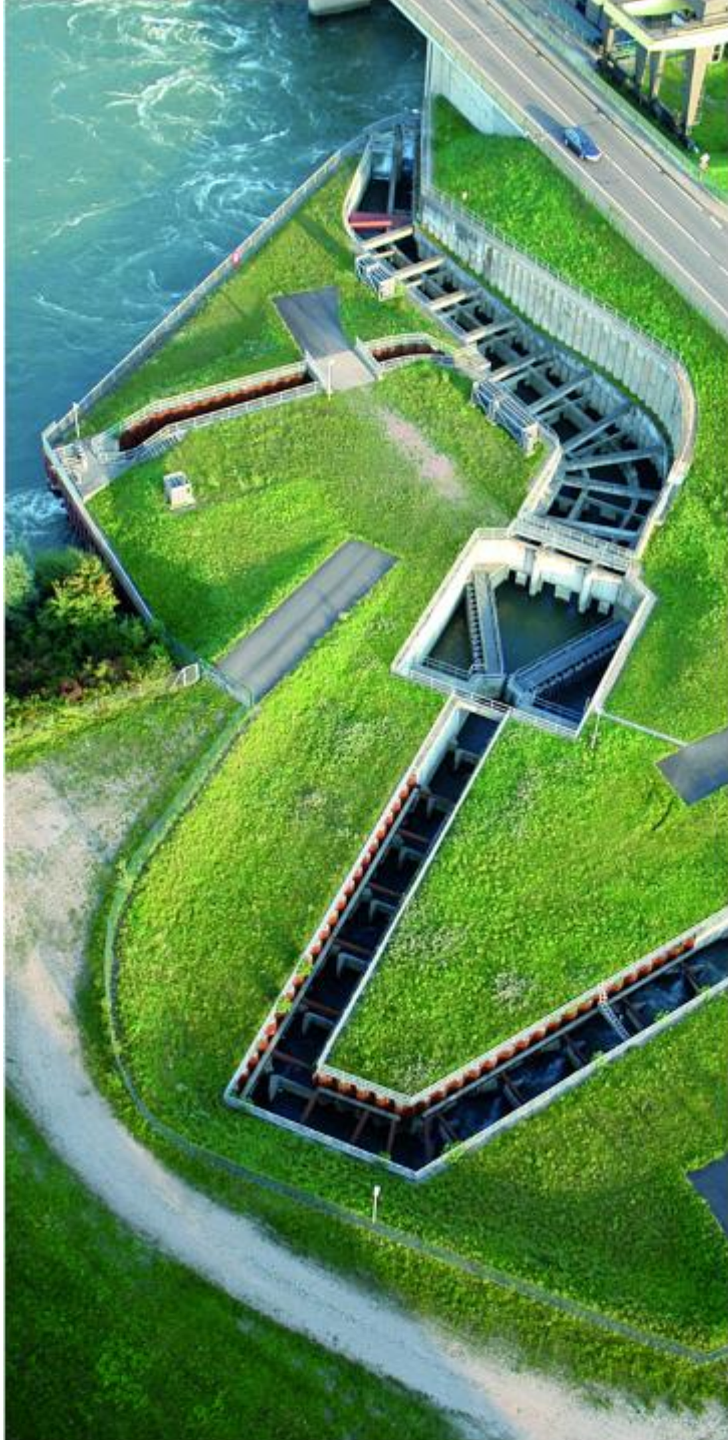
ROBUSTNESS ANALYSIS IN UNCERTAINTY PROPAGATION OF NUMERICAL MODELS

Bertrand IOOSS

Electricité de France (EDF) - R&D, Chatou, France

with Thibault DELAGE & Roman SUEUR (EDF R&D)

stresstest2019 workshop, May, 22, 2019, Paris

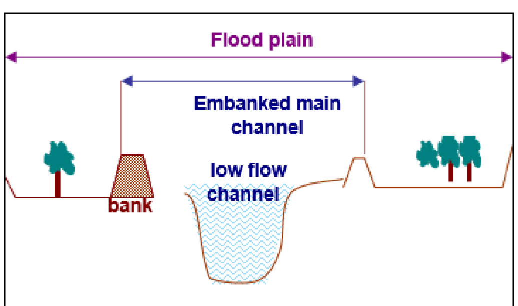


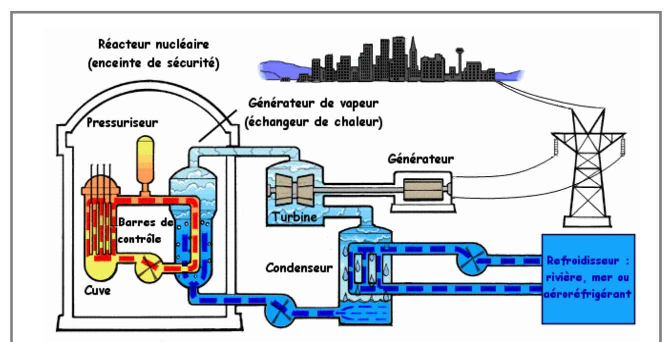
Risk management at EDF

Many uncertainties for the energy production and the safety due to:

- hazards (demand, weather, …),
- incomplete system knowledge (ageing, physics, …),
- internal agressions (failures, …)
- external agressions (earthquake, …)

In order to better understand, prove the safety and optimize its industrial processes, EDF R&D develops some physical numerical simulation codes





Uncertainty Quantification (UQ) in simulation-based studies

Exploratory studies: understand a phenomenon, an experim/indust. process

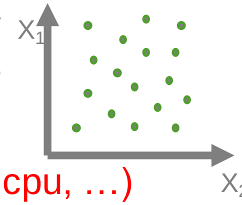
Safety studies: compute a failure risk and prioritize the risk indicators, with validated computer models

Design studies: optimize and manage the system performances

Uncertainties

- **Environnemental variables**
- **Physical parameters**
- **Parameters of process**

Design of experiments



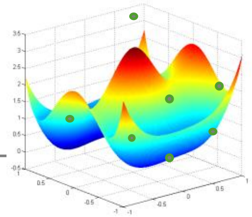
Computer code

$Y = G(X_1, \ldots, X_d)$

or experiment

Cost ($€, cpu, …) potentially large

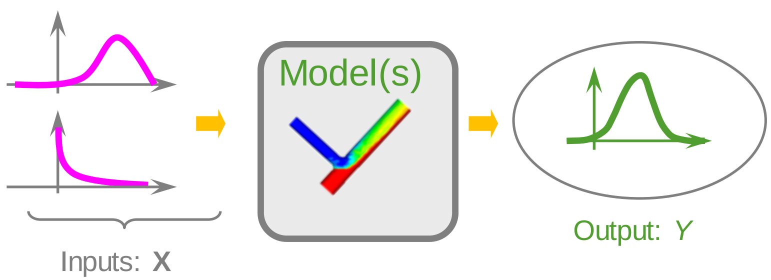
Possible needs of metamodel



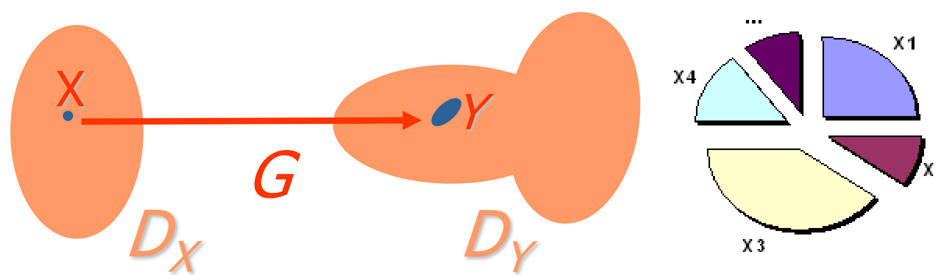
- **Distributions of outputs**
- **Probability of failure**
- **Most influential inputs**
- **Calibrating the input parameters**

UQ methods

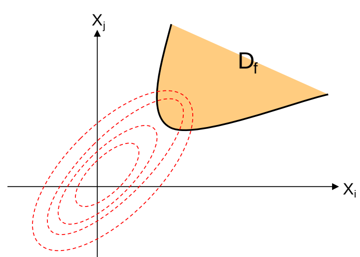
Uncertainty propagation



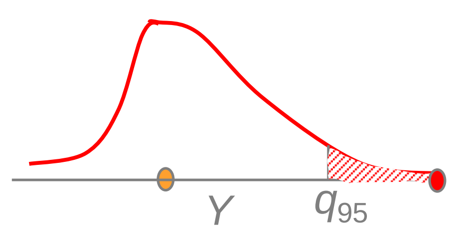
Global sensitivity analysis



Structural reliability (rare events)



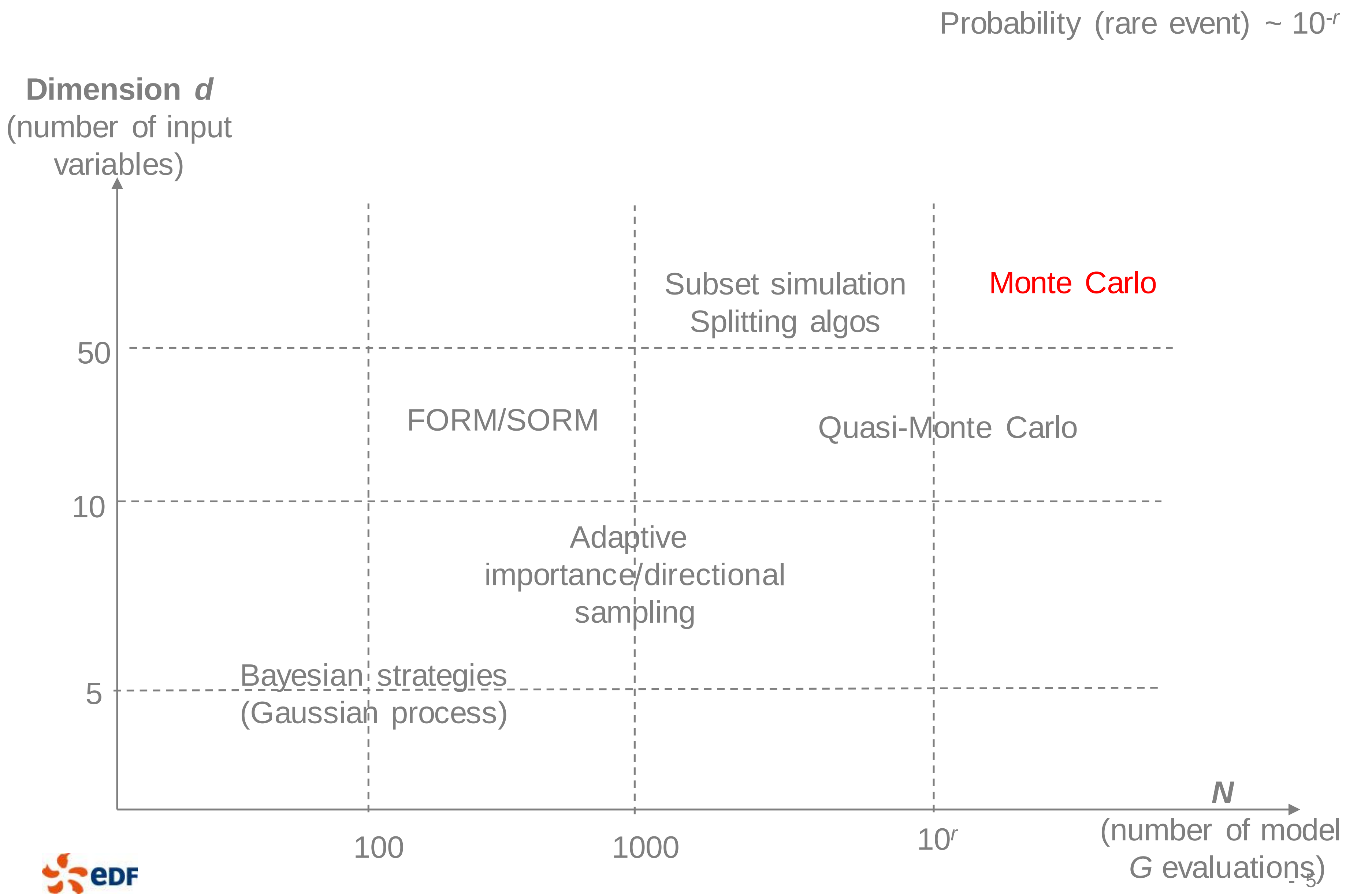
$$p_f = \int \mathbb{1}_{\{G(x)\leq 0\}}\, p(x)dx = \int_{\mathcal{D}_f} p(x)dx$$



