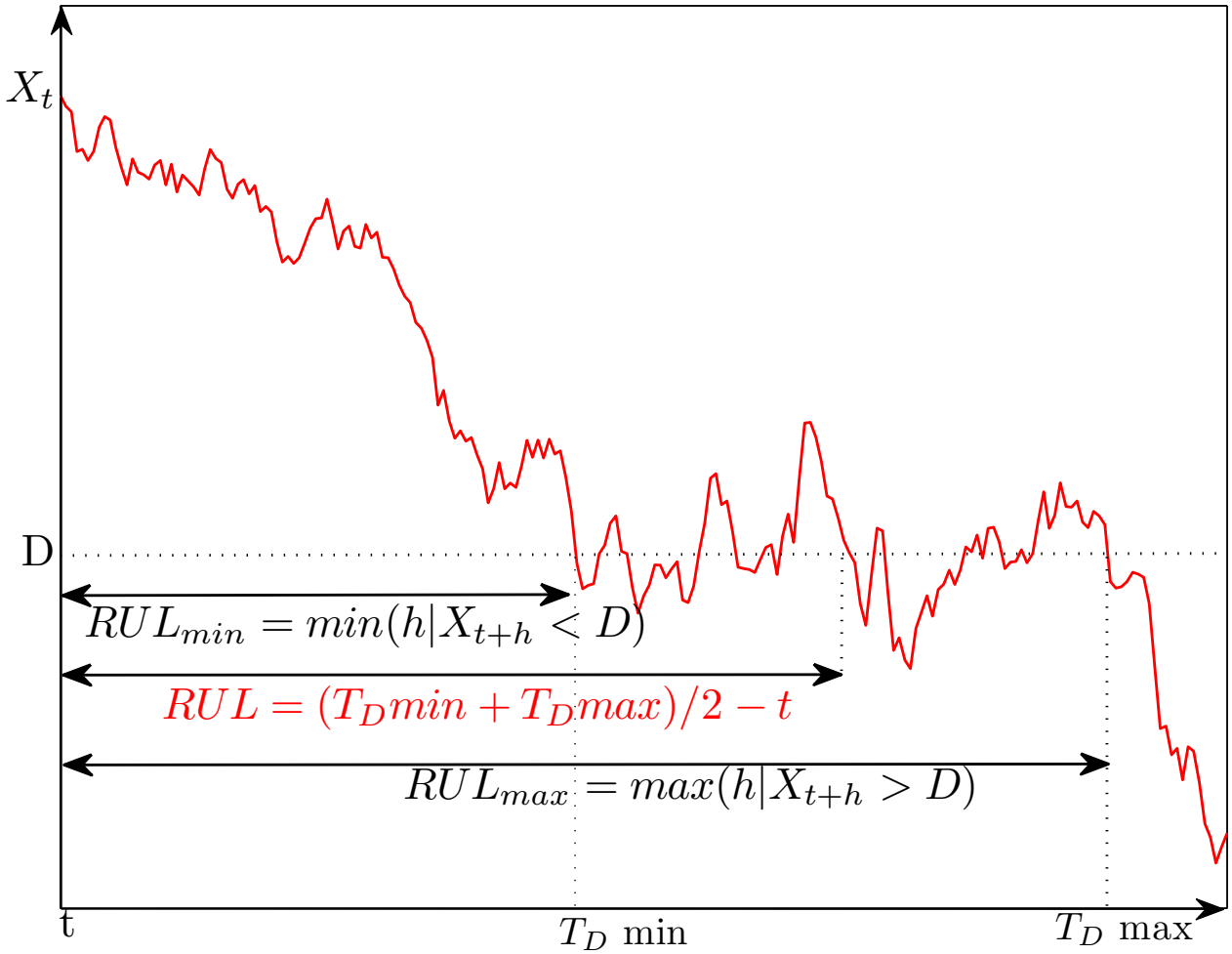
- Wiener parameters (η, α, σ) are estimated by maximizing the log-likelihood function:

$$l = \sum_{i=1}^{218} \sum_{j=1}^{T_i^{failure}-1} log(f_{\eta,\sigma,\alpha}(\triangle d_{i,j}))$$

- Weibull parameters are estimated independently also by log-likelihood function.

RUL estimation - Testing data set

- $D_{i',r}^{k(r)}$: degradation indicator of the unit i' at the last observed cycle r.
- $(\hat{\eta}, \hat{\sigma}, \hat{\alpha})$ are used to simulate the paths of Wiener process from the last cycle r of unit i'.



