OTICSCREAM: A Python module (and a cooking recipe) for the identification of penalizing configurations in computer experiments

JU-OT 2025 @ EDF Lab Saclay
June 13th, 2025

Vincent Chabridon[1], Joseph Muré[1], Elias Fekhari[1]

[1]EDF R&D, 6 quai Watier, 78401 Chatou, France.

Table of contents

Introduction

Context and motivations

- Industrial motivations: risk analysis in nuclear safety
  - Deterministic analyses ("worst cases")
    - Historical
    - Supposed to be "conservative" enough
    - **But** sometimes, far from reality!
  - Probabilistic analyses ("Best Estimate Plus Uncertainty" or **BEPU**)
    - Supposed to produce more realistic physics!
    - Needs to take (explicitly) uncertainties into account!
    - **But** safety margins need to be identified clearly!

- Simulation of an Intermediate-Break Loss-Of-Coolant Accident[1]
  - Modeling of a primary circuit (cold leg) of a PWR[2]
  - Simulation of a thermal-hydraulic transient scenario
  - Best-estimate computer model: the CATHARE2 code

[1] IBLOCA

[2] Pressurized Water Reactor

Context and motivations

❑ Simulation of an IBLOCA using the CATHARE2 code

- ➤ Uncertainty sources in input:
  - Type #1: Initial/boundary conditions ➜ probabilistic ($\mathcal{U}$, trunc. $\mathcal{N}$)
  - Type #2: Physical parameters ➜ probabilistic ($\mathcal{U}$, $\mathcal{LU}$, trunc. $\mathcal{N}$, $\mathcal{LN}$)
  - **Type #3: Scenario parameters ➜ not proba. (lower/upper bounds)**
- ➤ Scalar output quantity of interest (QoI):
  - Second peak of cladding temperature (PCT)

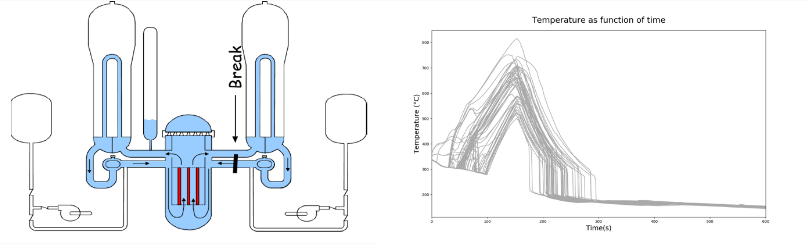


Figure 1: IBLOCA in a PWR (©CEA) / PCT trajectories from CATHARE2 (©EDF).

Context and motivations

- ❑ Main objectives
  - ➤ How to identify the most **penalizing configurations** for Type #3 inputs (scenario parameters) regardless the other sources of uncertainties (Type #1 and Type #2)?
  - ➤ **Penalizing configurations ➜ Leading to high PCT values**
- ❑ Main challenges/constraints
  - ➤ Large number of inputs ($\approx$ 100) ➜ Effective dimension might be lower!
  - ➤ Complex phenomena / strongly nonlinear response!
  - ➤ Computational cost ➜ Only $\approx 10^3$ simulations available!
  - ➤ Industrial/regulatory context ➜ crude Monte Carlo (i.i.d.)

Questions

- How to handle these challenges and constraints?
- Can we propose a generic methodology?

The ICSCREAM methodology

ICSCREAM: a four-step methodology

- Non-exhaustive list of available tools in the UQ literature
  - Global sensitivity analysis techniques to reduce the input dimension
  - Smart adaptive surrogate modeling strategies
  - Approximation tools based on low-rank or projection-based methods
  - Adaptive strategies for robust inversion [RB19]

    ➜ We probably need to combine all these things together, as in a cooking recipe!