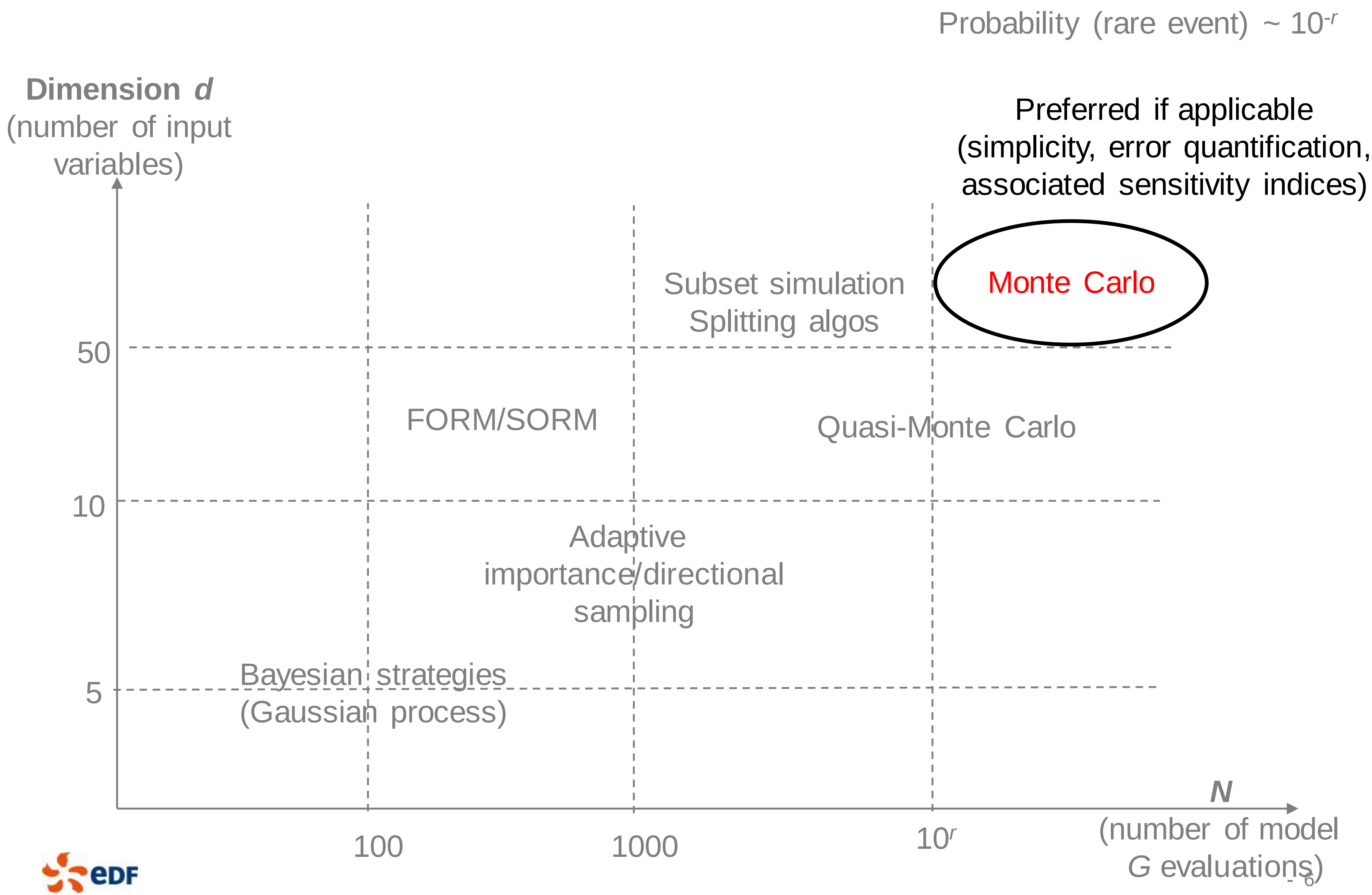
Parameters calibration (inverse problem)

Cas « classique » : Les paramètres à caler sont physiquement constants						Autre contexte : Les paramètres à caler sont intrinsèquement aléatoires
− Richesse de l'information +						
Peut-on faire une hypothèse sur la distribution d'incertitudes des paramètres ?	Cas particulier : « nombre de paramètres = nombre de données observées »	Oui		Oui		
Hypothèse de distribution	Aucune	Gaussienne		Non gaussienne		
Coût calcul		Faible	Important	Faible	Important	
Approche proposée	Inversion directe	Assimilation de données		Calage bayésien	Appel à expertise (métamodèles, …) → cf. section « pour aller plus loin »	Inversion probabiliste

Rare event inference methods: Classification for engineers



Rare event inference methods: Classification for engineers



Industrial objectives

Example: Simulation of IBLOCA accident

Pressurized Water Reactor scenario:
Loss of primary coolant accident due to a break in cold leg

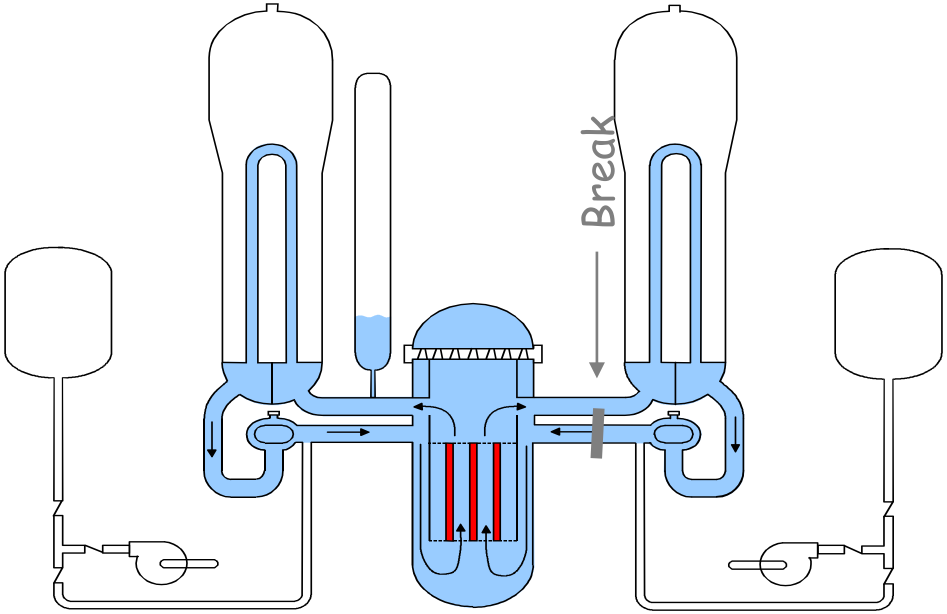
Variable of Interest :
Second peak of cladding temperature (PCT) = scalar output

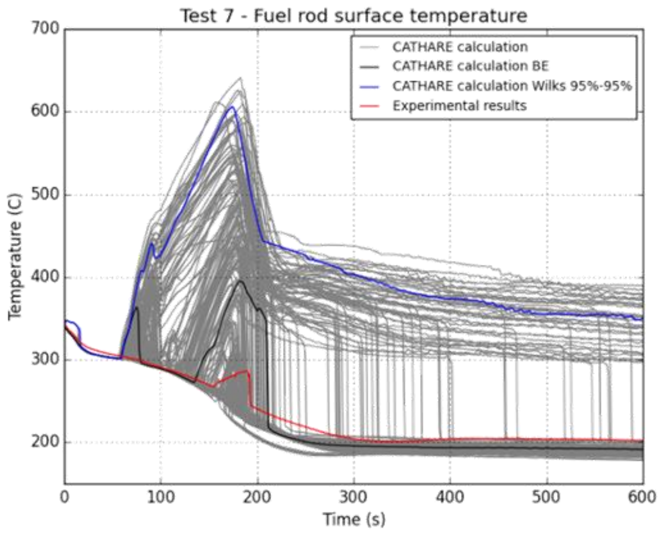
***p* (~ 100) uncertain input variables :**
Critical flowrates, initial/boundary conditions, phys. eq. coef., …

<u>Modelled using **CATHARE code**:</u>
(thermal-hydraulic phenomena)

CPU cost for one code run > 1 hour
Industrial studies consider ~ 2000 runs

Quantity of Interest (QoI) :
High values of the PCT





SAFETY ISSUES IN NUCLEAR ENGINEERING

Goals: Assess margins with regards to a regulatory criteria (the regulator will accept the safety approach if a sufficent margin remains, e.g. 100°C)

1) Historical approach

=> **conservative models** (e.g. without compensating physics) with **conservative inputs' values** (leading to the most penalizing calculation, corresponding to expert-based min. or max. value of each input)

2) Current approaches aims to take into account realistic/complex physics

=> **realistic models** (at the industrial level) with **conservative inputs**
Problems due to interactions and non-monotonicity of complex physics

3) Objectives: better assessment of the real margins

=> BEPU (Best Estimate Plus Uncertainties): **realistic models** & **inputs**

BEPU ISSUES

BEPU approaches are well (and naturally) developed in the **probabilistic framework** (needing to define probabilistic distributions of the inputs)

Turning deterministic to probabilistic studies induces large practical (and cultural) changes for the engineers: the idea is to have a large coverage of the possible situations; the worst-case situation is most often non physical

<u>Importance of the choice of the quantity of interest:</u>

- Probability of threshold exceedence

- High quantile (95% to 99%):

    - easier to compute,

    - model computations remain in the validity domain of the computer code,

    - for the regulator, it allows to keep its fundamental safety margin (by comparison with the threshold)

BEPU ISSUES

BEPU approaches are well (and naturally) developed in the **probabilistic framework** (needing to define probabilistic distributions of the inputs)

Turning deterministic to probabilistic studies induces large changes for the engineers: the idea is to have a large coverage of the possible situations and to show that the worst-case situation is most often non physical

Importance of the choice of the quantity of interest:

- Probability of threshold exceedence

- High quantile (95% to 99%)

Key point: Presence of so-called **epistemic uncertainties**: parameters which are uncertain due to a lack of knowledge (vs. stochastic uncertainties)

The French nuclear regulatory authority ask to justify the probabilistic approach

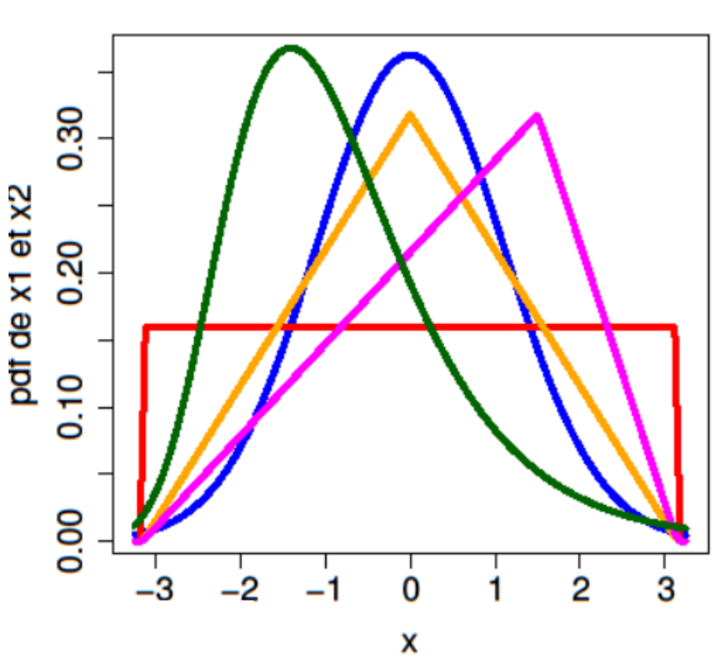
=> Robustness of the study results towards the input distributions

SCIENTIFIC MOTIVATION – IMPORTANCE OF INPUT PROBABILITY DISTRIBUTIONS IN UQ

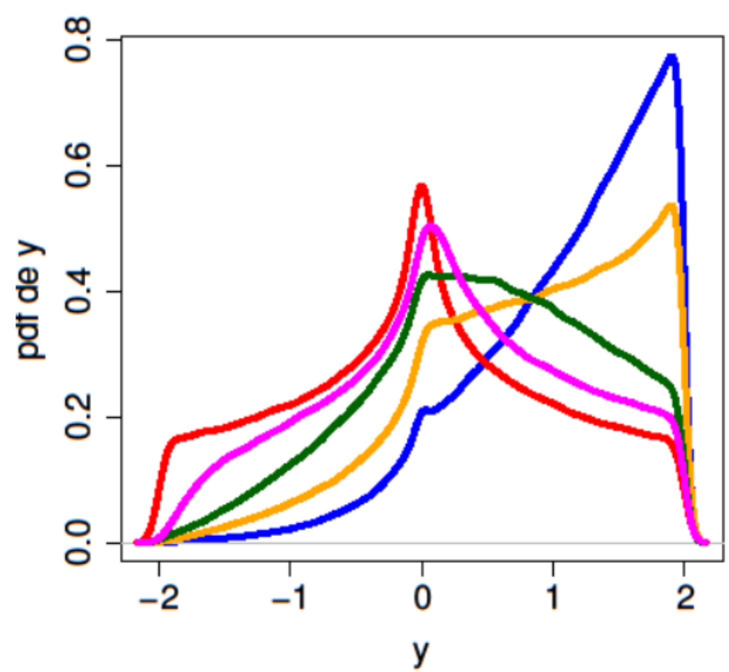
Let's consider $G(X_1, X_2) = \cos(X_1) + \cos(X_2)$ with X_1 and X_2 independent, following the same distribution

Strong impact of the choice of the input distributions on the output distribution, and particularly on some quantities of interest: probability of exceendance, quantile, …

pdf of X_1 and X_2



pdf of $G(\boldsymbol{X})$



Principles of PLI
(robustness indices)

PERTURBATED-LAW BASED INDICES (PLI) (1/2)

The motivation of the PLI indices was firstly to perform global sensitivity analysis on exceedance probability computations, as classical Sobol' indices focus on contributions of input on output variance (Paul Lemaitre's PhD work, 2014)

