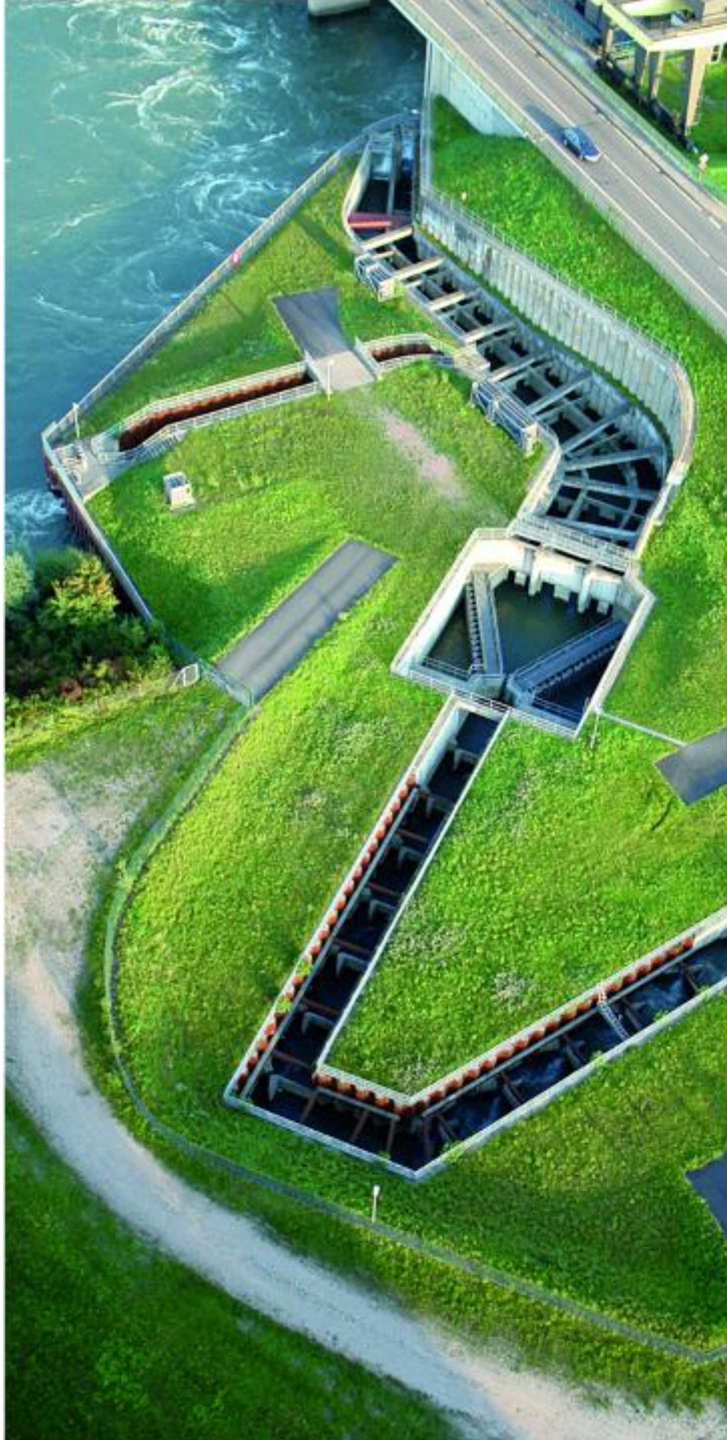
Increasing conservatism and robustness in uncertainty quantification-based safety studies using advanced statistical tools

Bertrand IOOSS

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

with Vincent Chabridon (EDF R&D) and Amandine Marrel (CEA/DES)

4th IPSMNA conference, November, 2d, 2022



Industrial motivations

Simulation of IB-LOCA accident for safety studies

IB-LOCA: *Intermediate break loss of coolant accident*

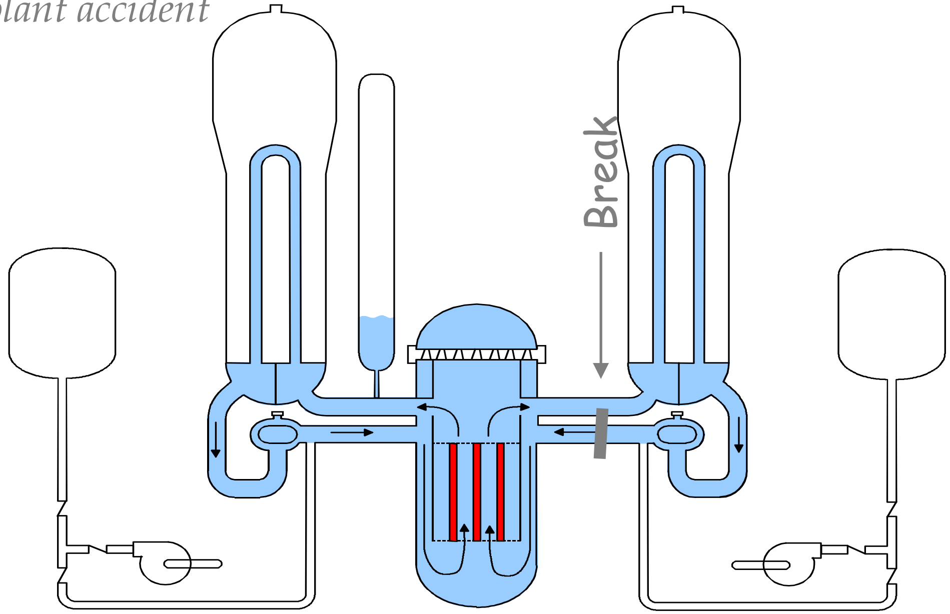
Pressurized Water Reactor scenario:
Loss of primary coolant accident due to a break in cold leg