- A possible answer: the ICSCREAM (ice-cream) methodology
  - Originally proposed as a **specific answer** in the nuclear safety context (see [MIC22])
  - **But** can still be seen as a **generic methodology** for a large panel of applications!
  - Based on state-of-the-art UQ techniques (and a few advanced tools) gathered in a **modular framework**!
  - Adapted to a **given-data** context (no access to the computer model)!

ICSCREAM: a four-step methodology

❏ The ICSCREAM methodology in a nutshell!

Step 1 — Learning sample generation (simulation-based / given-data)

Step 2 — Screening and ranking with HSIC-based indices

Step 3 — Surrogate model build-ing and validation

Step 4 — Identification of penalizing values

Step 1

Step 1: Monte Carlo sampling ("given-data" context)

❑ A few notations and preliminary hypotheses:

➤ Real-valued & deterministic black-box model:

$$\mathcal{M} : \left| \begin{array}{ccc} \mathcal{X} \subseteq \mathbb{R}^d & \longrightarrow & \mathcal{Y} \subseteq \mathbb{R} \\ \mathrm{X} & \longmapsto & Y = \mathcal{M}(\mathrm{X}) \end{array} \right. \tag{1}$$

➤ Probabilistic modeling of d **independent** input physical variables:

$$\mathrm{X} := (\underbrace{X_1, X_2, \ldots, X_p}_{\mathrm{X}_{\mathrm{pen}}}, \underbrace{X_{p+1}, \ldots, X_d}_{\mathrm{X}_{\mathrm{alea}}})^\top \sim P_{\mathrm{X}} \quad \text{over} \quad \mathcal{X} = \bigtimes_{i=1}^{d} \mathcal{X}_i \tag{2}$$

where:

- $\mathrm{X}_{\mathrm{pen}}$ ➜ candidate inputs to be **penalized** s.t. $\mathrm{X}_{\mathrm{pen}} \sim \prod_{i=1}^{p} \mathcal{U}\left([a_i, b_i]\right)$
- $\mathrm{X}_{\mathrm{alea}}$ ➜ aleatoric variables s.t. $\mathrm{X}_{\mathrm{alea}} \sim P_{\mathrm{X}_{\mathrm{alea}}}$

➤ **Learning sample** ➜ get a n-size **i.i.d.** sample of the couple (X, Y):

$$\left(\mathrm{X}^{(j)}, Y^{(j)}\right)_{(1 \leq j \leq n)} = \left(X_1^{(j)}, X_2^{(j)}, \ldots, X_d^{(j)}, Y^{(j)}\right)_{(1 \leq j \leq n)} \tag{3}$$

with $P_{\mathrm{X}^{(j)}} = P_{\mathrm{X}}$ and $Y^{(j)} = \mathcal{M}\left(X_1^{(j)}, X_2^{(j)}, \ldots, X_d^{(j)}\right), \forall j \in \{1, \ldots, n\}$

➤ **Risk analysis** ➜ a typical QoI is the α**-order quantile** estimated using the empirical estimator $\widehat{q}_\alpha(Y)$

Step 2

Step 2: Global and target sensitivity analyses

- Two usual settings in sensitivity analysis [DGIP21]
  - "Can we split the inputs into two classes $\{\text{noninfluential}, \text{influential}\}$?"
    ➜ Screening
  - "Is it possible to rank the inputs with respect to their influence?"
    ➜ Ranking
- Global (GSA) vs. target sensitivity analysis (TSA)
  - "Does an input X_i have an influence on the global variability of Y?"
    ➜ GSA study
  - "Does an input X_i have an influence on exceeding an output threshold (e.g., $\{Y > q_\alpha(Y)\}$) or not?"
    ➜ TSA study [RM18, SLRDV19, MC21]