Global sensitivity analysis: Classical view

1. **Understand the behaviour of the model** $Y = G(X)$

2. **Simplify the computer model (dimension reduction)** **Screening**
   - **Determine the non-influential variables** (that can be fixed)
   - Determine the non-influential phenomena (to skip in the analysis)
   - Build a simplified model, a metamodel

3. **Prioritize the uncertainty sources to reduce the model output uncertainty**

Quantitative partitioning

- Variables to be fixed to obtain the **largest output uncert. reduction**
- Most influential variables in a given output domain

For example: Sobol' indices $S_i = \frac{\mathrm{Var}(\mathrm{E}(Y|X_i))}{\mathrm{Var}(Y)}$ and $T_i = 1 - \frac{\mathrm{Var}(\mathrm{E}(Y|X_{-i}))}{\mathrm{Var}(Y)}$

Global sensitivity analysis: New view

1. **Understand the behaviour of the model** $Y = G(X)$

2. **Simplify the computer model (dimension reduction)** **Screening**
   - **Determine the non-influential variables** (that can be fixed)
   - Determine the non-influential phenomena (to skip in the analysis)
   - Build a simplified model, a metamodel

3. **Prioritize the uncertainty sources to reduce the model output uncertainty**

Quantitative partitioning

- Variables to be fixed to obtain the **largest output uncert. reduction**

- Most influential variables in a given output domain

4. **Analyze the robustness of the quantity of interest (QoI) with respect to the input uncertainty laws**

Robustness analysis

PERTURBATED-LAW BASED INDICES (PLI) (2/2)

The motivation of the PLI indices was firstly to perform global sensitivity analysis on exceedence probability computations, as classical Sobol' indices focus on contributions of input on output variance (Paul Lemaitre's PhD work, 2014)

The principle is to assess the influence of a perturbation on a parameter of the input distribution, on some quantity of interest of the model output

Recent interests:

- consideration of the quantile or superquantile as the quantity of interest
- use in industrial safety studies

NOTATIONS

We want to study a computer code G which :

- is deterministic
- is a « costly » numerical model (CPU time, memory,…)
- has d input variables
- allows calculating the value $G(X)$ for a given set of input values $X = (X_1, \dots, X_d)$

The input variables are uncertain, hence we denote

- $\mathbb{X} \subset \mathbb{R}^d$ the domain of variation of the random vector X
- $f = \prod_{i=1}^{d} f_i$ the probability density function of X
  - ► each f_i is the density of X_i, the i-th marginal of X
  - ► the uncertain input variables $X_1, \dots X_d$ are considered independent

CONTEXT

Some safety studies consider a α-order quantile $q^{\alpha} = \inf\{t \in \mathbb{R}, F_Y(t) \geq \alpha\}$

When estimating it, we need some conservatism and we often add a confidence level β (due to the estimation uncertainty) to obtain $\hat{q}_{\alpha,\beta}$:

$$P(\hat{q}_{\alpha,\beta} \geq q^{\alpha}) \geq \beta$$

Typical values: $\alpha = \beta = 0{,}95$

In our industrial applications, we evaluate $\hat{q}_{\alpha,\beta}$ by a Monte Carlo method (by the Wilks formula/order statistics or by bootstrap, for the β confidence level)

Problem: robustness of $\hat{q}_{\alpha,\beta}$ wrt uncertainty in some f_i

We focus in the following on the robustness of q^{α} (by its estimate) wrt uncertainty in some f_i

PLI INDICES : THE PRINCIPLE

We aim at quantifying the impact of a perturbation on the pdf of X_i

For example, what happens if we replace $\boldsymbol{E(X_i) = \mu_i}$ by $\boldsymbol{E(X_i) = \mu_i + \delta}$?

We then define the PLI-quantile indices as :

$$S_{i\delta} = \left(\frac{q_{i\delta}^{\alpha}}{q^{\alpha}} - 1\right)\mathbb{I}_{\{q_{i\delta}^{\alpha} > q^{\alpha}\}} + \left(1 - \frac{q^{\alpha}}{q_{i\delta}^{\alpha}}\right)\mathbb{I}_{\{\hat{q}_{i\delta}^{\alpha} < q^{\alpha}\}}$$

- $S_{i\delta} = 0$ when $q_{i\delta}^{\alpha} = q^{\alpha}$ i.e. when f_i has no impact on the quantile
- The sign of $S_{i\delta}$ indicates how the perturbation modifies the quantile
- This initial formulation comes from the PLI for probability of exceedence instead of quantile (interest to have a ratio wrt a probability of reference)
- Engineers suggest now to define the PLI-quantile as

$$S_{i\delta} = \left(\frac{q_{i\delta}^{\alpha}}{q^{\alpha}} - 1\right)$$

ESTIMATION: REVERSE IMPORTANCE SAMPLING

- Prerequisite :

  we have n obs. (the model calculations) : $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(n)}) \rightarrow (y^{(1)}, \dots, y^{(n)})$

- **We start from a classical Monte-Carlo estimator**

$$\hat{q}^{\alpha N} = \inf\{t \in \mathbb{R}, \hat{F}_Y^N(t) \geq \alpha\} \text{ where } \hat{F}_Y^N(t) = 1/N \sum_{n=1}^{N} \mathbb{I}_{\{y^{(n)} \leq t\}}$$

- **Let us note $f_{i\delta}$ the perturbed density of f_i by δ, we can estimate $F_{i\delta}$ by « reverse importance sampling »**

$$\hat{F}_{i\delta}^N = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{\{y^{(n)} \leq t\}} \frac{f_{i\delta}\left(x_i^{(n)}\right)}{f_i\left(x_i^{(n)}\right)}$$

► $\hat{q}^{\alpha N}$ **and** $\hat{q}_{i\delta}^{\alpha N}$ **can then be estimated with the same sample**

No need of new runs of G model

PLI INDICES ESTIMATION

We then estimate the PLI-quantile indices with the so-called plug-in estimator:

$$\hat{S}_{i\delta}^{N} = \left(\frac{\hat{q}_{i\delta}^{\alpha N}}{\hat{q}^{\alpha N}} - 1\right) \mathbb{I}_{\{\hat{q}_{i\delta}^{\alpha N} > \hat{q}^{\alpha N}\}} + \left(1 - \frac{\hat{q}^{\alpha N}}{\hat{q}_{i\delta}^{\alpha N}}\right) \mathbb{I}_{\{\hat{q}_{i\delta}^{\alpha N} < \hat{q}^{\alpha N}\}}$$

- Convergence and CLT of this estimator is under study but, as for a quantile estimate, confidence intervals should be easier to compute by using bootstrap
- For the PLI-probability $P_f = P(Y < 0)$, considering $\hat{P}_N = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}_{\{y^{(n)}<0\}}$ and $\hat{P}_{i\delta N}$:

Proposition

Si $P_f \neq P_{i\delta}$ et sous les conditions usuelles (i) $Supp(f_{i\delta}) \subseteq Supp(f_i)$ et (ii) $\int_{Supp(f_i)} \frac{f_{i\delta}^2(x)}{f_i(x)}\, dx < \infty$, on a

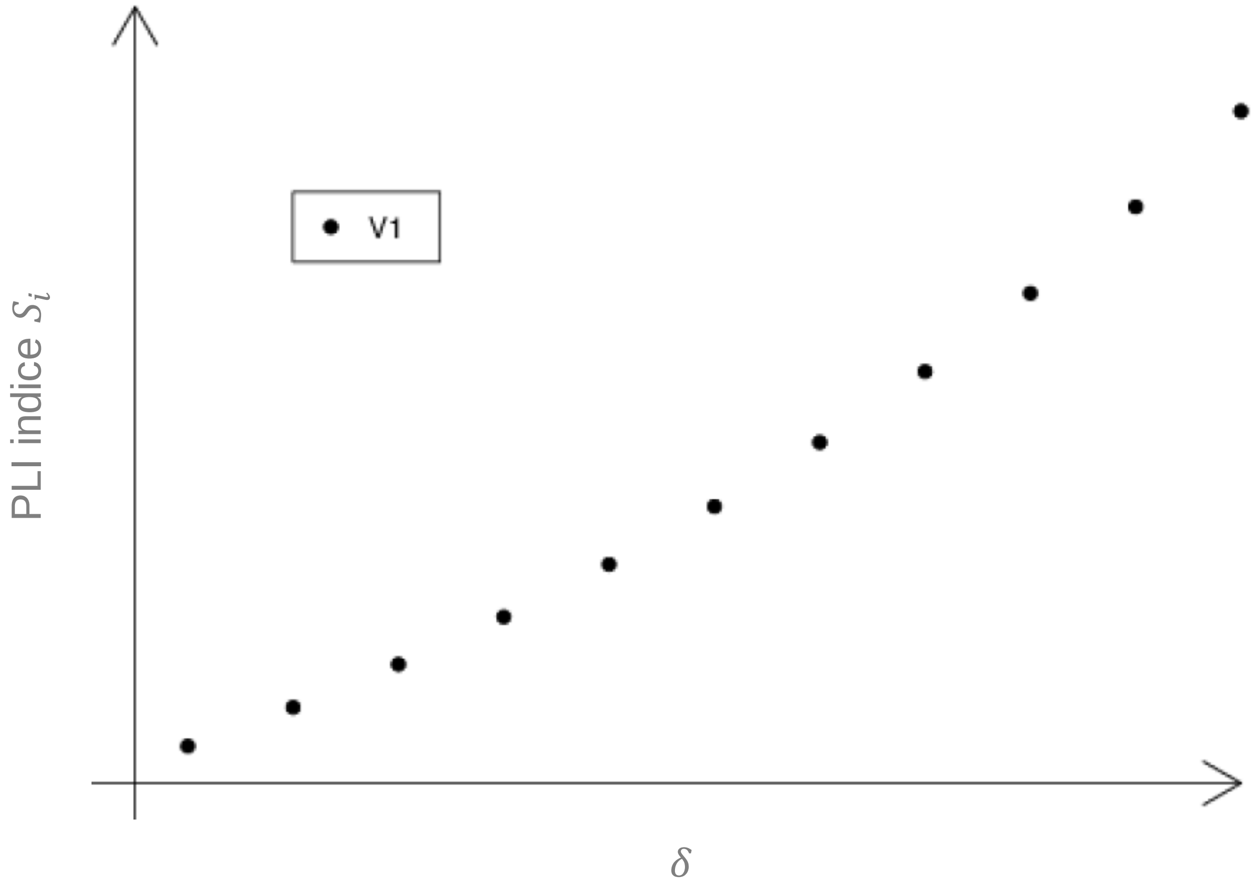
$$\sqrt{N}\left[\hat{S}_{i\delta N} - S_{i\delta}\right] \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}\left(0, d^T \Sigma_{i\delta} d\right)$$

avec $d = \left(\frac{\partial s}{\partial P_f}(P_f, P_{i\delta}), \frac{\partial s}{\partial P_{i\delta}}(P_f, P_{i\delta})\right)^T$ pour $P_f \neq P_{i\delta}$ et

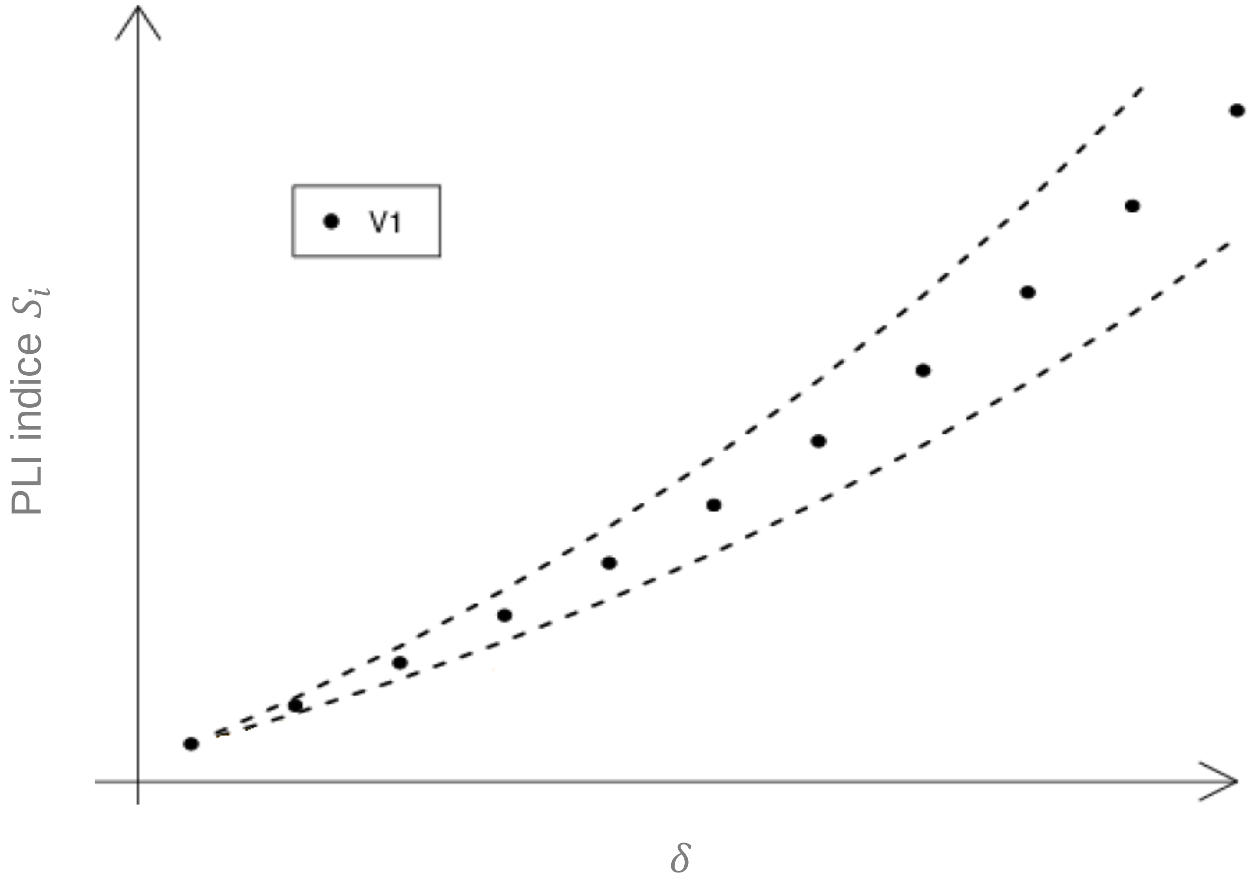
$$\hat{\Sigma}_{i\delta} = \begin{pmatrix} \hat{P}_N(1-\hat{P}_N) & \hat{P}_{i\delta N}(1-\hat{P}_N) \\ \hat{P}_{i\delta N}(1-\hat{P}_N) & \hat{\sigma}_{i\delta N}^2 \end{pmatrix}.$$