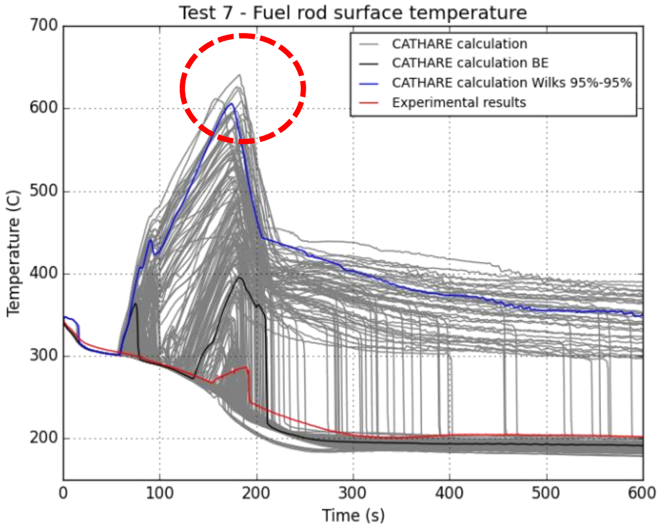
***d* (~ 100) uncertain input variables *X*:**
Critical flowrates, initial/boundary conditions, phys. eq. coef., …
Probabilistic modeling

Modelled using **CATHARE2 code**:
(thermal-hydraulic phenomena)

CPU cost for one code run > 1 hour
In industrial studies: *N* ~ O(1000) runs

Variables of Interest :
Cladding temperature (fct of time)
Peak Cladding Temperature (PCT)





100 output transients corresponding to 100 (Monte Carlo) runs

Safety issues in EDF engineering

Goals: Assess margins with regards to a regulatory criteria (the regulator will accept the safety approach if a sufficient margin remains)

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) New requirements:

Safety authorities: higher demands in terms of margin & realistic/complex physics

Operator: better control of margins (due to ageing) for better resources allocation & better maneuverability

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

3) Objectives: better assessment of the real margins

=> **UQ** (Uncertainty Quantification) approach: **realistic models & uncertain inputs**

CHALLENGES

UQ approaches are well (and naturally) developed in the **probabilistic framework** (needing to define pdf of the inputs) - The variable of interest (output) is then random

Importance of the choice of the quantity of interest (QoI):

- - High quantile (95% to 99%) allowing to compute a safety margin (by comparing the quantile value with the critical threshold)

Key points:

Limited number of model runs

High dimension issues

- large CPU time cost of the simulator

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

The French nuclear regulatory authority asks to **justify the probabilistic approach** for such inputs

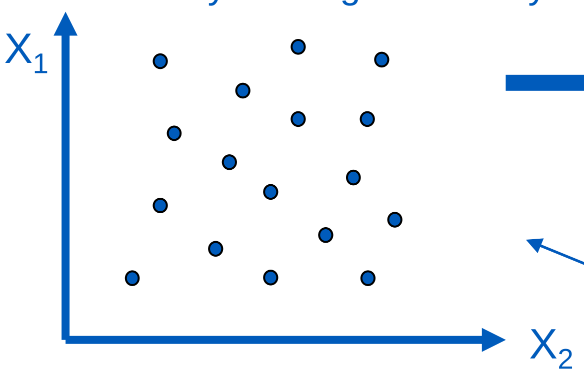
=> Robustness of the study results towards the input distributions

UQ and computer experiments

Computer experiments

Numerical design of experiments → Simulation → Output analysis

Uncertainty & design variability

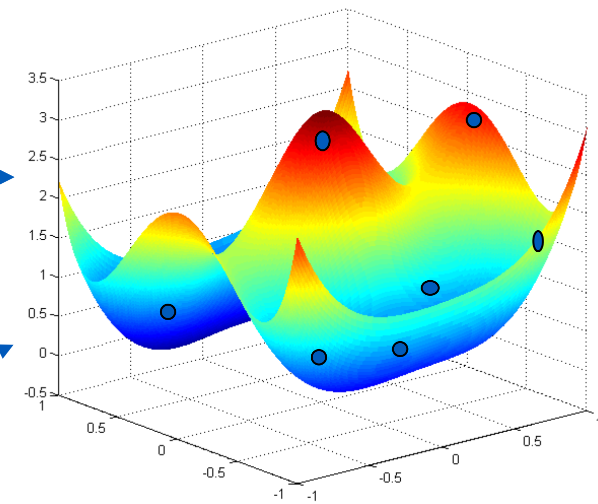


X is defined by its pdf

Computer code

$Z = G(X)$

(potential) Integration of data

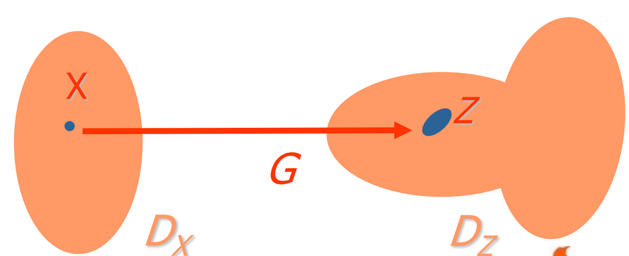


- Methods from proba/stats, numerics, scientific computing, machine learning, ...

