



HAL
open science

The ICSCREAM methodology: identification of penalizing configurations in computer experiments using screening and metamodel applications in thermal-hydraulics

Amandine Marrel, Bertrand Iooss, Vincent Chabridon

► To cite this version:

Amandine Marrel, Bertrand Iooss, Vincent Chabridon. The ICSCREAM methodology: identification of penalizing configurations in computer experiments using screening and metamodel applications in thermal-hydraulics. UQ22 - SIAM Conference on Uncertainty Quantification, SIAM, Apr 2022, Atlanta, United States. cea-03700747

HAL Id: cea-03700747

<https://cea.hal.science/cea-03700747>

Submitted on 21 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Speaker: Amandine Marrel

MS67 - The ICSCREAM Methodology: Identification of Penalizing Configurations in Computer Experiments using Screening and Metamodel Applications in Thermal-Hydraulics

Amandine Marrel

CEA, DES, IRESNE, DER, Cadarache F-13108 Saint-Paul-Lez-Durance,
amandine.marrel@cea.fr

Bertrand Iooss

EDF R&D, 6 quai Watier, 78400, Chatou, France, bertrand.iooss@edf.fr

Vincent Chabridon

EDF R&D, 6 quai Watier, 78400, Chatou, France, vincent.chabridon@edf.fr

In the framework of risk assessment in nuclear accident analysis, best-estimate computer codes, associated to a probabilistic modeling of the uncertain input variables, are used to estimate safety margins. A first step is often to identify the critical configurations (or penalizing, in the sense of a prescribed safety margin) of several input parameters (called “scenario inputs”), under the uncertainty on the other input parameters. However, the large CPUtime cost of most of the codes involves to develop highly efficient strategies. To achieve it with a very large number of inputs and from a small-size sample of simulations, a specific and original methodology, called ICSCREAM (Identification of penalizing Configurations using SCREening And Metamodel), has been proposed. The screening of influential inputs is based on an advanced global sensitivity analysis indices, namely the Hilbert-Schmidt Independence Criterion. Then, a Gaussian process metamodel is sequentially built and used to estimate, within a Bayesian framework, the conditional probabilities of exceeding a high-level threshold, according to the scenario inputs. The efficiency of this methodology is illustrated on a high-dimensional (hundred inputs) use case simulating an accident of primary coolant loss in a pressurized water reactor. The study focuses on the peak cladding temperature (PCT) and critical configurations are defined by exceeding the 90%-quantile of PCT.



ICSCREAM* METHODOLOGY

**Identification of penalizing Configurations
using SCREening And Metamodel*

DE LA RECHERCHE À L'INDUSTRIE

Amandine MARREL (CEA DES/IRENE/DER)

Bertrand IOOSS (EDF R&D)

Vincent CHABRIDON (EDF R&D)

SIAM UQ 2022

IRENE | DER | SESI | LEMS

Institut de recherche sur les systèmes nucléaires pour la production d'énergie bas carbone

Commissariat à l'énergie atomique et aux énergies alternatives - www.cea.fr



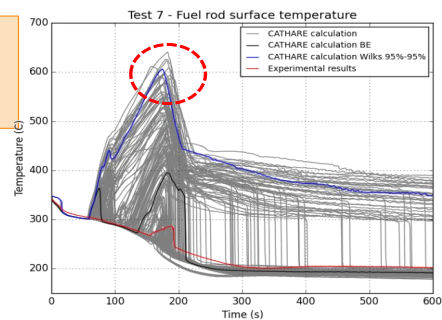
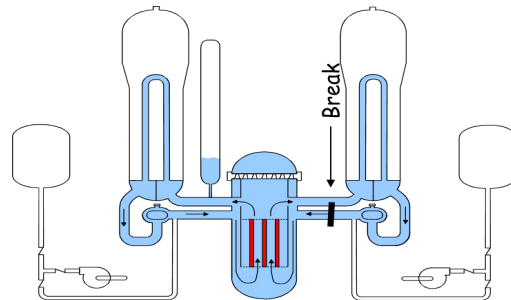
Simulation of IB-LOCA nuclear accident

