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

Amandine Marrel, Bertrand Iooss, Vincent Chabridon

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.

PROGRESS MEETING - CATHSBI ACTION

anr agence nationale de la recherche

ICSCREAM* METHODOLOGY

**Identification of penalizing Configurations using SCREening And Metamodel*

cea

DE LA RECHERCHE À L'INDUSTRIE

Amandine MARREL (CEA DES/IRESNE/DER)
Bertrand IOOSS (EDF R&D)
Vincent CHABRIDON (EDF R&D)

SIAM UQ 2022

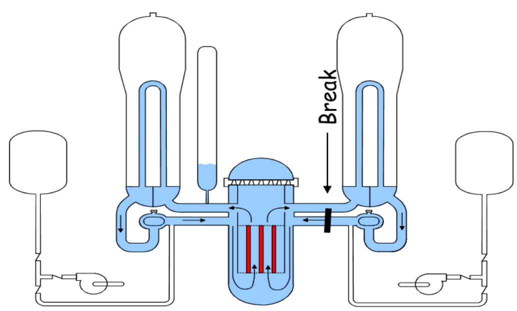
IRESNE */ DER / SESI / LEMS*

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

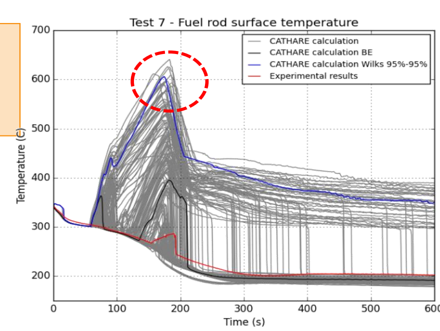
Simulation of IB-LOCA nuclear accident

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



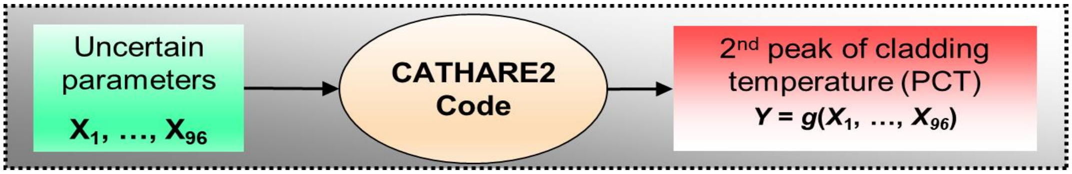
***d* (~ 100) input random variables:**
Critical flowrates, initial/boundary conditions, phys. eq. coef., …

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

Objective in IB-LOCA safety study



- 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*

Objective in IB-LOCA safety study

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*

STEP 1: Monte Carlo Sampling design

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

Uncertain inputs $\boldsymbol{X} = (X_1, \dots, X_{d'})$ with probability distributions + scenario inputs $\boldsymbol{X}_{pen}$

↓

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

Monte-Carlo design of n experiments $\boldsymbol{X_S} = \{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(n)}\}$ and associated CATHARE2 PCT outputs $\boldsymbol{Y_S}$

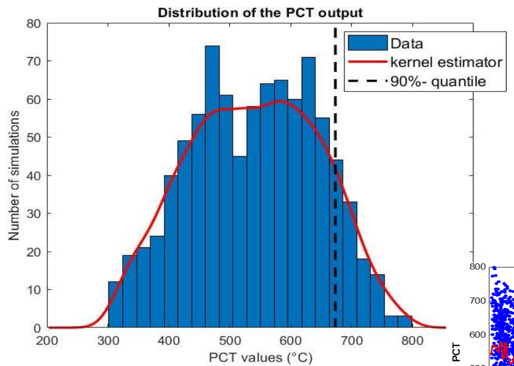
STEP 1: Monte Carlo Sampling design
Illustration on the IB-LOCA test case

▶ ***d* = 96 uncertain variables** with associated probability distributions (almost indep.)