RUL estimation - Testing data set

- RUL for each simulation path is defined as:

$$RUL^{i'}_{simulated} = \frac{(T^{i'}_L min + T^{i'}_L max)}{2} - r$$

  - $T^{i'}_L min = inf\{j > r, D^{k(j)}_{i',j} \leq L\}$
  - $T^{i'}_L max = inf\{j > r, \forall s > 0, D^{k(j+s)}_{i',j+s} \leq L\}$

- $RUL^{i'}_{estimated}$ of testing unit i' is defined as the empirical mean of $n = 10000$ simulated Wiener process paths.

Performance assessment of the first model

- Applying the stochastic degradation model to all 218 units of the testing data set, we obtained an estimated RULs set $RUL^{i'}_{estimated}$,for $i' = 1, ..., 218$.
- Penalty function criterion : provided by 2008 PHM Challenge, the penalty score for each testing unit is given by the following formula:

$$S_{i'} = \begin{cases} e^{-d_{i'}/13} - 1, & d_{i'} \leq 0 \\ e^{d_{i'}/10} - 1, & d_{i'} > 0 \end{cases} \qquad i' = 1, ..., 218$$

  where $d_{i'} = RUL^{i'}_{estimated} - RUL^{i'}_{actual}$ and the total score $S = \sum_{i=1}^{218} S_{i'}$
- Root mean squared error : $RMSE = \sqrt{\sum_{i'=1}^{218} (d_{i'})^2}$

Performance assessment of the first model

- Lifetime distribution model (Weibull distribution on the failure times) : $S = 9870$
- Total score of the different models on our degradation indicator $D_{i,j}^{k(j)}$:
  - Similarity-based prognostic approach proposed by Wang in the 2008 PHM conference : $S = 6690$
  - Wiener process model : $S = 5520$ and $RMSE = 438$
- The best results in the 2008 PHM Challenge:
  - Similarity-based prognostic model of Wang : $S = 5636$
  - Non-probabilistic models based on the neural networks of Peel (2008) and Heimes (2008):
    RMSE=519.8 and RMSE=984.

Conclusion for the first degradation model

- Good performance with reasonable calculation time on basic computer

- Do not get complete satisfaction because:
  - no conditional law for the RUL,
  - no monotone degradation.

$\Rightarrow$ Further work: there is a noise with the degradation indicator that must be “filtered” for better modeling and probabilistic calculation

Second degradation model (1)

- Note:
  - $\mathbf{Y^{(i)}} = (D_{i,1}^{k(1)}, ..., D_{i,n_i}^{k(n_i)})$: the observation vector for unit i.
  - $\mathbf{X^{(i)}} = (X_{i,1}^{k(1)}, ..., X_{i,n_i}^{k(n_i)})$: the non-observable actual random states of unit i.
- Our deterioration model:

$$D_{i,j}^{k(j)} = f(X_{i,j}^{(k(j)}, \epsilon_{i,j}^{k(j)})$$
$$D_{i,j}^{k(j)} = X_{i,j}^{(k(j)} + \epsilon_{i,j}^{k(j)}$$
$$Y^{(i)} = X^{(i)} + \epsilon^{(i)}$$

where :

- $\epsilon_{i,j}^{k(j)}, j = 1, ..., n_i$: the independent gaussian random variables with standard deviation $\sigma_j^{(i)}$ and mean equals to zero for unit i.

Second degradation model (2)

- Each component of the observations vector $\mathbf{Y}^{(i)}$, $i = 1, ..., 218$ is replaced by:

$$Y_{i,j}^{(k(j))} = D_{i,1}^{k(1)} - D_{i,j}^{k(j)},\ j = 1, ..., n_i$$

- The paths $Y^{(i)}$ are non-monotone increasing, with starting point at zero and with random failure threshold.
- Non-homogeneous Gamma process for $X_{i,j}^{k(j)}$
- One unique process for all the units and all the modes: X_j
- One unique law for the noise $\epsilon_{i,j}^{k(j)}$: $\mathcal{N}(0, \sigma)$

$\Rightarrow$ Y^i, X^i ϵ^i, for $i = 1, ..., 218$ are the realizations of the same stochastic process.

$\Rightarrow$ There will be noted Y, X, ϵ in the following.