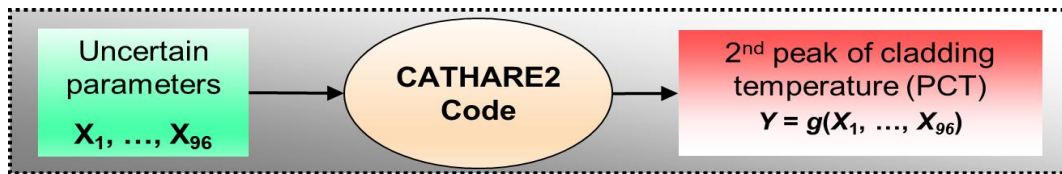
Accidental scenario on pressurized water reactor: IB-LOCA
LOss of primary **Coolant Accident** due to a **I**ntermediate **B**reak in cold leg

d (~ 100) input random variables:
Critical flowrates, initial/boundary conditions, phys. eq. coef., ...

Modelled with CATHARE2 code:
- Models complex thermal-hydraulic phenomena
- **Large CPU cost for one code run (> 1 hour)**

Variable of Interest:
2nd peak of cladding temperature (PCT)
= scalar output





- In IB-LOCA modeling framework, uncertain input parameters are:
 - ▶ (Type 1) Initial conditions, physical model parameters ⇒ Probabilistic ($\mathcal{U}, \mathcal{LU}, \mathcal{N}, \mathcal{LN}$)
 - ▶ (Type 2) **Scenario parameters (min / max bounds)** ⇒ **No probabilistic**

Objective in support of safety studies

Identify the most **penalizing configurations** for **Type 2** inputs, under the uncertainties of **type 1** inputs.

Penalizing configurations ⇔ leading to **high PCT values**

IB-LOCA: *Intermediate Break LOss of Coolant Accident*

Problems & constraints

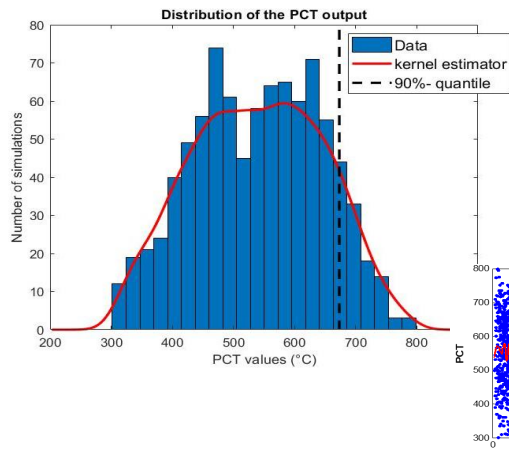
- **Very large number of inputs (~100)**, but **effective** dimension might be lower
 - Each CATHARE simulation ~ 1 hour ⇒ around 1000 simulations available
 - Phenomena involved are complex **with strong non-linearities**
 - Black-box model: intrusive methods not possible
- ⇒ **Monte Carlo sampling + advanced statistical tools for data analysis**
- ✓ **Screening and sensitivity analysis**
 - ✓ **Approximation with metamodel**
 - ✓ **Uncertainty propagation**
- ⇒ **Adapted to VERY HIGH DIMENSIONAL test case (~100 uncertain inputs)**

⇒ ICSCREAM* methodology in **4 Main Steps**

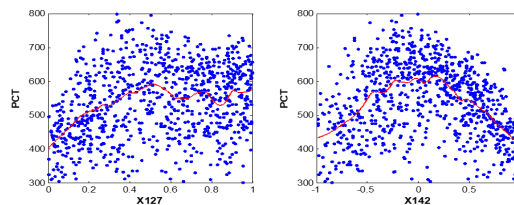
*Identification of penalizing Configurations using SCREening And Metamodel

Uncertainty quantification of uncertain inputs + scenario inputs to be penalized X_{pen} Uncertain inputs $X = (X_1, \dots, X_{d'})$ with probability distributions + scenario inputs X_{pen} **Step 1: Learning sample of n simulations (X_S, Y_S)** Monte-Carlo design of n experiments $X_S = \{x^{(1)}, \dots, x^{(n)}\}$ and associated CATHARE2 PCT outputs Y_S

- **$d = 96$ uncertain variables** with associated probability distributions (almost indep.)
- **$n = 889$ CATHARE2 simulations** : Monte-Carlo sample with inputs drawn following their assumed probability distributions