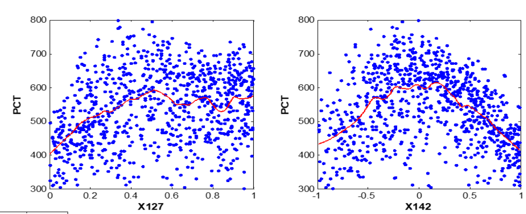
▶ ***n* = 889 CATHARE2 simulations** : Monte-Carlo sample with inputs drawn following their assumed probability distributions

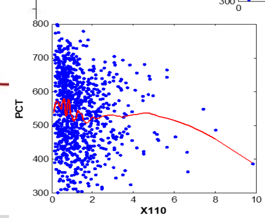
Empirical quantile 90%: $\boldsymbol{q_{0.9}} \approx \mathbf{673.18\ °C}$

Critical configurations are defined as: **PCT > $q_{0.9}$**



Scatter plots with 1-D local polynomials for trends





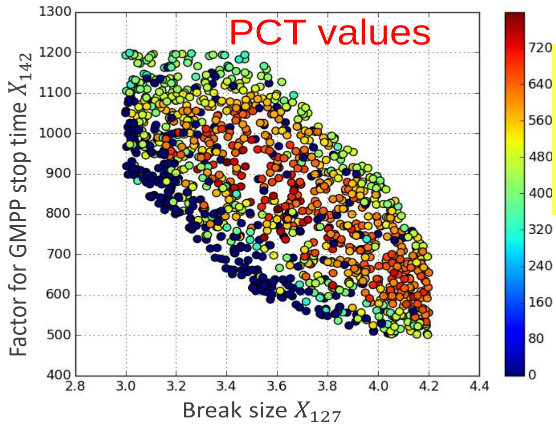
Complex relationships of **PCT** w.r.t. inputs

Metamodeling of PCT according to such a **large number of inputs** is a hard task

STEP 1: Monte Carlo Sampling design
Illustration on the IB-LOCA test case

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)

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

GMPP : *group of primary pumps*

STEP 2: Screening & ranking of inputs

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)

STEP 2: Screening & ranking of inputs

- **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)

STEP 2: Screening & ranking of inputs

Hilbert-Schmidt Independence Criterion (HSIC) compares the joint distribution between Y and X_i and the product of marginals of Y and X_i

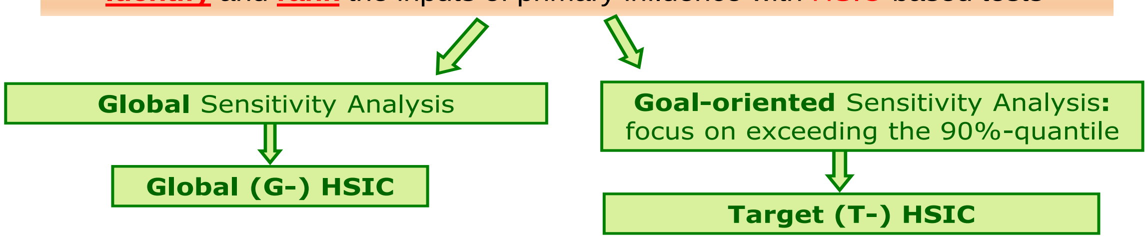
⇒ **"*Generalized cross-covariance*"** based on embeddings of probability distributions in **Reproducing Kernel Hilbert Space**

⇒ 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**

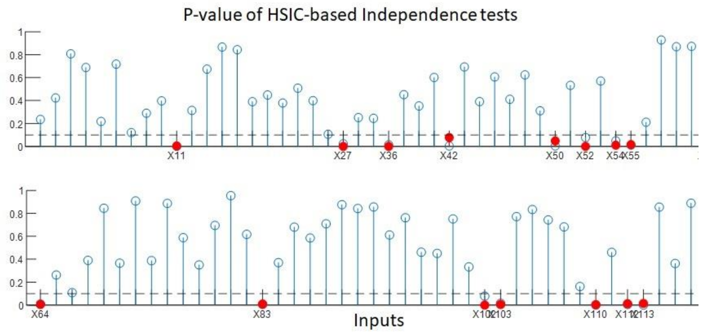
Identify and **rank** the inputs of primary influence with HSIC-based tests