[Lemaître 2014; Lemaître et al. 2015]

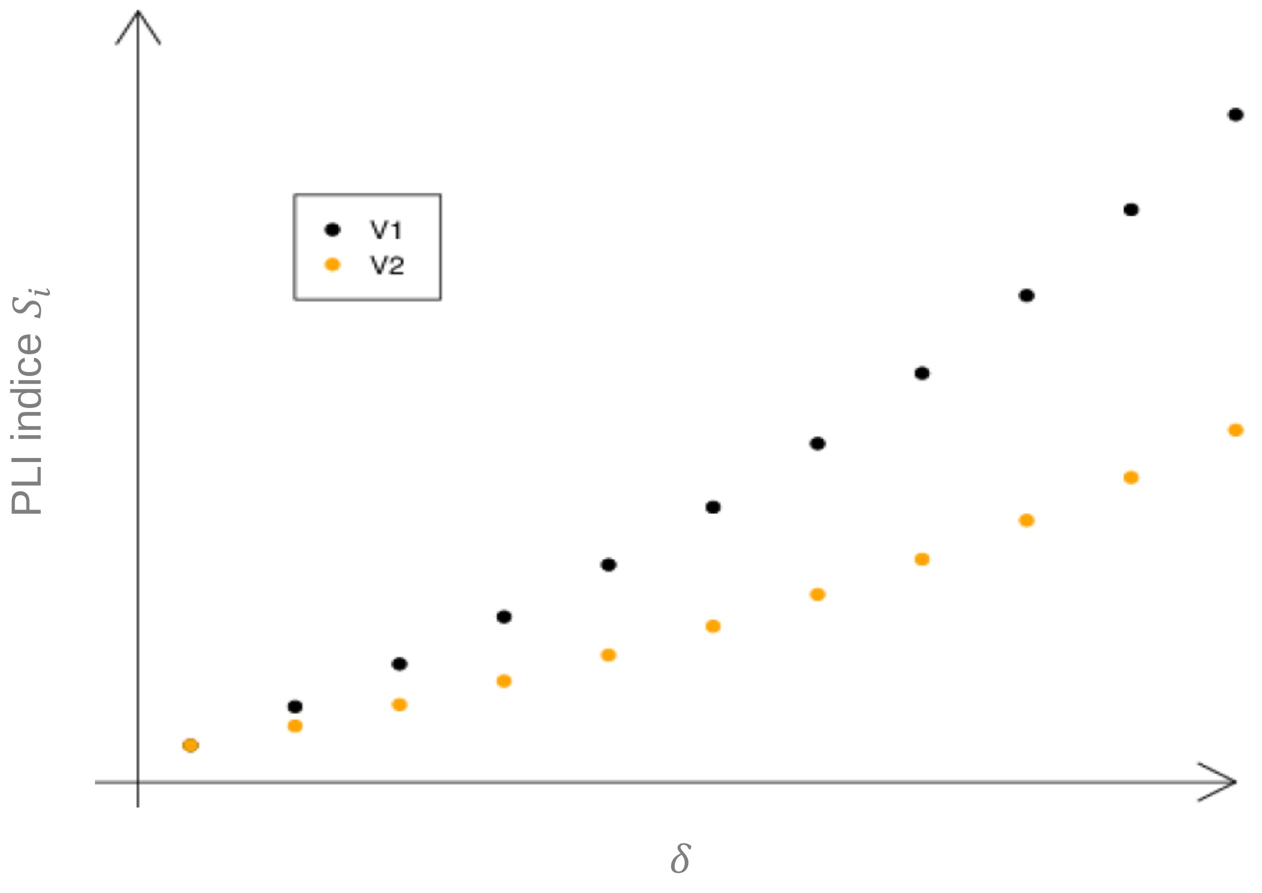
ILLUSTRATION



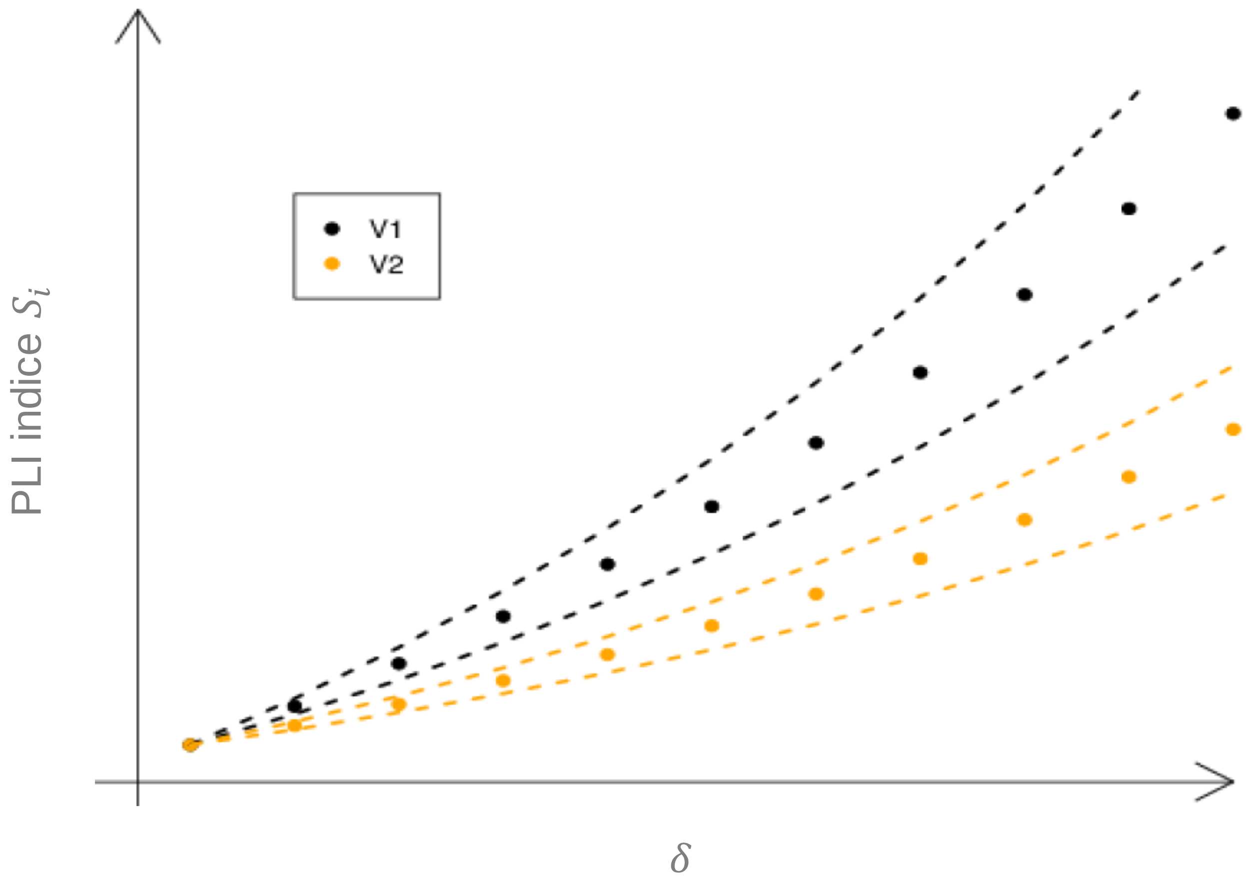
ILLUSTRATION



ILLUSTRATION



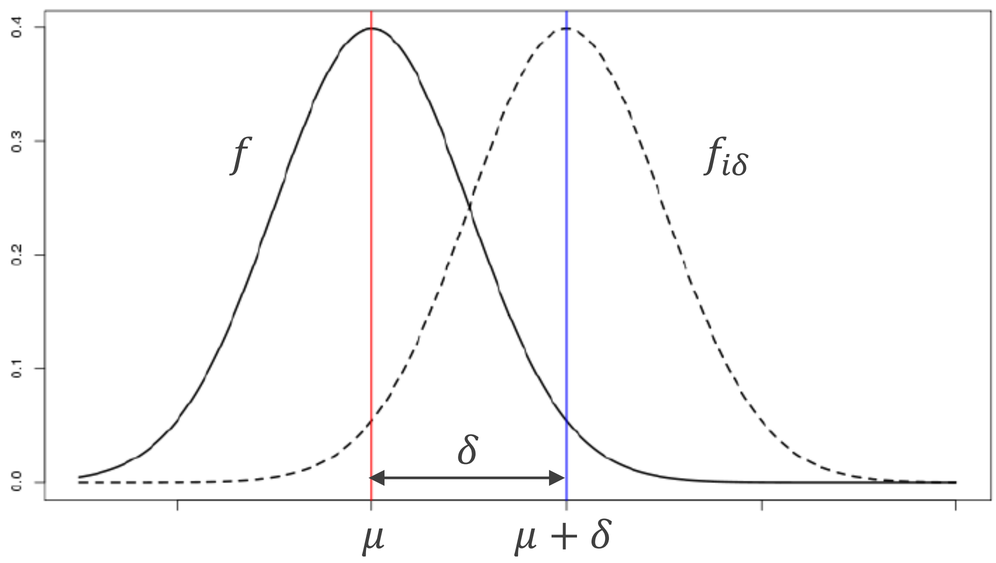
ILLUSTRATION



Density perturbation

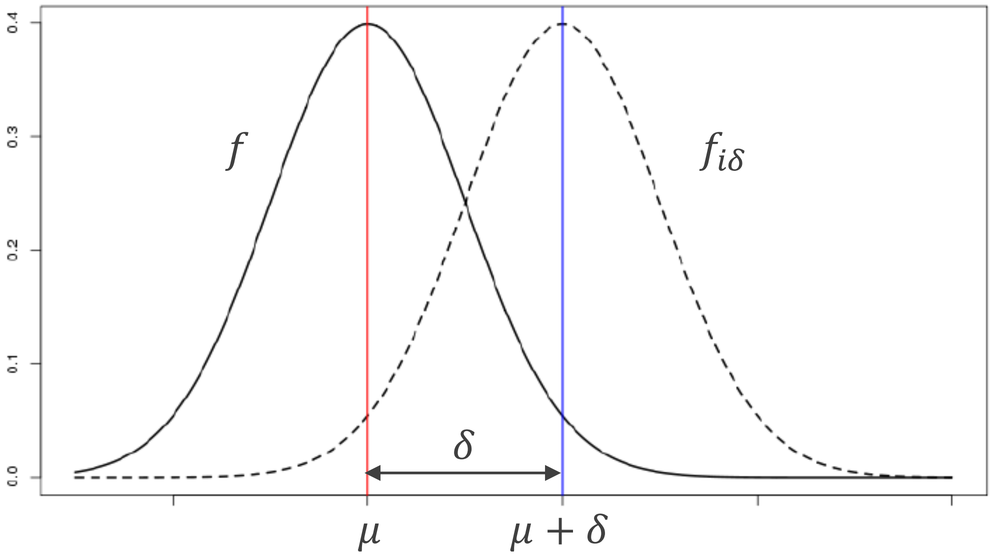
HOW TO DEFINE A DENSITY PERTURBATION ?

