# Thorough uncertain models rather than imperfect exact models

Mitra Fouladirad

ESRA SEMINAR

# Prediction Prediction and decision making problem

- Available data
- Quantity of interest for prediction
- Indicator construction
- Indicator evolution modelling
- Prediction taking into account model, operational and environmental conditions variability, parameters uncertainty

# Prediction and decision making problem

- Problem statement by expert
    - Available data
    - Quantity of interest for prediction
- Data science
    - Indicator construction: machine learning, clustering, CPA, etc.
    - Indicator evolution modelling: stochastic models, calibration
    - Prediction: filtering, bayesian methods, etc.

# Indicator evolution modelling

- Physic-based and data-based
- Data-based

## Modelling

What is going to be modelled?

- The quantity of interest
- An indicator related to the quantity of interest

How it is going to be modelled?

- physical model
- physical and statistical model:

## Modelling

- physical model : exact equation often deterministic
- physical and statistical model
  - random model (stochastic model)
  - deterministic physical model as average trend
  - uncertainty modelling through parameters, models or measurements

# Formalisation

- $X_t$ indicator at time $t$
- $X = (X_t)_{t \in T}$ take values in space $(E, \mathcal{E})$.
- $E = \mathcal{U} \cup \bar{\mathcal{U}}$, where $\mathcal{U}$ is the acceptable zone
- Probability of being in $\bar{\mathcal{U}}$, conditionnally to the past
- Observations $(Y_i)_{i \in \mathbb{N}}$

# One dimensional illustration

The *Remaining Useful Life* at time $t$:

$$\mathcal{RUL}_t = \inf\{s \geq t, \quad X_s \notin \mathcal{U}\} - t$$

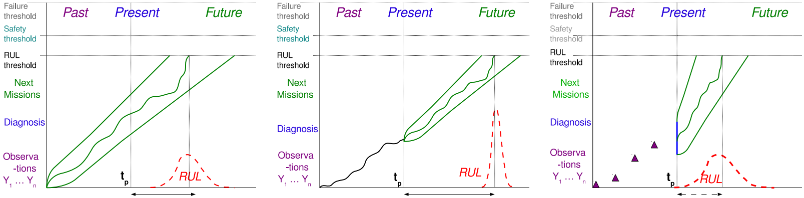
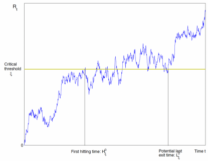
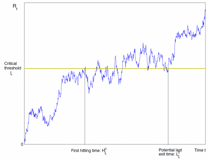
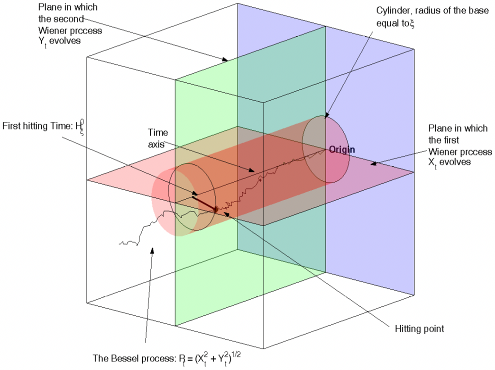



Figure: Illustration of the prognostic concept with different types of observation

# Examples $\Omega = \mathbb{R}$ or $\Omega = \mathbb{R}^2$



Aim:

$$\mathbb{P}(X_t \in \bar{\mathcal{U}} | \mathcal{F}_s, s < t)$$

At time $s$, give the distribution of $X_t$, $t > s$ .

Requirements:

- Have a model for $X = (X_t)_{t \in T}$
- Deal with parameters (known, unknown, uncertainty...)
- Deal with data

# Random models

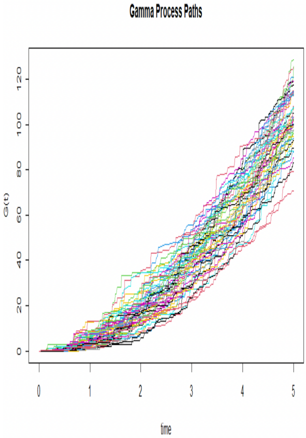
Stochastic processes good candidates to model $(X_t)_{t \in T}$.