Second degradation model (3)

$\Rightarrow$ Definition of non-homogeneous Gamma process

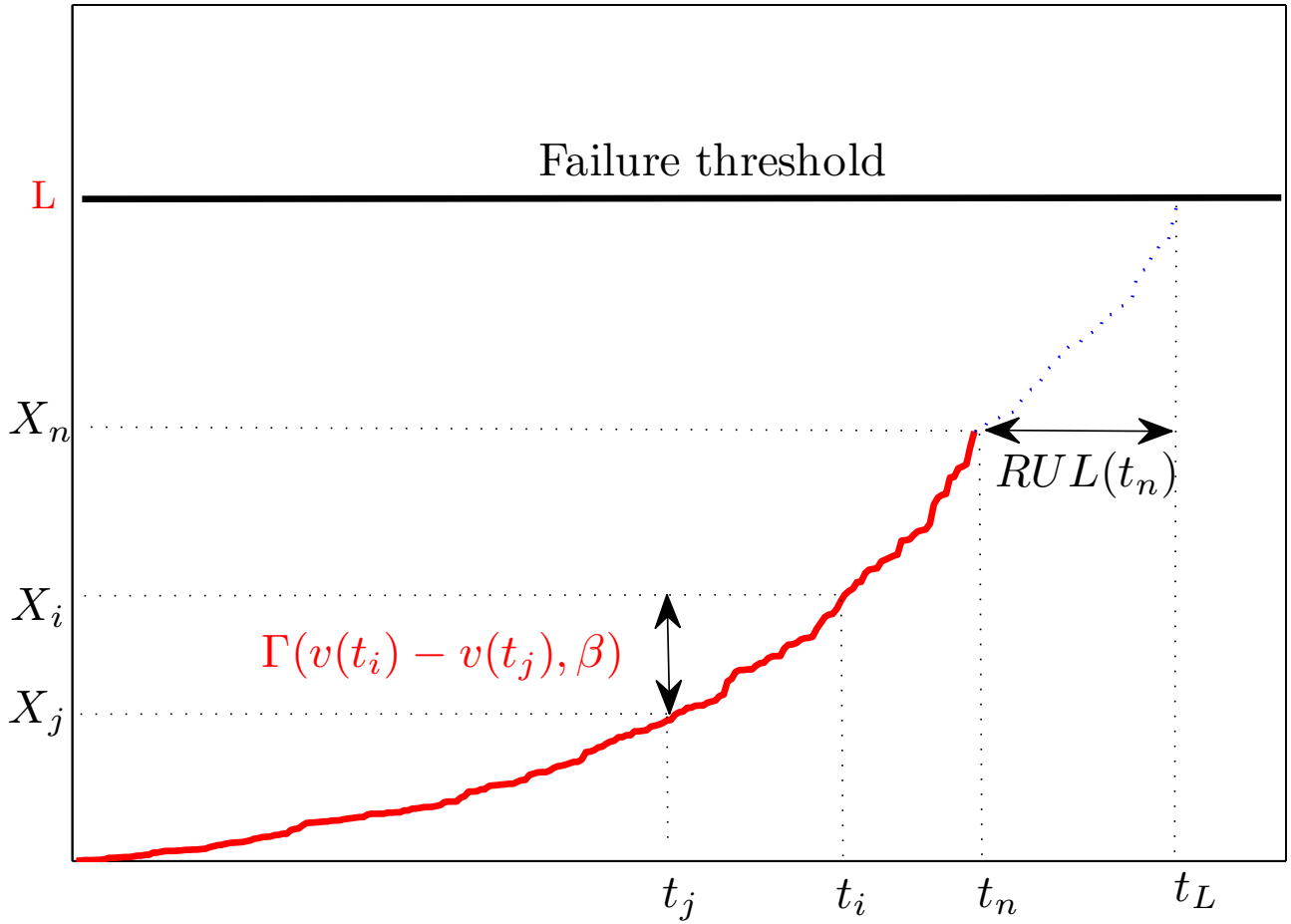
- The initial state $X_0 = 0$.
- $(X_j)_{j \geq 0}$ is supposed to be monotone, increasing.
- The increments $\delta_{X_j} = X_j - X_{j-1}$, $j = 1, 2, ..., n$ are independent and have the Gamma density:

$$f_{\delta_{X_j}}(\delta | v(t_j) - v(t_{j-1}), \beta) = \frac{\beta^{v(t_j)-v(t_{j-1})}}{\Gamma(v(t_j) - v(t_{j-1}))} \delta^{v(t_j)-v(t_{j-1})-1} e^{-\beta\delta} I_{(0,\infty)}(\delta)$$

- $\Gamma(u) = \int_{z=0}^{\infty} z^{u-1} e^{-z} dz$: Gamma function for $u > 0$.
- $I_A(\delta) = 1$ for $\delta \in A$, $I_A(\delta) = 0$ for $\delta \notin A$.
- Shape function $v(t) = \alpha t^b$ and scale parameter β.

where: $\Gamma(v) = \int_{z=0}^{\infty} z^{v-1} e^{-z} dz$ is the gamma function, .