- **Let's assume that the X_i input variable has a normal distribution $X_i \sim \mathcal{N}(\mu, \sigma^2)$**
- **What if the mean of X_i was not μ but $\mu + \delta$?**



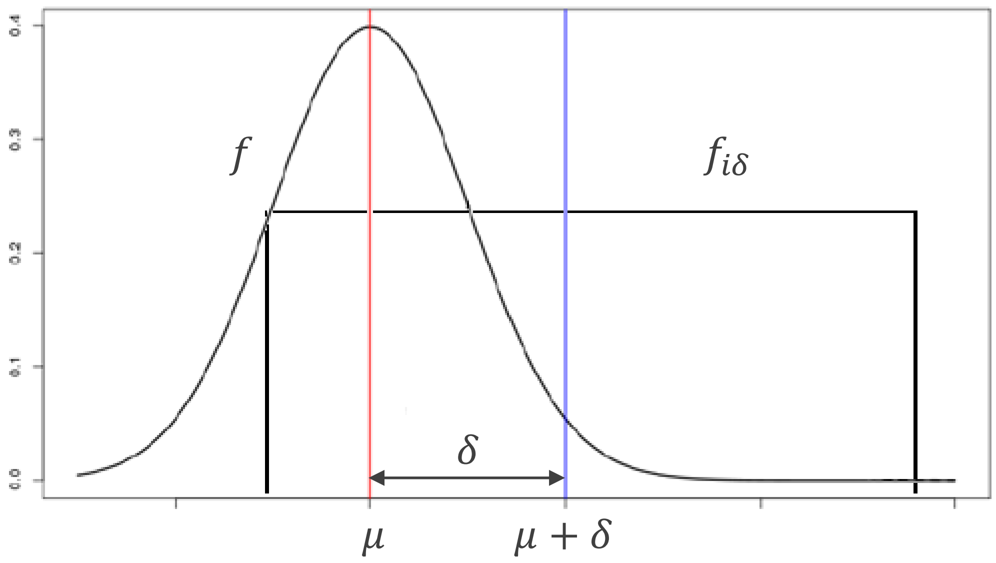
HOW TO DEFINE A DENSITY PERTURBATION ?

- **Let's assume that the X_i input variable has a normal distribution $X_i \sim \mathcal{N}(\mu, \sigma^2)$**
- **How to define $f_{i\delta}$ with the constraint $\int_{\mathbb{X}_i} x_i f_{i\delta}(x_i) dx_i = \mu + \delta$?**



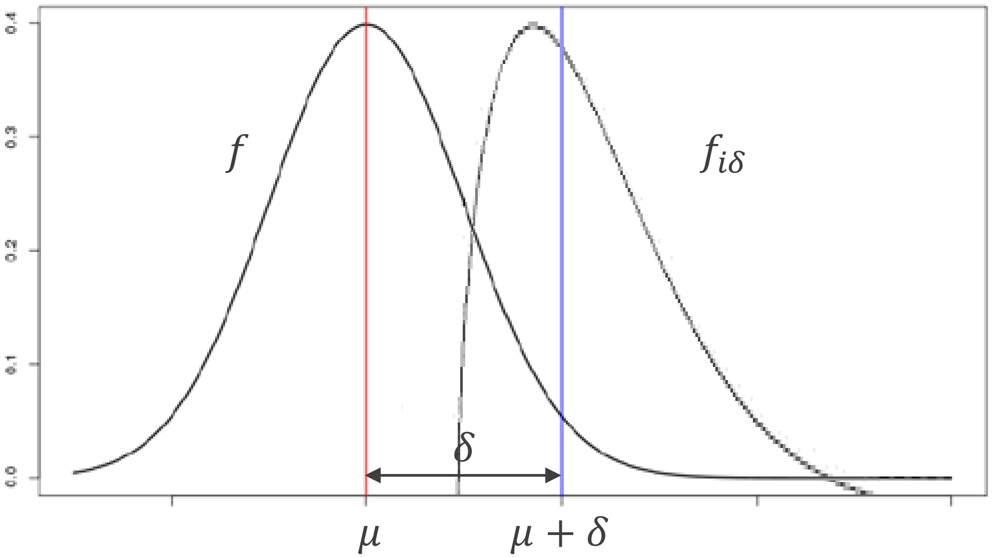
HOW TO DEFINE A DENSITY PERTURBATION ?

- **Let's assume that the X_i input variable has a normal distribution $X_i \sim \mathcal{N}(\mu, \sigma^2)$**
- **How to define $f_{i\delta}$ with the constraint $\int_{\mathbb{X}_i} x_i f_{i\delta}(x_i) dx_i = \mu + \delta$?**



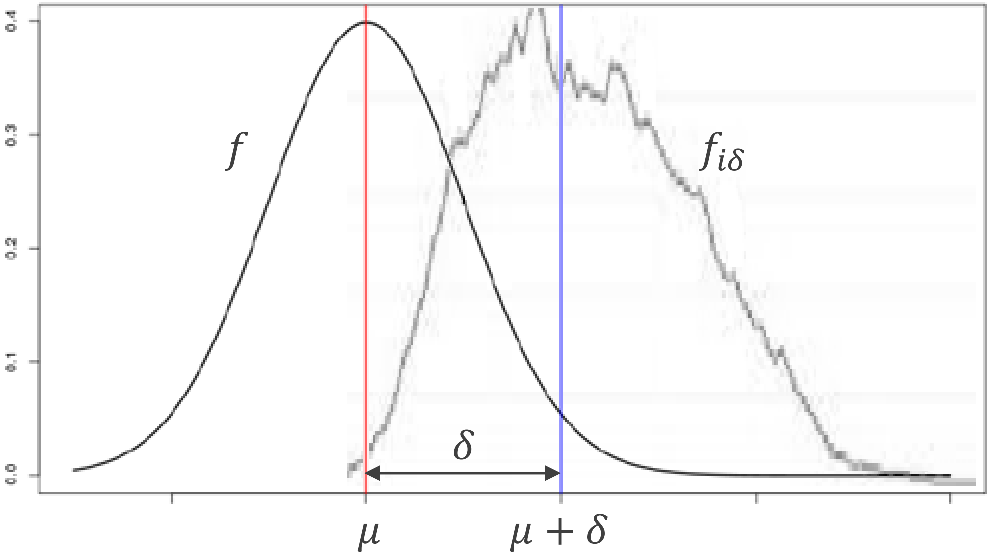
HOW TO DEFINE A DENSITY PERTURBATION ?

- **Let's assume that the X_i input variable has a normal distribution** $X_i \sim \mathcal{N}(\mu, \sigma^2)$
- **How to define $f_{i\delta}$ with the constraint** $\int_{\mathbb{X}_i} x_i f_{i\delta}(x_i) dx_i = \mu + \delta$ **?**



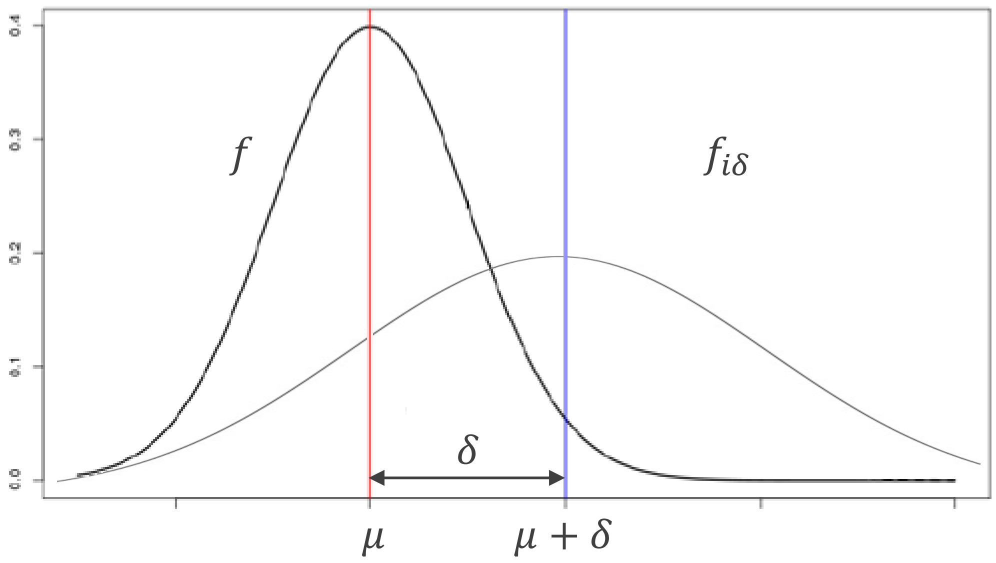
HOW TO DEFINE A DENSITY PERTURBATION ?

- **Let's assume that the X_i input variable has a normal distribution $X_i \sim \mathcal{N}(\mu, \sigma^2)$**
- **How to define $f_{i\delta}$ with the constraint $\int_{\mathbb{X}_i} x_i f_{i\delta}(x_i) dx_i = \mu + \delta$?**



HOW TO DEFINE A DENSITY PERTURBATION ?

- **Let's assume that the X_i input variable has a normal distribution $X_i \sim \mathcal{N}(\mu, \sigma^2)$**
- **How to define $f_{i\delta}$ with the constraint $\int_{\mathbb{X}_i} x_i f_{i\delta}(x_i) dx_i = \mu + \delta$?**



HOW TO DEFINE A DENSITY PERTURBATION ?

- We suggest to define the perturbed density $f_{i\delta}$ as the closest one from the initial f_i in the sense of the entropy, under the constraint of perturbation

- i.e. in the sense of Kullback-Leibler divergence :

$$KL(\pi_1, \pi_2) = \int_{-\infty}^{+\infty} \pi_1(x) \log\left(\frac{\pi_1(x)}{\pi_2(x)}\right) dx$$

- So we can give a general formal definition for $f_{i\delta}$ the following way :

$$f_{i\delta} = \underset{\substack{\pi \\ s.t.\ \mathbb{E}_\pi[g_k]=\delta_k \\ k=1,\dots,K}}{\operatorname{argmin}} KL(\pi, f_i)$$

where :

- $g_1, \dots, g_K$ are K linear constraints on the modified density
- and $\delta_1, \dots, \delta_K$ are the values for the perturbed parameters

EXAMPLES OF PERTURBED PDF

Mean μ ; Variance σ^2

Mean perturbation

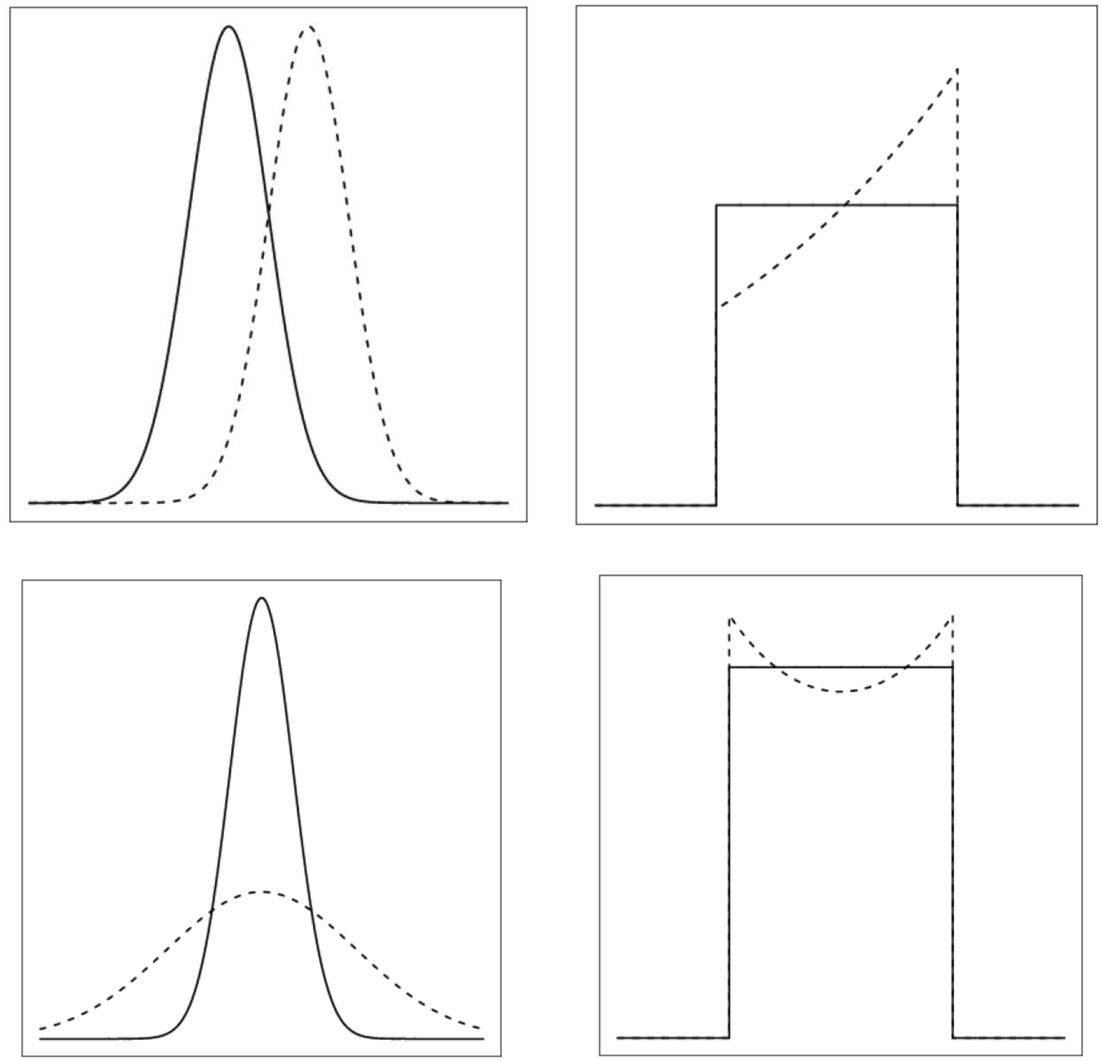
$$\mathbb{E}[X_i] = \mu + \delta$$

Variance perturbation

$$\mathbb{E}[X_i] = \mu$$

$$\mathrm{Var}[X_i] = \sigma^2 + \delta$$

Gaussian | Uniform



GENERAL DENSITY PERTURBATION

- In the case of some usual pdf, we have an analytical expression of $\frac{f_{i\delta}}{f_i}$
- e.g a perturbed Gaussian pdf is another Gaussian pdf of different mean or variance
- But it is not always possible! (e.g. lognormal pdf)
- By applying an iso-probabilistic transformation (e.g. Rosenblatt transformation), we switch to the **standard space** and then get Gaussian pdf for each inputs

