# Statistical developments for target and conditional sensitivity analysis: application on safety studies for nuclear reactor

Amandine Marrel, Vincent Chabridon

# Statistical developments for target and conditional sensitivity analysis: application on safety studies for nuclear reactor

**Amandine Marrel[a,b,*], Vincent Chabridon[c]**

[a] *CEA, DES, IRESNE, DER, F-13108 Saint-Paul-lez-Durance, France*
[b] *Institut de Mathématiques de Toulouse, F-31062 Toulouse, France*
[c] *EDF R&D, 6 Quai Watier, 78401 Chatou, France*

**Abstract**

Numerical simulators are essential for understanding, modeling and predicting physical phenomena. However, the available information about some of the input variables is often limited or uncertain. *Global sensitivity analysis* (GSA) then aims at determining (qualitatively or quantitatively) how the variability of the inputs affects the model output. However, from reliability and risk management perspectives, GSA might be insufficient to capture the influence of the inputs on a restricted domain of the output (e.g., a distribution tail). To remedy this, we define and use in this work *target* (TSA) and *conditional sensitivity analysis* (CSA), which aim respectively at measuring the influence of the inputs on the occurrence of the critical event, and on the output within the critical domain (ignoring what happens outside). As illustrated in the applications, these two notions can widely differ.

From existing GSA measures, we propose new operational tools for TSA and CSA. We first focus on the popular Sobol indices and show their practical limitations for both TSA and CSA. Then, the *Hilbert-Schmidt Independence Criterion* (HSIC), a dependence measure recently adapted for GSA purposes and well-suited for small datasets, is considered. TSA and CSA adaptations of Sobol and HSIC indices, and associated statistical estimators, are defined. Alternative CSA Sobol indices are thus defined to overcome the dependence of inputs induced by the conditioning. Moreover, to cope with the loss of information (especially when the critical domain is associated to a low probability) and reduce the variability of estimators, transformation of the output using weight functions is also proposed.

These new TSA and CSA tools are tested and compared on analytical examples. The efficiency of HSIC-based indices clearly appear, as well as the relevancy of smooth relaxation. Finally, these latter indices are applied and interpreted on a nuclear engineering use case simulating a severe accidental scenario on a pressurized water reactor.

*Keywords:* Sensitivity analysis, Reliability analysis, Sobol indices, Hilbert-Schmidt independence criterion.

*Corresponding author
*Email address:* amandine.marrel@cea.fr (Amandine Marrel)

## 1. Introduction

Nowadays, many phenomena are modeled by mathematical equations which are implemented to obtain complex numerical simulators. These computer codes are used to model and predict some underlying physical phenomena. Finally, the analysis of the simulation results can be helpful for decision-making, especially when decisions involve important financial, societal and safety stakes. However, these codes often take a large number of input parameters characterizing the studied phenomenon or related to its physical and numerical modeling. The available information about some of these parameters is often limited or uncertain. The uncertainties mainly arise from the lack of knowledge about the underlying physics and about the characterization of the input parameters of the model. There are also additional sources of uncertainty arising from the particular choice of conception or scenario parameters. Consequently, many of the input parameters are uncertain (or considered as such) and it is important to assess how these uncertainties can affect the model output. Sensitivity analysis (SA) methods are performed to evaluate how input uncertainties contribute, qualitatively or quantitatively, to the variation of the output. In a probabilistic framework, the uncertain parameters are modeled by random variables characterized by probability distributions. In this paper, no distinction is made between the modeling of epistemic nor aleatory uncertainties as the proposed methods remain blind to this kind of distinction (as soon as the probabilistic framework holds).

Sensitivity analysis aims at determining how the variability of the input parameters affects the value of the output or the quantity of interest [1, 2, 3]. It thus allows to identify and perhaps quantify, for each input parameter or group of parameters, its contribution to the variability of the output. Many authors agree to distinguish several purposes for SA called "SA-settings" in the literature [1, 4]. The purpose can be to prioritize input parameters in order of influence on the output variability, or to separate the inputs into two groups: those which mostly influence the output uncertainty and those whose influence can be neglected. This task is known as "screening". Another one, called "factor mapping", aims at getting a finer identification of functional relationship between some specific values in input and output regions of interest. This last use of SA consists in determining which values of these inputs are responsible of the occurrence of the phenomenon in a given domain. More generally, SA results provide valuable information for the impact of uncertain inputs, the comprehension of the model and the underlying physical phenomenon. It can also be used for various purposes: reducing uncertainties by targeting characterization efforts on most influential inputs, simplifying the model by setting non-influential inputs to reference values, or validating the model with respect to (w.r.t.) the phenomenon under study.

In this work, we focus on specific domains of values of the phenomenon and we want to determine which inputs contribute the most in the occurrence of the phenomenon in a given domain. For this, we first define the notion of *target sensitivity analysis* (TSA) which aims at measuring the influence of the inputs over a restricted domain of the studied phenomenon, and in particular over the *occurrence* of the phenomenon in

this restricted domain. Such domain of interest would usually be extreme and relatively rare, constituting a risk or an opportunity. It will be called a *critical domain* in this paper. Alternatively, we also define the *conditional sensitivity analysis* (CSA) which evaluates the influence of the inputs on the output *within* this critical domain only, ignoring what happens outside. Let us underline that those two notions can widely differ. Note that this point will be further illustrated by the numerical applications proposed in this paper.

In this framework, we aim at proposing new methods and tools for both TSA and CSA purposes. Global SA (GSA) has been an active research field for several decades but, to the best of our knowledge, it seems that target SA, in the sense that we understand, has only been deeply studied in the reliability community (see, e.g., [5] for a review) but still remains an open field in the SA community. Unlike a previous working report by one of the authors [6], which provides much more theoretical arguments and a more extensive analysis, the goal here is to propose operational tools for both TSA and CSA and to illustrate how these tools can be efficiently used to treat industrial applications. Some new estimators are also proposed. In addition, the practical characteristics of TSA and CSA tools are further investigated, such as convergence according to the size of the sample. Besides, the applications we have to deal with, are mostly involving expensive-to-evaluate complex simulators with a large number of inputs (e.g., from a few dozens to a hundred). Thus, one often has only access to a limited number of code simulations (e.g., from a hundred to a thousand samples). Consequently, these core constraints have to be taken into account when selecting and proposing dedicated tools. Finally, we also want to illustrate how the information provided can be used in a complementary way for physical interpretation.

In the next section, we first propose a brief review on GSA approaches before focusing on existing tools for TSA. Our contributions in this framework are then introduced. Then, in Section 3, we propose a dedicated framework for TSA and CSA and then introduce several dedicated extensions of usual GSA measures. More precisely, we focus on the usual and widely used Sobol indices and a dependence measure, namely the *Hilbert-Schmidt independence criterion* (HSIC). The new proposed TSA and CSA tools are then tested and compared on two analytical examples in Section 4. Finally, in Section 5, a further application of the most relevant and adapted tools is proposed, on a use case simulating a severe accidental scenario on nuclear reactor.

Before that, we introduce a few notations. Mathematically, the numerical simulator (or model) can be modeled by assuming a deterministic input-output function $\mathcal{M}(\cdot)$ given by:

$$\mathcal{M} : \left| \begin{array}{ccc} \mathcal{X} & \longrightarrow & \mathcal{Y} \\ \mathbf{X} & \longmapsto & Y = \mathcal{M}(\mathbf{X}) \end{array} \right. \tag{1}$$

where $Y$ is the output variable of interest (considered as a single scalar output here). It is assumed that the methodology proposed in this paper is non-intrusive w.r.t. the model. The uncertain inputs are supposed to be independent and are treated in a probabilistic framework by assuming, first, a probability space

$(\Omega, \mathcal{A}, \mathbb{P})$. The inputs are gathered in a $d$-dimensional random vector $\mathbf{X} := (X_1, X_2, \ldots, X_d)^\top$ with finite second moments ($\mathbf{X} \in L_2(\mathbb{P})$) and distributed according to a continuous joint probability distribution $P_{\mathbf{X}} := \prod_{i=1}^{d} P_{X_i}$ over a measurable space $\mathcal{X} := \prod_{i=1}^{d} \mathcal{X}_i$ with $\mathcal{X} \subseteq \mathbb{R}^d$. For each realization of the input vector $\mathbf{X}(\omega)$, denoted by $\mathbf{x} := (x_1, x_2, \ldots, x_d)^\top \in \mathbb{R}^d$, an observed scalar output value $y = \mathcal{M}(\mathbf{x})$ is obtained. Thus, by propagating the uncertainties through $\mathcal{M}(\cdot)$, one can assume a probabilistic structure for the output which is a random variable characterized by a distribution $P_Y$ over a measurable space $\mathcal{Y} \subseteq \mathbb{R}$. We also assume that $Y \in L_2(\mathbb{P})$. Finally, the restricted or critical domain of interest previously mentioned is noted $\mathcal{C} \subset \mathcal{Y}$ and associated to a *critical probability* $\mathbb{P}(Y \in \mathcal{C}) = P_Y(\mathcal{C})$.

## 2. From global to target sensitivity analysis: a brief review

### *2.1. Global sensitivity analysis: from Sobol indices to HSIC*

To assess and quantify the global impact of each input uncertainty on the output, statistical methods have been developed for GSA purposes ([2, 3]). These methods are mostly based on the use of Monte Carlo simulations obtained from the model, i.e., on a random sampling of inputs according to their probability distributions. Common GSA methods include the *Derivative-based Global Sensitivity Measures* (known as "DGSM indices", see [7] for a review). The construction of these indices is based on a generalization of local sensitivity measures by averaging partial derivatives w.r.t. each input over its range of variation. However, estimating these indices requires a large number of code calls, which considerably limits their use in the case of expensive models. To overcome this disadvantage, efficient estimation strategies based on the use of metamodels have been proposed in the literature (see, e.g., [8] for the use of polynomial chaos expansions and [9] for the use of Gaussian process regression). Another widely used approach for GSA relies on the decomposition of the output variance (called the "ANOVA decomposition" for ANalysis Of VAriance), originally introduced by [10], where each term of the decomposition represents the part of the contribution of an input (or a group of inputs) to the output variance. Sensitivity indices, namely the *Sobol indices*, are then directly derived from this decomposition ([11, 12]). On the one hand, these indices are easily interpretable which made them very popular in many research fields. On the other hand, their expressions involve multidimensional integrals whose estimation by Monte Carlo methods requires, in practice, a large number of model simulations (several tens of thousands). Their direct estimation is thus intractable for expensive-to-evaluate simulators under strict budget constraints. Several studies have proposed improvements, e.g., either by using quasi Monte Carlo sampling schemes or by constructing more efficient estimators (see [13] for a review). Despite that, the number of model calls using these methods is still rather high and again, a possible option is to estimate these indices using metamodels (see, e.g., [14, 15, 16]).

*Dependence measures*, recently introduced in the GSA community by [17] and [18], enable to overcome several of the limitations listed above. First, these measures quantify, from a probabilistic point of view, the

dependence between each input and the output of interest. Reciprocally, they allow (under some assumptions that will be detailed further) to fully characterize the independence between the two variables considered (the nullity of the measure being in this case equivalent to the independence). These measures can be used quantitatively to prioritize the inputs in order of influence on the output, as well as qualitatively to perform the screening of inputs, for instance by using statistical tests like those proposed by [19] or [20]. Among the existing dependence measures in the literature, we can first mention the *dissimilarity measures* (between distributions) introduced by [21]. The underlying idea consists in comparing the probability distribution of the output with its distribution when a given input is fixed. These measures actually belong to a broader class based on Csiszár $f$-divergence ([22]). This latter includes several older notions of dissimilarity between distributions, such as the Hellinger distance ([23]), the Kullback-Leibler divergence ([24]) or the total variation distance ([25]). Moreover, in [17], the author also highlights the links between Csiszár $f$-divergences and the mutual information introduced by [26] as well as with the least-squares mutual information ([27]). Despite their interesting theoretical properties, the estimation of measures based on Csiszár $f$-divergences is, in practice, costly in terms of the number of simulations, particularly for large dimensional problems. Note that, even a first-order Sobol index can also be defined as a very simple dissimilarity measure ([17]).

Finally, other dependence measures whose estimation suffers less from the so-called "curse of dimensionality" have been investigated by [17]. Among them, one can mention the *distance correlation* which is based on characteristic functions ([28]). It has been shown that this measure has good properties for testing the independence between two random variables for large dimensional problems ([29, 30]). Moreover, this measure turns out to be a special case of a larger class of dependence measures ([29]), built upon the use of mathematical objects called "*characteristic kernels*" ([31]). These characteristic-kernel-based dependence measures are highly effective for testing the independence between random variables of various types (e.g., scalars, vectors, categorical variables). Among them, the *Hilbert-Schmidt Independence Criterion*, denoted "HSIC" ([32]), generalizes the notion of covariance between two random variables and thus enables to capture a very wide spectrum of forms of dependence between the variables. For this reason, [17], then [20] investigated the use of HSIC for GSA purposes and compared it to Sobol indices. Note that the HSIC is identical to the distance covariance for a particular choice of kernels ([29]). As illustrated by [20], HSIC also has a twofold advantage in terms of estimation: first, a low estimation cost (in practice, a few hundred simulations compared to several tens of thousands for Sobol indices) and second, its estimation for all inputs does not depend on the number of inputs $d$. In addition, HSIC-based statistical independence tests have also been developed by [19], in an asymptotic framework. More recently, extensions to a non-asymptotic framework and aggregated versions of these tests have been proposed, respectively by [20] and [33]. These works have also shown the effectiveness and highlighted the interest of using HSIC-based statistical tests for screening purposes. For all these reasons, a strong focus will be put on HSIC-based indices in the present work.

### 2.2. Regional, quantile-oriented and reliability sensitivity analysis

If GSA methods are not originally designed to achieve TSA (or CSA), it appears that a range of methods have already been proposed in literature to achieve similar goals. This subsection aims at reviewing a few of them and discussing their known advantages and limits.