$\Rightarrow$ 4 parameters to be optimised: α, β, b, σ.


Failure threshold
L
X_n
X_i
X_j
$\Gamma(v(t_i) - v(t_j), \beta)$
$RUL(t_n)$
t_j
t_i
t_n
t_L

Joint distribution of system state

- For estimating the RUL, the joint conditional density of $\mathbf{X}$ figured out the observation vector $\mathbf{Y}$ is calculated as follows:

$$\mu_{\mathbf{X}/\mathbf{Y}}(x_1, ..., x_n) = K_1 e^{-\beta x_n} \prod_{j=1}^{n} (x_j - x_{j-1})^{\alpha(t_j^b - t_{j-1}^b)-1} e^{(-\frac{g^2(x_j, Y_j)}{2\sigma^2})} |g\prime(x_j, Y_j)|$$

where $g\prime(., y) = \frac{\partial g(.,y)}{\partial y}$ and K_1 is the coefficient defined as follows:

$$\frac{1}{K_1} = \int ... \int e^{-\beta x_n} \prod_{j=1}^{n} (x_j - x_{j-1})^{\alpha(t_j^b - t_{j-1}^b)-1} e^{(-\frac{g^2(x_j, Y_j)}{2\sigma^2})} |g\prime(x_j, Y_j)| dx_1 ... dx_n$$

- It's difficult to calculate the coefficient K_1 $\Rightarrow$.

Remaining Useful Lifetime estimation

- Remaining Useful Lifetime (RUL) estimation is based on the failure probability at the next inspection given the n observations $Y_1, ..., Y_n$.

- The distribution function of $RUL(t_n)$ figured out the observations is defined as follows:

$$\begin{aligned} F_{RUL(t_n)}(h) &= P(X_{t_n+h} > L | X_n > L, Y_1, ..., Y_n) \\ &= \int\int \bar{F}_{\alpha((t_n+h)^b - t_n^b),\beta}(l - x_n).f_L(l).\mu_{X_n/Y_1,...,Y_n} dl dx_n \end{aligned}$$

  - $\bar{F}_{\alpha((t_n+h)^b - t_n^b),\beta}$: the reliability function of Gamma process with shape function $\alpha((t_n + h)^b - t_n^b)$ and scale parameter β.
  - $\mu_{X_n/Y_1,...,Y_n}$: the conditional density of X_n.
  - $f_L(l)$: the density function of the failure threshold.

Methodology for RUL estimation

- With the training set: $\hat{\alpha}, \hat{\beta}, \hat{b}, \hat{\epsilon}$
- Propose an estimator of $F_{RUL(t_n)}(h)$ using $\hat{\alpha}, \hat{\beta}, \hat{b}, \hat{\epsilon}$ and the testing data set to approximate $\mu_{X_n/Y_1,\ldots,Y_n}$

$\Rightarrow$ Gibbs algorithm (MCMC)

- With the training set, for each unit:
  - Simulate a Markov Chain whose stationnary law is $\mu_{X/Y}$.
  - Each state Z_j of the Markov Chain is an approximation of X_j .
  - Use the outputs of the algorithm to create a data set that follow $\mu_{X/Y}$.
  - Estimate the parameters $\alpha, \beta, b, \epsilon$ with SEM algorithm.
- With the testing set, for each unit:
  - Simulate a Markov Chain whose stationary law is $\mu_{X_n/Y_1,\ldots,Y_n}$.
  - One state of the Markov Chain: Z_n.
  - Use the outputs of the algorithm to approximate X_n by Z_n
  - Use the estimated parameters and Z_n to approximate $F_{RUL(t_n)}(h)$.