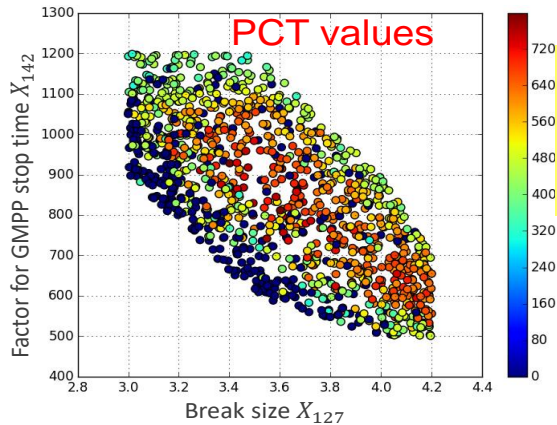
Empirical quantile 90%: $q_{0.9} \approx 673.18 \text{ }^\circ\text{C}$ **Critical configurations** are defined as: **PCT > $q_{0.9}$** 

Scatter plots with 1-D local polynomials for trends

Complex relationships of **PCT** w.r.t. inputs**Metamodeling of PCT** according to such a large number of inputs is a hard task

Among 96 inputs, 2 scenario inputs to be penalized (here dependent):

- ▶ X_{127} (break size): uniform distribution on [3, 4.2] inches
- ▶ X_{142} (factor for GMPP stop time): uniform random variable whose range of variation depends on the value of X_{127}



▶ **Objective:**

Precisely capture **critical configurations of (X_{127}, X_{142})** which lead to the **highest probability of PCT exceeding $q_{0,9}$ (≈ 673.18 °C)**

$$X_{pen} = \{X_{127}, X_{142}\} \subset X$$

GMPP : group of primary pumps

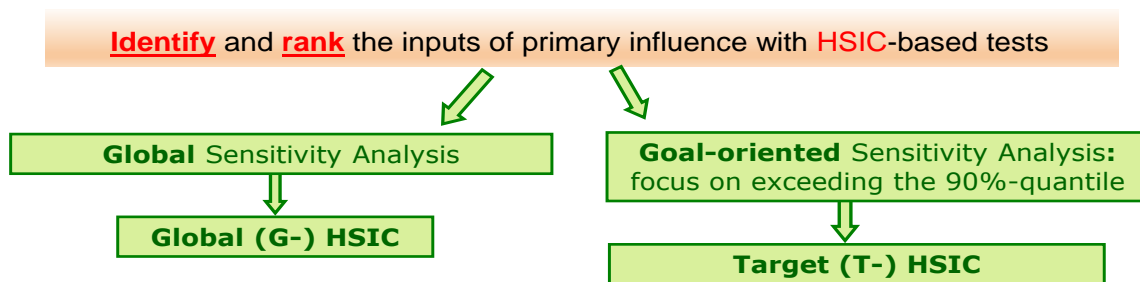
Uncertainty quantification of uncertain inputs + scenario inputs to be penalized X_{pen}

Step 1: Learning sample of n simulations (X_S, Y_S)

Step 2: Screening and ranking with HSIC-based independence tests from (X_S, Y_S)

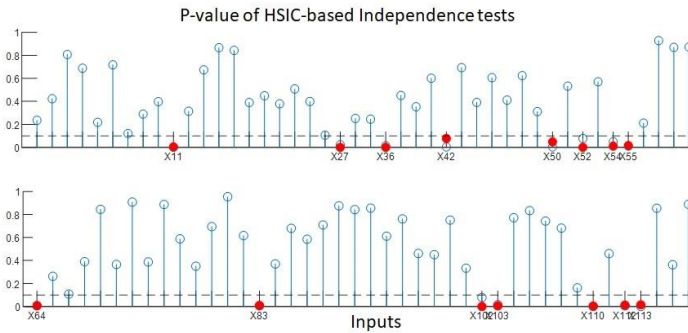
- **Main aims of Sensitivity Analysis (SA):** Understand the behavior of the model output w.r.t. inputs, quantify how the variability of the input parameters influences the output
- **For Screening purpose:** Separate the inputs into influential and non-influential
 - ⇒ Reduction of the model
 - Here: build a simplified model, referred to as metamodel
- **For Ranking purpose:**
 - **Ranking the inputs by influence degree**
 - Use in the building process of the metamodel (sequential building process)