STEP 2: Screening & ranking of inputs
Illustration on the IB-LOCA test case

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



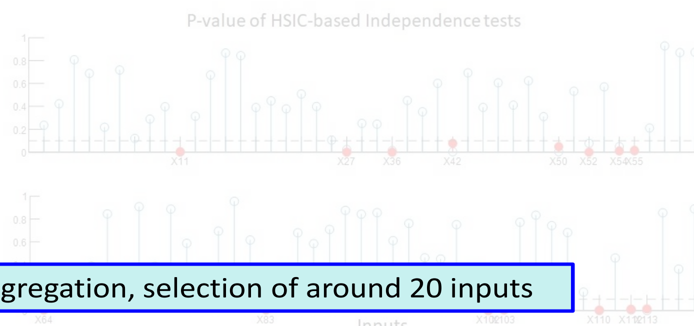
T-HSIC ⇨ Impact on exceeding the 90%-quantile $\hat{q}_{0.9}(Y)$

~ 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

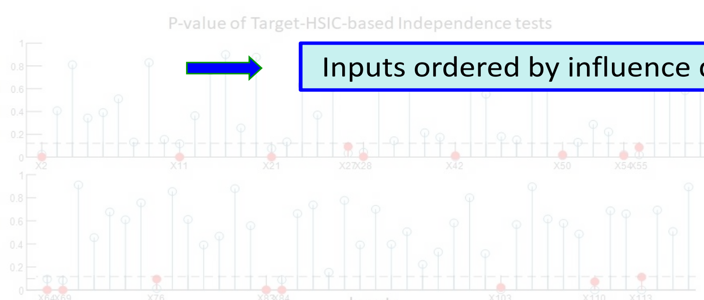
STEP 2: Screening & ranking of inputs
Illustration on the IB-LOCA test case

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*

~ 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}

↓

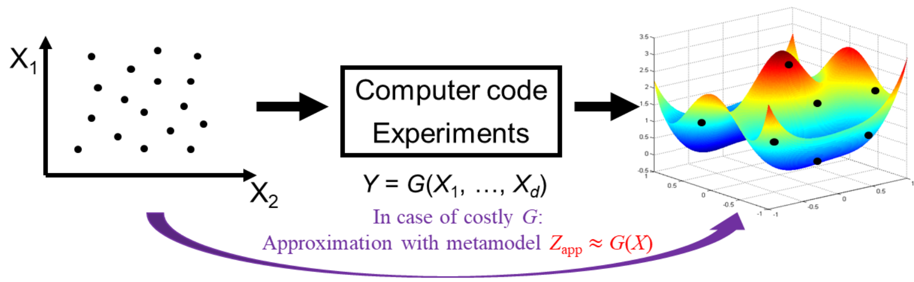
Step 1: Learning sample of *n* simulations (X_S, Y_S)

↓

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)



For details on GP, See Rasmussen & Williams [2005], Gramacy [2021]

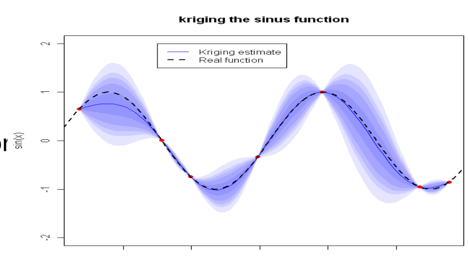
Goal: replace code by a mathematical function called metamodel

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~100) ?