Step 2: Global and target sensitivity analyses

- A brief overview of the literature
  - Sobol' indices [PT17, PD19]
  - Indices based on dissimilarity measures (e.g., Borgonovo indices, Csiszár f-divergences, etc.) [BP16, Rah16]
  - **Hilbert-Schmidt Independence Criterion** (**HSIC**) [GBSS05, DV15]
    - ➜ Allows to measure influence beyond the variance
    - ➜ Adapted to the "given-data" context
    - ➜ Allows to perform GSA and TSA without having much trouble
    - ➜ Indices come together with a rigorous statistical framework

Step 2: Global and target sensitivity analyses

- A brief overview about HSIC indices
  - Basic idea: to assess the dependency between the random variables X_i and Y by comparing an infinite collection of features of both $P_{X_i,Y}$ and $P_{X_i}P_Y$
  - We consider a dependence measure, the **maximum mean discrepancy** (**MMD**) between $P_{X_i,Y}$ and $P_{X_i}P_Y$, which is expressed as the distance between their mean embeddings in some RKHS[3].

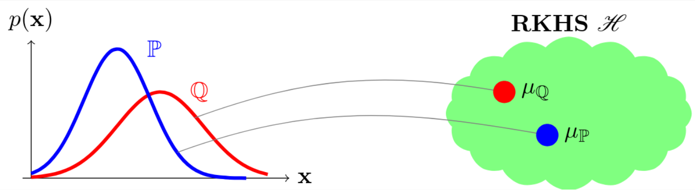


Figure 2: Embedding of marginal distributions (source: [MFSS17]).

[3] Reproducing Kernel Hilbert Space.

Step 2: Global and target sensitivity analyses

- A brief overview about HSIC indices
  - A first mathematical expression:

$$\begin{aligned} \mathrm{HSIC}_{\kappa_i\kappa}(X_i, Y) &= \mathrm{MMD}^2_{\kappa_i\kappa}(P_{(X_i,Y)}, P_{X_i}P_Y) && \text{(4a)} \\ &= \mathbb{E}[\kappa_i(X_i, X_i^{'})\kappa(Y, Y^{'})] + \mathbb{E}[\kappa_i(X_i, X_i^{'})]\mathbb{E}[\kappa(Y, Y^{'})] && \text{(4b)} \\ &- 2\mathbb{E}[\mathbb{E}[\kappa_i(X_i, X_i^{'})|X_i]\ \mathbb{E}[\kappa(Y, Y^{'})|Y]] \end{aligned}$$

  - A fundamental property ➜ if the kernels κ_i and κ are **characteristic**

$$\mathrm{HSIC}_{\kappa_i\kappa}(X_i, Y) = 0 \Leftrightarrow X_i \perp Y \qquad (5)$$

  - In practice ➜ one uses various types of **estimators** and a **statistical hypothesis testing** in order to decide whether an input is noninfluential or not

Step 3

Step 3: Gaussian process surrogate modeling

- **Gaussian process (GP) regression** [RW06]
  - Goal: build a predictor $\widetilde{\mathcal{M}}(\cdot)$ that mimics the behavior of $\mathcal{M}(\cdot)$ using a **dataset** $\mathcal{D} = \{(\mathrm{x}^{(i)}, y^{(i)}),\ i = 1, \ldots, n\}$ (supervised learning)
  - Probabilistic surrogate model ➜ assume

$$Y(\mathrm{x}) \sim GP(\mu(\mathrm{x}), \Sigma(\mathrm{x}', \mathrm{x})) \tag{6}$$

  - GP conditioning on data (universal kriging predictor):

$$\widehat{Y}(\mathrm{x}) = [Y(\mathrm{x})|\mathcal{D}, \sigma^2, \boldsymbol{\theta}] \sim \mathcal{N}_1(\mu_{\widehat{Y}}(\mathrm{x}), \sigma^2_{\widehat{Y}}(\mathrm{x})) \tag{7}$$

  - Estimation strategies for $(\sigma^2, \boldsymbol{\theta})$ ➜ MLE [MIVDV08]