- Hilbert-Schmidt Independence Criterion (HSIC)** compares the joint distribution between Y and X_j and the product of marginals of Y and X_j
- ⇒ “**Generalized cross-covariance**” based on embeddings of probability distributions in **Reproducing Kernel Hilbert Space**
 - ⇒ High-dimensional method for global sensitivity analysis (GSA)
 - ⇒ Computation from **a unique random sample**, of reasonable size (in practice few hundreds)
 - ⇒ HSIC-based **statistical independence tests for a more robust screening**

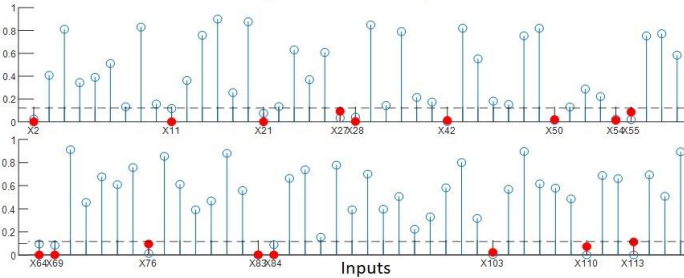


Global-HSIC tests

~ 18 influential inputs in GSA
Influence ++ : X142 (GMPP time)
Influence + : X127 (break size)
Influence : X113, X110, X11
 Lower influence : X50, X42, X112, X83, X64, X125, X55, X103, X36, X27, X54, X102, X52



P-value of Target-HSIC-based Independence tests

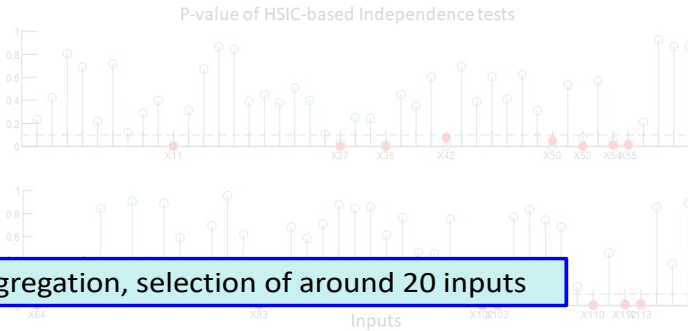


T-HSIC ⇒ Impact on exceeding the 90%-quantile $\hat{q}_{0.9}(Y)$

~ 19 influential inputs in TSA
Influence ++ : X142 (GMPP time)
Influence + : X113, X110, X127, X125, X83
 Lower influence : X42, X103, X76, X50, X55, X54, X2, X27, X28, X21, X84, X64, X11

Global-HSIC tests

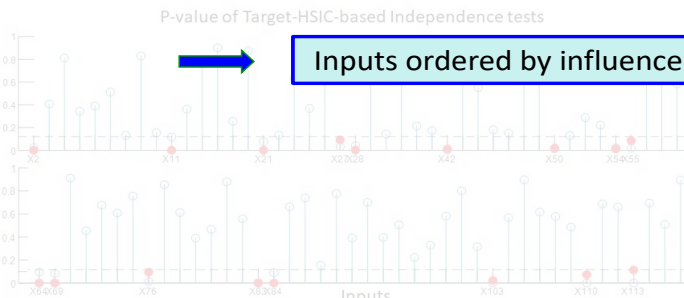
~ 18 influential inputs in GSA
Influence ++ : X142
Influence + : X127
Influence : X113, X110, X11
 Lower influence : X50, X42, X112, X83, X64, X125, X55, X103, X36, X27, X54, X102, X52



From aggregation, selection of around 20 inputs



Inputs ordered by influence d° , using P -values



Impact on exceeding the 90%-quantile $\hat{q}_{0.9}(Y)$

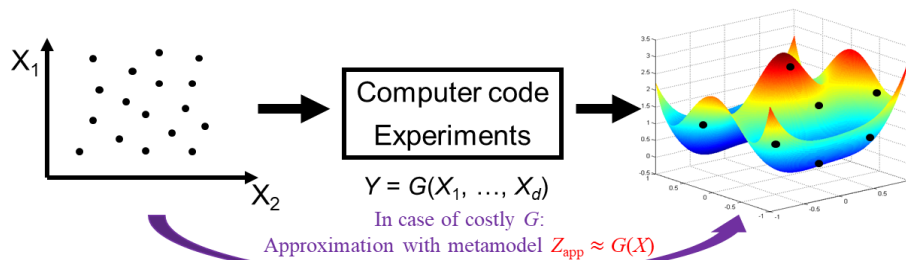
~ 19 influential inputs in TSA
Influence ++ : X142
Influence + : X113, X110, X127, X125, X83
 Lower influence : X42, X103, X76, X50, X55, X54, X2, X27, X28, X21, X84, X64, X11