#### 2.2.1. Regional sensitivity analysis

One of the first contributions in a TSA-like purpose comes from [34], motivated by environmental applications. The proposed methodology compares the distribution of the inputs within a critical domain against their distribution outside. The authors proposed to use the *Kolmogorov distance* as follows:

$$\sup_{x \in \mathcal{X}} |F_{X|Y \in \mathcal{C}}(x) - F_{X|Y \in \mathcal{Y} \setminus \mathcal{C}}(x)|$$

where $F_{X|A}$ is the cumulative distribution function (CDF) of $X$ conditioned by an event $A \in \mathcal{A}$ of nonzero probability. Such an approach is called "*regional SA*" by the authors. Note that using a similar terminology would not fully meet our purpose here because of the fact that this method focuses on SA within the critical domain (which is a CSA purpose) rather than its occurrence. Therefore, the term "regional" does not clearly make a distinction between TSA and CSA.

Using a comparison between CDFs conditionally to the critical domain seems a good choice for CSA as it involves only two conditionings, which facilitates the estimation, for instance with Monte Carlo simulations. However, one difficulty, mentioned by the authors and common to all TSA methods, arises when the critical probability $P_Y(\mathcal{C})$ is low. Another deficiency pointed out by the authors is the difficulty to study inputs in interaction. From this viewpoint, one can note that a metric comparing CDFs can be extended to a multidimensional framework, which would allow to regroup several inputs. However, the particular metric used here, namely the infinity norm over the differences, is sensitive to outliers. Both aspects make it particularly unsuitable for categorical inputs (e.g., a binary event such as a threshold being exceeded).

#### 2.2.2. Quantile-oriented sensitivity analysis

Recently, [35] proposed a generic framework called "*goal-oriented SA*". The idea is to generalize the Sobol indices, whose definition involves the calculation of the variance of a conditional expectation (e.g., , for the first-order index, one has $\mathrm{S}_1(X,Y) = \mathrm{Var}\left[\mathbb{E}\left[Y|X\right]\right] / \mathrm{Var}\left[Y\right]$). However, Sobol indices can be seen as involving a distance between expectations, simply by noticing that $\mathrm{Var}\left[\mathbb{E}\left[Y|X\right]\right] = \mathbb{E}\left[\left(\mathbb{E}\left[Y|X\right] - \mathbb{E}\left[Y\right]\right)^2\right]$. Therefore, one can extend such a definition to any other statistic defined by means of a *contrast function* $\psi(y,\theta)$. For instance, the generalization of the Sobol index to a statistic defined by another contrast function $\psi$ becomes: $\min_{\theta\in\mathbb{R}} \mathbb{E}\left[\psi(Y,\theta)\right] - \mathbb{E}\left[\min_{\theta\in\mathbb{R}} \mathbb{E}\left[\psi(Y,\theta)|X\right]\right]$, provided that the random variable $\min_{\theta\in\mathbb{R}} \mathbb{E}\left[\psi(Y,\theta)|X\right]$ is well defined. Note that the first-order Sobol index is obtained with the contrast function $\psi(y,\theta) = (y-\theta)^2$. Moreover, in order to study critical quantities of interest, the authors focus on quantiles and consider the

contrast function $\psi(y,\theta) = (y-\theta)(1_{y\leq\theta} - \alpha)$, for a given level $\alpha \in ]0,1[$. However, resulting indices turn out to be difficult to estimate in practice, as shown by [36] and [37].

Still considering quantile-oriented SA, one can mention the recent adaptation of Sobol indices proposed by [38] where expectations are replaced by quantiles: $\mathbb{E}\left[\left(F^{-1}_{Y|X}(\alpha) - F^{-1}_{Y}(\alpha)\right)^2\right]$ where $F^{-1}_{Y}$ is the generalized inverse of a CDF. The efficiency of different estimation strategies is investigated (brute force Monte Carlo and double-loop reordering approach) and some analytical comparisons with Sobol indices are proposed.

Unfortunately, these methods have some limitations w.r.t. the objectives of the present paper: first, a rather high cost of estimation; second, a less straightforward interpretation than the sensitivity of the occurrence of a phenomenon, or of the variation of a phenomenon in a critical domain.

*2.2.3. Reliability-oriented sensitivity analysis*

If TSA has not been explicitly defined in the GSA community, it appears that such a similar problem has been intensively studied in the reliability community. These methods have been often gathered under the terminology "*reliability SA*" ([39]) or "*reliability-oriented SA*" ([5, 40]).

In the first place, one can mention the ones developed in the context of *approximation methods* such as the First-/Second-Order Reliability Methods (FORM/SORM). The aim of these algorithms is to find the *most probable failure point* and then to construct a first- or second-order approximation of the failure domain boundary around this point [41]. These algorithms compute the failure probability (or a *reliability index*, which is obtained as a transformation of the failure probability) by solving an optimization problem. In such a context, gradients of the failure probability w.r.t. input distribution parameters can be directly obtained from the FORM approximation ([42]). In addition, variance-based indices (called *importance factors*) can also be obtained as a by-product of the FORM analysis ([43]). Similar indices have been recently proposed in the SORM context by [44]. Finally, one can mention the *omission sensitivity factors* introduced by [45] which consist in estimating the reliability index w.r.t. to a an input fixed at its median value.

As a second category of TSA-like methods, one can mention the ones developed in the context of *simulation methods* such as crude Monte Carlo sampling, importance sampling, subset sampling and other variants (see, e.g., [46] for a review of these algorithms). A large panel of local sensitivity indices based on derivatives of the failure probability w.r.t. various input modeling choices (e.g., distribution parameters or model parameters), have been proposed for several simulation algorithms (see, e.g., [47, 48, 49]). These algorithms rely on the calculation of *score functions* as by-products of the reliability estimation [50, 51]. If these approaches are suited for rare event estimation purposes, they remain local (in the sense of the limited part of the input space which is explored by the use of derivatives) and very specific as they only apply to failure probabilities, or rely on some strong assumptions about the phenomenon of interest (linear or quadratic behavior at failure). Thus, they do not really fulfill the TSA or CSA purposes as pursued in this paper.

Finally, a third category could be mentioned. This last one gathers a set of GSA methods which have been

adapted to the reliability context. One can mention the various extensions of the Sobol indices considering other quantities of interests than the model output (see, e.g., [52] for the failure probability case and [53] for the indicator function case). Considering the case of the Sobol indices adapted to the indicator function of the critical event (as proposed by [53] and then studied by [39]), one gets for the first-order index:

$$\mathrm{S}_1(X, 1_{\mathcal{C}}(Y)) = \frac{\mathrm{Var}\left[\mathbb{E}\left[1_{\mathcal{C}}(Y)|X\right]\right]}{\mathrm{Var}\left[1_{\mathcal{C}}(Y)\right]} \tag{2}$$

where $1_{\mathcal{C}}(y) = 1$ if $y \in \mathcal{C}$, 0 otherwise. Similar extensions can be defined to get higher order (interactions) and total Sobol indices. Efficient estimation strategies of these indices have been proposed in [54, 40]. However, as pointed out in [39, 5], these indices reach some limits in terms of interpretability when considering rare events. This is due to the fact that, most of the time, reaching the critical domain (or the failure domain) can be seen as a consequence of a specific combination of the inputs. Therefore, one can get low values for the first-order indices and total indices close to unity. Finally, as illustrated in [5, Chap. 7], estimating the total indices can be a challenging task as input dimension gets larger. To finish with, one can mention that other indices have been proposed in the literature (see, e.g., [55] for moment-independent indices and [56] for the *perturbed-law indices*, which are dedicated to robustness analysis), however, the corresponding indices are not within the scope of this paper.

As a conclusion, it appears that most of the indices presented above do not fully meet the TSA/CSA purposes. Thus, in the following, we focuse on two specific measures to provide tools for TSA and CSA: first, adapted versions of the Sobol indices on the indicator function (which will serve as a reference); second, adapted versions of the HSIC. The first choice is motivated by the popularity of the indices and their recent adaptations which have provided a first step towards TSA purposes. As for the second choice, it is justified by the good theoretical and practical properties of the HSIC and associated estimators (further detailed in subsection 3.3), even for a small learning dataset. This last point is particularly important in the case of our expensive simulators.

## 3. Proposed tools and measures for target and conditional sensitivity analysis

This section focuses on some of the tools introduced in [6], but further justifies their construction and introduces new estimators for some of them. For more theoretical considerations and discussions, the interested reader should refer to the report [6].

The model $\mathcal{M}(\cdot)$ is considered to be a best-estimate black box model for which only input-output observations (or realizations) are available. Thus, in the following, one assumes that a finite $n$-sample $\left(\mathbf{X}^{(i)}, Y^{(i)}\right)_{1\leq i\leq n}$ of the inputs/output couple, with $Y^{(i)} = \mathcal{M}(\mathbf{X}^{(i)})$, for $i = 1, \ldots, n$, is available. Moreover, the input samples $\left(\mathbf{X}^{(i)}\right)_{1\leq i\leq n}$ are *independent and identically distributed* (i.i.d.) according to the law of the inputs $P_{\mathbf{X}}$. To obtain such a sample, a crude Monte Carlo procedure is used (neither adaptive sampling nor

importance sampling methods are considered here). From a pragmatic point of view, one should note that this work falls within a more general methodological framework about uncertainty management in industrial numerical simulation (as presented in [57]). The aim is to focus on typical applications for which only a single i.i.d. input-output sample is available. This sample has to be used for all the traditional steps of the uncertainty treatment process. As an example, this sample can be used for both GSA purposes, building a metamodel and during the uncertainty propagation phase, as illustrated in [58, 59]. However, if the objective is only to perform TSA or CSA and if the sampling design can be well chosen, it is obvious that goal-oriented sampling methods (e.g., importance sampling), which would add simulations in the critical area, would be better suited and relevant to improve the estimation of TSA and CSA indices. These considerations are left for future work.

Before proposing extensions of Sobol indices and HSIC for TSA and CSA, a first subsection focuses on providing formal explanations about the core concepts of TSA and CSA and introducing a few common tools and notations extensively used in the following parts.

### *3.1. What is behind target and conditional SA?*

As introduced in Section 1, we are interested in this work to determine which inputs contribute the most to a critical phenomenon such that "$Y$ belongs to a given critical domain $\mathcal{C}$". Behind this general objective, two questions can arise:

- The most straightforward is "Which inputs influence the occurrence of the event $\{Y \in \mathcal{C}\}$?" We define it as *target sensitivity analysis* (TSA) which aims at measuring the influence of the inputs over the restricted domain of the studied phenomenon $\mathcal{C}$, and in particular over the *occurrence* of the phenomenon in this restricted domain. Naturally, this occurrence can be defined with the indicator function $1_{\mathcal{C}}(\cdot)$ of the critical domain.

- Complementary, this first problem can be completed by another type of question about the influence of the inputs on $Y$ *within* the critical domain only, ignoring what happens outside. We define this as *conditional sensitivity analysis* (CSA). In other words, knowing that we are in a given critical configuration, which input variables will drive the phenomenon.

For the first objective, the natural idea is to directly extend sensitivity measures to the binary variable $1_{\mathcal{C}}(Y)$. However, this transformation might result in a significant loss of the information conveyed by the relative values of $Y$, especially for samples which are close (but outside) the critical domain. Indeed, all the data outside $\mathcal{C}$ are summed up to zero whereas a sample very close to the critical border is much more informative than a distant one. This is all the more unfavorable when the critical probability $P_Y(\mathcal{C})$ is low and when a limited number of observations is assumed (as very often, e.g., in the context of nuclear safety

applications). Note that, even if this problem is not considered here, a "hard thresholding" is all the more questionable in the case of a noisy simulator.

To overcome such a limitation, we propose to use a *weighted thresholding transformation* (or "smooth thresholding") in order to relax the binary assumption. For this, we consider a decreasing distance $d_{\mathcal{C}} : \mathcal{Y} \to \mathbb{R}_+$ between each point and the critical domain $\mathcal{C}$. The closer is an observation to the critical domain, the more likely it is to convey similar information. By doing this, one obviously assumes some kind of regularity of the phenomenon's statistical properties. More generally, when $\mathcal{Y}$ lies in an Euclidean space, we propose to consider the weight function $w_{\mathcal{C},\exp}(y) := \exp(-d_{\mathcal{C}}(y)/s)$, where $d_{\mathcal{C}}(y) := \inf_{y' \in \mathcal{C}} \|y - y'\|$. Here, the exponential function encodes multiplicative contributions, and $s \in \mathbb{R}$ is a smoothing parameter depending typically on a measure of dispersion of the values of $Y$. Note that, other kind of relaxation function based on logistic function has been proposed by [60]), still for SA purposes, but in an optimization context. In the following, the generic notation $w : \mathcal{Y} \to [0, 1]$ will be used to denote any kind of weight functions (here, functions $1_{\mathcal{C}}(\cdot)$ or $w_{\mathcal{C},\exp}(\cdot)$). As a result, any sensitivity measure between a group of inputs $X$ and $w(Y)$ yields a "target" sensitivity measure (or index).