→ Use the information of screening and ranking from HSIC
$\Rightarrow$ **Sequential estimation of GP hyperparameters**

STEP 3: Approximation with a GP Metamodel

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

Inputs in GP covariance:

- ✓ Main influential inputs in a tensorized anisotropic covariance
- ✓ Inputs of 2ndary 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

STEP 3: Approximation with a GP Metamodel

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

Inputs in GP covariance:

- ✓ Main influential inputs in a tensorized anisotropic covariance
- ✓ Inputs of 2ndary 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

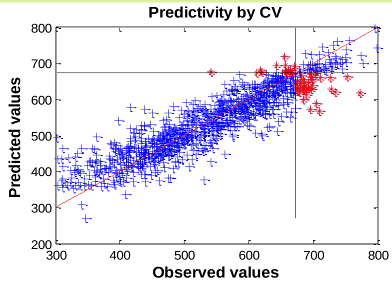
- **Quality of GP predictor $\widehat{Y}_{GP}$**: predictivity coefficient Q^2, analysis of residuals
- **Quality of GP prediction variance and intervals**: Predictive Variance adequacy (PVA) coefficient, α-α plot

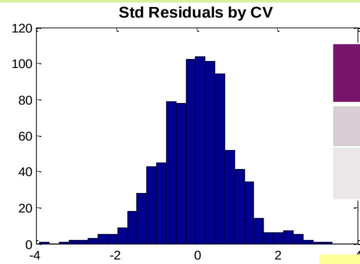
⇨ In practice: Estimation by **Cross Validation**

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

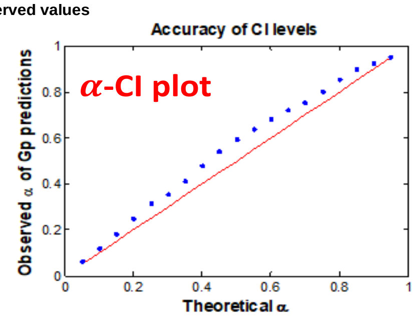
STEP 3: Approximation with a GP Metamodel
Illustration on the test case

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





Accuracy criterion	Sequential GP
Q^2	0.82
PVA	0.15



⇒ Q^2 : 82 % of PCT variance explained by the GP built with the 20 selected 96 inputs

⇒ 18% of variance unexplained: inaccuracy of the GP + total effect of the 76 neglected inputs

⇒ **Low PVA and good α-CI plot:** accurate confidence intervals in prediction

STEP 4: Uncertainty propagation with the GP

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 with Gaussian process (GP) from (X_S, Y_S)

↓

Step 4: Uncertainty propagation with GP metamodel
⇒ **Identify penalizing values of X_{pen}** under the uncertainty of the other inputs $\{X \backslash X_{pen}\}$

STEP 4: Uncertainty propagation with the GP

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

⇒ 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)

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

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

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

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

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

Φ : CDF of standard Gaussian distribution

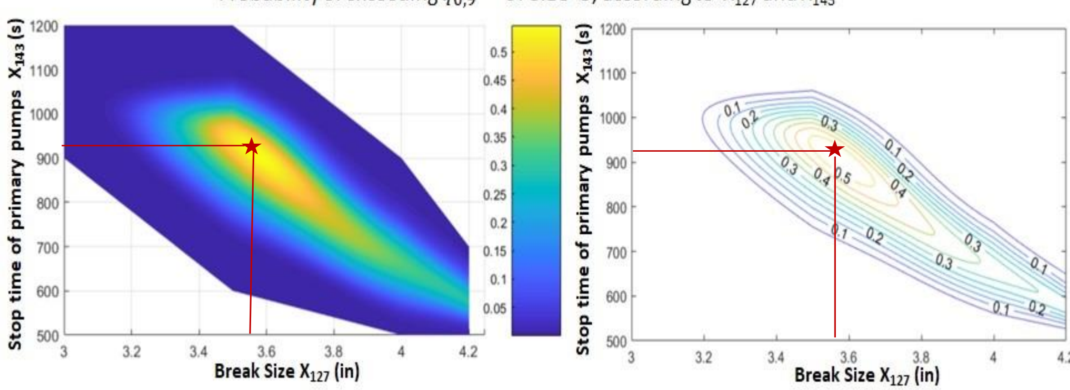
- In practice, for each value of $\boldsymbol{X_{pen}} = \{X_{127}, X_{143}\}$, $\hat{P}(\boldsymbol{X_{pen}})$ is estimated by intensive Monte-Carlo computation (here integral in dimension 18 in the use-case)