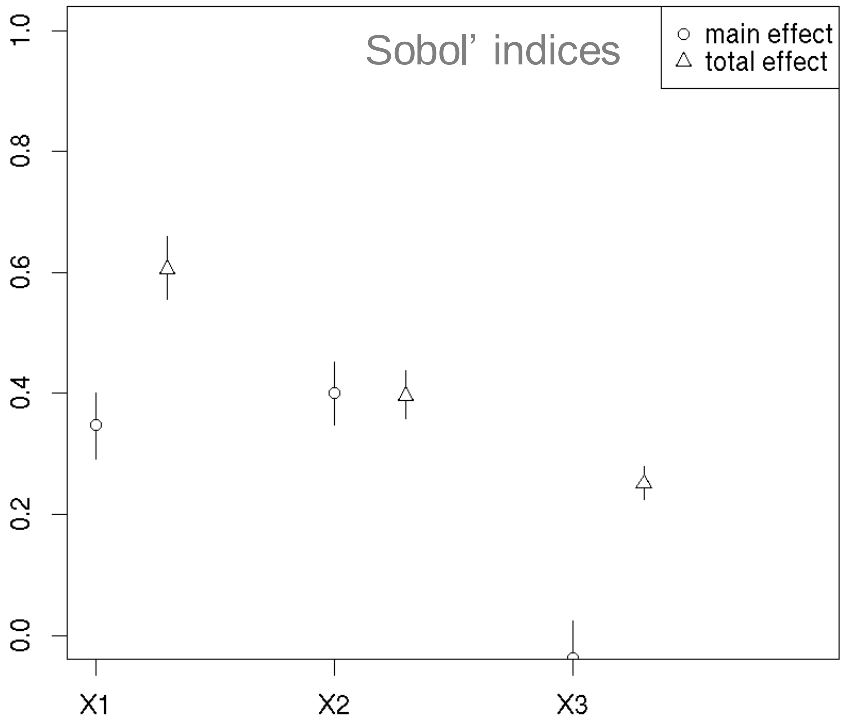
$$\Phi^{-1} \circ F_{X_i}\left(x_i^{(1)}, \dots, x_i^{(N)}\right) = \left({x'}_i^{(1)}, \dots, {x'}_i^{(N)}\right) \sim \mathcal{N}(0,1)$$

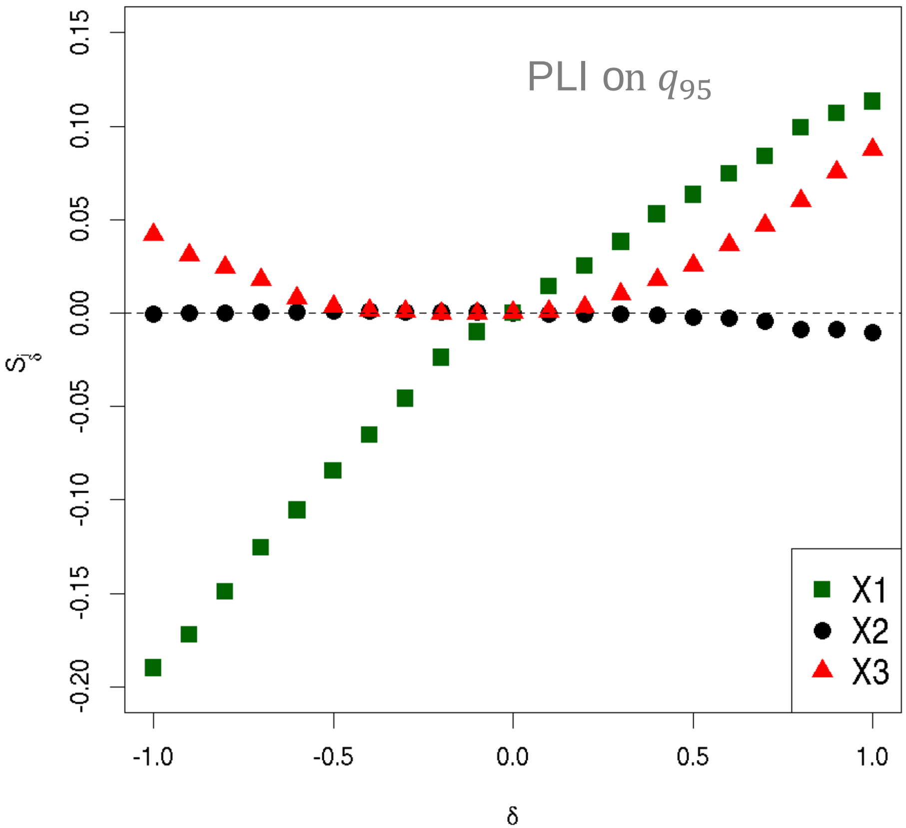
- PLI indices can then be easily determined

Applications

A1) ANALYTICAL EXAMPLE

$$G(X) = sin(X_1) + 7 * sin^2(X_2) + 0,1 * X_3^4 * sin(X_1)\ ; \quad X_i \sim U(-\pi, \pi)$$ **independent**

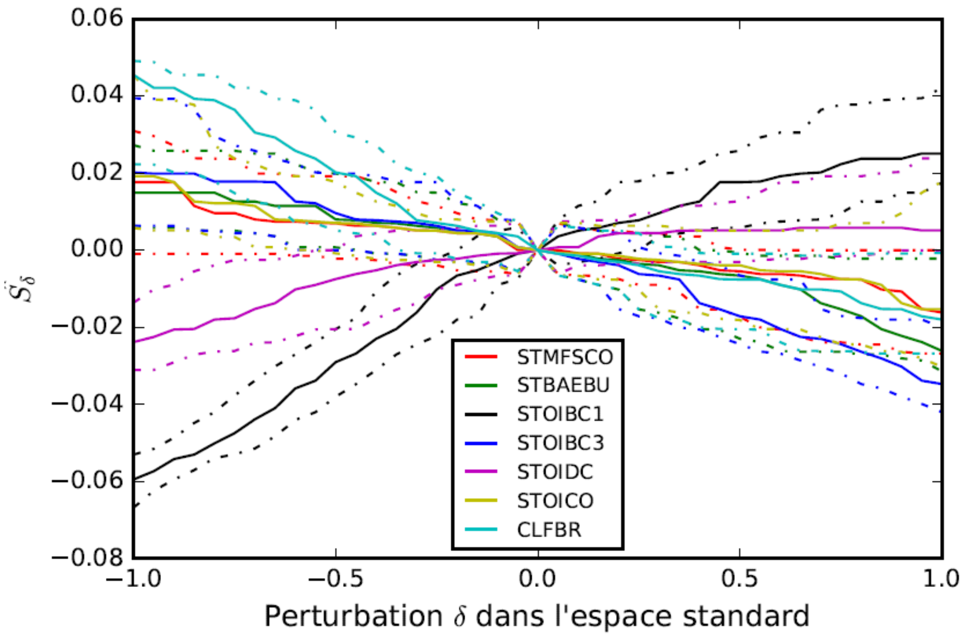




The provided information are different

A2) THERMAL-HYDRAULIC MODEL

- 27 inputs with truncated Gaussian, log-normal, uniform, log-uniform, triangular pdf
- Monte-Carlo sampling of 1000 runs
- Perturbation on the mean between [-1;1] in the standard space (each input ~$\mathcal{N}(0,1)$)
- Graphs show the PLI of the 7 most influential variables
- 90%-confidence intervals are obtained by bootstrap



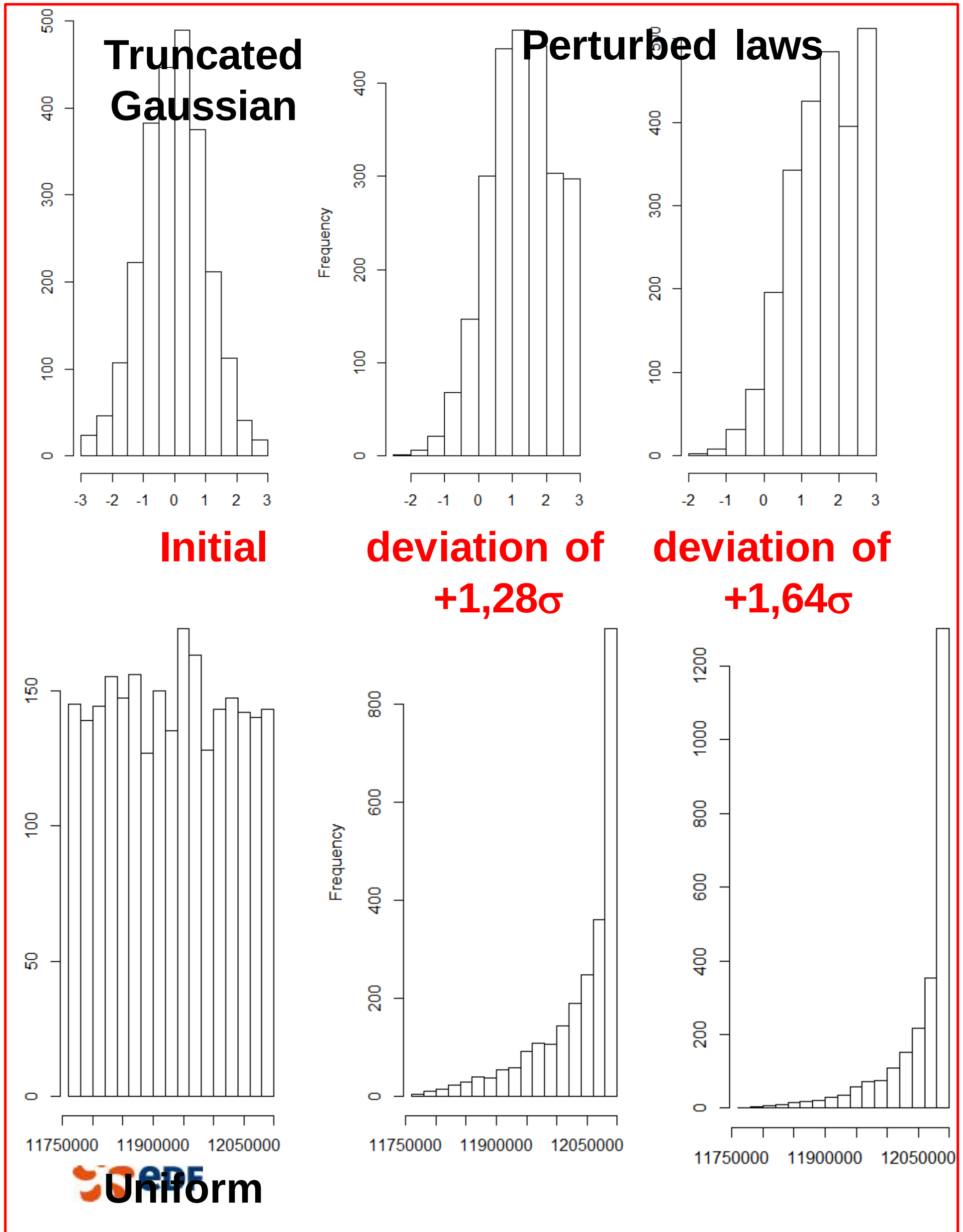
Results

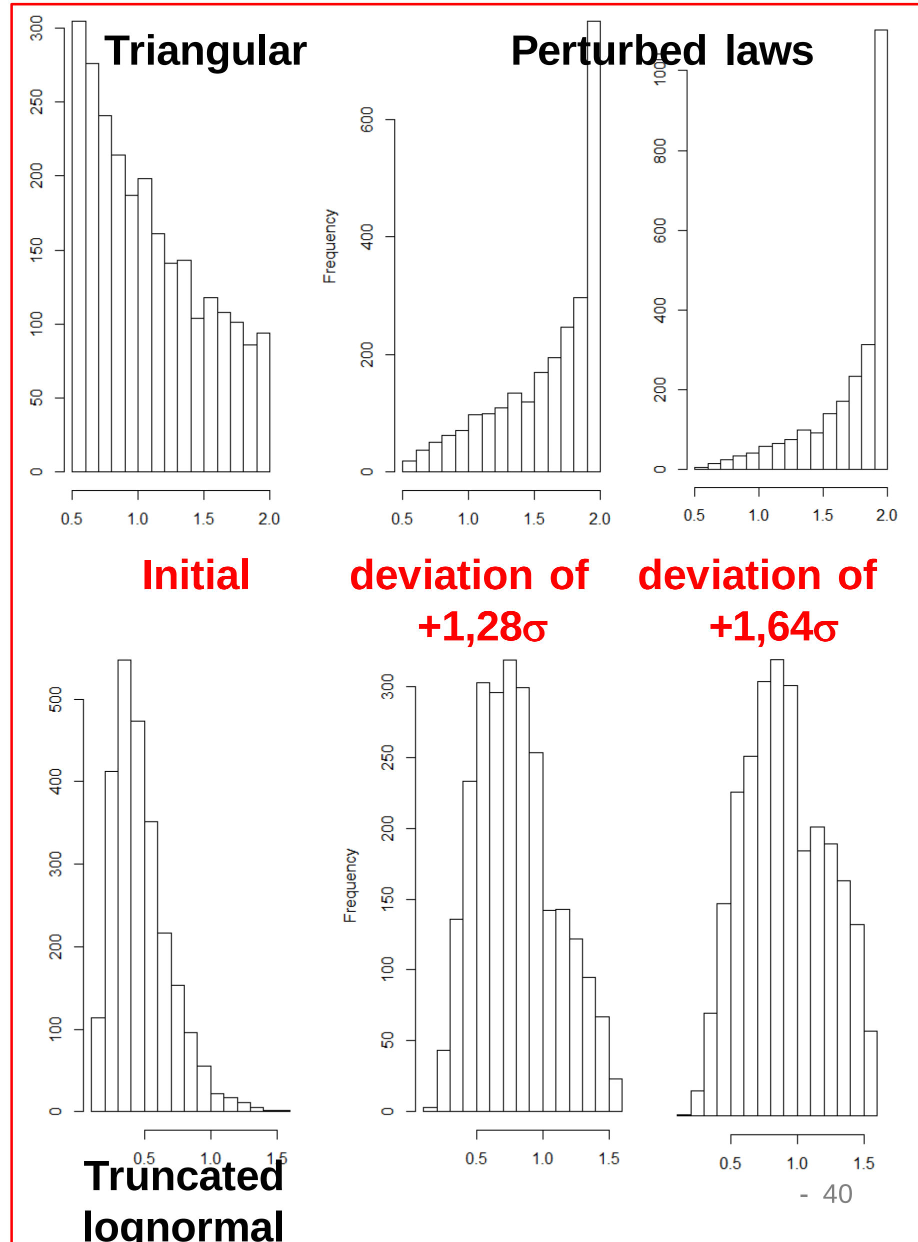
- Quantile seems to be robust towards the pdf: less than 5% variation