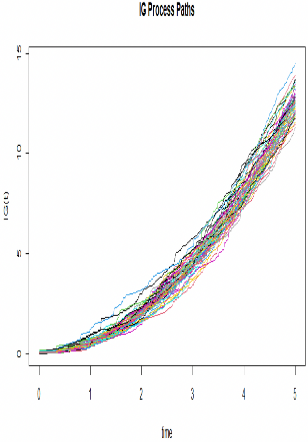
**Definition**

A random or stochastic process is a time series of random variables or vectors.

# $\Omega = \mathbb{R}^d$, $d$-dimensional Lévy process

$$X_0 = 0, \quad X_t - X_s \perp\!\!\!\perp \{X_u : u \leq s\}, \forall s < t, \;\; X_{t+s} - X_s \sim X_t - X_0$$

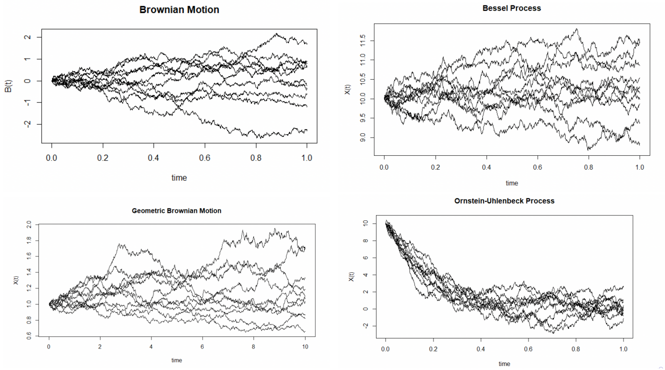


# $\Omega = \mathbb{R}$, Diffusion process

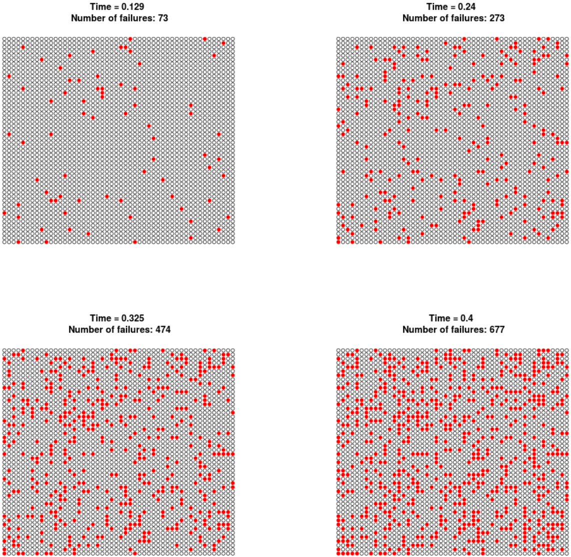
$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$$

$(W_t)_{t \geq 0}$ a standard Brownian Motion



# Spatial-temporal process

# Spatial-temporal process



# Examples

- Markov processes
- Semi Markov processes
- Jump processes
- Self-exited processes
- ...

Different choices according to :

- the dependence to the past,
- the distribution,
- event arrivals,
- $\Omega$, $T$,
- ....

SiC MOSFET threshold Voltage Instability

# Threshold voltage measure

- Silicon Carbide Metal Oxide Semiconductor Field Effect Transistor: SICMOSFET
- SiC MOSFET threshold Voltage Instability: Accelerated Degradation Test
- Instability of the voltage threshold is an aging mechanism
- Threshold voltage measure,
- Variable usage profile
- Time Dependent Dielectric Breakdown :

$$T_{BD} = k \frac{N_{BD}^{1/m}}{R_G} e^{\frac{E_A}{k_B T}}$$

$N_{BD}$ is the critical density of trap in oxyde, $R_G$ default creation rate,$m$ non-linearity coefficient of trap creation and $k$ is a constant.