Now, concerning CSA, a natural idea to study the behavior of $Y$ within the critical domain consists in conditioning $Y$ by the event $\{Y \in \mathcal{C}\}$. For a given initial probability space $(\Omega, \mathcal{A}, \mathbb{P})$, if $A \in \mathcal{A}$ is an event of nonzero probability, then conditioning by $A$ simply means providing the measurable space $(\Omega, \mathcal{A})$ with the probability measure $P_{|A}$, defined as $P_{|A}(B) := \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$ for all $B \in \mathcal{A}$. Applied to $Y$ conditioned by $\{Y \in \mathcal{C}\}$, one gets: $P_{Y|\{Y \in \mathcal{C}\}}(B) = \dfrac{\int_B 1_{\mathcal{C}}(y) dP_Y}{\int_{\mathcal{Y}} 1_{\mathcal{C}}(y) dP_Y}$. The probability density of $Y|\{Y \in \mathcal{C}\}$ is therefore $\dfrac{1_{\mathcal{C}}(y) dP_Y}{P_Y(\mathcal{C})}$. The resulting probability distribution is therefore the probability distribution absolutely continuous w.r.t. $P_Y$ whose density is proportional to the indicator function, ignoring a normalization factor. Let $P_Y^{1_{\mathcal{C}}}$ denote this *probability measure $P_Y$ weighted by $1_{\mathcal{C}}$*. Now, just as for TSA where a smooth relaxation of the indicator function has been proposed, it might be useful to consider extensions of conditioning allowing to take into account some of the information outside (but closed to) the critical domain. For this, one can generalize the previous weighted probability to any weight function $w(\cdot)$ such that:

$$P_Y^w(B) = \frac{\int_B w(y) dP_Y}{\int_\Omega w(y) dP_Y} = \int_B \frac{w(y)}{\mathbb{E}\left[w(Y)\right]} dP_Y \tag{3}$$

provided that $\mathbb{E}\left[w(Y)\right]$ is not zero (i.e., $w(Y)$ is a positive nonzero random variable over $(\Omega, \mathcal{A}, \mathbb{P})$). Note that, under this formalism and provided that the expectations exist, we have for conditional expectations:

$$\mathbb{E}\left[Y | Y \in \mathcal{C}\right] = \mathbb{E}_{Y \sim P_Y^w}[Y] = \frac{\int_{\mathcal{Y}} Y w(Y) dP_Y}{\int_{\mathcal{Y}} w(Y) dP_Y} = \mathbb{E}\left[\frac{W(Y)}{\mathbb{E}\left[W(Y)\right]} Y\right]. \tag{4}$$

Similarly, any sensitivity measure is defined depending on a (usually implicit) probability space on $\mathbf{X}$ and $Y$. Moreover, we have $Y = \mathcal{M}(\mathbf{X})$. Therefore, when conditioning by $\{Y \in \mathcal{C}\}$, we change the underlying probability measures: $Y \sim P_Y^w$, $\mathbf{X} \sim P_{\mathbf{X}}^w$, and $(\mathbf{X}, Y) \sim P_{(\mathbf{X},Y)}^w$.

### 3.2. TSA and CSA from Sobol indices

#### 3.2.1. Reminders about Sobol indices and their estimators

Assumed that inputs are independent, and following the ANOVA decomposition, first-order Sobol indices can be defined for any input $(X_i)_{1\leq i\leq d}$ by:

$$\mathrm{S}_1(X_i, Y) = \frac{\mathrm{Var}\left[\mathbb{E}\left[Y|X_i\right]\right]}{\mathrm{Var}\left[Y\right]}. \tag{5}$$

Any higher order index can be defined similarly, as well as any total index, usually denoted $\mathrm{S}_T$ ([13]).

In the general framework, the estimation of Sobol index involves the expensive estimation of conditional expectation (e.g., $\mathbb{E}\left[\mathbb{E}\left[Y|X_i\right]^2\right]$). To overcome this limitation, well-known *pick-freeze* approaches have been proposed ([13]). By denoting $\mathbf{X}_{-i} = \{X_j\}_{1\leq j\leq d} \setminus X_i$ and remembering that $Y = \mathcal{M}(\mathbf{X}) = \mathcal{M}(X_i, \mathbf{X}_{-i})$, pick-freeze estimators are based on the judicious following decomposition:

$$\mathrm{Var}\left[\mathbb{E}\left[Y|X_i\right]^2\right] = \mathrm{Cov}\left[\mathcal{M}(X_i, \mathbf{X}_{-i}), \mathcal{M}(X_i, \mathbf{X}'_{-i})\right] \tag{6}$$

where $\mathbf{X}'_{-i}$ is an i.i.d. copy of $\mathbf{X}_{-i}$. This decomposition is valid if $Y = \mathcal{M}(X_i, \mathbf{X}_{-i})$ and $Y' = \mathcal{M}(X_i, \mathbf{X}'_{-i})$ are of the same law and independent conditionally to $X_i$. This condition is ensured if $X_i$ and $\mathbf{X}_{-i}$ are independent. From this decomposition, natural estimators can be deduced, considering an i.i.d. $n$-sample of the corresponding outputs $(Y, Y')$. Note that this method yields a total cost of model evaluation in $\mathcal{O}(nd)$ to compute the full set of first-order and total indices (more details and associated references can be found in [13]).

#### 3.2.2. Adaptation for TSA

As mentioned previously, Sobol indices can be directly extended to the binary transformation (see Eq. (2)) and more generally with any transformation $w(\cdot)$. For any input $(X_i)_{1\leq i\leq d}$, the first-order target Sobol index, denoted T-$\mathrm{S}_{1,w}$, is given by:

$$\text{T-}\mathrm{S}_{1,w}(X_i, Y) = \frac{\mathrm{Var}\left[\mathbb{E}\left[w(Y)|X_i\right]\right]}{\mathrm{Var}\left[w(Y)\right]}. \tag{7}$$

Any higher order index can be defined similarly, as well as any total index.

#### 3.2.3. Adaptation for CSA

Based from the notations introduced in subsection 3.1, the conditional first-order Sobol index is given by considering $\mathrm{S}_1(X_i, Y)$ under $(X_i, Y) \sim P^w_{(X_i,Y)}$ (and similarly for any higher order index). Even if the definition of this index is theoretically possible[1], this induces somes problems in practice in terms of estimation. Indeed, even if the inputs are initially independent under $P_{\mathbf{X}}$, they are (usually) not anymore under

[1] Putting aside the fact that such an index could lead sometimes to a wrong interpretation, as the ANOVA and the unicity of the decomposition are not ensured anymore.

$P^w_{\mathbf{X}}$. Consequently, usual pick-freeze estimators cannot be used anymore. Solutions proposed for dependent inputs could be adapted ([61]), but they would probably require to explicit the conditional probability $P^w_{\mathbf{X}}$ and especially the conditional covariance which is almost impossible in practice.

To overcome this limitation and avoid being under $P^w_{\mathbf{X}}$, we propose to consider an alternative transformation of $Y$, while staying under probability $P_{\mathbf{X}}$. For this, a first idea could be to consider the transformation $w(Y)Y$ to account for the belonging to the critical region but also the value of $Y$. However, the fact that $w(Y)Y$ vanishes away from the critical domain seems arbitrary: the value zero might not be meaningful w.r.t. the possible values of $Y$. Since Sobol index is a measure of variance, it still seems relevant to replace the value zero by a constant value over, but equal to the expectation of the resulting transformation. Consequently, the regions where $w(Y)$ vanishes, would then not contribute to the variance of the phenomenon. To obtain this, we define the following transformation:

$$\widetilde{Y}_w = w(Y)Y + (1 - w(Y))y_0 \quad \text{where} \quad y_0 = \frac{\mathbb{E}\left[w(Y)Y\right]}{\mathbb{E}\left[w(Y)\right]}, \quad \text{in order to have} \quad y_0 = \mathbb{E}\left[\widetilde{Y}_w\right]. \tag{8}$$

Observe that $y_0$ is also $\mathbb{E}_{Y\sim P^w_Y}[Y]$. For the particular case $w(\cdot) = 1_{\mathcal{C}}(\cdot)$, we logically obtain $y_0 = \mathbb{E}\left[Y|Y \in \mathcal{C}\right]$.

Building Sobol indices from $\widetilde{Y}_w$, we thus obtain what we call *hybrid conditional* Sobol indices:

$$\text{C-S}_{1,w}(X_i, Y) = \frac{\text{Var}\left[\mathbb{E}\left[\widetilde{Y}_w|X_i\right]\right]}{\text{Var}\left[\widetilde{Y}_w\right]}. \tag{9}$$

Any higher order index, as well as total index can be defined similarly.

Contrary to the initial formulation of conditional index, the hybrid index $\text{C-S}_{1,w}(X_i, Y)$ is defined under probability $P_{\mathbf{X}}$ and can be estimated with any usual method, such as the usual pick-freeze one.

### *3.3. TSA and CSA from HSIC*

The Hilbert-Schmidt Independence Criterion (HSIC) proposed by [62] rests upon kernel-based approaches for detecting dependence, and more particularly, on cross-covariance operators in *Reproducing Kernel Hilbert Spaces* (RKHS).

#### *3.3.1. Reminders and statistical inference about HSIC*

Let $\mathcal{F}_i : \mathcal{X}_i \to \mathbb{R}$ and $\mathcal{G} : \mathcal{Y} \to \mathbb{R}$ be two universal RKHS equipped with their kernels $\kappa_i(\cdot,\cdot)$ and $\kappa(\cdot,\cdot)$, and dot products $\langle\cdot,\cdot\rangle_{\mathcal{F}_i}$ and $\langle\cdot,\cdot\rangle_{\mathcal{G}}$ respectively. Under these hypotheses, the generalized *cross-covariance* operator $C_{(X_i,Y)}[\cdot]$ is defined, for any functions $f \in \mathcal{F}_i$ and $g \in \mathcal{G}$, by:

$$\langle f, C_{(X_i,Y)}[g]\rangle_{\mathcal{F}_i} = \text{Cov}\left(f(X_i), g(Y)\right) \tag{10}$$

which generalizes the notion of covariance between $X_i$ and $Y$. Therefore, a larger panel of input-output dependency can be captured by this operator. The HSIC is then defined as as the square Hilbert-Schmidt

norm of the cross-covariance operator:

$$\mathrm{HSIC}(X_i, Y)_{\mathcal{F}_i,\mathcal{G}} = ||C_{(X_i,Y)}||^2_{\mathrm{HS}} = \sum_{l,m} \langle u_l, C_{(X_i,Y)}[v_m] \rangle_{\mathcal{F}_i} \tag{11}$$

where $(u_l)_{l\geq 0}$ and $(v_m)_{m\geq 0}$ are orthonormal bases of, respectively, $\mathcal{F}_i$ and $\mathcal{G}$.

Due to RKHS properties (known as the "kernel trick"), [62] show that HSIC can be expressed only using kernels:

$$\begin{aligned}\mathrm{HSIC}(X_i, Y)_{\mathcal{F}_i,\mathcal{G}} = \mathbb{E}\left[\kappa_i(X_i, X_i^{'})\kappa(Y, Y^{'})\right] + \mathbb{E}\left[\kappa_i(X_i, X_i^{'})\right]\mathbb{E}\left[\kappa(Y, Y^{'})\right] \\ - 2\mathbb{E}\left[\kappa_i(X_i, X_i^{'})\kappa(Y, Y^{''})\right]\end{aligned} \tag{12}$$

where $(X_i^{'}, Y^{'})$ and $(X_i^{'}, Y^{''})$ are i.i.d. copies of $(X_i, Y)$. This last equation highlights interesting estimation properties since it only involves expected values which are simpler to estimate than variances of conditional expected values (such as in the Sobol indices).

Authors of [62] propose to estimate each $\mathrm{HSIC}(X_i, Y)$ by the following estimator:

$$\widehat{\mathrm{HSIC}}(X_i, Y) = \frac{1}{n^2}\sum_{1\leq k,j\leq n}(L_i)_{k,j}L_{k,j} + \frac{1}{n^4}\sum_{1\leq k,j,q,r\leq n}(L_i)_{k,j}L_{q,r} - \frac{2}{n^3}\sum_{1\leq k,j,r\leq n}(L_i)_{k,j}L_{j,r} \tag{13}$$

where $L_i$ and $L$ are the Gram matrices defined for all $k, j \in \{1, \ldots, n\}$ by $(L_i)_{k,j} = \kappa_i\left(X_k^{(k)}, X_k^{(j)}\right)$ and $(L)_{k,j} = \kappa\left(Y^{(k)}, Y^{(j)}\right)$.

This V-statistic estimator can also be written in the following more compact form (see [62]):

$$\widehat{\mathrm{HSIC}}(X_i, Y) = \frac{1}{n^2}\mathrm{Tr}(L_i H L H) \tag{14}$$

where $\mathrm{Tr}(\cdot)$ is the usual trace operator and $H$ the matrix whose components are defined for all $k, j \in \{1, \ldots, n\}$ by $H_{k,j} = \delta_{k,j} - 1/n$, with $\delta_{k,j}$ the Kronecker symbol between $k$ and $j$ which is equal to 1 if $i = j$ and 0 otherwise.

Note that neither the HSIC definition nor its estimator requires the independence of inputs.

From this HSIC, [17] defines a normalized (with values belonging to $[0, 1]$) sensitivity index which makes its interpretation easier:

$$R^2_{\mathrm{HSIC}}(X_i, Y) = \frac{\mathrm{HSIC}(X_i, Y)}{\sqrt{\mathrm{HSIC}(X_i, X_i)\ \mathrm{HSIC}(Y, Y)}}. \tag{15}$$

In practice, $R^2_{\mathrm{HSIC}}(X_i, Y)$ is estimated via a plug-in approach from $\widehat{\mathrm{HSIC}}(X_i, Y)$.

Finally, a fundamental property of HSIC is that $\mathrm{HSIC}(X_i, Y)_{\mathcal{F}_i,\mathcal{G}} = 0$ if and only if $X_i$ and $Y$ are independent, as long as the associated RKHS are universal. For example, Gaussian and Laplace kernels generate universal RKHSs. The Gaussian kernel, which is classically used for real variables, is defined by:

$$\kappa(x, x^{'}) = \exp\left(-\lambda\|x - x'\|_2^2\right) \tag{16}$$

where $\lambda$ is a fixed positive real parameter, also called the "bandwidth parameter" of the kernel. Usually, one uses in practice $\lambda = 1/\sigma^2$ with $\sigma^2$ being the empirical variance of the sample of $X$. The property of formally characterizing the independence makes the HSIC naturally relevant for GSA purpose. Moreover, statistical independence tests based on HSIC can be built to make the interpretation more robust and perform an objective screening despite the potential fluctuations of the estimates ([20]).