Uncertainty quantification of uncertain inputs + scenario inputs to be penalized X_{pen}

Step 1: Learning sample of n simulations (X_S, Y_S)

Step 2: Screening and ranking with HSIC and T-HSIC independence tests from (X_S, Y_S)

Step 3: Sequential Metamodeling → Gaussian process (GP) regression from (X_S, Y_S)



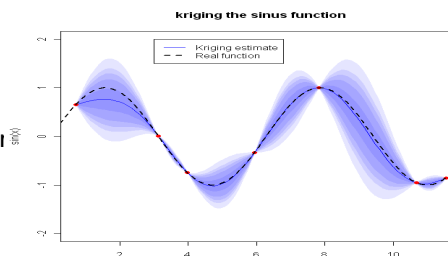
Goal: replace code by a mathematical function called metamodel

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)



→ Challenge to be addressed here: how to build the GP in large dimension ($d \sim 100$) ?

→ Use the information of screening and ranking from HSIC
⇒ **Sequential estimation of GP hyperparameters**

⇒ Use of Step 2 results (HSIC-based ranking and screening)

Inputs in GP covariance:

- ✓ Main influential inputs in a tensorized anisotropic covariance
- ✓ Inputs of 2^{ndary} influence in an isotropic covariance (coarser way)
- ✓ Non-significant influential inputs ⇒ effect only captured by an additional variance (nugget effect)

Likelihood-based “forward” estimation of GP hyperparameters:

- ✓ Robust Sequential building: successive inclusion of ordered influential inputs

⇒ Use of Step 2 results (HSIC-based ranking and screening)

Inputs in GP covariance:

- ✓ Main influential inputs in a tensorized anisotropic covariance
- ✓ Inputs of 2^{ndary} influence in an isotropic covariance (coarser way)
- ✓ Non-significant influential inputs ⇒ effect only captured by an additional variance (nugget effect)

Likelihood-based “forward” estimation of GP hyperparameters:

- ✓ Robust Sequential building: successive inclusion of ordered influential inputs