# Threshold voltage measure

- Silicon Carbide Metal Oxide Semiconductor Field Effect Transistor: SICMOSFET
- SiC MOSFET threshold Voltage Instability: Accelerated Degradation Test
- Instability of the voltage threshold is an aging mechanism
- Threshold voltage measure,
- Variable usage profile
- Time Dependent Dielectric Breakdown :

$$T_{BD} = k \frac{N_{BD}^{1/m}}{R_G} e^{\frac{E_A}{k_B T}} + W_t$$

$N_{BD}$ is the critical density of trap in oxyde, $R_G$ default creation rate, $m$ non-linearity coefficient of trap creation and $k$ is a constant.

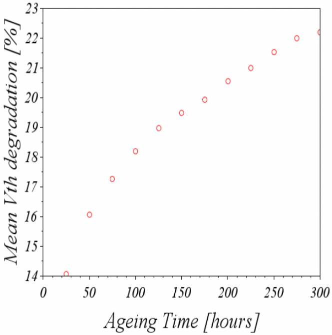
# Threshold voltage measure

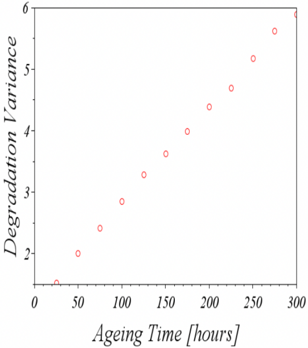
- Silicon Carbide Metal Oxide Semiconductor Field Effect Transistor: SICMOSFET
- SiC MOSFET threshold Voltage Instability: Accelerated Degradation Test
- Instability of the voltage threshold is an aging mechanism
- Threshold voltage measure,
- Variable usage profile
- Time Dependent Dielectric Breakdown :

$$T_{BD} = k\frac{N_{BD}^{1/m}}{R_G}e^{\frac{E_A}{k_B}T}$$

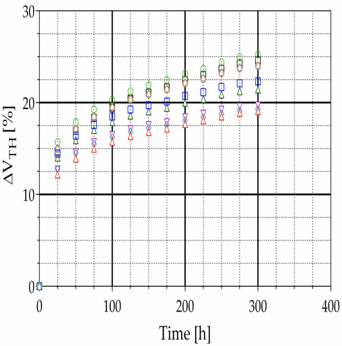
$N_{BD}$ is the critical density of trap in oxyde, $R_G$ default creation rate, $m$ non-linearity coefficient of trap creation and $k$ is a constant.

# Threshold voltage measure



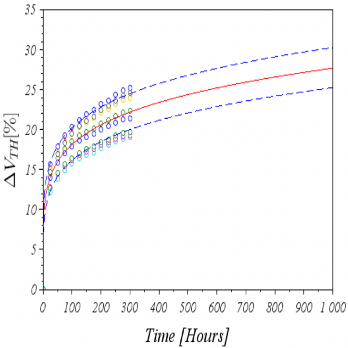



# Threshold voltage measure



# Threshold voltage measure

An **extended gamma process** is used where the average trend in non linear and the variance is increasing with time



# Threshold voltage measure

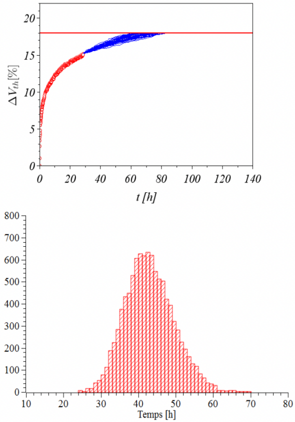



Figure: Crossing time of a given threshold is studied

# Threshold voltage measure: Gamma process with covariates

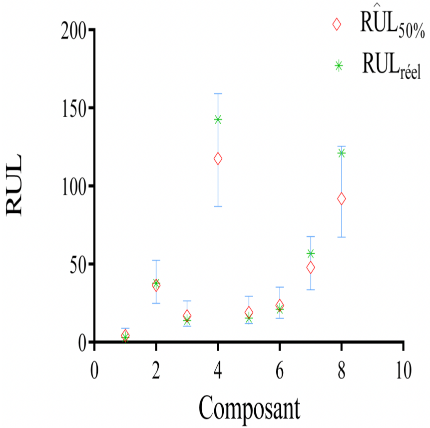



Figure: The influence of usage parameters on the remaining time before crossing the threshold