- Validation, Verification and Uncertainty Quantification

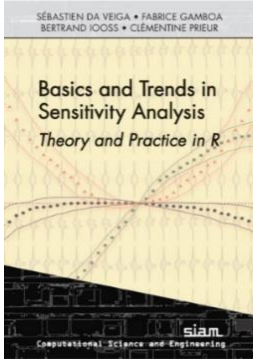
- Global Sensitivity Analysis (GSA)

How an input or a group of input apportion into the uncertainty of the output?



EDF

Global sensitivity analysis: Main settings



1. **Understand the behaviour of the model output wrt inputs**

Exploration

2. **Simplify the computer model (dimension reduction)**

Determine the non-influential variables (that can be fixed)

Screening

3. **Prioritize the uncertainty sources**

Quantitative partitioning

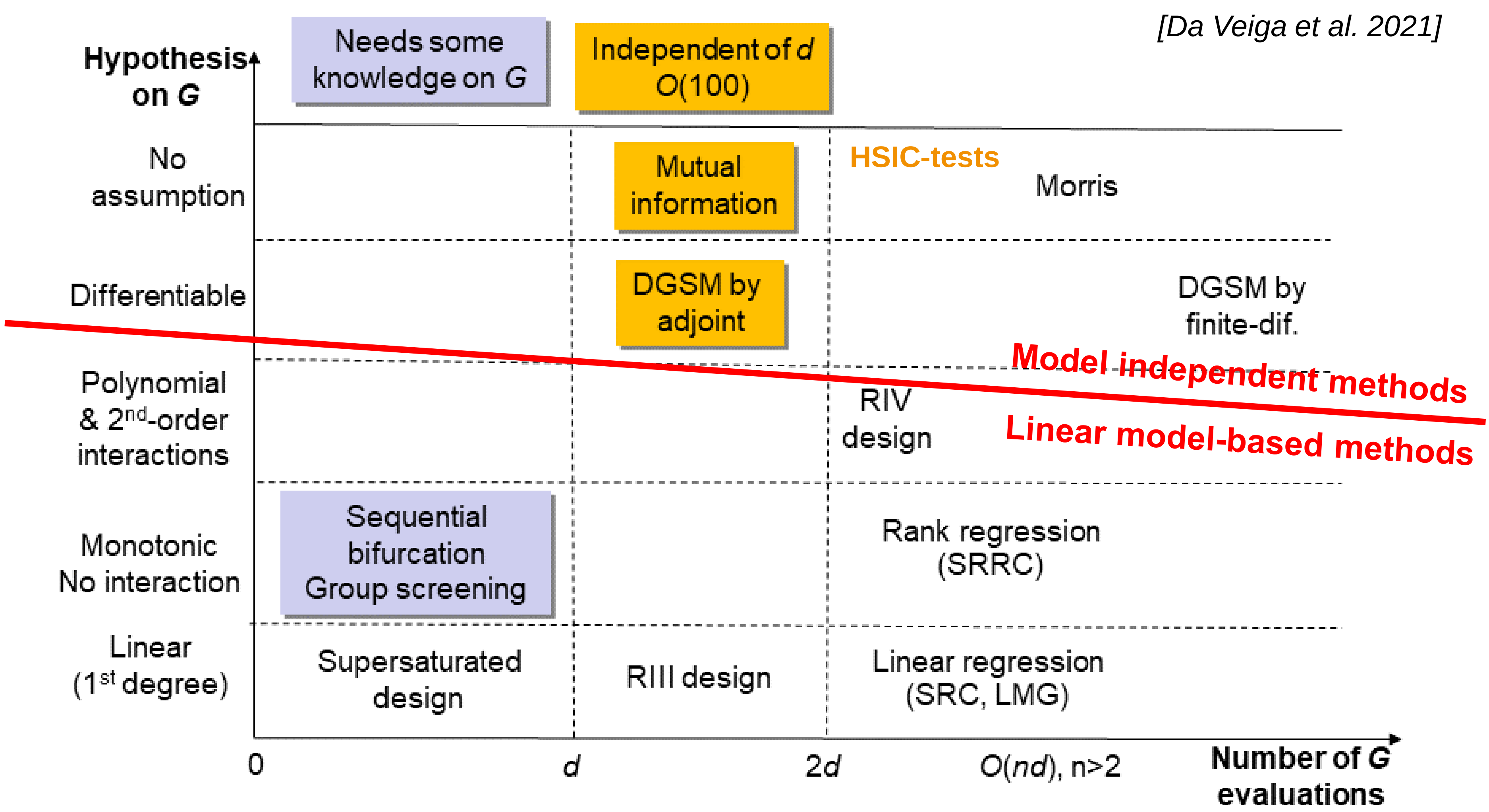
Variables to be fixed to obtain the **largest output uncert. reduction**