Gibbs algorithm (1)

- For $j = 1$,

$$\mu_{\mathbf{X}/\mathbf{Y}}(x_1/x_2, ..., x_n) = K_{2,1} x_1^{\alpha(t_1^b)-1} (x_j - x_{j-1})^{\alpha(t_j^b - t_{j-1}^b)-1}$$
$$e^{-\frac{g^2(X_1,Y_1)}{2\sigma^2}} |g\prime(x_1, Y_1)| 1_{(0<x_1<x_2)}$$

- For $2 \leq j \leq n-1$,

$$\mu_{\mathbf{X}/\mathbf{Y}}(x_j/x_1, ..., x_{j-1}, x_{j+1}, ..., x_n) = K_{2,j} (x_j - x_{j-1})^{\alpha(t_j^b - t_{j-1}^b)-1}$$
$$(x_{j+1} - x_j)^{\alpha(t_{j+1}^b - t_j^b)-1} e^{(-\frac{g^2(x_j,Y_j)}{2\sigma^2})} |g\prime(x_j, Y_j)| 1_{(x_{j-1}<x_j<x_{j+1})}$$

- For $j = n$,

$$\mu_{\mathbf{X}/\mathbf{Y}}(x_n/x_1, ..., x_{n-1}) = K_{2,n} e^{-\beta x_n} (x_n - x_{n-1})^{\alpha(t_n^b - t_{n-1}^b)-1}$$
$$e^{-\frac{g^2(x_n,Y_n)}{2\sigma^2}} |g\prime(x_n, Y_n)| 1_{(x_{n-1}<x_n)}$$

where $K_{2,j}$ are tractable constants dependent on $x_1, ..., x_{j-1}, x_{j+1}, ..., x_n$ and $y_1, ..., y_n$

Gibbs algorithm (2)

$\Rightarrow$ A random variable from $\mu_{\mathbf{X}/\mathbf{Y}}(x_1, ..., x_n)$ is generated as follows:

- Given a values set: $(z_1(k), ..., z_n(k))$
- The next generated values set $(z_1(k+1), z_2(k+1), ..., z_n(k+1))$ is defined as follows:
  - $z_1(k+1)$ is generated from $\mu_{Z_1}(k+1)$=$\mu_{\mathbf{X}/\mathbf{Y}}(z_1|z_2(k), ..., z_n(k))$.
  - $z_j(k+1)$ is generated from $\mu_{Z_j}(k+1)$=$\mu_{\mathbf{X}/\mathbf{Y}}(z_j|z_1(k+1), ..., z_{j-1}(k+1), z_{j+1}(k), z_n(k))$, $j = 2, ..., n-1$.
  - $z_n(k)$ is generated from $\mu_{Z_n}(k+1)$=$\mu_{\mathbf{X}/\mathbf{Y}}(z_n|z_1(k+1), ..., z_{n-1}(k+1))$.

  $\Rightarrow$ Algorithm stops at $k = Q_0$ when the stationary state of the Markov chain is got.

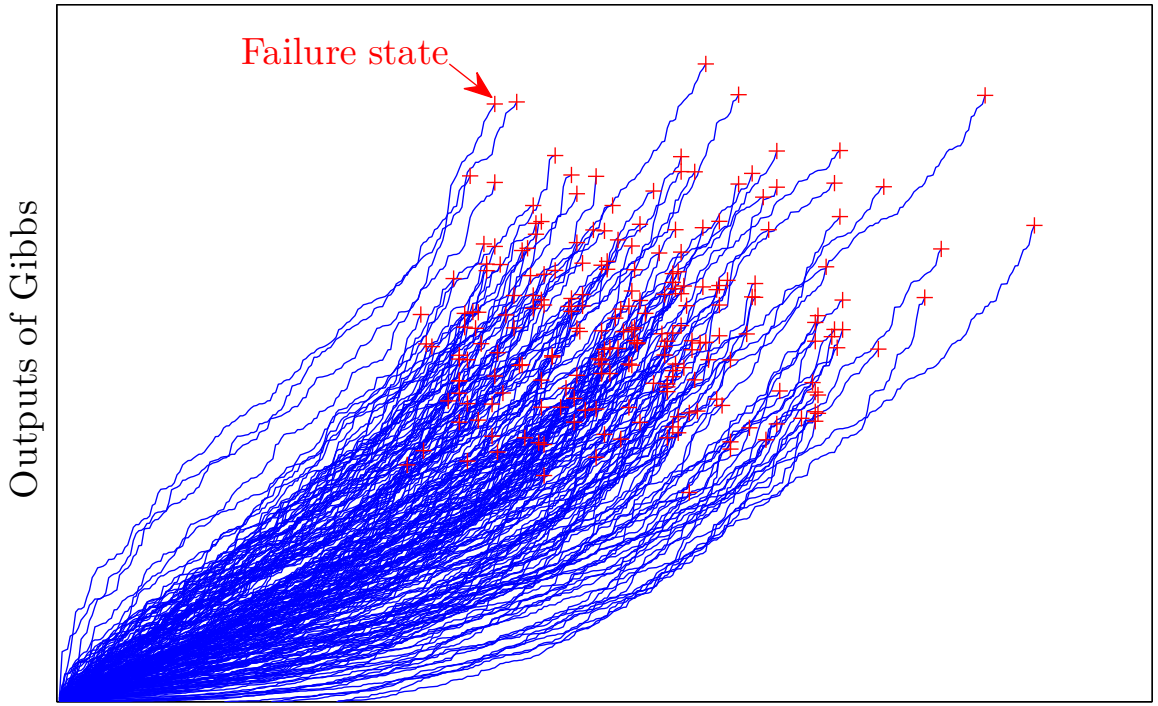
  $\Rightarrow$ Each observation vector $Y^{(i)}$ gives us an approximation vector $Z^{(i)}$ of $X^{(i)}$.