Crack propagation

# Crack propagation

- For M components, a crack size $X_{k,t}$, observed at inspection times $t \in \{t_{1,k}, \dots, t_{n_k,k}\}$, $k \in \{1, \dots, M\}$.
- The fatigue damage variable $D$ is a function of time or the number of cycles $t$, the parameters $C$ and $m$ which are depending on the property of material, $K$ is the stress intensity factor and the stress intensity range $\Delta K$:

$$\frac{dD}{dt} = C(\Delta K)^m$$

  where $\Delta K$ is dependent on $D$ through the equation $\Delta K(D) = \beta\sqrt{D}$ and $\beta$ is a constant in stable environment.

# Crack propagation

$$\frac{dD}{dt} = C(\Delta K)^m + W_t \tag{1}$$

or

$$\frac{dD}{dt} = C(\Delta K)^m$$

$$\frac{dD}{dt} = C(\Delta K)^m$$

Stochastic process: Gamma process

$$D_t - D_s \sim \Gamma(\alpha(t-s), \beta)$$

where uncertainty is also propagated through parameters

$$\pi(\alpha, \beta) \propto \frac{1}{\beta}\sqrt{\alpha\Psi_1(\alpha) - 1} \quad \beta|\alpha, \mathrm{d} \sim \mathcal{IG}\left(\alpha \sum_{k=1}^{M} t_{n_k,k}, \sum_{k=1}^{M}\left[S_k + \sum_{i=r_k+1}^{n_k} z_{k,i}\right]\right).$$

where $\Psi_1(.)$ is the trigamma function.

# Crack propagation: bayesian estimation for a gamma process with censored data

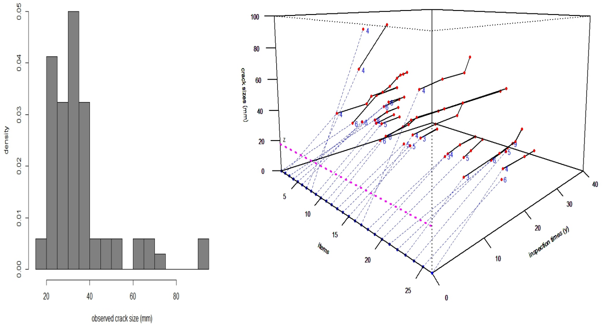



Figure: Histogram of truly observed crack sizes. All observations are above the detection threshold $z = 18mm$ (Left).Observed trajectories of crack size over 26 similar components. The dashed blue lines correspond to censored parts of the trajectories, and the numbers of missing data per observed trajectory are written in blue. (Right)

# Crack propagation: parameter uncertainty

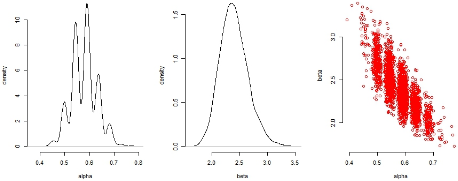
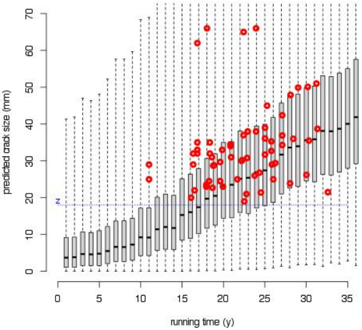


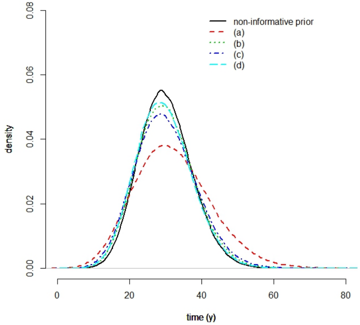

Figure: Marginal posterior distributions of $(\alpha, \beta)$ and joint posterior sampling, given a non-informative prior.