4. Analyze the robustness of the quantity of interest (QoI) wrt the input probability distributions

Robustness analysis

Screening for high-dimension issue

Classification of screening methods



1st (advanced) tool: HSIC for screening

[Gretton et al. 2005]

Nonparametric dependence measure: relying on a dissimilarity measure between **joint** probability distribution of (X_i,Z) and **product of marginals**

Fundamental property: $\boldsymbol{HSIC(X_i,Z)_{\mathcal{F}_{X_i},\mathcal{G}} = 0 \;\Leftrightarrow\; X_i \perp Z}$

➢ **Hilbert-Schmidt independence criterion (HSIC):** based on RKHS framework

$$HSIC(X_i,Z)_{\mathcal{F}_{X_i},\mathcal{G}} = \sup_{f_{X_i}\in\mathcal{F}_{X_i},\, g\in\mathcal{G}} Cov(f_{X_i}(X_i), g(Z))$$

➢**Use of the kernel trick** (considering RKHS $\mathcal{F}_{X_i}$ (resp. $\mathcal{G}$) with characteristic kernel k_{X_i} (resp. k_Z) associated to each X_i (resp. to Z), and the RKHS product for (X_i, Z)

$$HSIC(X_i,Z)_{\mathcal{F}_{X_i},\mathcal{G}} = E\left[k_{X_i}(X_i,X_i')k_Y(Z,Z')\right] + E\left[k_{X_i}(X_i,X_i')\right]E[k_Z(Z,Z')] - 2E[k_{X_i}(X_i,X_i')k_Z(Z,Z'')]$$

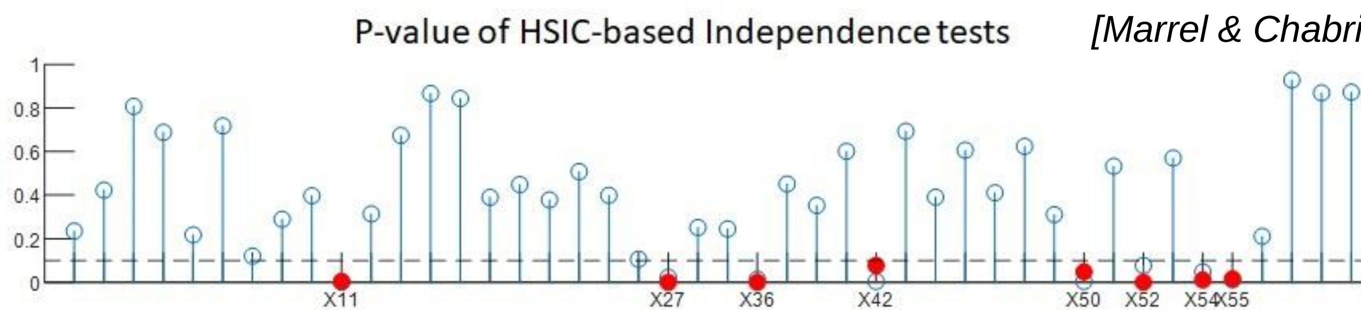
provided that (X_i',Z') is i.i.d. to (X_i,Z), and Z'' is i.i.d. to Z and Z'

In practice: Monte Carlo-based estimators + asymptotic law under $X_i \perp Z$

=> HSIC-based **statistical independence tests for screening**

[Marrel & Chabridon 2021]

Example



Robustness analysis for epistemic uncertainty issue

2nd tool: PLI (Perturbed Law-based Indices)

[Lemaître et al., 2015]
[Iooss et al. 2022]

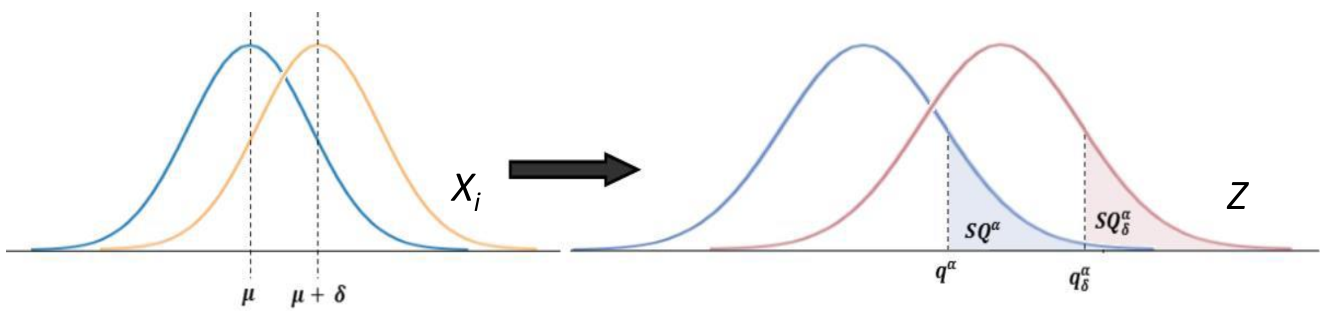
We aim at quantifying the impact of a perturbation on the pdf of X_i

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

We define the PLI-quantiles as : $S_{i\delta} = \left(\frac{q_{i\delta}^{\alpha}}{q^{\alpha}} - 1\right)$ (with q^{α} and $q_{i\delta}^{\alpha}$ the α-quantile of Z and the perturbed quantile)