*3.3.2. Adaptation for TSA*

HSIC (and $R^2_{\text{HSIC}}$) can be directly extended to any transformation $w(Y)$. For any input $(X_i)_{1\leq i\leq d}$, the target indices are given by:

$$\text{T-HSIC}_w(X_i, Y) = \text{HSIC}(X_i, w(Y)); \tag{17}$$

$$\text{T-}R^2_{\text{HSIC},w}(X_i, Y) = \frac{\text{HSIC}(X_i, w(Y))}{\sqrt{\text{HSIC}(X_i, X_i)\ \text{HSIC}(w(Y), w(Y))}}. \tag{18}$$

Estimators similar to Eq. (14) can be directly built. The only precaution lies in the adaptation of the kernel for the transformation of $Y$ if the binary transformation $w(\cdot) = 1_{\mathcal{C}}(\cdot)$ is used. In this case, the use of the categorical kernel $\kappa(z, z^{'}) = \delta_{zz'}$ defined for a binary variable $z$, is recommended, as underlined by [17] and [63]. On the other hand, for the transformation $w_{\mathcal{C},\exp}(\cdot)$, any characteristic kernel dedicated to real variables, such as Gaussian or Laplace kernels, can be used.

*3.3.3. Adaptation for CSA*

For its conditional version, $\text{HSIC}(X_i, Y)$ must be estimated under $(X_i, Y) \sim P^w_{(X_i,Y)}$. Contrary to Sobol indices, the problem of non-independence of the inputs under $P^w_{\mathbf{X}}$ no longer exits, since independence of the inputs is not assumed in the HSIC definition. Regarding its estimation, since HSIC can be defined through kernel distances and expressed as expectations of kernels (Eq. (12)), we obtain from Eq. (4) (Section 3.1):

$$\begin{aligned}
\text{HSIC}_{(X_i,Y)\sim P^w_{(X_i,Y)}}(X_i, Y) &= \mathbb{E}_{(X_i,Y)\sim P^w_{(X_i,Y)}}[\kappa_i(X_i, X_i^{'})\kappa(Y, Y^{'})] + \mathbb{E}_{X_i\sim P^w_{X_i}}[\kappa_i(X_i, X_i^{'})]\mathbb{E}_{Y\sim P^w_Y}[\kappa(Y, Y^{'})] \\
&\quad - 2\mathbb{E}_{(X_i,Y)\sim P^w_{(X_i,Y)}}[\kappa_i(X_i, X_i^{'})\kappa(Y, Y^{''})] \\
&= \mathbb{E}\left[\kappa_i(X_i, X_i^{'})\kappa(Y, Y^{'})\bar{w}(Y)\bar{w}(Y')\right] \\
&\quad + \mathbb{E}\left[\kappa_i(X_i, X_i^{'})\bar{w}(Y)\bar{w}(Y')\right]\mathbb{E}\left[\kappa(Y, Y^{'})\bar{w}(Y)\bar{w}(Y')\right] \\
&\quad - 2\mathbb{E}\left[\kappa_i(X_i, X_i^{'})\kappa(Y, Y^{''})\bar{w}(Y)\bar{w}(Y'')]\right]
\end{aligned} \tag{19}$$

where $\bar{w}(Y) = \dfrac{w(Y)}{\mathbb{E}[w(Y)]}$ (remember that $(X_i^{'}, Y^{'})$ and $(X_i^{'}, Y^{''})$ are still i.i.d. copies of $(X_i, Y)$).

The two conditional indices obtained are, respectively, the conditional HSIC (denoted by $\text{C-HSIC}_w(X_i, Y)$) and the conditional $R^2_{\text{HSIC}}$ (denoted by $\text{C-}R^2_{\text{HSIC},w}(X_i, Y)$).

Similarly to Eq. (14), we propose the following estimator for $\text{C-HSIC}_w(X_i, Y)$:

$$\widehat{\text{C-HSIC}}_w(X_i, Y) = \frac{1}{n^2}\text{Tr}\left(WL_iWH_1LH_2\right), \tag{20}$$

where:

- $W$ is the matrix of empirical normalized weights defined by $W = \text{Diag}\left(\widehat{\bar{w}}(Y^{(j)})\right)_{1\leq j\leq n}$ with $\widehat{\bar{w}}(Y) = \dfrac{w(Y)}{\Sigma_{i=1}^{n} w(Y^{(i)})}$;
- $H_1 = I_n - \dfrac{1}{n}UW$ and $H_2 = I_n - \dfrac{1}{n}WU$, with $I_n$ the identity matrix of size $n$ and $U$ a matrix filled with 1.

The proof of the trace formulation of the estimator is similar than those provided in [33] (Appendix A), for the estimation of HSIC under alternative distribution of the input. One has simply to consider $dP_{\mathbf{X}}$ and $dP_{\mathbf{X}}^{w} = \dfrac{w(Y(x))}{\mathbb{E}\left[w(Y)\right]} dP_{\mathbf{X}}$ as the alternative and prior distributions, which are denoted $\tilde{f}$ and $f$ in [33] respectively.

## 4. Numerical tests

In this section, numerical illustrations and comparisons of the proposed Sobol and HSIC-based tools for TSA and CSA purposes are investigated. For this, some examples already introduced in [6] are considered. However, more extensive numerical tests are realized. These examples also demonstrate that TSA and CSA explore aspects of a model which are both different from GSA and valuable for practitioners.

For this, we consider here the following indices:

- First-order target Sobol indices: T-$\text{S}_{1,w}(X_i, Y)$, given by Eq. (7) and estimated by pick-freeze method;
- First-order hybrid conditional Sobol indices: C-$\text{S}_{1,w}(X_i, Y)$, given by Eq. (9) also estimated by pick-freeze method;
- Normalized target HSIC: T-$R^2_{\text{HSIC},w}(X_i, Y)$ given by Eq. (18) and estimated from formula (14) applied with $w(Y)$ and plug-in approach;
- Normalized conditional HSIC: C-$R^2_{\text{HSIC},w}(X_i, Y)$ defined from Eq. (19) and estimated with formula (20) and plug-in approach.

Note that the index $w$ denotes the weight function which is, in the following, either $1_{\mathcal{C}}$ or the smooth relaxation $w_{\mathcal{C}}$ defined in Eq. (21).

We suppose here that the critical domain $\mathcal{C}$ is defined by $Y$ exceeding a given critical value such that: $\mathcal{C} = \{y \in \mathcal{Y} \mid y \geq c\}$, with $c$ chosen as the 90%-quantile of $Y$, $c = F_Y^{-1}(0.9)$. Consequently, we propose to use the following $w_{\mathcal{C}}(\cdot)$ function:

$$w_{\mathcal{C}}(y) = \exp\left(-\frac{\max(c-y, 0)}{s\,\sigma_Y}\right) \tag{21}$$

where $\sigma_Y$ is an estimation of the standard deviation of $Y$, and $s = 1/5$ a tuning parameter of the smoothness, chosen so that $w_{\mathcal{C}}(\cdot)$ almost vanishes one standard deviation away from $\mathcal{C}$. Note that this value of $s$ is empirically chosen as offering a reasonable numerical trade-off. Its impact could be further investigated in future studies.

**Remark 1.** *The choice of a quantile as a threshold value is arbitrary and is used here for illustrative purposes. Another level of quantile, or more generally, another threshold value could be considered. All the tools remain generic and usable, as long as the critical domain is of nonzero probability. However, practical limits must be kept in mind if the critical domain is associated to extremely rare events: it is necessary to have a reasonable number of simulations inside or closed to the critical domain. In the case of rare events, adaptive sampling (not developed here, as explained in Section 3) will have to be considered to ensure a sufficient number of samples falling into the critical domain.*

**Remark 2.** *The relaxation function $w_{\mathcal{C}}(\cdot)$ allows a compromise between "brute" TSA and a better exploitation of the available information. It provides some kind of interpolation between GSA and TSA. This compromise is represented in Eq. (21) by the smoothing parameter $s$. As explained, we propose the reasonable choice $s = 1/5$, confirmed from our practical experience feedback. However, it might also be possible to choose $s$ so as to control that a certain percentage of data is taken into account in the TSA. For example, for a 90%-quantile, $s$ can be chosen such as $w_{\mathcal{C}} = 0.5$ for $y$ equals to the 80%-quantile. $s$ can also be adapted according to the number of simulations. Higher value of $s$ can be chosen for smaller size sample, but while keeping in mind that notion of "target" SA will be all the more relaxed. Anyway, deeper studies about the practical value of $s$ could be further investigated.*

**Remark 3.** *For the HISC, Gaussian kernels will be used for $X$ as for $Y$, with the usual parametrization $\lambda = 1/\sigma^2$ with $\sigma^2$ being the empirical variance of the sample of $X$ and $Y$, respectively. An exception will be made for the T-HSIC$_w$ when $w = 1_{\mathcal{C}}$: in this case, the categorical kernel ([63]) will be used for $w(Y)$, as previously explained in Section 3.3.2.*

### *4.1. Presentation of test case analytical functions*

To illustrate TSA and CSA and compare the different indices and estimators, we first consider two analytical models. The first one, defined in dimension $d = 2$, is given as follows:

$$\mathcal{M}_1 : \mathbf{X} \mapsto Y = \min(X_1, X_2),$$

where inputs $X_1$ and $X_2$ are two independent random variables, following respectively a standard normal distribution, and a uniform distribution over $[0, 1]$. Despite a very simple formulation, $\mathcal{M}_1$ exhibits a very strong non-linearity, as shown by Figure 1(a). Conditional distributions are also illustrated by Figure 1(c) and 1(d). The critical value, namely the 90%-quantile of $Y$, is $c \simeq 0.62$ for $\mathcal{M}_1$.

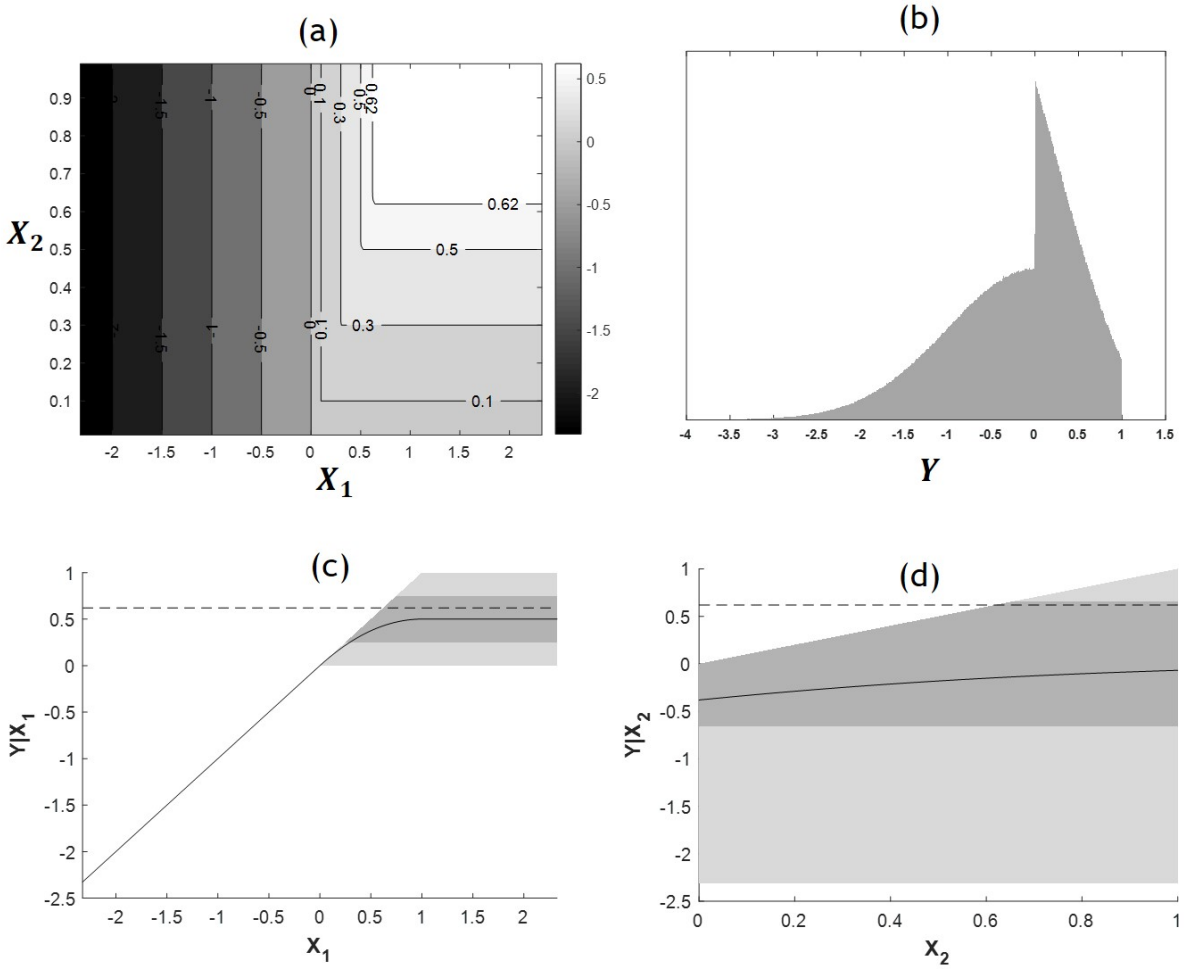



Figure 1: Model $\mathcal{M}_1$ – Contour plot (a), probability distribution of the output $Y$ (b), conditional distributions (c and d) with conditional expectation in solid line, 90%-quantile in dotted line ($c \simeq 0.62$), [25%-quantile; 75%-quantile] area in dark grey and [1%-quantile; 99%-quantile] area in light grey.

We then explore the more complicated Ishigami function which is well-known from the sensitivity analysis community. This model, illustrated by Figure 2, is defined in dimension $d = 3$ by:

$$\mathcal{M}_2 : \mathbf{X} \mapsto Y = \sin(X_1) + a \sin^2(X_2) + bX_3^4 \sin(x_1),$$

where $a, b \in \mathbb{R}_+$; all inputs $X_1, X_2$ and $X_3$ are independent and uniformly distributed over $[-\pi, \pi]$.