# Crack propagation: parameter uncertainty propagation




Boxplots of the posterior distributions (given an non-informative prior on $(\alpha, \beta)$) of each predicted crack size and real observations (in red). The boxes represent the interquartile intervals (25-75%-ordered percentiles) around the median, and the lines indicate the intervals 1-99%. 5000 posterior samples produced by the Gibbs algorithm are used to compute these box plots.

# Crack propagation: prediction uncertainty propagation



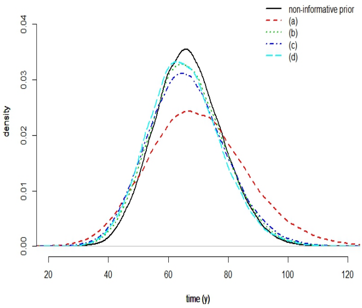




Figure: Posterior predictive density functions of the visiting time $T_{40}$ given non-informative and informative priors (Left). Posterior predictive density functions of the failure time $T_{90}$ given non-informative and informative priors (Right).

# Taking into account the physical model

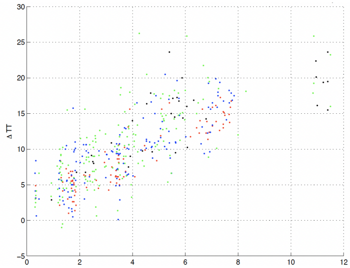
Ageing of steel impacted by external factors

# Steel ageing

- Ageing of steel due to an external factor $\phi$.
- Measurement change of ductile-brittle transition temperature $\Delta TT_\phi$.
- Physical model

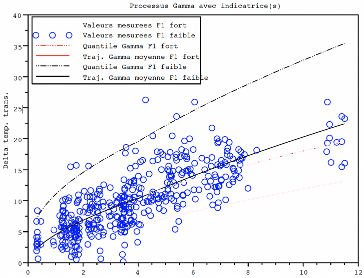
$$\mathbb{E}(\Delta TT_\phi) = F_1(c_{Cu}, c_P, c_{Ni})\phi^\alpha$$

where $F_1(c_{Cu}, c_P, c_{Ni}) = A(1 + a_1(c_p - 0.04)^+ + a_2(c_{Cu} - 1.2)^+ + a_3 c_{Ni}^2 c_{Cu})$ and $A, a_1, a_2, a_3 \in \mathbb{R}$



# Steel ageing

Modelling with a non-homogeneous gamma process



# Cavitation

# Cavitation

Hydraulic Francis turbine runners: erosive cavitation

$$X_{t_i} - X_{t_{i-1}} = (t_i - t_{i-1})c_i$$

where $X_{t_i}$ and $c_i$ are the mass and cavitation intensity at time $t_i$ respectively (2 order differential equation 2).

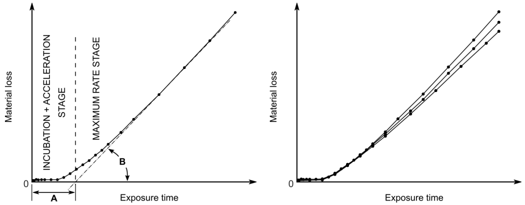



Figure: (Left) Characteristic stages of the cumulative erosion-time curve, (Right) Cavitation erosion paths ($m = 3$) from laboratory.

# Cavitation

Modelling with a non-homogeneous gamma process

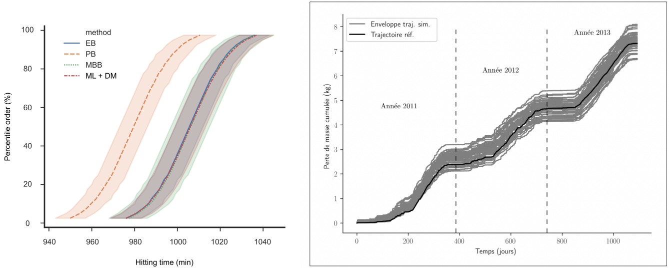
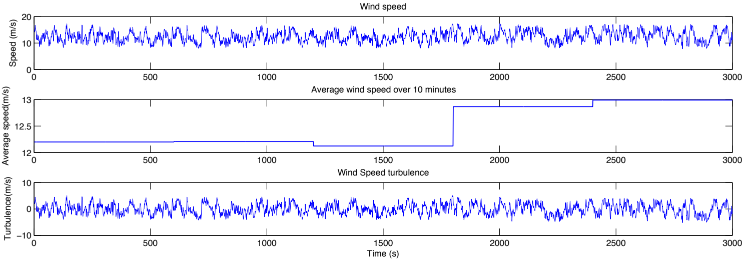



Figure: Hitting time distributions of $T_\rho$, Path prediction.