STEP 4: Uncertainty propagation with the GP

Illustration on the test case

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

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) ⇒ $\hat{P} \approx 0.55$
- ► 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

Conclusion and prospects

- **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 progagation

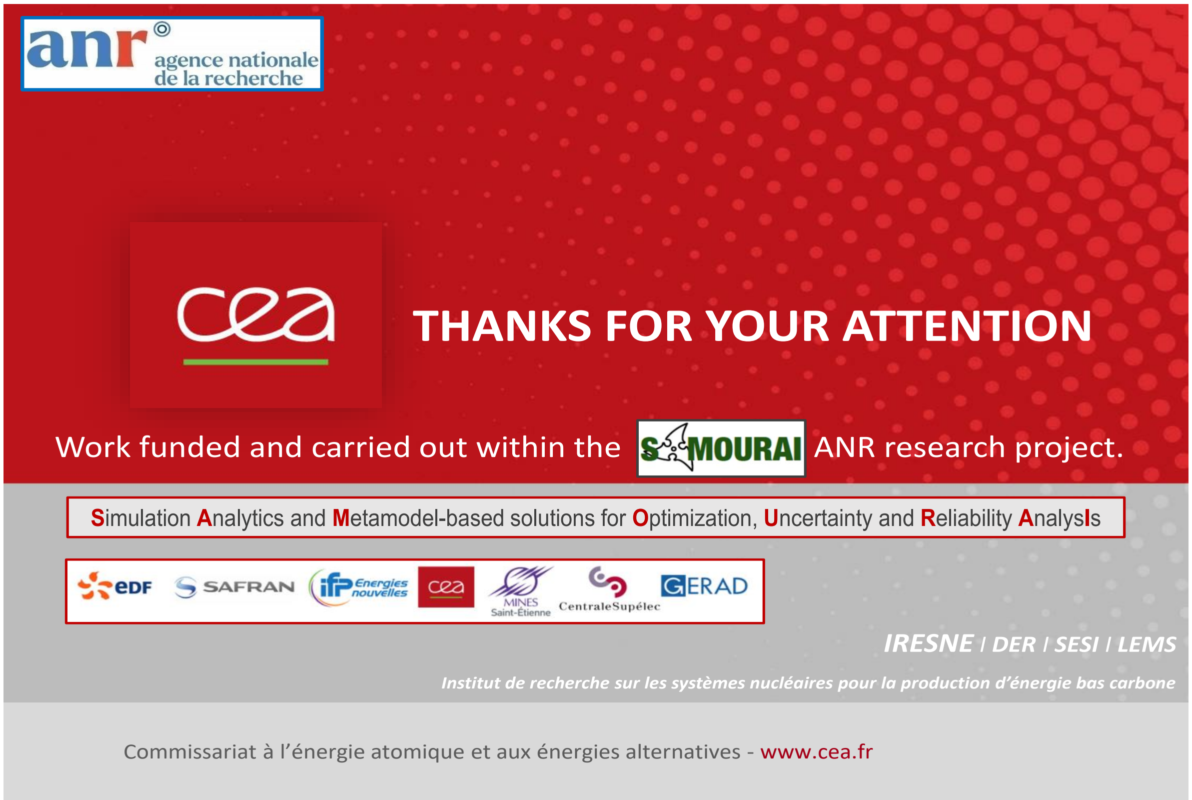
- **Identification of penalizing configurations** with conditional probabilities estimated by GP-metamodel
- Other applications on IB-LOCA datasets with 10 inputs to be penalized (*See Marrel et al. [2021]).* Industrialization in EDF Uncertainty platform OPENTurns
- Planned improvements in the framework of the ANR SAMOURAI Project

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

anr agence nationale de la recherche

Some References

- 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.


anr agence nationale de la recherche
cea
THANKS FOR YOUR ATTENTION
Work funded and carried out within the SAMOURAI ANR research project.
Simulation Analytics and Metamodel-based solutions for Optimization, Uncertainty and Reliability AnalysIs
EDF
SAFRAN
ifp Energies nouvelles
cea
MINES Saint-Étienne
CentraleSupélec
GERAD
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