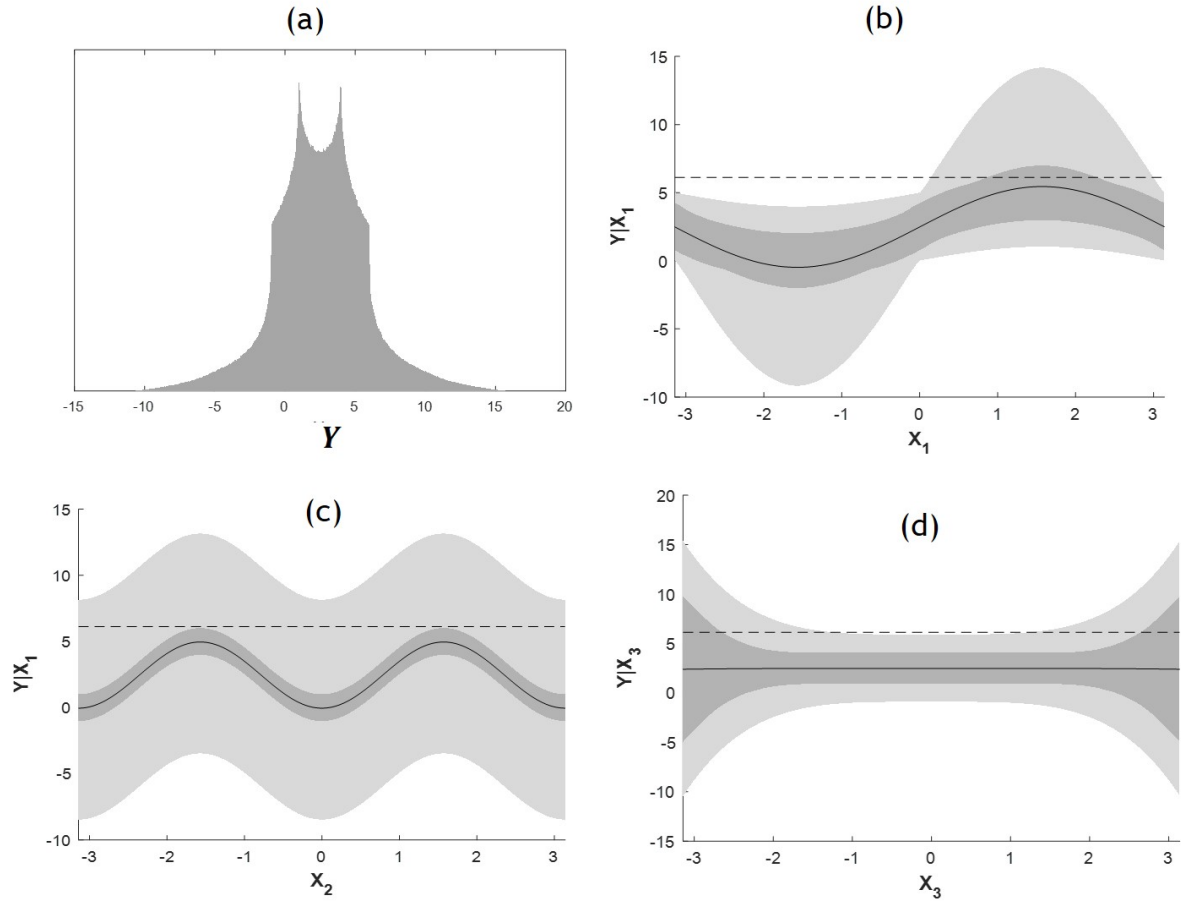



Figure 2: Model $\mathcal{M}_2$ – Contour plot (a), probability distribution of the output $Y$ (b), conditional distributions (c and d) with conditional expectation in solid line, 90%-quantile in dotted line ($c \simeq 6.12$), [25%-quantile; 75%-quantile] area in dark grey and [1%-quantile; 99%-quantile] area in light grey.

The influence of the factor $X_2$ is purely additive, its importance being modulated by the parameter $a$. The influence of the factor $X_1$ includes an additive part and an interaction with the factor $X_3$, the balance being tuned by parameter $b$. We set here the parameters $a = 5$ and $b = 0.1$. In this case, and for reference, the first-order Sobol indices are analytically known: $\mathrm{S}_1(X_1, Y) = 0.40$, $\mathrm{S}_1(X_2, Y) = 0.29$ and $\mathrm{S}_1(X_3, Y) = 0$, while total-order ones are $\mathrm{S}_T(X_1, Y) = 0.71$, $\mathrm{S}_T(X_2, Y) = 0.29$, and $\mathrm{S}_T(X_3, Y) = 0.31$. Finally, the critical value, namely the 90%-quantile of $Y$, is $c \simeq 6.12$ for $\mathcal{M}_2$.

*4.2. Numerical experiments and results*

For each test case, i.i.d. input samples of size $n = 1000$ are first simulated and propagated through the model. Then, Sobol and HSIC-based TSA and CSA indices listed at the beginning of Section 4 are estimated. The process is repeated hundred times to capture the variance due to sampling. The reference values for each indices are computed from a sample of size $n = 10^5$ (asymptotic convergence of estimators being allowed for this size). Results are given by Figures 3 and 4 for $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. Note that Sobol-based and HSIC-based indices are compared here, for one input, from a sample of same size $n$. However, it is important to keep in mind that Sobol-based indices are computed by the pick-freeze method with $n$ simulations for each input which yields a total of $nd$ simulations, while HSIC-based indices are computed with the same sample for all the inputs. This is a significant advantage of HSIC-based indices, especially for large dimensional cases.

The model $\mathcal{M}_1$ clearly illustrates that the information provided by GSA, TSA and CSA can sometimes differ significantly. Indeed, if we consider GSA, the input $X_1$ (standard normal variable) is much more important than $X_2$ (uniform variable). This is not surprising, since $X_1$ presents more variability and takes values far below and above the minimum of $X_2$. Sobol and $R^2_{\text{HSIC}}$ indices clearly reflect that.

TSA results indicate that the ordering of the inputs is the same, although the relative importance difference is lower. This can be explained by the fact that $X_1$ has a lower probability to be above the critical value $c = 0.62$ than $X_2$ (probability around 0.27 and 0.4 for $X_1$ and $X_2$, respectively), hence $X_1$ is still determining again the outcome of having $\{y \geq c\}$. But, in the same time, the variability of $X_1$ below the threshold has no influence anymore. Moreover, as highlighted by the works of [39], Sobol indices lead to high values of interaction indices (sum of first indices around 0.5), which complicates the interpretation in terms of ranking. Furthermore, the estimators of Sobol indices are much less precise than the HSIC-based indices for GSA, also and above all, for TSA. Even in the case where hard thresholding $w = 1_{\mathcal{C}}$ is considered (significant loss of information), T-$R^2_{\text{HSIC},w}$ still clearly identifies the correct order of the inputs. It also can be noted that the smoothed versions of T-$\text{S}_{1,w}$ and T-$R^2_{\text{HSIC},w}$, obtained with $w = w_{\mathcal{C}}$, present less variability while still ordering correctly the inputs. In practice, this makes possible to exploit in a more intensive manner the available information, which reduces the variance of statistical estimators. In return, smoothed indices make some kind of interpolation (or compromise) between TSA and GSA. Consequently, the tuning parameter $s$ should not be chosen too large to stay close to the TSA purpose (as stated in Remark 2, Section 4).

CSA results provide a whole different information: now $X_2$ is more important than $X_1$. Indeed, conditionally to both $X_2$ and $X_1$ being no less than $c$, $X_2$ varies in $[c, 1]$ while $X_1$ varies in $[c, +\infty]$, in such a way that the former has more chance to determine the value of their minimum. This is clearly captured by C-$\text{S}_{1,w}$ and C-$R^2_{\text{HSIC},w}$ indices, but the difference is again less marked for the former, whose accuracy

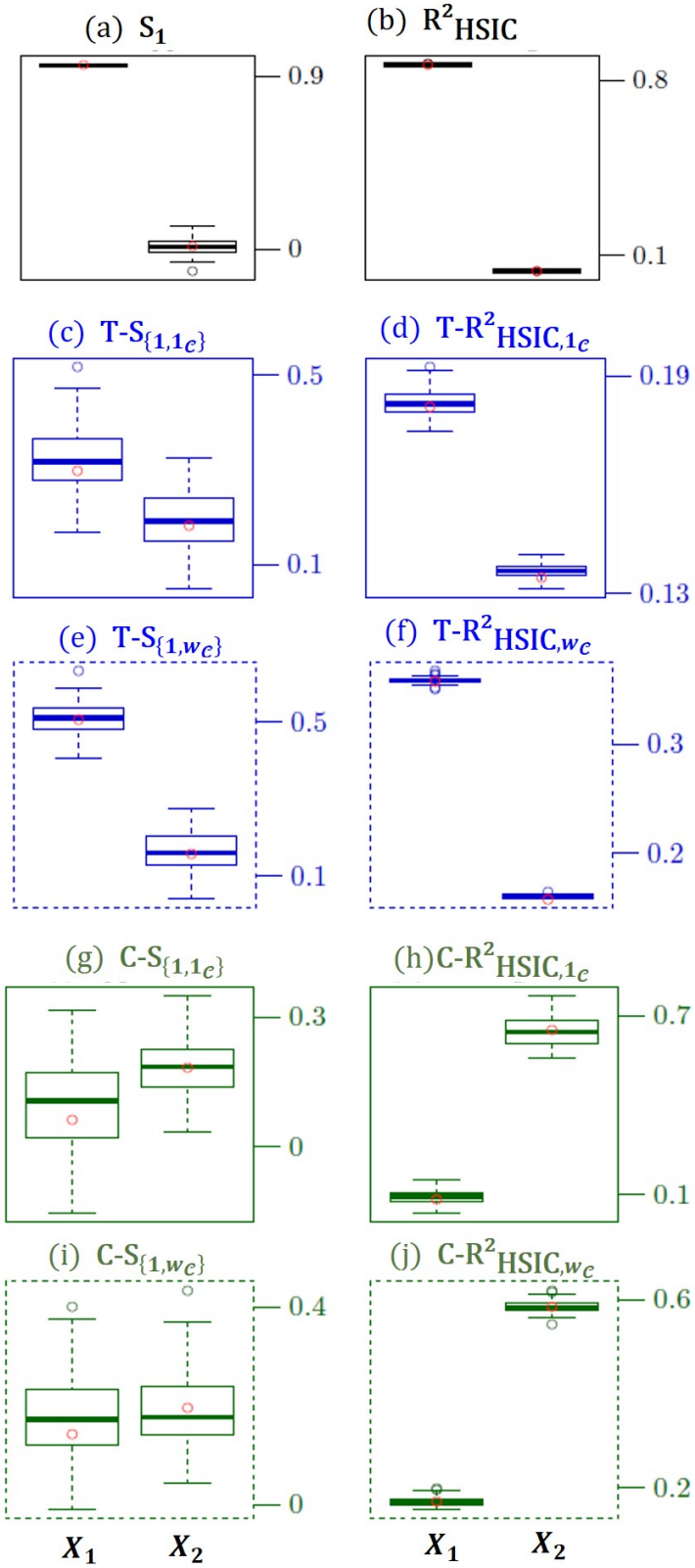



Figure 3: Model $\mathcal{M}_1$ – Global (black), target (blue) and conditional (green) SA indices, estimated from samples of size $n = 1000$. Reference values are given by red circles.

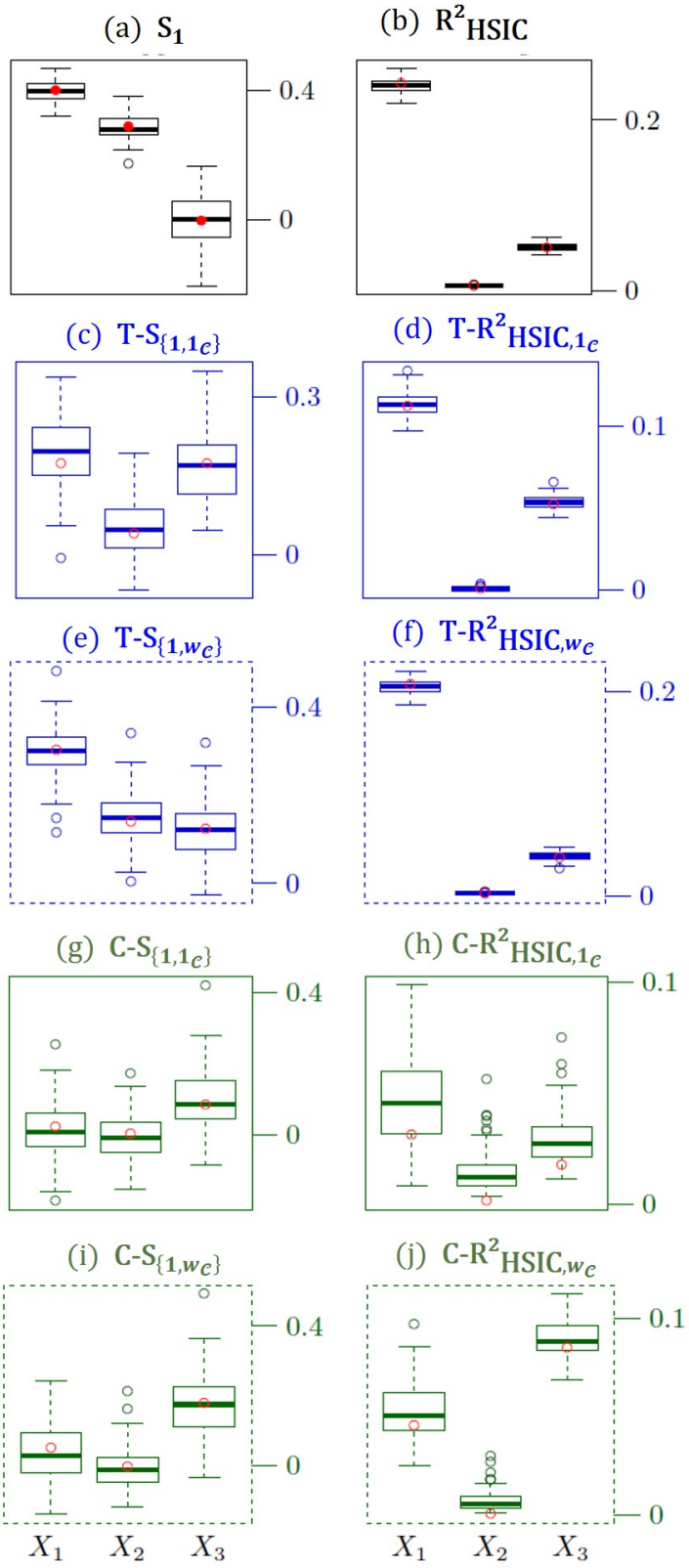



Figure 4: Model $\mathcal{M}_2$ – Global (black), target (blue) and conditional (green) SA indices, estimated from samples of size $n = 1000$. Reference values are given by red circles.