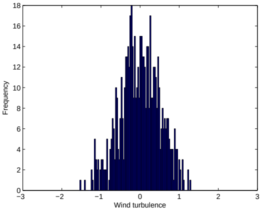
# Data-based models

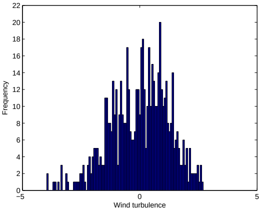
Wind speed modelling

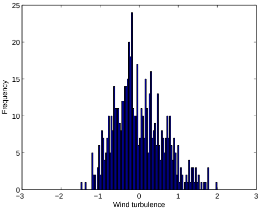
# wind


Wind speed
Speed (m/s)
20
10
0
0
500
1000
1500
2000
2500
3000
Average wind speed over 10 minutes
Average speed(m/s)
13
12.5
12
Wind Speed turbulence
Turbulence(m/s)
10
0
−10
Time (s)

# wind: clustering



# Wind speed statistical modelling

- Aerodynamic models: considering turbulence etc.,
- Navier-Stokes equations
- Sensitivity analysis difficult to implement
- uncertainty propagation time consuming.

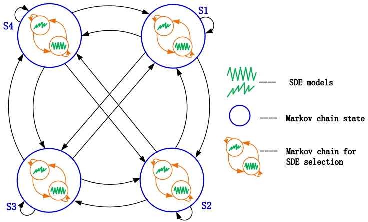
# Wind speed statistical modelling

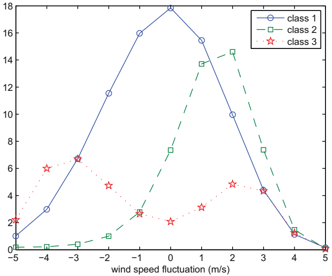
The Ornstein-Uhlenbeck process is a stochastic process that satisfies the following stochastic differential equation:

$$dX_t = a(c - X_t)dt + bdW_t,\ t \in [0, T],\quad Y(0) = 0. \tag{2}$$

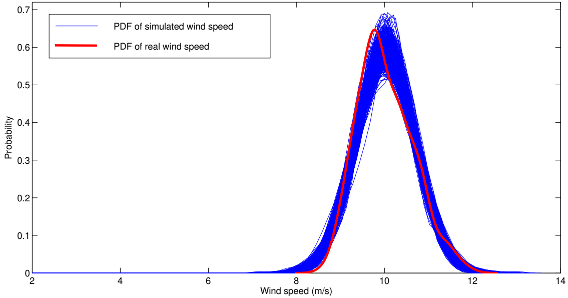
where $(W_t)_{t \geq 0}$ is a standard Brownian motion. The parameters $a >$ is the rate of mean reversion, $c > 0$ is the long-term mean of the process and $b > 0$ is the volatility or average magnitude, per square-root time, of the random fluctuations that are modelled as Brownian motions.