Step 3: Gaussian process surrogate modeling

- GP building in practice
  - GP metamodel will be a function of several kind of inputs:
    - Explanatory inputs: $\mathsf{X}_{\text{exp}} = \{\mathsf{X}_{\text{PII}} \cup \mathsf{X}_{\text{pen}}\}$
    - Secondary influential inputs: X_{SII}
    - Neglected inputs: $\mathsf{X}_{\epsilon} = \{\mathsf{X} \setminus \{\mathsf{X}_{\text{exp}} \cup \mathsf{X}_{\text{SII}}\}\}$
  - GP mean: either constant or linear basis
  - X_{exp} ➜ tensorized anisotropic covariance function (Matérn ν)

$$R_{\boldsymbol{\theta}_{\text{exp}}}(\mathsf{x}_{\text{exp}}, \mathsf{x}'_{\text{exp}}) = \prod_{i=1}^{d_{\text{exp}}} R_{\theta_i}(x_{\text{exp}}^{(i)} - x_{\text{exp}}^{\prime (i)}) \tag{8}$$

  - X_{SII} ➜ isotropic covariance function (Matérn ν)

$$R_{\theta_{\text{SII}}}(\mathsf{x}_{\text{SII}}, \mathsf{x}'_{\text{SII}}) = R_{\theta_{\text{SII}}}(||\mathsf{x}_{\text{SII}} - \mathsf{x}'_{\text{SII}}||_2) \tag{9}$$

  - X_{ϵ} ➜ homoscedastic nugget effect
  - Resulting tensorized covariance:

$$\Sigma(\mathsf{x}, \mathsf{x}') = \sigma^2 \left(R_{\boldsymbol{\theta}_{\text{exp}}}(\mathsf{x}_{\text{exp}}, \mathsf{x}'_{\text{exp}}) R_{\theta_{\text{SII}}}(\mathsf{x}_{\text{SII}}, \mathsf{x}'_{\text{SII}}) + \gamma \delta_{(\mathsf{x}, \mathsf{x}')}\right) \tag{10}$$

  - GP metamodel either built directly or **sequentially**
  - Optimization of hyperparameters ➜ multi-start + LHS + optim. algo.

Step 3: Gaussian process surrogate modeling

- Cross-validation metrics
  - Validation of the GP predictor:
    - Analysis of the standardized residuals
    - Computation of the **predictivity coefficient** Q^2

$$Q^2 = 1 - \frac{\sum_{j=1}^{n_{\mathrm{val}}} \left(y^{\mathrm{val},(j)} - \hat{y}^{\mathrm{val},(-j)} \right)^2}{\sum_{j=1}^{n_{\mathrm{val}}} \left(y^{\mathrm{val},(j)} - \bar{y}^{\mathrm{val}} \right)^2} \tag{11}$$

    - QQ-plot, rate of good classification, ...
  - Validation of the GP predictive variance:
    - Computation of the **Predictive Variance Adequacy** (PVA) [Bac13]

$$\mathrm{PVA} = \left| \log \left(\frac{1}{n_{\mathrm{val}}} \sum_{j=1}^{n_{\mathrm{val}}} \frac{\left(y^{\mathrm{val},(j)} - \hat{y}^{\mathrm{val},(-j)} \right)^2}{\widehat{\sigma^2}_{(-j)}} \right) \right| \tag{12}$$

    - Comparison of theoretical levels of GP-based confidence intervals and observed levels
    - Many other metrics can be considered for validation [DILGM22]
- Validation based on subsampling techniques
  - Computation of Q^2 using **kernel herding** [FIM+22]