Real state estimation

$\Rightarrow$ Approximation of the non-observable state X_j:

$$\bar{z}_j = \frac{1}{Q} \sum_{q=1}^{Q} z_j(Q_0 + q) \tag{1}$$

- Q_0 : the number of sequences to get the convergence state.
- Q : the number of sequences to give sufficient precision to the empirical distribution of interest.



Model parameters estimation

- Parameters of model are estimated based on the outputs of Gibbs algorithm and by using the Stochastic EM (SEM) method.
- The observations set $\mathbf{Y}^{(i)} = (Y_{i,j}^{k(1)}, ..., Y_{i,n_i}^{k(n_i)})$, $i = 1, ..., 218$ with component independently observed at inspection times $0 < t_1 < ... < t_{n_i}$.
- Maximizing as follows by SEM method to estimate the parameters of model:

$$L(\alpha, b, \beta, \sigma) = \sum_{i=1}^{218} \sum_{j=1}^{n_i} [(\alpha((t_j)^b - (t_{j-1})^b) - 1) ln(X_{i,j}^{k(j)} - X_{i,j-1}^{k(j-1)})$$

$$-\beta(X_{i,j}^{k(j)} - X_{i,j-1}^{k(j-1)}) - \frac{g^2(X_{i,j}^{k(j)}, Y_{i,j}^{k(j)})}{2\sigma^2} + ln(|g'(X_{i,j}^{k(j)}, Y_{i,j}^{k(j}|) - \ln(\sigma\sqrt{2\pi})$$

$$+ \alpha((t_j)^b - (t_{j-1})^b) \ln(\beta) - \ln(\Gamma(\alpha((t_j)^b - (t_{j-1})^b)))]$$