# Diffusion embedded Markov chain



# wind: clustering



Wind turbine health indicator modelling

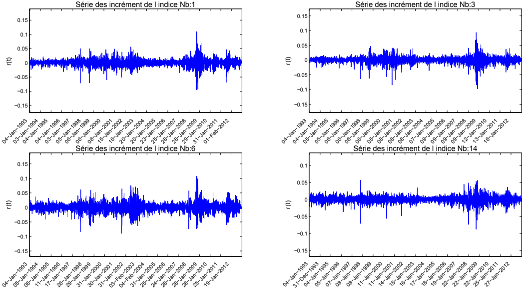



Figure: Four indicator increments

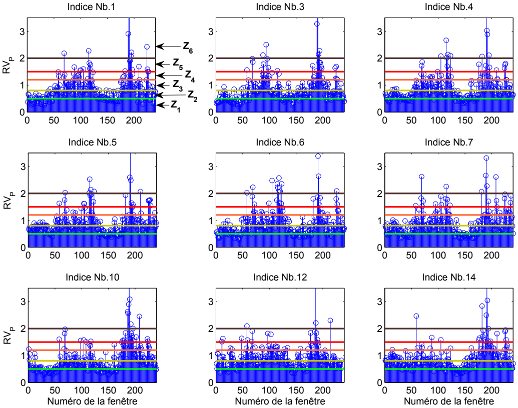



Figure: Volatility evolution in time

# Stochastic model

Jump diffusion process

$$dX_t = X_t\left(\mu_d dt + \sigma_d dW_t + \xi_J dN_t\right)$$

$\mu_D \in \mathbb{R}$, $\sigma_D \in \mathbb{R}^+$, $\xi_J$ is a r.v, $(W_t)_{t\geq 0}$ is a standard brownian motion and $(N_t)_{t\geq 0}$ is a Poisson process.

- Dynamic environment

$$\frac{dX_t}{X_t} = \mu_d dt + \sqrt{V_t}dW_t^{(1)} + (e^J - 1)dN_t$$

$$dV_t = \kappa_V(\theta_V - V_t)dt + \sigma_V\sqrt{V_t}dW_t^{(2)}$$

where $(W_t^{(1)})_{t\geq 0}$ and $(W_t^{(2)})_{t\geq 0}$ are two independent standard brownian motions, $J$ is a real random variable, $\sigma_V \in \mathbb{R}^+$, $\theta_V \in \mathbb{R}$ and $\kappa_V \in \mathbb{R}^+$.

# Clustering

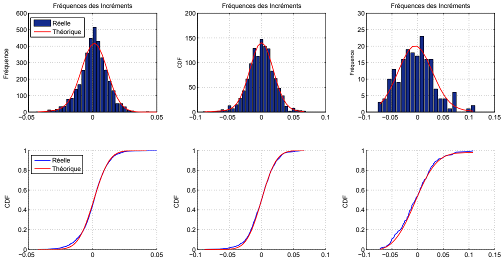
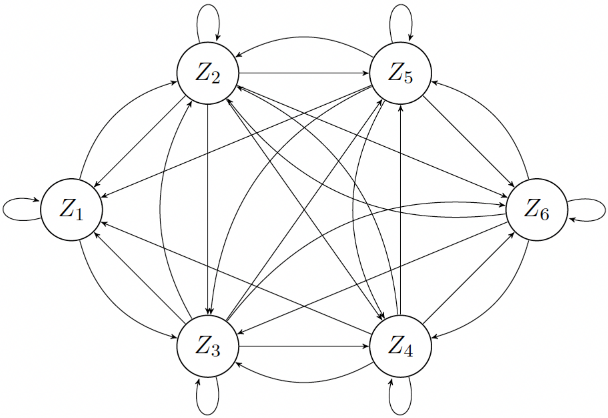



Figure: Increments of indicator in three different classes .

# Regime switching model



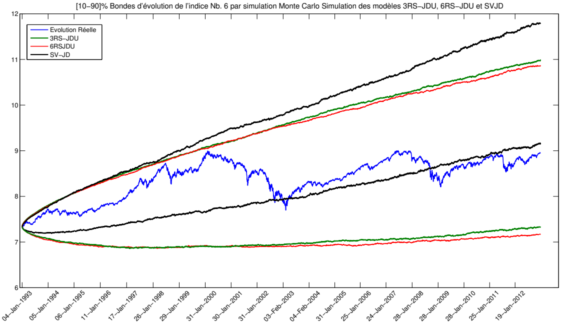



Figure: Intervalles de confiance de $[10, 90]\%$ one indicator with 3RS-JDU, 6RS-JDU et SV-JD.