Step 4

Step 4: Identification of penalizing configurations

- Identification of the penalizing values wrt. the QoI
  - Goal: perform an uncertainty propagation with the GP surrogate model in order to compute

$$\widehat{P}(\mathsf{X}_{\mathrm{pen}}) = \mathcal{P}(Y_{\mathrm{GP}}(\mathsf{X}_{\mathrm{exp}}, \mathsf{X}_{\mathrm{SII}}) > \widehat{q}_\alpha(Y) \mid \mathsf{X}_{\mathrm{pen}} = \mathsf{x}_{\mathrm{pen}}) \tag{13a}$$

$$= \ldots \text{(more details in [MIC22])} \tag{13b}$$

$$= 1 - \int_{\widetilde{\mathcal{X}}_{\mathrm{exp}}} \Phi\left(\frac{\widehat{q}_\alpha - \mu_{\widehat{Y}}(\mathsf{x})}{\sigma_{\widehat{Y}}(\mathsf{x})}\right) \mathrm{d}P_{\widetilde{\mathsf{X}}_{\mathrm{exp}}}(\widetilde{\mathsf{x}}_{\mathrm{exp}}) \tag{13c}$$

  where $\widetilde{\mathsf{X}}_{\mathrm{exp}} = \{\mathsf{X}_{\mathrm{exp}} \cup \mathsf{X}_{\mathrm{SII}}\} \setminus \mathsf{X}_{\mathrm{pen}}$
  - Assumption: $\widetilde{\mathsf{X}}_{\mathrm{exp}}$ and $\mathsf{X}_{\mathrm{pen}}$ are **independent**
  - In practice ➜ intensive Monte Carlo sampling to estimate this multidimensional integral

oticscream

Numerical illustration with 'oticscream'

- oticscream: a Python module based on OpenTURNS
  - A full Python package that uses/relies on OpenTURNS
  - Implements the full methodology proposed in [MIC22] but with a few more extensions!
  - Available ➜ https://github.com/vchabri/oticscream

Numerical illustration with ‘oticscream’

- **Let’s imagine a role-playing game!**
  - You’re an R&D engineer. Your colleague physicist gives you:
    - A Monte Carlo (i.i.d.) sample of size $n = 250$
    - A list of input variables (15 random ones vs. 5 scenario parameters)
    - A target risk measure for its studies: $q_{0.90}(Y)$

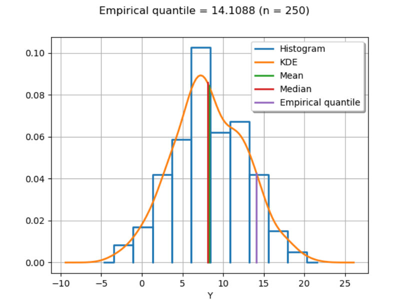
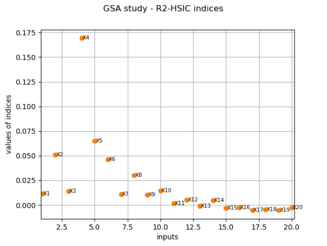


Figure 3: Empirical distribution of the output.

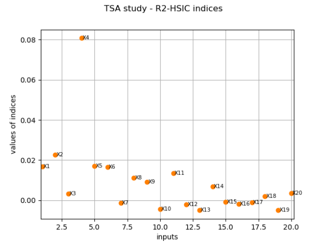
- **Step 1:**
  - Estimate the QoI: $\widehat{q}_{0.90}(Y) \approx 14.11$

Numerical illustration with ‘oticscream’

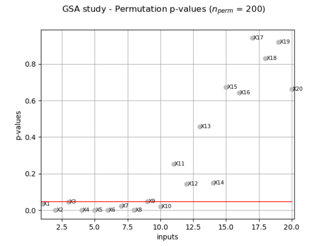
❑ Step 2: perform GSA and TSA using HSIC-based indices



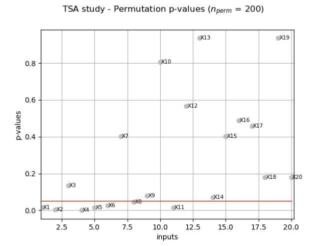
(a)



(b)