- It gives results in terms of percentage of perturbations
- $S_{i\delta} = 0$ 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

Example

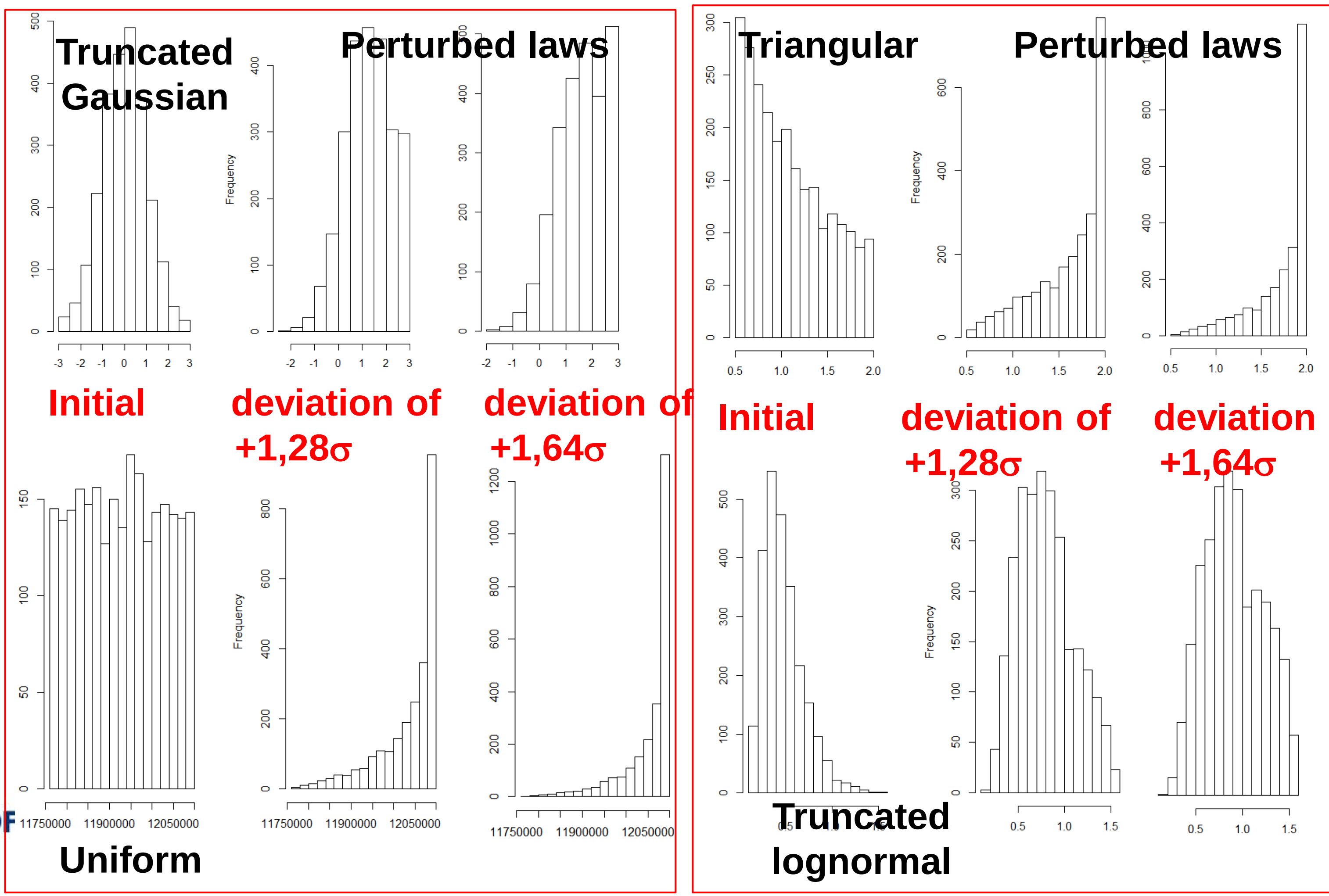


Perturbation of the mean of the pdf of the input X_i

Computation of the deviation of the QoI (quantile of the output Z, …)