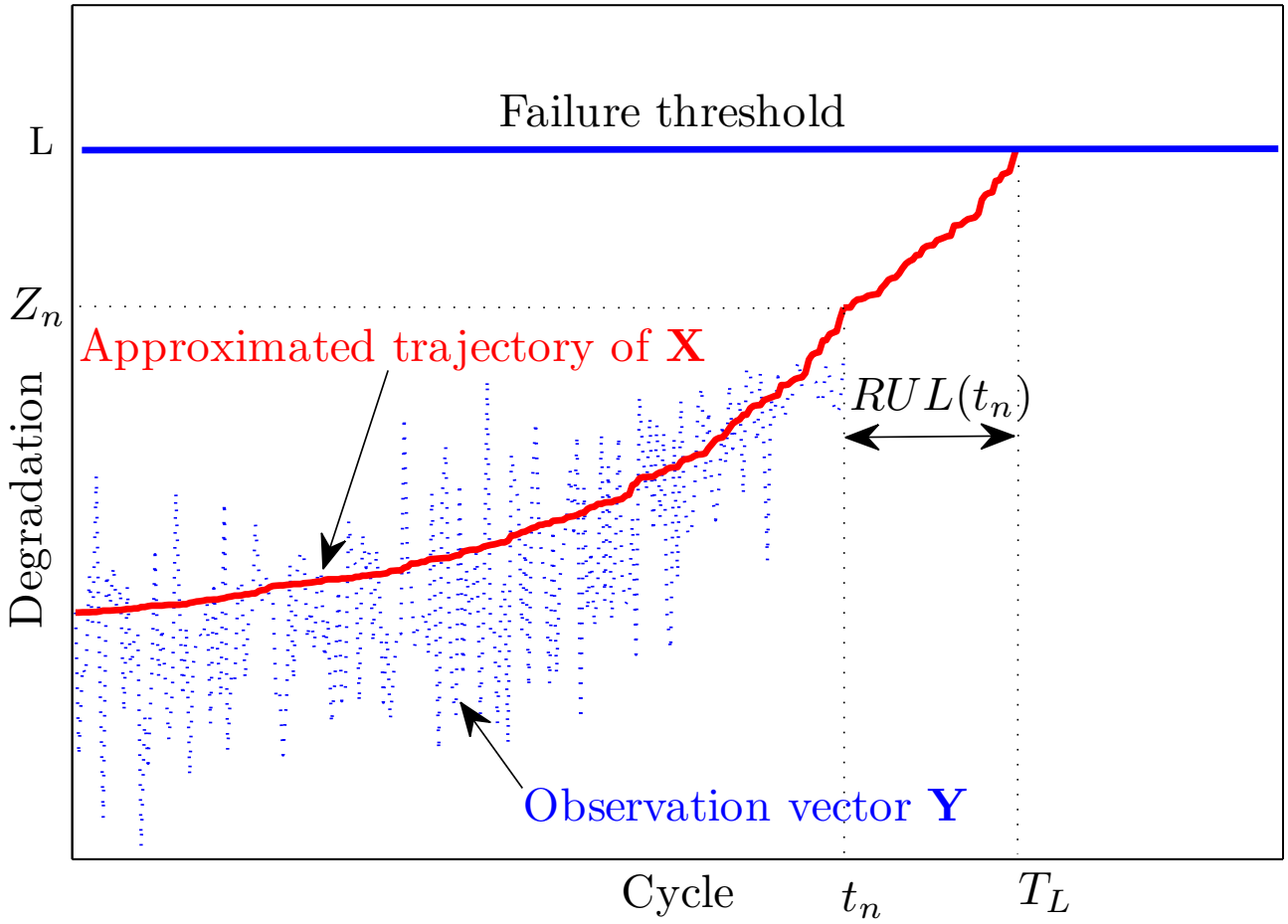
RUL estimation

- The conditional distribution $F_{RUL(t_n)}(h)$:

$$\begin{aligned} F_{RUL(t_n)}(h) &= P(X_{t_n+h} > L | X_n > L, Y_1, ..., Y_n) \\ &= \int \int \bar{F}_{\alpha((t_n+h)^b - t_n^b),\beta}(l - x_n).f_L(l).\mu_{X_n/Y_1,...,Y_n} dl dx_n \end{aligned}$$

can be estimated by Gibbs algorithm as follows:

$$\hat{F}_{RUL(t_n)}(h) = \frac{1}{Q} \sum_{q=Q_0+1}^{Q_0+Q} \int \bar{F}_{\hat{\alpha}((t_n+h)^{\hat{b}} - t_n^{\hat{b}}),\hat{\beta}}(l - z_n(q)).f_L(l) dl$$


Failure threshold
L
Z_n
Approximated trajectory of **X**
$RUL(t_n)$
Degradation
Observation vector **Y**
Cycle
t_n
T_L

Performance assessment of the second model

- Estimation quality: close to Wiener process model
- Calculation time (PC Dell Core i5-2500 3.3GHz, 4Go RAM, et Matlab 2010)
  - Wiener process
    - For parameter estimation with the training data set: 4h
    - For the RUL estimation of one unit on the testing data set: 11 min.
  - Gamma process with noise
    - For parameter estimation with the training data set: 3 days
    - For the RUL estimation of one unit on the testing data set: 30 min.

Perspectives for the second model

- Size of the observation window

- Performance of maintenance policies based on:

$$\begin{aligned} F_{RUL(t_n)}(h) &= P(X_{t_n+h} > L | X_n > L, Y_1, ..., Y_n) \\ &= \int \int \bar{F}_{\alpha((t_n+h)^b - t_n^b),\beta}(l - x_n).f_L(l).\mu_{X_n/Y_1,...,Y_n} dl dx_n \end{aligned}$$