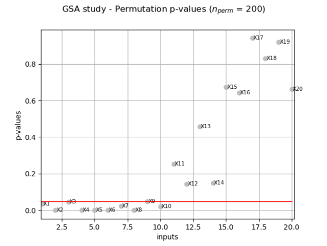
(c)



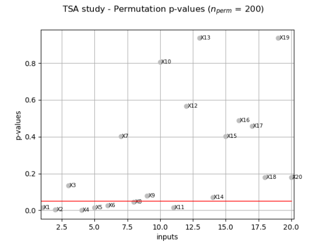
(d)

Numerical illustration with ‘oticscream’

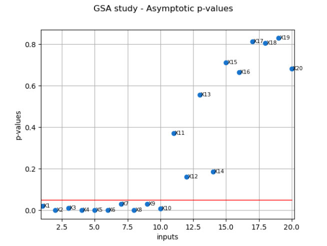
- Step 2: perform GSA and TSA using HSIC-based indices



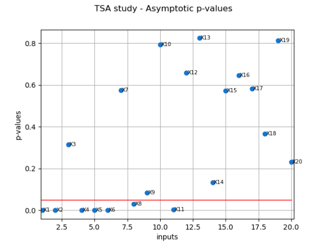
(e)



(f)

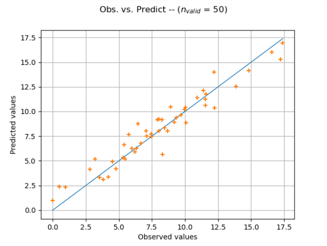


(g)

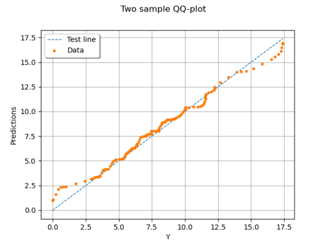


(h)

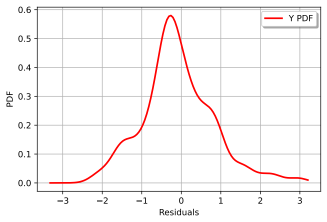
- Step 3: you build the GP metamodel and validate it

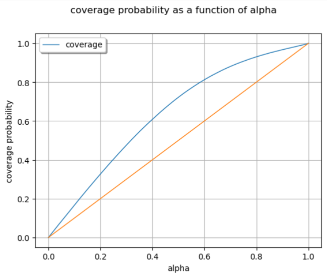


(i)

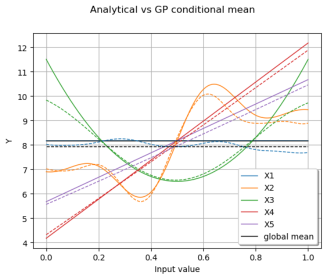


(j)

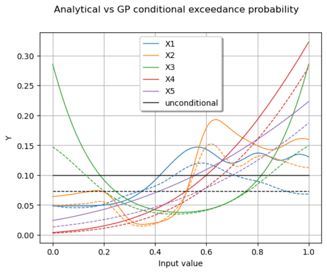




- Step 4: use the conditional GP metamodel



(m)



(n)

Conclusion

Conclusion

- Summary
  - Main goal: find penalizing configurations for some scenario parameters
  - ICSCREAM: a four-step methodology
  - Uses both state-of-the-art UQ techniques and more refined tools

    ➜ Talk by M. Keller and F. Delcoigne (JU-OT 2025)

Conclusion

❑ Limitations and perspectives

- ➤ This is a first attempt to “automate” a recipe for finding penalizing values in a **given-data** context!
- ➤ **However**, in practice, it is difficult to provide global guarantees
  ➜ every step needs to be carefully controlled in practice
- ➤ About GSA / TSA using HSIC:
  - Use advanced multiple-testing strategies for robust selection
  - Add and use HSIC-ANOVA indices more adapted to ranking purposes!
- ➤ About metamodel training and validation:
  - Gu’s robust approaches for parameter estimation in GP regression [GWB18]
  - Take benefit from the large panel of validation metrics [DILGM22]
  - Work on kernel-herding-based validation metrics
- ➤ Enhance the testing and documentation of the Python module!
- ➤ Publish it on PyPI!