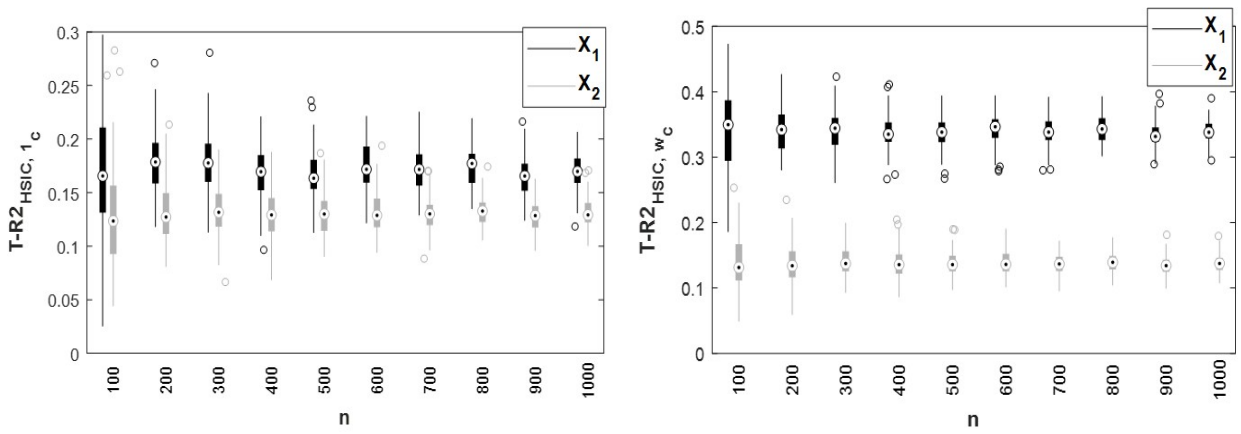
of estimators and provided information are less precise. Note that the transformation proposed by Eq. (8) which yields the hybrid Sobol index (Eq. (9)), allows to capture the CSA information, while remaining under the comfortable assumption of independence. Finally, once again smooth relaxation significantly reduces the variance of estimators, but it can also distort the information conveyed by CSA, especially for imprecise indices or for which the ranking is not very pronounced (case of hybrid Sobol indices). Contrariwise, relaxation is very relevant for the indices with a high screening power, such as HSIC-based ones.

**Remark 4.** *Although $R^2_{HSIC}$ is a normalized index which lies in $[0,1]$, its value can only be interpreted in terms of ranking of the inputs, its value depending on the considered kernels (and underlying RKHS). This explains why the values T-$R^2_{HSIC,w_c}$ and T-$R^2_{HSIC,1_c}$ quantitatively differ. Even between two characteristic kernels, one can moreover have a greater power of discrimination than the other and more quickly identify the good ranking. To perform a robust screening, it is necessary to use statistical tests of independence built upon HSIC. This point will be discussed in the perspectives.*

If we now observe the results of Ishigami model $\mathcal{M}_2$ given by Figure 4, it appears that the relative influences of the inputs are different in each analysis case. In GSA, the input $X_1$ is the most important, and the inputs $X_2$ and $X_3$ have lower importance, being ranked differently according to Sobol and HSIC-based indices. In TSA, $X_3$ now has similar importance to that of $X_1$, while $X_2$ has much less. Indeed, the combined effect of $X_1$ and $X_3$ easily exceeds the critical value $c = 6.12$, while the isolated action of $X_2$ can only approach $c$ (since $a = 5$ is significantly less than $c$). As previously, T-$R^2_{\mathrm{HSIC},w}$ indices offer more precision than T-$\mathrm{S}_{1,w}$ and lead to a more pronounced and clearer ranking. As for CSA in $\mathcal{M}_1$, the smooth relaxation is relevant for T-$R^2_{\mathrm{HSIC},w}$ (reduction of the variability of the estimators) but modifies too much the results of T-$\mathrm{S}_{1,w}$ indices.

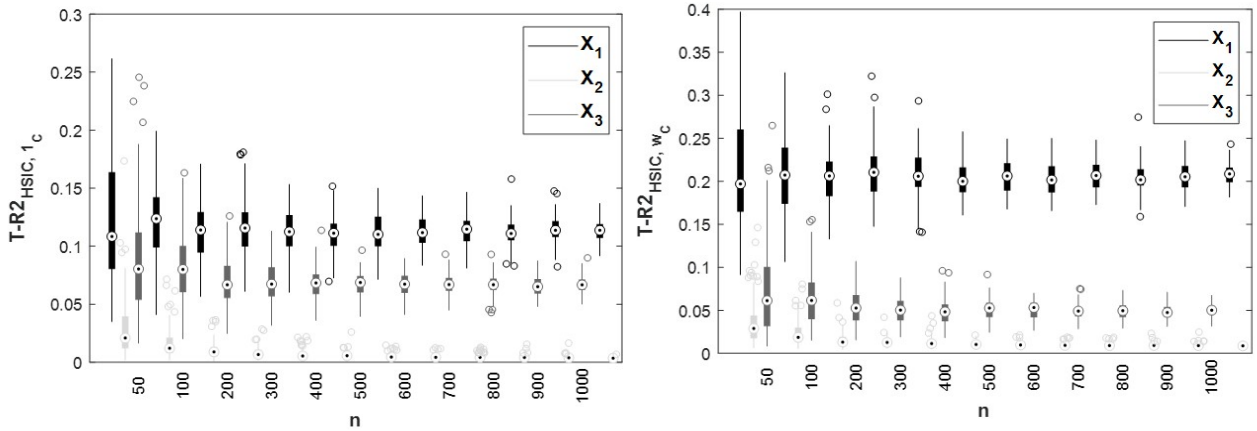
Considering now the CSA results, $X_3$ becomes the more dominant input: the term $X_3^4$ presents steep derivatives in the area of high values and strongly influences the $Y$ value, within the critical domain. Moreover, this chaotic influence is difficult to capture: estimators C-$\mathrm{S}_{1,w}$ and C-$R^2_{\mathrm{HSIC},w}$ are thus very variables. Note that the influence of $X_3$ is better captured by C-$R^2_{\mathrm{HSIC},w}$ if smoothing transformation is used.

Finally, the relevance of smoothing relaxation for HSIC-based indices is confirmed and illustrated by the convergence plots given by Figures 5 and 6, for $\mathcal{M}_1$ and $\mathcal{M}_2$ respectively. From $n = 200$ and for both models, T-$R^2_{\mathrm{HSIC},w_c}$ correctly order the inputs while this distinction is sometimes not yet clearly observed for T-$R^2_{\mathrm{HSIC},1_c}$, due to the large estimation error. Similar results are obtained for conditional HSIC-based indices, not provided here for the sake of brevity.




(a) T-$R^2_{\mathrm{HSIC},1_{\mathcal{C}}}$ (b) T-$R^2_{\mathrm{HSIC},w_{\mathcal{C}}}$

Figure 5: Model $\mathcal{M}_1$ – Convergence plots of T-$R^2_{\mathrm{HSIC}}$ indices, with the indicator function $1_{\mathcal{C}}$ (a) and the smooth relaxation $w_{\mathcal{C}}$ (b).




(a) T-$R^2_{\mathrm{HSIC},1_{\mathcal{C}}}$ (b) T-$R^2_{\mathrm{HSIC},w_{\mathcal{C}}}$

Figure 6: Model $\mathcal{M}_2$ – Convergence plots of T-$R^2_{\mathrm{HSIC}}$ indices, with the indicator function $1_{\mathcal{C}}$ (a) and the smooth relaxation $w_{\mathcal{C}}$ (b).

## 5. Application on a nuclear safety use case

We consider here a nuclear safety use-case, treated in [58] and based on thermal-hydraulic computer experiments. A HSIC-based GSA was performed by the authors in order to identify the non-influential inputs and rank the influential ones. This step is preliminary to the sequential building of a joint Gaussian process metamodel, with the group of influential inputs as explanatory variables. This metamodel is then used to accurately estimate Sobol sensitivity indices and high-order quantiles of the output.

We propose here to apply our TSA and CSA HSIC-based indices on the same learning sample to identify the inputs influencing the exceeding of a given quantile and compare to GSA results.

### *5.1. Description of the use case*

In support of regulatory work and nuclear power plant design and operation, safety analysis considers the so-called "Loss Of Coolant Accident" which takes into account a double-ended guillotine break with a specific size piping rupture. The use-case under study does not focus on a full realistic model of a reactor but a simplified one (e.g., regarding both physical phenomena and dimensions of the system). The numerical model is based on code CATHARE2 (V2.5_3mod3.1) which simulates the time evolution of physical quantities during a thermal-hydraulic transient. It simulates a test carried out on the Japanese mock-up "Large Scale Test Facility" in the framework of the OECD/ROSA-2 project. The model used is representative of an *Intermediate Break Loss Of Coolant Accident* (IBLOCA) [64]. The mock-up represents a reduced scale Westinghouse PWR (1/1 ratio in height and 1/48 in volume), with two loops (instead of three or four loops on an actual reactor) and an electric powered heating core (10 MWe), see Figure 7. It operates at the same pressure and temperature values as the reference PWR. The simulated accidental transient is an IBLOCA with a break on the cold leg and no safety injection on the broken leg.

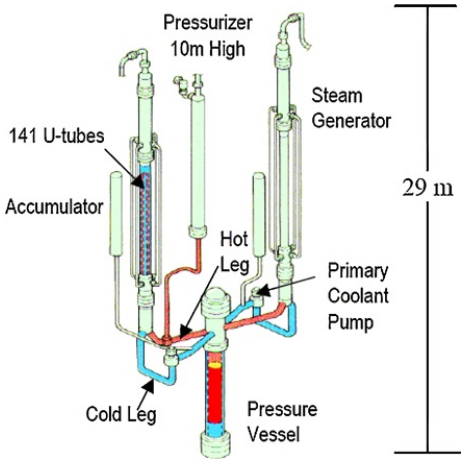



Figure 7: IBLOCA model – Hot and cold legs of the Large Scale Test Facility.

In this use-case, $d = 27$ scalar input variables of CATHARE2 code are uncertain, defined by their probability density function (pdf). These pdf are uniform, log-uniform, normal or log-normal. They correspond to various system parameters such as, for instance, boundary conditions, critical flow rates, interfacial friction coefficients, condensation coefficients and heat transfer coefficients. Only uncertainties related to physical parameters are considered here and no uncertainty on scenario variables (initial state of the reactor before the transient) is taken into account. Table 1 provides more details about these uncertain inputs and their probability density functions. The nature of these uncertainties appears to be epistemic since they arise from a lack of knowledge on the true value of these parameters. The output variable of interest is a single scalar which is the maximal *peak cladding temperature* (PCT) during the accident transient (as an example, see the peak in Figure 8). This quantity is derived from the physical outputs provided by CATHARE2 code.

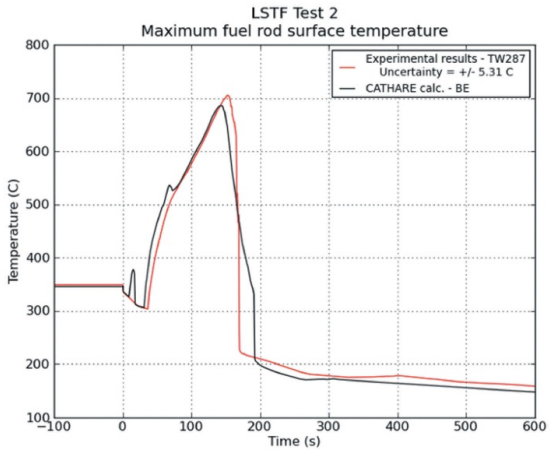



Figure 8: IBLOCA model – Physical simulation output of the model with CATHARE2 code: maximal rod cladding temperature during the transient.

### *5.2. GSA, TSA and CSA results*

As detailed in [58], a sample of $n = 500$ CATHARE2 simulations is available built from a space-filling design and following the prior distributions of inputs defined in Table 1. The histogram of the obtained values for the output of interest, namely the PCT, is given by Figure 9 (temperature is in °C). A kernel density estimator of the data is also added on the plot. We focus on the 90%-quantile of the PCT which is empirically estimated here to $q_{0.9} \simeq 722.9$.