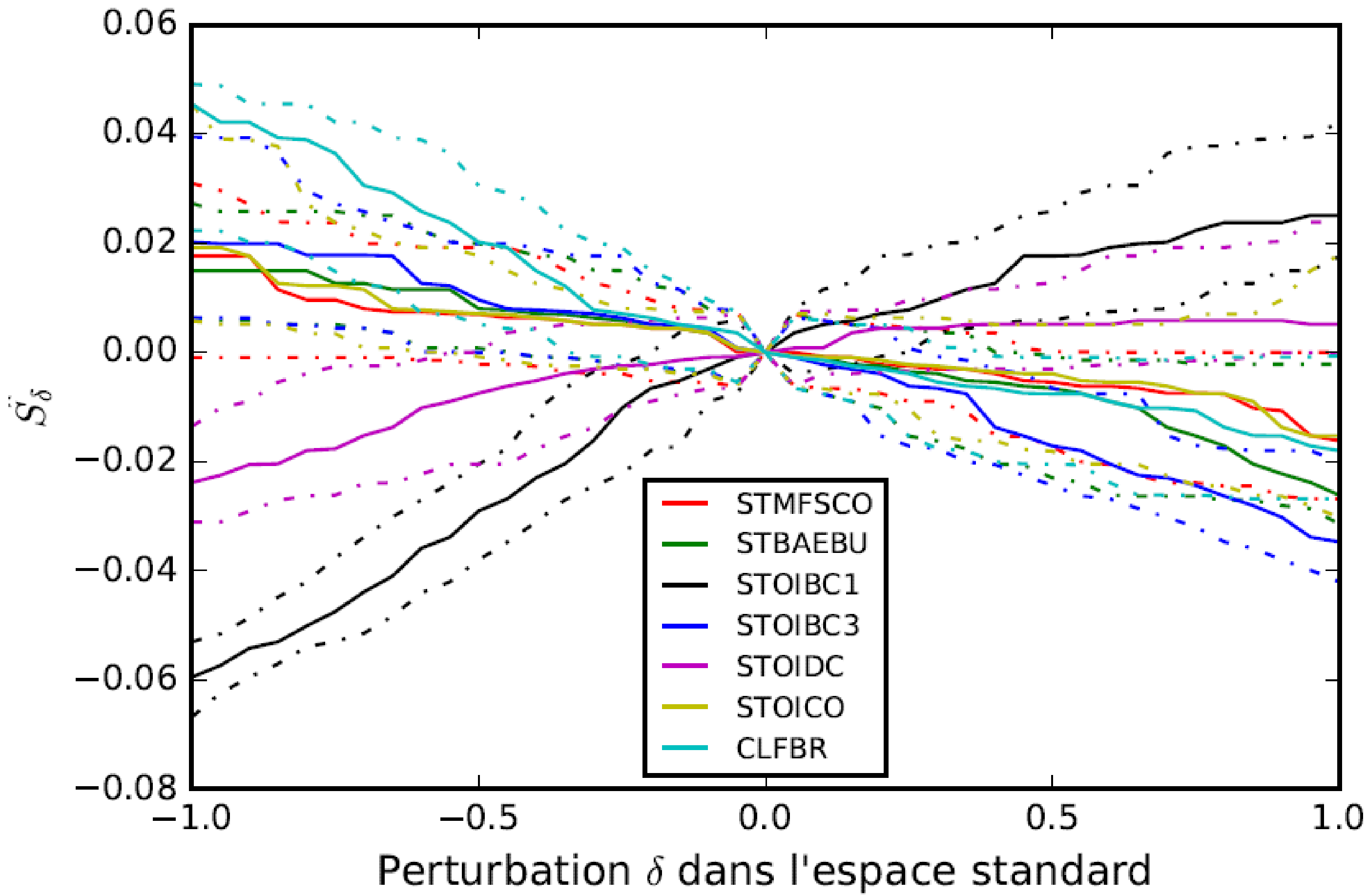
Example: Themal-hydraulic model of a mockup

- 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 $\sim\mathcal{N}(0,1)$)



RESULTS

- Graphs show the PLI of the 7 most influential variables
- 90%-confidence intervals are obtained by bootstrap



Observations: 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

In the statistical part of a safety study, EDF has proposed 3 steps:

1. HSIC for screening 100 inputs to keep around 10 in the next step,
2. PLI for identifying the few (3 or 4) most influential inputs on the QoI,
3. Penalization of these inputs before the final computation of the QoI *(the penalization of all the inputs is avoided)*

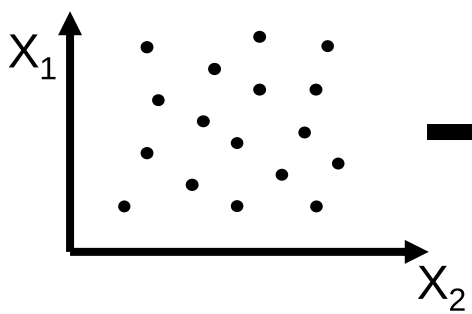
Metamodel for the limited number of model runs issue

Approximation with a metamodel

Replace the code by a statistical function called metamodel

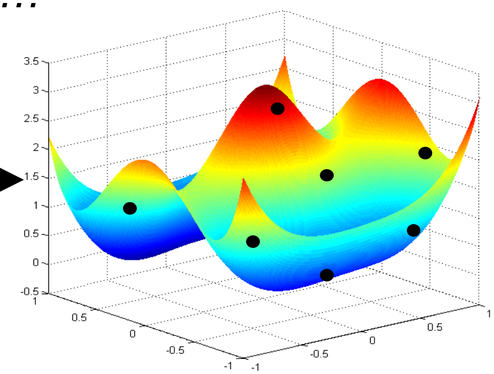
- ✓ With good approximation and prediction capabilities ⇒ to be controled
- ✓ With a negligible cpu cost for prediction
- ✓ Built from a Monte Carlo sample of n experiments ($n \sim 10\ d$)