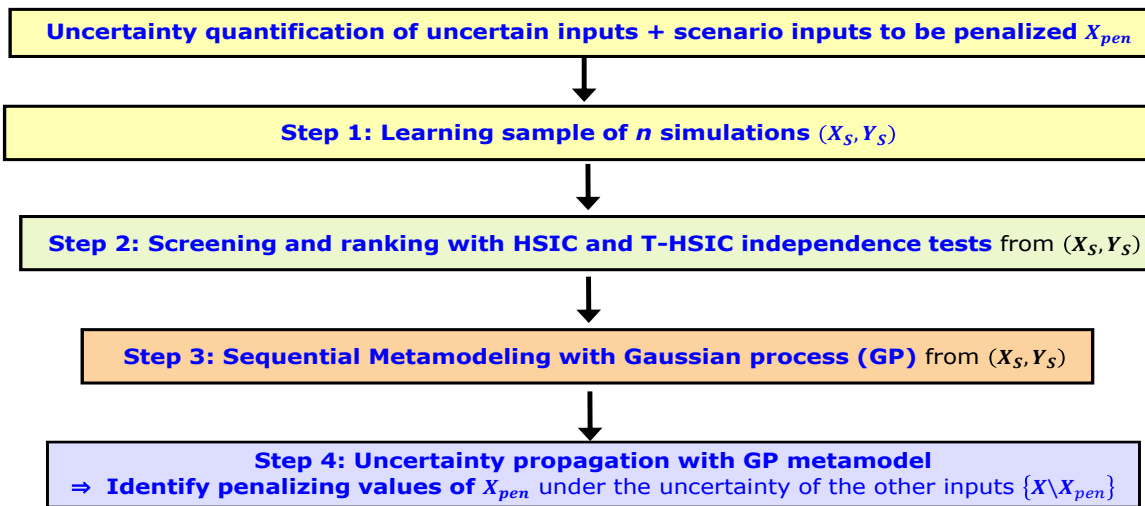
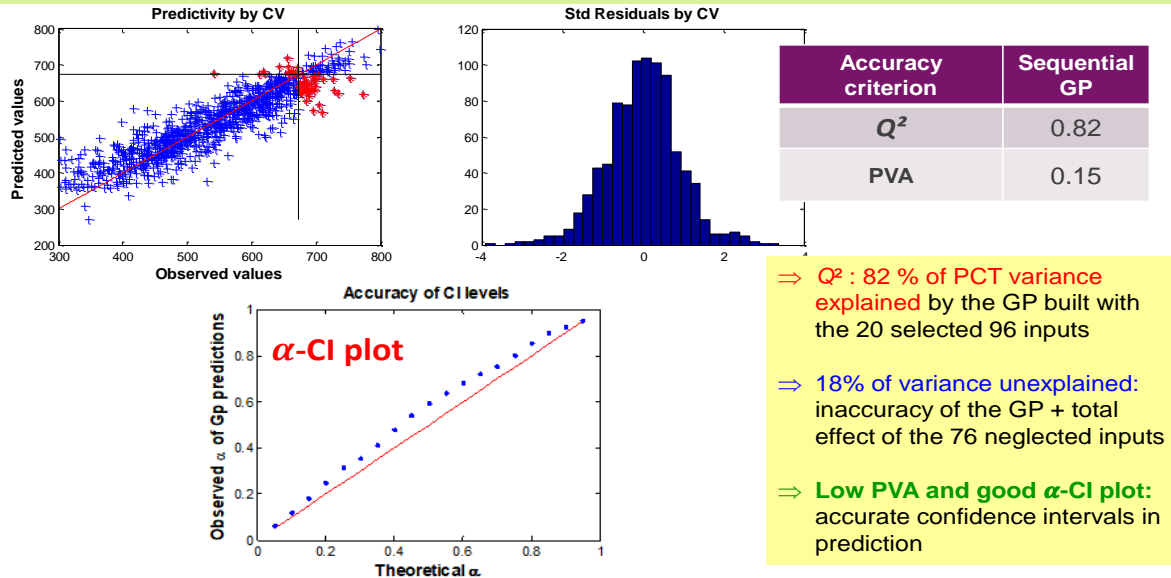
Assessment of the accuracy of GP metamodel

- **Quality of GP predictor** \hat{Y}_{GP} : predictivity coefficient Q^2 , analysis of residuals
- **Quality of GP prediction variance and intervals**: Predictive Variance adequacy (PVA) coefficient, α - α plot

⇒ In practice: **Estimation by Cross Validation**

Full description of GP validation criteria in Demay et al. [2021]

Assessment of accuracy and predictivity of final GP metamodel built on $N = 889$ simulations



Step 4: Uncertainty propagation with GP metamodel to identify the penalizing values of X_{pen} under the uncertainty of the other inputs $\{X \setminus X_{pen}\}$

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

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

X_{exp} : explanatory inputs of the GP
 $\tilde{X}_{exp} = X_{exp} \setminus X_{pen}$

\tilde{X}_{exp} and X_{pen} are independent (necessary condition)

Φ : CDF of standard Gaussian distribution

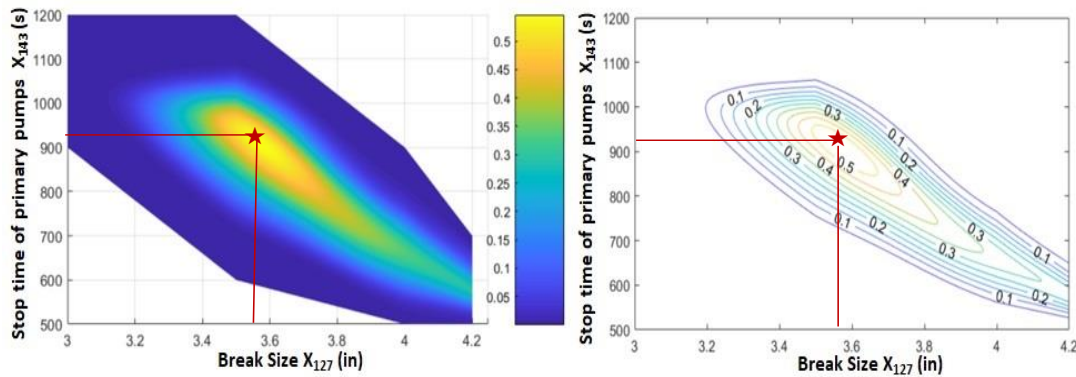
Variation domain of \tilde{X}_{exp} Joint distribution of \tilde{X}_{exp}