Thank your for your attention!

Any question?

F. Bachoc.
Cross Validation and Maximum Likelihood estimations of hyper-parameters of Gaussian processes with model misspecification.
Computational Statistics and Data Analysis, 66:55–69, 2013.

E. Borgonovo and E. Plischke.
Sensitivity analysis: a review of recent advances.
European Journal of Operational Research, 248:869–887, 2016.

S. Da Veiga, F. Gamboa, B. Iooss, and C. Prieur.
Basics and Trends in Sensitivity Analysis. Theory and Practice in R.
Society for Industrial and Applied Mathematics, Philadelphia, PA, 2021.

References ii

C. Demay, B. Iooss, L. Le Gratiet, and A. Marrel.
Model selection based on validation criteria for Gaussian process regression: An application with highlights on the predictive variance.
Quality and Reliability Engineering International, 38:1482–1500, 2022.

S. Da Veiga.
Global sensitivity analysis with dependence measures.
Journal of Statistical Computation and Simulation, 85(7):1283–1305, 2015.

- E. Fekhari, B. Iooss, J. Muré, L. Pronzato, and M.-J. Rendas.
  Studies in Theoretical and Applied Statistics, chapter "Model predictivity assessment: incremental test-set selection and accuracy evaluation".
  Springer Proceedings in Mathematics & Statistics 406. Springer, 2022.
- A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf.
  Measuring statistical dependence with hilbert-schmidt norms.
  In *International conference on algorithmic learning theory*, volume 16, pages 63–78. Springer, 2005.
- M. Gu, X. Wang, and J. O. Berger.
  Robust Gaussian stochastic process emulation.
  The Annals of Statistics, 46(6A):3038 – 3066, 2018.

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, 214:107711, 2021.

K. Muandet, K. Fukumizu, B. Sriperumbudur, and B. Schölkopf.
Kernel Mean Embedding of Distributions: A Review and Beyond.
Foundations and Trends® in Machine Learning, 46(1-2):1–141, 2017.

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(3):301–321, 2022.

A. Marrel, B. Iooss, F. Van Dorpe, and E. Volkova.
An efficient methodology for modeling complex computer codes with Gaussian processes.
Computational Statistics and Data Analysis, 52:4731–4744, 2008.

G. Perrin and G. Defaux.
Efficient Evaluation of Reliability-Oriented Sensitivity Indices.
Journal of Scientific Computing, pages 1–23, 2019.

C. Prieur and S. Tarantola.
Variance-based sensitivity analysis: theory and estimation algorithms.
In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Handbook of Uncertainty Quantification*, chapter 35, pages 1217–1239. Springer International Publishing, 2017.

S. Rahman.
The f-Sensitivity Index.
SIAM/ASA Journal of Uncertainty Quantification, 4:130–162, 2016.

Y. Richet and V. Bacchi.
Inversion Algorithm for Civil Flood Defense Optimization: Application to Two-Dimensional Numerical Model of the Garonne River in France.
Frontiers in Environmental Science, 7:160, 2019.

H. Raguet and A. Marrel.
Target and conditional sensitivity analysis with emphasis on dependence measures.
ArXiv e-prints, pages 1–48, 2018.

C. E. Rasmussen and C. K. I. Williams.
Gaussian Processes for Machine Learning.
The MIT Press, 2006.

A. Spagnol, R. Le Riche, and S. Da Veiga.
Global sensitivity analysis for optimization with variable selection.
SIAM/ASA Journal of Uncertainty Quantification, 7(2):417–443, 2019.