Ex : Polynomials, splines, neural networks, regression trees…



Computer code Experiments

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

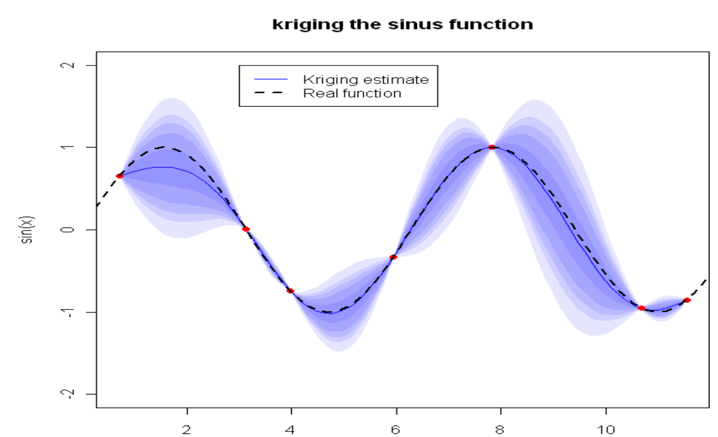


Choice: Gaussian process (GP) metamodel

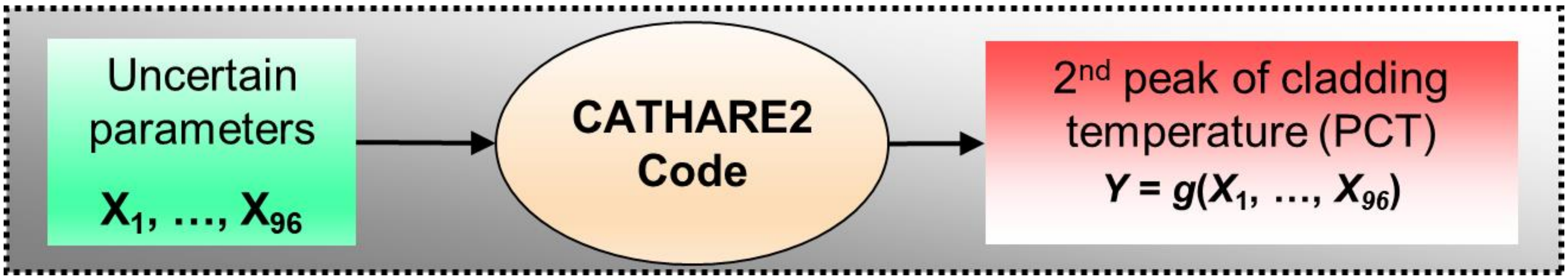
see *Rasmussen & Williams [2005]*

Part of *Supervised Machine Learning*

Advantage: gives a prediction with an associated error bound (Gaussian distribution at each point)



ICSCREAM: Identification of penalizing configurations using Screening and Metamodel



- In IB-LOCA modeling framework, uncertain input parameters are:
  - (Type 1) initial/boundary conditions ⇒ probabilistic $(\mathcal{U}, \mathcal{N})$
  - (Type 2) Parameters of physical models ⇒ probabilistic $(\mathcal{U}, \mathcal{LU}, \mathcal{N}, \mathcal{LN})$
  - (Type 3) **Scenario parameters (min / max bounds)** ⇒ **no probabilistic**

ICSCREAM final objective

Identify the most **penalizing configurations** for **Type 3** inputs, regardless to the uncertainties of **Types 1 & 2** inputs.

Penalizing configurations ⇔ leading to high PCT values

3[d] tool: Uncertainty propagation with GP

[Marrel et al. 2022]

A GP metamodel Y_{Gp} has been fitted as a function of the HSIC-detected influential inputs from a 1000-size learning sample

Solving a difficult inversion problem: Uncertainty propagation with GP metamodel to identify the penalizing values of $\boldsymbol{X_{pen}}$ under the uncertainty of the other inputs $\{\boldsymbol{X} \backslash \boldsymbol{X_{pen}}\}$

⇒ Precisely capture critical configurations of $\boldsymbol{X_{pen}} = \{X_{127}, X_{143}\}$ which lead to the highest probability of PCT > $\hat{q}_{0.9}(Y)$ (under randomness of the other variables)

$$
\begin{aligned}
\hat{P}(\mathbf{X_{pen}}) &= P[Y_{Gp}(\mathbf{X_{exp}}) > \hat{q}_{0.9} \,|\mathbf{X_{pen}}] \\
&= 1 - \mathbb{E}(1_{Y_{Gp}(\mathbf{X_{exp}}) \leq \hat{q}_{0.9}} |\mathbf{X_{pen}}) \\
&= 1 - \mathbb{E}(1_{Y_{Gp}(\tilde{\mathbf{X}}_{\mathbf{exp}}, \mathbf{X_{pen}}) \leq \hat{q}_{0.9}} |\mathbf{X_{pen}}) \\
&= 1 - \mathbb{E}(\mathbb{E}(1_{Y_{Gp}(\tilde{\mathbf{X}}_{\mathbf{exp}}, \mathbf{X_{pen}}) \leq \hat{q}_{0.9}} |\tilde{\mathbf{X}}_{\mathbf{exp}}) |\mathbf{X_{pen}}) \\
&= 1 - \int_{\mathscr{X}_{exp}} \Phi\left(\frac{\hat{q}_{0.9} - \hat{Y}_{Gp}(\tilde{\mathbf{x}}_{\mathbf{exp}}, \mathbf{X_{pen}})}{\sqrt{MSE[\hat{Y}_{Gp}(\tilde{\mathbf{x}}_{\mathbf{exp}}, \mathbf{X_{pen}})]}} \right) d\mathbb{P}_{\tilde{\mathbf{X}}_{\mathbf{exp}}}(\tilde{\mathbf{x}}_{\mathbf{exp}})
\end{aligned}
$$