- Sign of the PLI allows to know which value allows us to be conservative

- Non-monotonic behaviour (STOIDC)

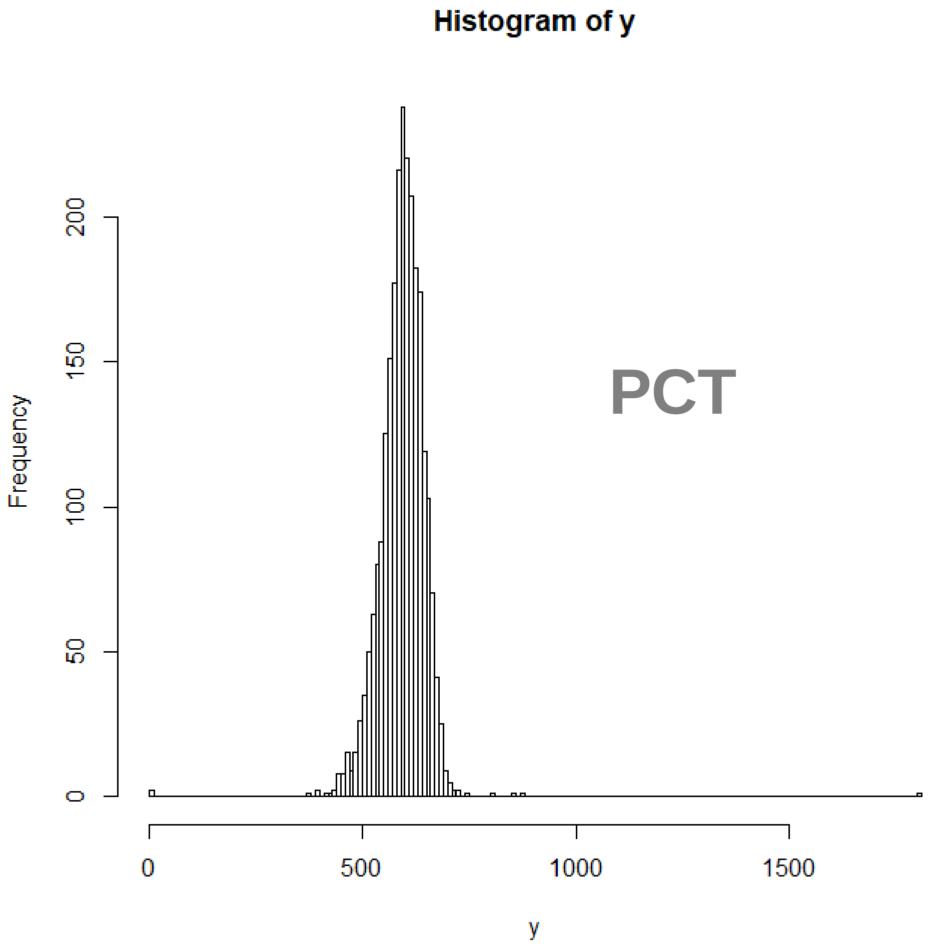
EXAMPLE OF PDF WITH MEAN PERTURBATIONS





A3) REACTOR CASE STUDY

2500 Monte Carlo runs

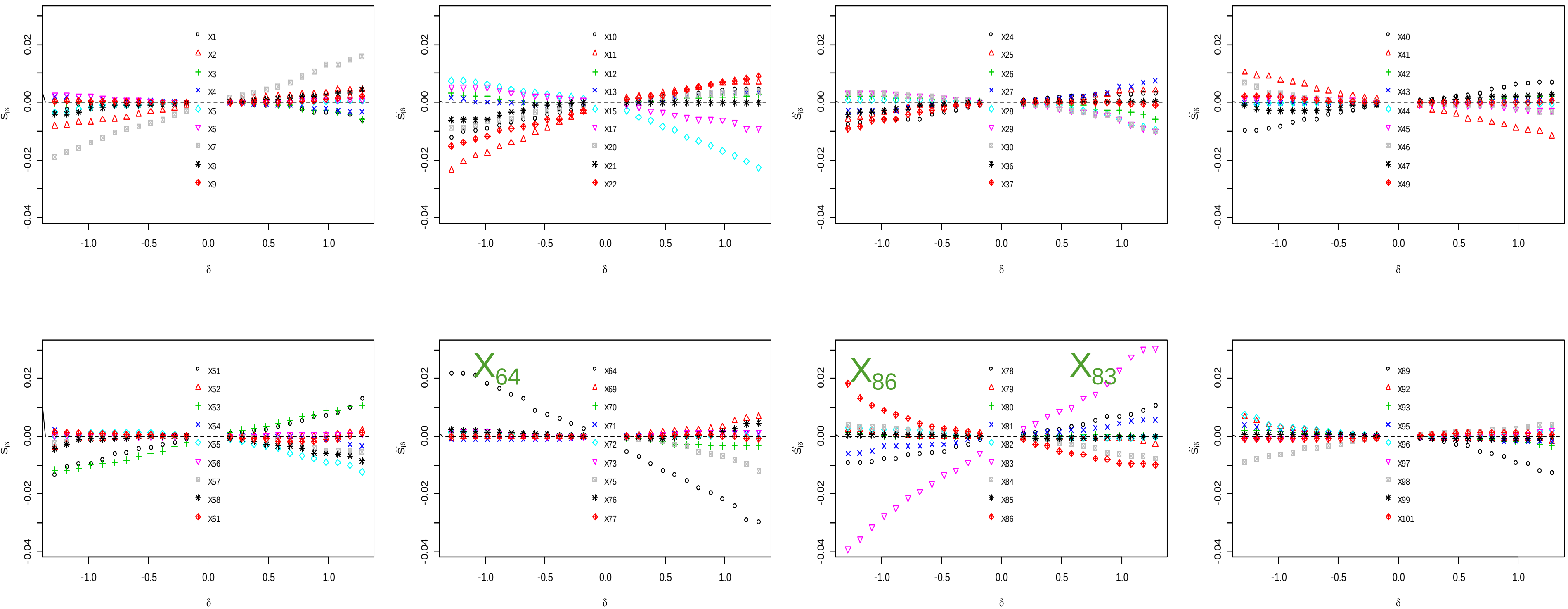


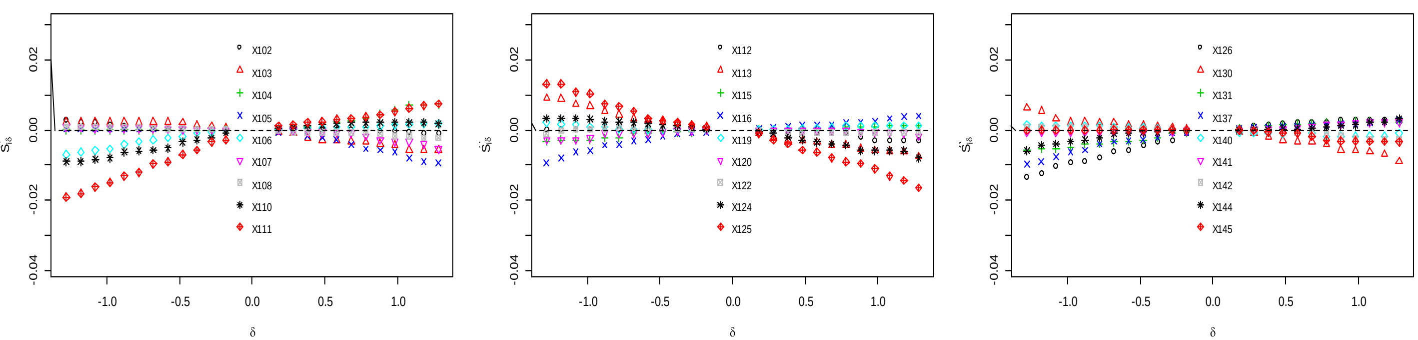
Quantiles:

q_{95} = 665,6°C

$q_{95/95}$ = 667,4°C

PLI-q95 with perturbation of the mean (=> 1,28σ)





X_{64} = clad heat transfer coef.
$=X_{65}$ $=X_{66}$
X_{83} = residual power
X_{86} = creep velocity

Conclusions: - robustness is demonstrated (2% of max. deviation)
- identification of influential inputs on q_{95}

Conclusion

CONCLUSION: BENEFITS OF PLI

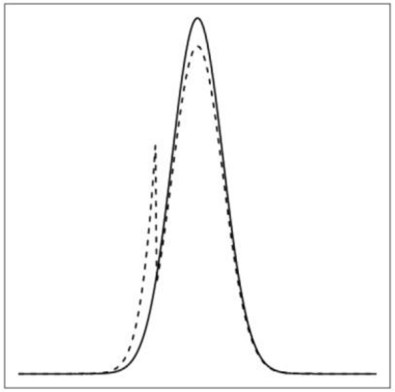
- Allows to quantify the robustness of a quantile of a model output wrt uncertainty on inputs' pdf parameters (mean and variance)
- Confidence Intervals (CLT for probability of exceedance, bootstrap for quantile) allows to adjust the calculation budget (number of runs of the G code)
- No need of new runs of the G code and the input dimension is not an issue
- Easy to perturb several inputs at the same time
- Easy to develop PLI for other risk measures, e.g. superquantile $E(Y|Y \geq q_\alpha)$: promising because it seems more stable

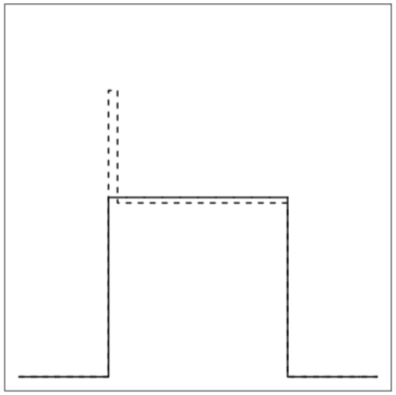
CONCLUSION: CURRENT WORKS AND PERSPECTIVES

- Interpretation of the results (standard space → physical space), maybe by redefining δ as a distance metric between pdf
- CLT for quantile-PLI and improving the estimation of $F_{i\delta}$
- Perturbation in case of statistically dependent inputs
- Perturbation of other quantities than the min and variance:
  - minimal and maximal bounds of the pdf
  - quantile

CONCLUSION: CURRENT WORKS AND PERSPECTIVES

- Interpretation of the results (standard space → physical space), maybe by redefining δ as a distance metric between pdf
- CLT for quantile-PLI and improving the estimation of $F_{i\delta}$
- Perturbation in case of statistically dependent inputs
- Perturbation of other quantities than the min and variance:
  - minimal and maximal bounds of the pdf
  - quantile