From the learning sample of $n = 500$ simulations, $R^2_{\text{HSIC}}$ indices are estimated with $\lambda = 1/\sigma^2$ and $\sigma^2$ being the empirical variance of the sample. Significant values are given by Table 2[2]. Analyzing these results

[2]Note that GSA HSIC-based independence tests ([20]), not detailed here, are performed to screen the significantly influential inputs

Table 1: IBLOCA model – List of the 27 uncertain input parameters and associated physical models in CATHARE2 code.

| **Type of inputs** | **Inputs** | **pdf**[a] | **Physical models** |
|---|---|---|---|
| Heat transfer | $X_1$ | $\mathcal{N}$ | Departure from nucleate boiling |
| in the core | $X_2$ | $\mathcal{U}$ | Minimum film stable temperature |
| | $X_3$ | $\mathcal{LN}$ | HTC[b] for steam convection |
| | $X_4$ | $\mathcal{LN}$ | Wall-fluid HTC |
| | $X_5$ | $\mathcal{N}$ | HTC for film boiling |
| Heat transfer in the steam | $X_6$ | $\mathcal{LU}$ | HTC forced wall-steam convection |
| generators (SG) U-tube | $X_7$ | $\mathcal{N}$ | Liquid-interface HTC for film condensation |
| Wall-steam friction in core | $X_8$ | $\mathcal{LU}$ | |
| Interfacial friction | $X_9$ | $\mathcal{LN}$ | SG outlet plena and crossover legs together |
| | $X_{10}$ | $\mathcal{LN}$ | Hot legs (horizontal part) |
| | $X_{11}$ | $\mathcal{LN}$ | Bend of the hot legs |
| | $X_{12}$ | $\mathcal{LN}$ | SG inlet plena |
| | $X_{13}$ | $\mathcal{LN}$ | Downcomer |
| | $X_{14}$ | $\mathcal{LN}$ | Core |
| | $X_{15}$ | $\mathcal{LN}$ | Upper plenum |
| | $X_{16}$ | $\mathcal{LN}$ | Lower plenum |
| | $X_{17}$ | $\mathcal{LN}$ | Upper head |
| Condensation | $X_{18}$ | $\mathcal{LN}$ | Downcomer |
| | $X_{19}$ | $\mathcal{U}$ | Cold leg (intact) |
| | $X_{20}$ | $\mathcal{U}$ | Cold leg (broken) |
| | $X_{27}$ | $\mathcal{U}$ | Jet |
| Break flow | $X_{21}$ | $\mathcal{LN}$ | Flashing (undersaturated) |
| | $X_{22}$ | $\mathcal{N}$ | Wall-liquid friction (undersaturated) |
| | $X_{23}$ | $\mathcal{N}$ | Flashing delay (undersaturated) |
| | $X_{24}$ | $\mathcal{LN}$ | Flashing (saturated) |
| | $X_{25}$ | $\mathcal{N}$ | Wall-liquid friction (saturated) |
| | $X_{26}$ | $\mathcal{LN}$ | Global interfacial friction (saturated) |

[a] $\mathcal{U}$, $\mathcal{LU}$, $\mathcal{N}$ and $\mathcal{LN}$ respectively stands for uniform, log-uniform, normal and log-normal probability distributions.
[b] Heat Transfer Coefficient.

leads to notice the large influence of $X_{10}$ (the interfacial coefficient in the horizontal part of the hot legs), followed by $X_{12}$ (the interfacial friction coefficient in the steam generator inlet plena) and $X_2$ (the minimum stable film temperature in the core). Then, a last group of six inputs have a lower influence (but rather similar in terms of order of magnitude): namely $X_{14}$, $X_9$, $X_{22}$, $X_{15}$, $X_6$ and $X_{13}$. Finally, the other 18 input variables are non-influential. If the estimates fluctuate a little bit due to the small dataset, these results are

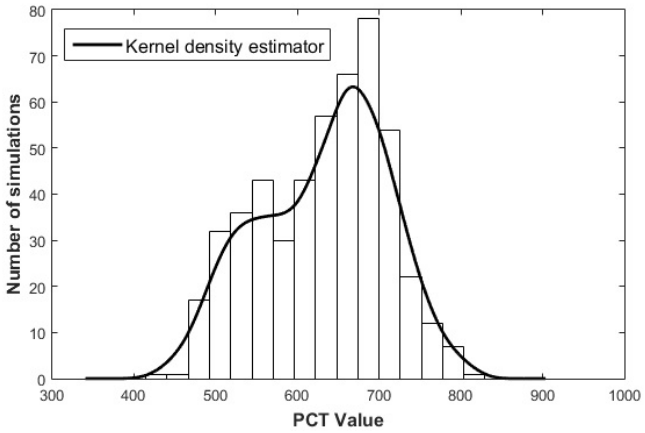



Figure 9: IBLOCA model – Histogram of the PCT from the learning sample of $n = 500$ simulations.

relevant regarding precedent studies (see, e.g., in [58]).

Table 2: IBLOCA model – Normalized HSIC-based sensitivity indices $R^2_{\text{HSIC}}$ for the influential inputs.

| Inputs | $X_{10}$ | $X_{12}$ | $X_2$ | $X_{14}$ | $X_9$ | $X_{22}$ | $X_{15}$ | $X_6$ | $X_{13}$ |
|---|---|---|---|---|---|---|---|---|---|
| $R^2_{\text{HSIC}}$ | **0.43** | **0.04** | **0.03** | 0.02 | 0.02 | 0.02 | 0.01 | 0.01 | 0.01 |

We now compute normalized target and conditional HSIC-based indices regarding the critical event of exceeding the 90%-quantile $q_{0.9}$ of the PCT, with the smooth relaxation and similar parametrization as in Section 4 ($s = 1/5$). Results are given in Table 3. Two interesting features can be highlighted from such an analysis:

- First, one can notice that TSA confirms the crucial underlying role played by $X_{10}$ and $X_{12}$ in the global phenomenon of interest. Nonetheless, it also reveals the role played by both $X_{14}$ (the interfacial friction coefficient of the core) and $X_9$ (the interfacial friction coefficient of the steam generator outlet plena and crossover legs together) on the occurrence of the critical event. Finally, it appears that $X_6$ (the HTC forced wall-steam convection coefficient), which had a nonnegligible influence from the GSA point of view, now has a minimal influence from a TSA point of view;

- Second, one can see that CSA results are, not only very informative, but also highlights complex underlying behaviors. As one may notice, by conditioning w.r.t. the critical event, $X_{22}$ (the wall-liquid friction coefficient of the undersaturated break flow) appears to have a similar influence as $X_{10}$ and $X_{12}$. Moreover, $X_{13}$ (the interfacial friction coefficient of the downcomer) becomes more influential within the critical domain.

First attempts of physical interpretation can be based on these results, supplemented by some scatterplots

Table 3: IBLOCA model – Normalized target and conditional HSIC-based sensitivity indices for the influential inputs.

| Inputs | $X_{10}$ | $X_{12}$ | $X_2$ | $X_{14}$ | $X_9$ | $X_{22}$ | $X_{15}$ | $X_6$ | $X_{13}$ |
|---|---|---|---|---|---|---|---|---|---|
| T-$R^2_{\text{HSIC},w_\mathcal{C}}$ | **0.30** | **0.05** | 0.02 | **0.03** | **0.03** | 0.02 | 0.01 | **0** | 0.02 |
| C-$R^2_{\text{HSIC},w_\mathcal{C}}$ | **0.09** | **0.07** | **0.03** | 0.02 | 0.02 | **0.10** | 0.01 | 0.01 | **0.04** |

of the PCT w.r.t to the key inputs identified by GSA, TSA and CSA. These scatterplots are given by Figure 10. The predominant global influence of $X_{10}$ is clearly observed. Moreover, as already explained in [58], larger values of $X_{10}$ in the horizontal part of the hot legs lead to larger values of the PCT. This can be explained by the increase of vapor which brings the liquid in the horizontal part of hot legs, leading to a reduction of the liquid water return from the rising part of the U-tubes of the SG to the core (through the hot branches and the upper plenum). Since the amount of liquid water available to the core cooling is reduced, higher PCTs are observed. However, beyond a value ($X_{10} > 3$), the water nonreturn effect seems to have been reached and $X_{10}$ appears to be less influential. This explains its predominance in GSA and TSA, but less in CSA. As revealed by CSA results, the wall to liquid friction (in under-saturated break flow conditions) in the break line $X_{22}$ and the interfacial friction coefficient in the SG inlet plena $X_{12}$ are very influential on the highest PCT values, beyond $q_{0.9}$. This is consistent with the previous interpretation of [58] and scatterplots: low values of $X_{22}$ lead to higher break flow rates, resulting in a loss of the primary water mass inventory at the higher break, and thus a more significant core uncovering (then higher PCT). For $X_{12}$, its higher values lead to a greater supply (by the vapor) of liquid possibly stored in the water plena to the rest of the primary loops (then lower PCT).

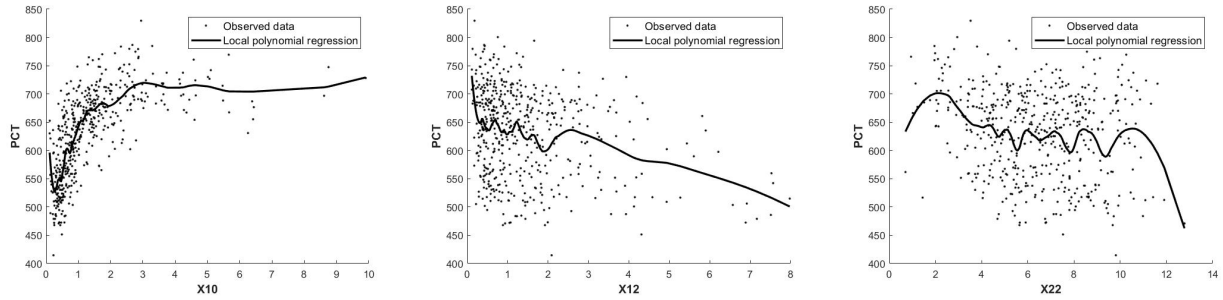



Figure 10: IBLOCA model – Scatterplots with local polynomial regression of PCT according to $X_{10}$, $X_{12}$ and $X_{22}$.

As a consequence, this application clearly illustrates how TSA and CSA results provide complementary insights about the phenomenon under study. Coupled with a standard GSA analysis, they enable to understand more deeply the underlying behavior of the code (and, at least, of the physics). One of the best advantage of the provided tools remain in their ability to be used through a single learning sample, without requiring the construction and use of a metamodel. On the contrary, the provided information can be

judiciously used to then build a metamodel from a limited number of inputs, e.g., for the identification of penalizing configurations as in [59].

To strengthen these results and study their robustness, the next subsection proposes a numerical convergence analysis.

### *5.3. Convergence analysis*

In this subsection, a numerical convergence analysis is performed using 100 *bootstrap* repetitions. Note that, for the sake of clarity in the convergence plots, a focus has been put on the minimal set of influential variables (i.e., $X_{10}$, $X_{12}$, $X_{22}$ and $X_9$). Results are given by Figure 11.

They confirm the predominance of $X_{10}$ in terms of both global and target influence, which is detected from very small sample size ($n = 100$). However, looking closely at CSA results, it appears that CSA is more difficult to estimate and much less discriminating: the coefficient of variations are rather large, especially for small-size learning samples. First, the difficult estimate is explained by the more complex information that CSA considers, and which suffers more from the loss of information. Second, the weak discriminating power is due to the fact that the behavior within the critical domain is mainly driven by a combination of the inputs rather than single identifiable influences. In other words, these variables might be highly dependent within the critical domain. However, we observe see that $X_{22}$ clearly stands out from the larger size samples.

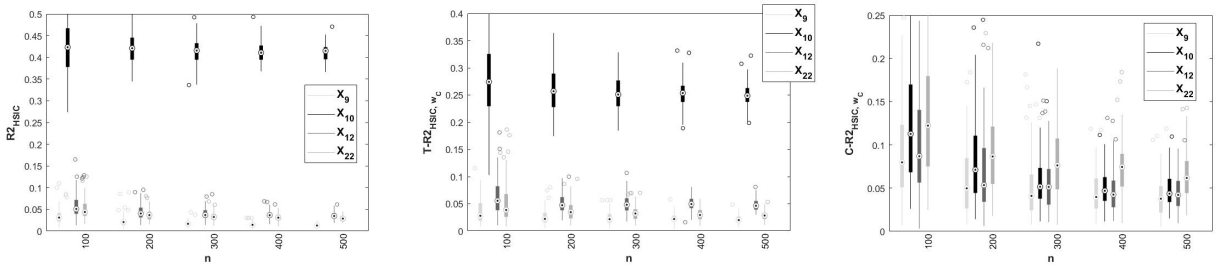



Figure 11: IBLOCA model – Convergence plots of GSA, TSA and CSA HSIC indices.

Note that the quantile estimation error was not discussed here. For large sample sizes and for not too extreme quantiles, we argue that this error will have little influence on the TSA and CSA results. But it will be less the case for small samples. In this case, the estimation error should be taken into account with bootstrap approaches.

## 6. Conclusion

As part of sensitivity analysis (SA) of numerical simulators in presence of uncertain inputs, this work focuses the analysis towards a restricted domain of the output distribution (e.g., a distribution tail). As illustrated in a safety nuclear use case, this domain may correspond in practice to a critical domain of

the studied phenomenon. To capture the influence of the uncertain inputs on this restricted domain, the notions of target (TSA) and conditional (CSA) sensitivity analyses are defined. TSA aims at measuring the influence of the inputs on the occurrence of the critical event, while CSA measures its influence within the critical domain (ignoring what happens outside). Both specific sensitivity analyses have numerous applications, particularly in safety and risk assessment studies. Furthermore, as underlined and illustrated in our numerical applications, these two notions can also widely differ.

Starting from existing global SA (GSA) measures, we propose new operational tools dedicated to TSA and CSA. For this, we first focus our attention on the usual and popular sensitivity indices based variance decomposition, namely Sobol indices. Then, the *Hilbert-Schmidt Independence Criterion* (HSIC), a dependence measure recently adapted for GSA purposes and well-suited for a small learning dataset, is also considered. Adapted versions of Sobol and HSIC are proposed for TSA and CSA, as well as associated statistical estimators. More specifically, alternative Sobol indices are defined for CSA to overcome the dependence of inputs induced by the conditioning. Furthermore, to cope with the loss of information (especially when the critical event is associated to a low probability) and reduce the variability of estimators, a transformation of the output using smooth weight functions is also proposed for all the TSA and CSA indices.