## Data

- Noise: with or without noise
- Type: high frequency, periodic or non-periodic measurement intervals, outliers, extreme values
- Size: small, large

# Formalisation with observations without noise

- Direct measurements at time $(t_k)_{k\in\mathbb{N}}$,
- $X_{t_k}$ indicator at $t_k$

$$\mathbb{P}(X_{t_k+h} \in \bar{\mathcal{U}} | X_{t_1}, ..., X_{t_k})$$

Such that $X_{t_i} \in \mathcal{U}$ , $i \in \{1, \cdots, k\}$

Challange:

The calculation of the conditional distribution

$$\mathcal{L}(X_{t_k+h} | X_{t_1}, ..., X_{t_k})$$

# Formalisation with noisy observations

- The system is inspected at times $t_k$, $k \in \mathbb{N}$,
- $Y_{t_k}$ observation at $t_k$,
- $X_{t_k}$ system state at $t_k$
- $Y_i = g_i(X_{t_i}, \epsilon_i)$, $\epsilon_i$ r.v. considered as noise, $i \in \mathbb{N}^*$
- Aim: find

$$\mathbb{P}(X_{t_k+h} \in \bar{\mathcal{U}} | Y_{t_1}, \ldots, Y_{t_k})$$

for $h > 0$

Challange:

The calculation of the conditional distribution

$$\mathcal{L}(X_{t_k+h} | Y_{t_1}, \ldots, Y_{t_k})$$

# Simplest case: $\Omega = \mathbb{R}$ and memoryless process

$$\mathbb{P}(X_{t_k+h} > L | Y_{t_1}, \ldots, Y_{t_k}) = \int \underbrace{\bar{F}_{X_{t_k+h}}(L-x)}_{\text{Survival function}} . \underbrace{\mu_{Y_{t_1},\ldots,Y_{t_k}}(t_k)}_{\text{conditionnally to observations}} \quad dx$$

Filtration methods: MCMC, Particle filter

$$\widehat{F}_{t_k}(h) = \frac{1}{Q} \sum_{q=Q_0+1}^{Q_0+Q} \bar{F}_{X_{t_k+h}}(L - \widehat{x}_k^{(q)})$$

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

Turbofan Engine Degradation

# Data

- Turbofan Engine Degradation Simulation Data
- Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dynamical model.
- Run-to-failure trajectories for a small fleet of aircraft engines under realistic flight conditions.

## Data

PHM data challenge

- 218 components
- 24 sensors
- 3 operational conditions
- time series

Two sets

- learning set: data until failure
- test set : data before failure

# Health indicator

Steps

- Sensor selection
- Operational mode distinction
- CPA
- health indicator proposition

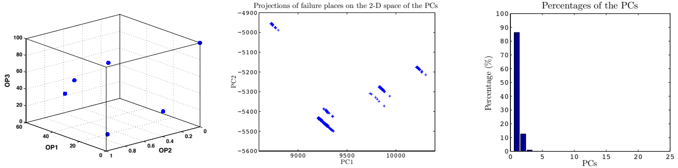
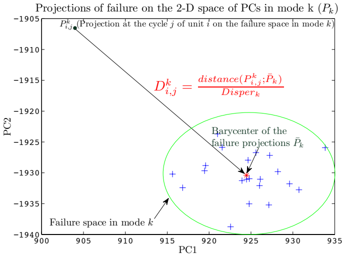
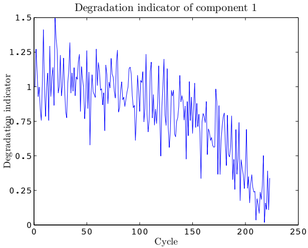
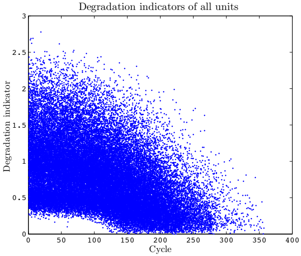



Figure: 6 modes and ACP per mode

# Health indicator construction



# Health indicator construction

Stochastic processes

- Non monotonic degradation: Wiener process with drift
- Monotonic degradation with noise: non homogeneous gamma process with additive noise

# NEXT STEP

Let's talk and work together