=> better behaviour with other risk measures as superquantile, expectile, …

REFERENCES

P. Lemaître. *Sensitivity analysis in structural reliability*, PhD Thesis, Université de Bordeaux I, 2014

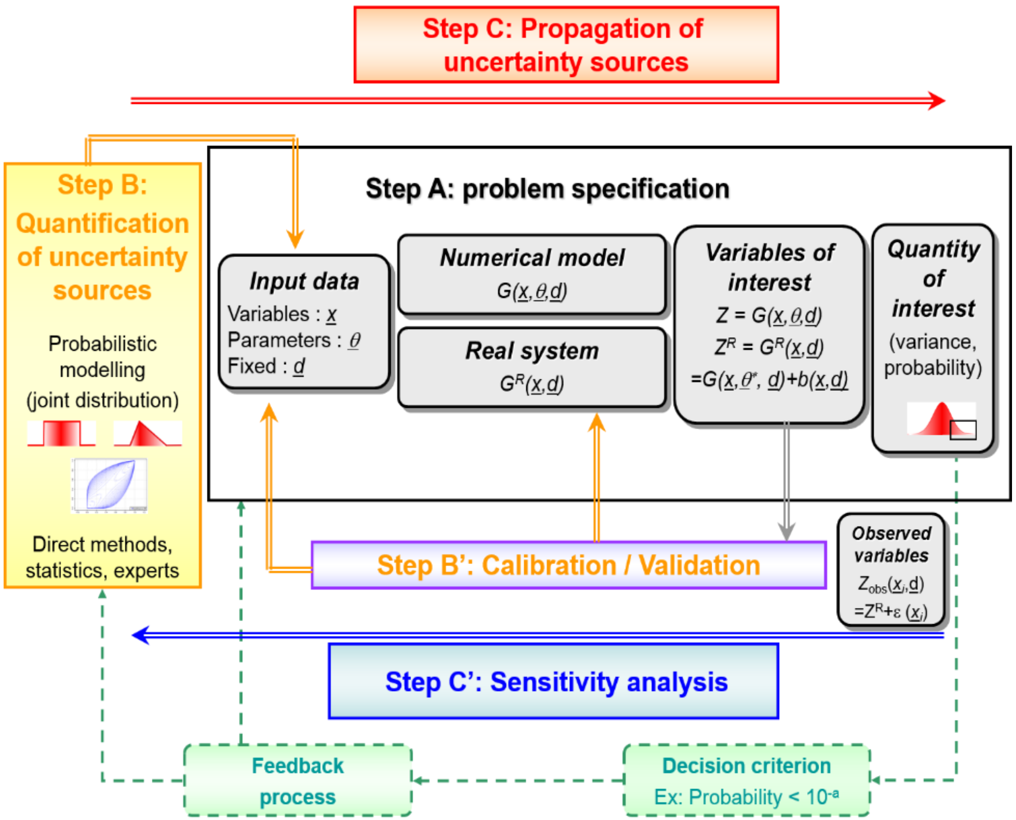
P. Lemaître, E. Sergienko, A. Arnaud, N. Bousquet, F. Gamboa and B. Iooss. Density modification based reliability sensitivity analysis. *Journal of Statistical Computation and Simulation*, 85 :1200-1223, 2015

J. Bect, R. Sueur, A. Gerossier, L. Mongellaz, S. Petit and E. Vazquez (2015), Echantillonnage préferentiel et métamodèles : méthodes bayésiennes optimale et défensive, *47emes Journees de Statistique de la SFdS*, Lille, France

R. Sueur, B. Iooss and T. Delage. Sensitivity analysis using perturbed-law based indices for quantiles and application to an industrial case, *10th International Conference on Mathematical Methods in Reliability (MMR 2017),* Grenoble, France, July 2017

Software implementation: R package 'sensitivity' , which also includes all global sensitivity analysis methods

Thanks - OpenTURNS software for UQ



Baudin, Dutfoy, I. and Popelin. Open TURNS: An industrial software for uncertainty quantification in simulation. In: Handbook of uncertainty quantification, Springer, 2017

Features

- Licence : LGPL
- Linux, Windows
- B, B', C, C'
- Metamodels & optimization framework
- Multi-thread evaluation of an analytical formula
- Distributed and multi-thread evaluation of a Python function
- www.openturns.org
- Programming: Python module, C++ library, GUI via SALOME

End

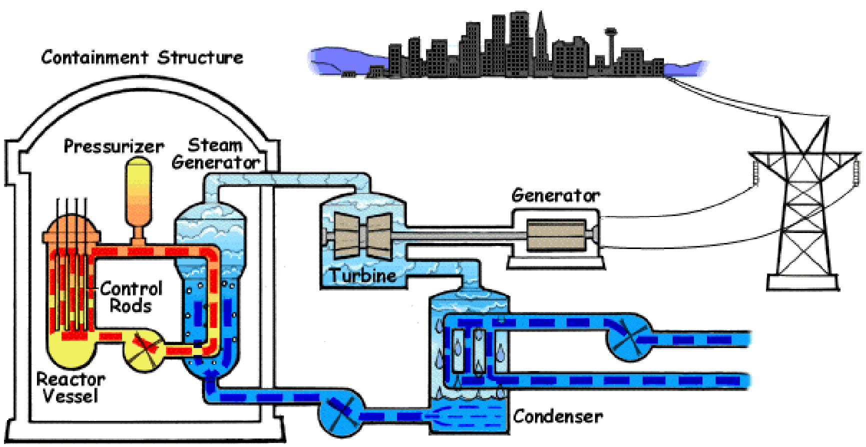
A4) NON DESTRUCTIVE CONTROL (NDC) MODEL

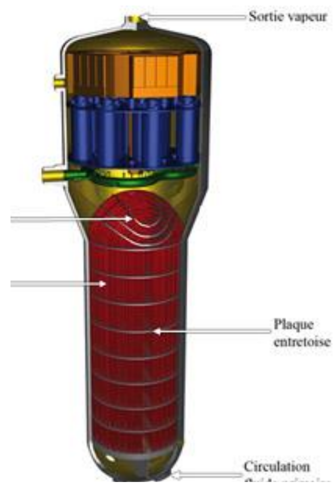
Problem: evolution of the performance qualification of NDC process
* including additional statistical information
*** time saving (experimental and engineering)**

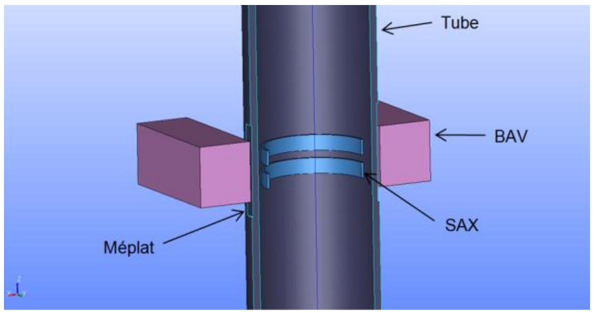
Results: methodology of evaluation and exploitation of PoD (Proba of Detection) curves

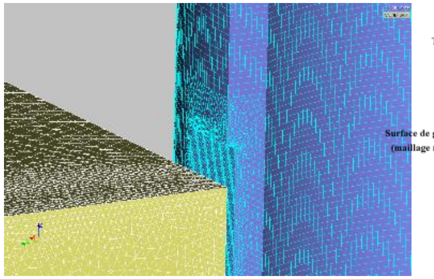
Strategy: « best-estimate » simulations + statistical methods

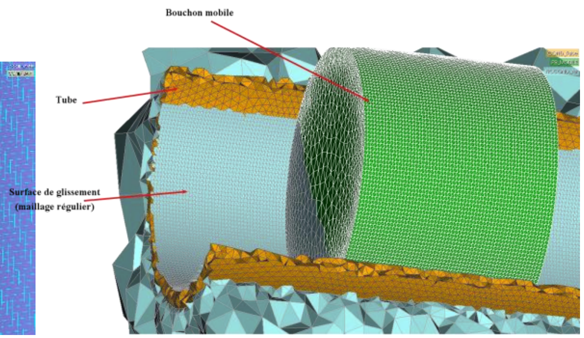
Industrial application: steam generator tubes controlled by eddy current NDE

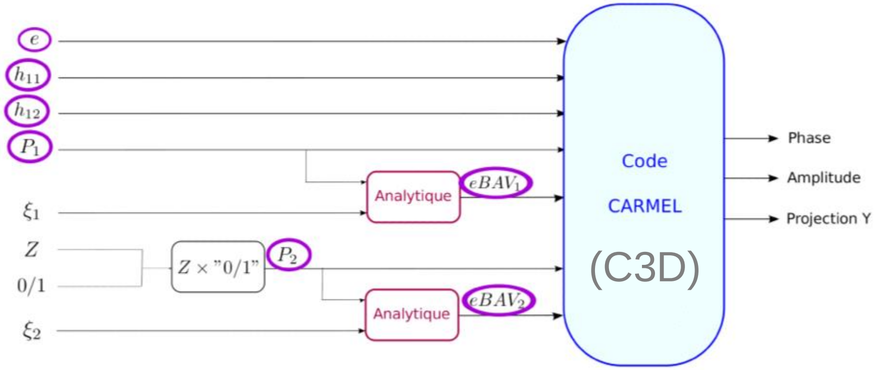










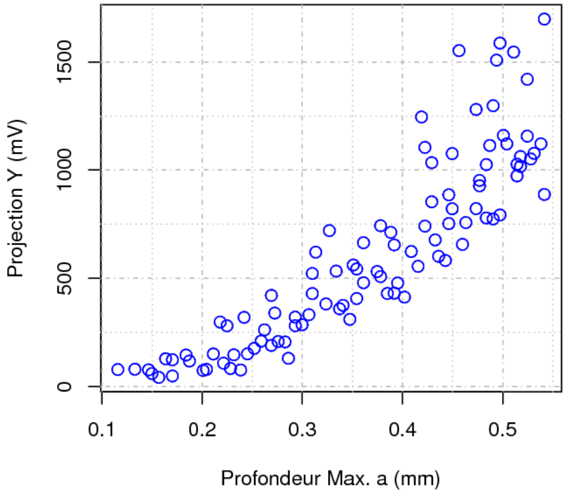


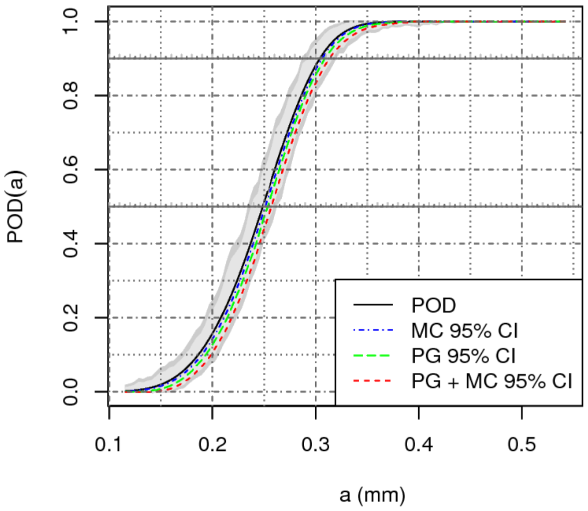
FUNCTIONAL RISK CURVES

A PoD curve gives an event probability (here the detection proba) in function of a characteristic parameter *a* (here the size of the flaw)

[Le Gratiet et al. 2017]

<u>Strategy:</u> Estimation with qualified computer models (C3D)





<u>Methodology:</u>

1) **Expert meetings: Identify potentially influential uncertain parameters *X* and their pdf**

2) **Exp. design (« space filling » & sequential): 100 code runs**

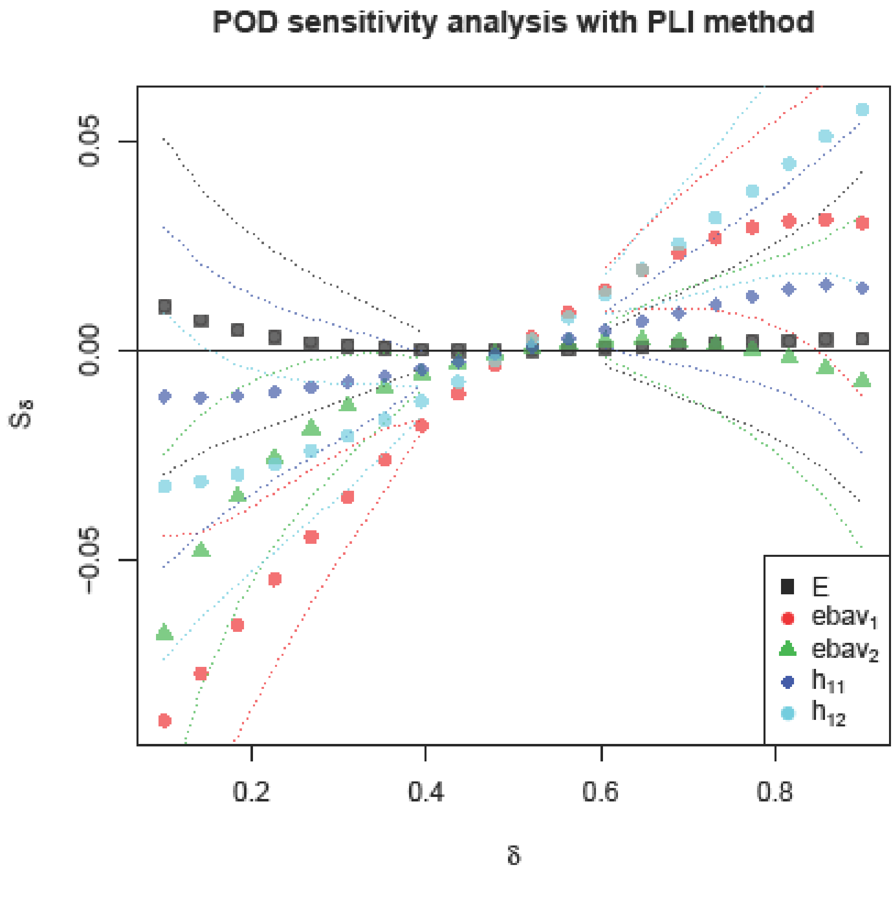
3) **Metamodels fitting, PoD curves estimation, sensitivity analysis**

PLI-PROBABILITY

Robustness analysis with respect to the pdfs of the input parameters

Translation of the mean of each input

***a* is fixed**



PLI OF FUNCTIONAL RISK CURVES