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


STEP 4: Uncertainty propagation with the GP
Illustration on the test case

Computation of $\hat{P}(X_{pen})$

Probability of exceeding $\hat{q}_{0.9} = 673.18^\circ\text{C}$, according to X_{127} and X_{143}



- ▶ Strong interaction between the two scenario parameters
- ▶ Worst case: (3.57 inches, 907.8 seconds) ⇒ $\hat{P} \approx 0.55$
- ▶ 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 $X_{127} < 3.3$ ⇒ the water inventory does not degrade too much (whatever GMPP)
 - If $X_{127} > 3.9$ ⇒ break tends to be prevailing and reduces the impact of stop time of GMPP

- **Innovative Best-Estimate-Plus-Uncertainty approach for Industrial application with a large number of uncertain inputs**
 - ⇒ **Several advanced statistical tools efficiently combined**
 - ⇒ **Operational methodology able to deal with real complex cases with large number of inputs**
- **Global methodology based on ONE single Monte Carlo inputs-output sample:**
 - ⇒ Screening + global & target sensitivity analysis + Metamodel + Uncertainty propagation
- **Identification of penalizing configurations** with conditional probabilities estimated by GP-metamodel
- Other applications on IB-LOCA datasets with 10 inputs to be penalized (See Marrel et al. [2021]). Industrialization in EDF Uncertainty platform OPENTurns
- Planned improvements in the framework of the ANR  Project

Simulation Analytics and Metamodel-based solutions for Optimization, Uncertainty and Reliability Analysis



- F. Bachoc, Cross Validation and Maximum Likelihood estimations of hyper-parameters of Gaussian processes with model misspecification, *Computational Statistics and Data Analysis*, 2013.
- S. Da Veiga, Global sensitivity analysis with dependence measures, *Journal of Statistical Computation and Simulation*, 85:1283-1305, 2015.
- M. De Lozzo and A. Marrel, New improvements in the use of dependence measures for sensitivity analysis and screening, *Journal of Statistical Computation and Simulation*, 86:3038-3058, 2016.
- C. Demay, B. Iooss and L. Le Gratiet, A. Marrel. Model selection for Gaussian Process regression: applications on the characterization of chemical segregation, *Quality and Reliability Engineering International*, 2021.
- R. El Amri and A. Marrel. HSIC-based independence tests with optimal sequential permutations: application for sensitivity analysis of numerical simulators, *Quality and Reliability Engineering International*, 2021.
- B. Gramacy. Gaussian Process Modeling, Design, and Optimization for the Applied Sciences. Chapman and Hall/CRC, 2021.
- B. Iooss B., and P. Lemaître. A review on global sensitivity analysis methods. In *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, Springer, 2015.
- B. Iooss and A. Marrel, Advanced methodology for uncertainty propagation in computer experiments with large number of inputs, *Journal of Nuclear Technology*, 205:1588-1606, 2019.
- Gretton, Bousquet, Smola and Schölkopf. Measuring statistical dependence with Hilbert-Schmidt norms. In: *Proceedings Algorithmic Learning Theory*, 2015.
- **A. Marrel, B. Iooss and V. Chabridon, Statistical identification of penalizing configurations in high-dimensional thermalhydraulic numerical experiments: The ICSCREAM methodology, Nuclear Science & Engineering, 2021.**
- A. 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), 2021.
- P. Mazgaj, J-L. Vacher and S. Carnaveli. Comparison of CATHARE results with the experimental results of cold leg intermediate break LOCA obtained during ROSA-2/LSTF test 7, *EPJN*, 2, 2016.
- C.E. Rasmussen and C.K.I. Williams. Gaussian processes for machine learning. MIT Press, 2006.



THANKS FOR YOUR ATTENTION

Work funded and carried out within the  ANR research project.

Simulation Analytics and Metamodel-based solutions for Optimization, Uncertainty and Reliability Analysis



IRESNE | DER | SESI | LEMS

Institut de recherche sur les systèmes nucléaires pour la production d'énergie bas carbone

Commissariat à l'énergie atomique et aux énergies alternatives - www.cea.fr