$\tilde{\boldsymbol{X}}_{\boldsymbol{exp}} = \boldsymbol{X_{exp}} \setminus \boldsymbol{X_{pen}}$

$\tilde{\boldsymbol{X}}_{\boldsymbol{exp}}$ and $\boldsymbol{X_{pen}}$ are independent

Variation domain of $\tilde{\boldsymbol{X}}_{\boldsymbol{exp}}$

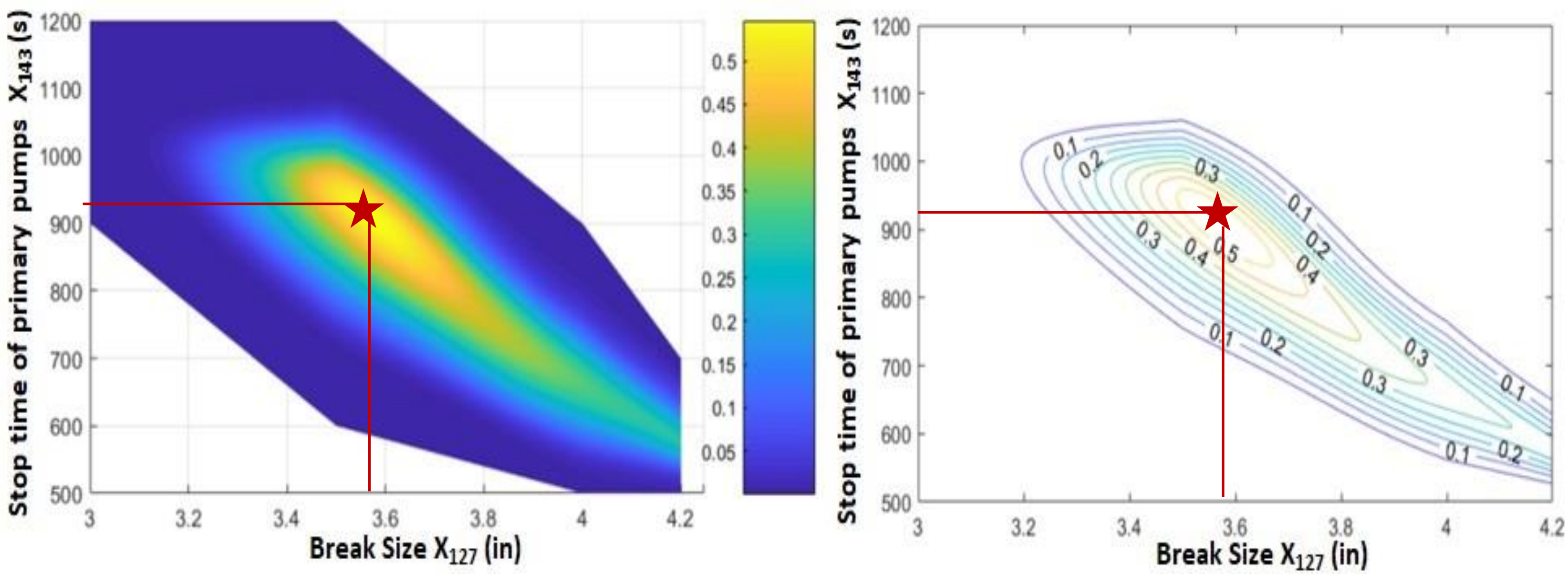
Joint distribution of $\tilde{\boldsymbol{X}}_{\boldsymbol{exp}}$

- In practice, for each value of $\boldsymbol{X_{pen}} = \{X_{127}, X_{143}\}$, computation of $\hat{P}(\boldsymbol{X_{pen}})$ by intensive Monte-Carlo computation (integral in dimension 18)

Application on IBLOCA test case

Computation of $\hat{P}(\boldsymbol{X}_{pen})$

Probability of exceeding $\hat{q}_{0,9} = 673.18°C$, according to X_{127} and X_{143}



► Strong interaction between the two scenario parameters

► Worst case: (3.57 inches, 907.8 seconds) $\Rightarrow \hat{P} \approx 0.55$

► Physical explanation: these two parameters drive the degradation of the water inventory

- The smaller X_{127}, the longer the pump will have to run for the same inventory degradation
  - If $X_{127} < 3.3 \Rightarrow$ the water inventory does not degrade too much (whatever GMPP)
- If $X_{127} > 3.9 \Rightarrow$ break tends to be prevailing and reduces the impact of stop time of GMPP

References

Marrel and V. Chabridon, Statistical developments for target and conditional sensitivity analysis: application on safety studies for nuclear reactor, *Reliability Engineering and System Safety*, 214, 107711, 2021

S. Da Veiga, F. Gamboa, B. Iooss and C. Prieur. *Basics and trends in sensitivity analysis: Theory and practice in R*, SIAM, 2021

B. Iooss, V. Vergès and V. Larget, BEPU robustness analysis via perturbed-law based sensitivity indices, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 236:855-865, 2022

A. Marrel, B. Iooss and V. Chabridon, The ICSCREAM methodology: Identification of penalizing configurations in computer experiments using screening and metamodel – Applications in thermal-hydraulics, *Nuclear Science and Engineering*, 196:301-321, 2022

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