Then, the proposed tools are illustrated and compared on several analytical test cases. These experiments clearly illustrate the interest and the complementarity of the information provided by TSA and CSA, relative to that provided by GSA. Results also show the efficiency of HSIC-based indices which are relevant tools for GSA, TSA and CSA. Their estimators offer a good compromise between bias and variance, while presenting efficient ranking performance (from few hundred of simulations) and requiring much less simulations than estimators of Sobol-based indices. The relevancy of smooth relaxation also clearly appears for TSA and CSA HSIC-based indices. It offers a compromise between "brute" goal-oriented SA and a better exploitation of the available information, which yields a significant reduction of the variance of statistical estimators. This compromise is tuned by a smoothing parameter. But, in return, the relaxation should not be set too large in order to remain close to the notion of TSA. Finally, TSA and CSA HSIC-based indices are applied on a nuclear engineering use case which simulates a severe accidental scenario on a pressurized water reactor. With several tens of uncertain inputs, this provides a practical illustration, more realistic and challenging. This application also shows how TSA and CSA, coupled with a standard GSA analysis, enable to further understand the behavior of the code and modeled physics.

In perspective of this work, it is planned to further study the impact of the smoothing parameter used in the relaxation. A reasonable choice is proposed here but it might also be possible to choose it so as to control that a certain percentage of data is taken into account. It can also be adapted according to the number of simulations, in particular for very small samples.

More ambitiously, we believe that it is essential to develop the use of statistical independence tests associated to our TSA and CSA indices. This will provide a more rigorous and accurate statistical framework, and will allow a more objective conclusion on significantly influential inputs. The adaptation of the available HSIC-based tests ([20]) to TSA-HSIC is relatively direct, contrary to the extension to CSA-HSIC which is subject to a modification of the asymptotic law under independence. Moreover, the impact of the smooth relaxation on the power of independence tests should also be assessed. Finally, if the objective is only to perform a TSA or CSA of the model and if the choice of sampling is possible, it is obvious that goal-oriented sampling methods such as importance sampling would be of great interest. Indeed, by adding simulations in the critical area, these methods would significantly improve the estimation of TSA and CSA indices. This may be the subject of future work.

## Acknowledgments

This work was initiated as part of Hugo Raguet's post-doctorate funded by CEA (Commissariat à l'Energie Atomique et aux Energies Alternatives), and continued afterwards. The authors are grateful to Henri Geiser & Thibault Delage who performed the simulations using the CATHARE2 code. CATHARE2 code is developed under the collaborative framework of the NEPTUNE project, supported by CEA, EDF (Electricité de France), Framatome and IRSN (Institut de Radioprotection et de Sûreté Nucléaire). The authors are also grateful to Reda El Amri & Anouar Meynaoui for their help in the implementation in the *sensitivity* R package of some of the tools used in this paper.

## References

[1] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global Sensitivity Analysis. The Primer, Wiley, 2008.

[2] B. Iooss, P. Lemaître, A Review on Global Sensitivity Analysis Methods, in: G. Dellino, C. Meloni (Eds.), Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications, Springer US, Boston, MA, 2015, Ch. 5, pp. 101–122.

[3] P. Wei, Z. Lu, S. Song, Variable importance analysis: a comprehensive review, Reliability Engineering and System Safety 142 (2015) 399–432.

[4] E. Borgonovo, Sensitivity Analysis, International Series in Operations Research & Management Science, Springer International Publishing, 2017.

[5] V. Chabridon, Reliability-oriented sensitivity analysis under probabilistic model uncertainty – Application to aerospace systems, Ph.D. thesis, Université Clermont Auvergne, (in English) (2018).

[6] H. Raguet, A. Marrel, Target and conditional sensitivity analysis with emphasis on dependence measures, Preprint hal-01694129 (2018).

[7] S. Kucherenko, B. Iooss, Derivative-Based Global Sensitivity Measures, in: R. Ghanem, D. Higdon, H. Owhadi (Eds.), Handbook of Uncertainty Quantification, Springer International Publishing, 2017, Ch. 36, pp. 1241–1263.

[8] C. V. Mai, B. Sudret, Computing derivative-based global sensitivity measures using polynomial chaos expansions, Reliability Engineering and System Safety 134 (2015) 241–250.

[9] M. De Lozzo, A. Marrel, Estimation of the Derivative-Based Global Sensitivity Measures Using a Gaussian Process Metamodel, SIAM/ASA Journal of Uncertainty Quantification 4 (1) (2016) 708–738.

[10] W. Hoeffding, A class of statistics with asymptotically normal distribution, The Annals of Mathematical Statistics 19 (3) (1948) 293–325.

[11] I. M. Sobol, Sensitivity estimates for nonlinear mathematical models, Mathematical modelling and computational experiments 1 (4) (1993) 407–414.

[12] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation 55 (2001) 271–280.

[13] C. Prieur, S. Tarantola, Variance-based sensitivity analysis: theory and estimation algorithms, in: R. Ghanem, D. Higdon, H. Owhadi (Eds.), Handbook of Uncertainty Quantification, Springer International Publishing, 2017, Ch. 35, pp. 1217–1239.

[14] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliability Engineering and System Safety 93 (2008) 964–979.

[15] A. Marrel, B. Iooss, B. Laurent, O. Roustant, Calculations of Sobol indices for the Gaussian process metamodel, Reliability Engineering and System Safety 94 (2009) 742–751.

[16] K. Konakli, B. Sudret, Global sensitivity analysis using low-rank tensor approximations, Reliability Engineering and System Safety 156 (2016) 64–83.

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

[18] S. Rahman, The $f$-Sensitivity Index, SIAM/ASA Journal of Uncertainty Quantification 4 (2016) 130–162.

[19] A. Gretton, K. Fukumizu, C. H. Teo, L. Song, B. Schölkopf, A. J. Smola, A kernel statistical test of independence, in: Advances in Neural Information Processing Systems, 2008, pp. 585–592.

[20] M. De Lozzo, A. Marrel, New improvements in the use of dependence measures for sensitivity analysis and screening, Journal of Statistical Computation and Simulation 86 (15) (2016) 3038–3058.

[21] M. Baucells, E. Borgonovo, Invariant Probabilistic Sensitivity Analysis, Management Science 59 (11) (2013) 2536–2549.

[22] I. Csiszár, A class of measures of informativity of observation channels, Periodica Mathematica Hungarica 2 (1-4) (1972) 191–213.

[23] E. Hellinger, Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen., Journal für die reine und angewandte Mathematik 136 (1909) 210–271.

[24] S. Kullback, R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22 (1) (1951) 79–86.

[25] L. I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D: nonlinear phenomena 60 (1-4) (1992) 259–268.

[26] C. E. Shannon, A mathematical theory of communication, Bell system technical journal 27 (3) (1948) 379–423.

[27] T. Suzuki, M. Sugiyama, T. Kanamori, J. Sese, Mutual information estimation reveals global associations between stimuli and biological processes, BMC bioinformatics 10 (1) (2009) S52.

[28] G. J. Székely, M. L. Rizzo, N. K. Bakirov, Measuring and testing dependence by correlation of distances, The annals of statistics 35 (6) (2007) 2769–2794.

[29] G. J. Székely, M. L. Rizzo, The distance correlation t-test of independence in high dimension, Journal of Multivariate Analysis 117 (2013) 193–213.

[30] S. Yao, X. Zhang, X. Shao, Testing mutual independence in high dimension via distance covariance, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80 (3) (2018) 455–480.

[31] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, G. R. Lanckriet, Hilbert space embeddings and metrics on probability measures, Journal of Machine Learning Research 11 (2010) 1517–1561.

[32] A. Gretton, R. Herbrich, A. J. Smola, O. Bousquet, B. Schölkopf, Kernel Methods for Measuring Independence, Journal of Machine Learning Research 6 (2005) 2075–2129.

[33] A. Meynaoui, M. Albert, B. Laurent, A. Marrel, Aggregated test of independence based on hsic measures, Preprint hal-02020084v3 (2020).

[34] R. C. Spear, G. M. Hornberger, Eutrophication in Peel Inlet. II. Identification of critical uncertainties via generalized sensitivity analysis, Water Research 14 (1980) 43–49.

[35] J.-C. Fort, T. Klein, N. Rachdi, New sensitivity analysis subordinated to a contrast, Communications in Statistics - Theory and Methods 45 (15) (2016) 4349–4364.

[36] T. Browne, J.-C. Fort, B. Iooss, L. Le Gratiet, Estimate of quantile-oriented sensitivity indices, Preprint hal-01450891 (2017).

[37] V. Maume-Deschamps, I. Niang, Estimation of quantile oriented sensitivity indices, Statistics and Probability Letters 134 (2018) 122–127.

[38] S. Kucherenko, S. Song, L. Wang, Quantile based global sensitivity measures, Reliability Engineering and System Safety 185 (2019) 35–48.

[39] P. Lemaître, Analyse de sensibilité en fiabilité des structures, Ph.D. thesis, Université de Bordeaux, (in English) (2014).

[40] G. Perrin, G. Defaux, Efficient Evaluation of Reliability-Oriented Sensitivity Indices, Journal of Scientific Computing (2019) 1–23.

[41] R. Rackwitz, Reliability analysis – a review and some perspectives, Structural Safety 23 (2001) 365–395.

[42] P. Bjerager, S. Krenk, Parametric Sensitivity in First Order Reliability Theory, Journal of Engineering Mechanics 115 (1989) 1577–1582.

[43] M. Hohenbichler, R. Rackwitz, Sensitivity and importance measures in structural reliability, Civil Engineering Systems 3 (1986) 203–209.

[44] A. Kouassi, Propagation d'incertitudes en CEM. Application à l'analyse de fiabilité et de sensibilité de lignes de transmission et d'antennes, Ph.D. thesis, Université Clermont Auvergne, (in French) (2017).

[45] H. O. Madsen, Omission sensitivity factors, Structural Safety 5 (1) (1988) 35–45.

[46] J. Morio, M. Balesdent, D. Jacquemart, C. Vergé, A survey of rare event simulation methods for static input-output models, Simulation Modelling Practice and Theory 49 (2014) 297–304.

[47] Y.-T. Wu, Computational Methods for Efficient Structural Reliability and Reliability Sensitivity Analysis, AIAA Journal 32 (8) (1994) 1717–1723.

[48] Z. Lu, S. Song, Z. Yue, J. Wang, Reliability sensitivity method by line sampling, Structural Safety 30 (6) (2008) 517–532.

[49] S. Song, Z. Lu, H. Qiao, Subset simulation for structural reliability sensitivity analysis, Reliability Engineering and System Safety 94 (2) (2009) 658–665.

[50] R. Y. Rubinstein, The score function approach for sensitivity analysis of computer simulation models, Mathematics and Computers in Simulation 28 (1986) 351–379.

[51] V. Chabridon, M. Balesdent, J.-M. Bourinet, J. Morio, N. Gayton, Reliability-based sensitivity estimators of rare event probability in the presence of distribution parameter uncertainty, Reliability Engineering and System Safety 178 (2018) 164–178.

[52] J. Morio, Influence of input PDF parameters of a model on a failure probability estimation, Simulation Modelling Practice and Theory 19 (10) (2011) 2244–2255.

[53] L. Li, Z. Lu, F. Jun, W. Bintuan, Moment-independent importance measure of basic variable and its state dependent parameter solution, Structural Safety 38 (2012) 40–47.

[54] P. Wei, Z. Lu, W. Hao, J. Feng, B. Wang, Efficient sampling methods for global reliability sensitivity analysis, Computer Physics Communications 183 (2012) 1728–1743.

[55] L. Cui, Z. Lu, X. Zhao, Moment-independent importance measure of basic random variable and its probability density evolution solution, Science China Technical Sciences 53 (10) (2010) 1138–1145.

[56] P. Lemaître, E. Sergienko, A. Arnaud, N. Bousquet, F. Gamboa, B. Iooss, Density modification-based reliability sensitivity analysis, Journal of Statistical Computation and Simulation 85 (6) (2015) 1200–1223.

[57] E. De Rocquigny, N. Devictor, S. Tarantola, Uncertainty in industrial practice: a guide to quantitative uncertainty management, Wiley, 2008.

[58] B. Iooss, A. Marrel, Advanced methodology for uncertainty propagation in computer experiments with large number of inputs, Nuclear Technology 205 (12) (2019) 1588–1606.

[59] A. Marrel, B. Iooss, V. Chabridon, Statistical identification of penalizing configurations in high-dimensional thermalhydraulic numerical experiments: The ICSCREAM methodology, Preprint hal-02535146 (2020).

[60] A. Spagnol, R. Le Riche, S. Da Veiga, Global Sensitivity Analysis for Optimization with Variable Selection, SIAM/ASA Journal of Uncertainty Quantification 7 (2) (2019) 417–443.

[61] G. Chastaing, F. Gamboa, C. Prieur, Generalized Sobol sensitivity indices for dependent variables: numerical methods, Journal of Statistical Computation and Simulation 85 (7) (2015) 1306–1333.

[62] A. Gretton, O. Bousquet, A. Smola, B. Scholkopf, Measuring statistical dependence with hilbert-schmidt norms, in: International Conference on Algorithmic Learning Theory, Vol. 16, 2005, pp. 63–78.

[63] L. Song, A. Smola, A. Gretton, J. Bedo, K. Borgwardt, Feature selection via dependence maximization, Journal of Machine Learning Research 13 (May) (2012) 1393–1434.

[64] P. Mazgaj, J.-L. Vacher, S. Carnevali, Comparison of CATHARE results with the experimental results of cold leg intermediate break LOCA obtained during ROSA-2/LSTF test 7, EPJ Nuclear Sciences & Technology 2 (1) (2016).