Kernel-based sensitivity analysis for (excursion) sets

Noé Fellmann, Christophette Blanchet-Scalliet, Céline Helbert, Adrien Spagnol, Delphine Sinoquet

Kernel-based Sensitivity Analysis for (excursion) sets

N. Fellmann[1,2], C. Blanchet-Scalliet[1], C. Helbert[1], A. Spagnol[2], and D. Sinoquet[2]

[1]École Centrale de Lyon, CNRS UMR 5208, Institut Camille Jordan, 36 Avenue Guy de Collongue, 69134 Écully, France
[2]IFP Energies Nouvelles
{noe.fellmann,christophette.blanchet,celine.helbert}@ec-lyon.fr
{noe.fellmann,delphine.sinoquet,adrien.spagnol}@ifpen.fr

February 15, 2024

Abstract

In this paper, we aim to perform sensitivity analysis of set-valued models and, in particular, to quantify the impact of uncertain inputs on feasible sets, which are key elements in solving a robust optimization problem under constraints. While most sensitivity analysis methods deal with scalar outputs, this paper introduces a novel approach for performing sensitivity analysis with set-valued outputs. Our innovative methodology is designed for excursion sets, but is versatile enough to be applied to set-valued simulators, including those found in viability fields, or when working with maps like pollutant concentration maps or flood zone maps.

We propose to use the Hilbert-Schmidt Independence Criterion (HSIC) with a kernel designed for set-valued outputs. After proposing a probabilistic framework for random sets, a first contribution is the proof that this kernel is *characteristic*, an essential property in a kernel-based sensitivity analysis context. To measure the contribution of each input, we then propose to use HSIC-ANOVA indices. With these indices, we can identify which inputs should be neglected (*screening*) and we can rank the others according to their influence (*ranking*). The estimation of these indices is also adapted to the set-valued outputs. Finally, we test the proposed method on three test cases of excursion sets.

Keywords: HSIC-ANOVA indices, random sets, screening, ranking

1 Introduction

In many fields, it is essential to understand the input/output relationships of models that simulate complex physical systems. This knowledge can be used to simplify and optimize the model and provide valuable insights to experts. Sensitivity analysis (SA) is one response to this challenge. It quantifies how variations in inputs translate into variations in outputs, precisely measuring the impact of each input on the output. However, in certain domains, models may have highly complex outputs that traditional SA methods, originally designed for scalar outputs, may not be suitable for. Specifically, we are interested in models with set-valued outputs, where each evaluation of the model produces a subset of a larger space. Our interest lies in the need to quantify the influence of parameters on the excursion sets of optimization problems, with the goal of simplifying constrained robust optimization problems. Set-valued output models are also prevalent in other domains, including image processing, map modeling, and viability fields. Therefore, we aim to measure the impact of each input, on a set-valued output through an adapted SA approach.

Global sensitivity analysis (GSA) is a methodology used to assess the impact of input variations on the output of a system or model across the entire parameter space. A detailed review of GSA methods can be found in the book Da Veiga et al. [2021] or in Iooss and Lemaître [2015]. These methods can be distinguished into two types:

- *screening*-oriented SA techniques are devoted to identifying influential and non-influential inputs;
- *ranking*-oriented SA aims to compute sensitivity indices or importance measures, which are scalars representing the effect of an input or group of inputs on the output. These indices are then used to rank the inputs by their influence.

One well-known screening technique is the Morris method (Morris [1991]), which is based on one-at-a-time designs, i.e., where each input varies while the others are held fixed. Regarding ranking methods, the most commonly used indices are the Sobol' indices, which are variance-based sensitivity measures (Sobol' [2001]). These indices quantify the portion of the output variance that can be attributed to one or a group of inputs. This variance decomposition is called the ANOVA (ANalysis Of VAriance) decomposition. However, these indices have several drawbacks. One main issue is that their estimation cost in terms of the number of model evaluations is very high. For instance, the simulation budget needed for the 'pick and freeze' method (Sobol' [1993]') increases linearly with the input dimension. Rank-based estimators have recently been used to circumvent this problem (Gamboa et al. [2022]), but at the cost of not being able to estimate the total-order Sobol' indices. Another major drawback is that Sobol' indices quantify the input contributions to the output variance, but not to the entire output distribution. That is why other types of indices examine how the entire output distribution is affected by the input parameters. This is done by comparing the input and output probability distributions. For instance, Borgonovo indices compare the density functions (Borgonovo [2007]), Cramér von Mises indices look at the cumulative distribution function (Gamboa et al. [2018]), and kernel-based sensitivity indices compare the embedding of the distributions (Da Veiga [2021]). The former method relies on embedding distributions in Reproducing Kernel Hilbert Spaces (RKHS) using kernel functions, which makes it easier to compute distances. By embedding the input and output distributions in the RKHS, indices based on the Hilbert-Schmidt Independence Criterion (HSIC) (Gretton et al. [2005b]) can be defined. These indices quantify the dependence between an input and the output and can be used for screening.

They also have a low cost because they can be estimated using a single sample. Under certain assumptions, an ANOVA-like decomposition of HSIC exists, as shown in Da Veiga [2021] which makes them usable for ranking.

Several works in the literature have addressed SA adapted to complex outputs. However, to the best of our knowledge, the specific case of set-valued outputs has not yet been studied. For instance, Higdon et al. [2008], Marrel et al. [2011], and Perrin et al. [2021] examined spatial outputs, while De Lozzo and Marrel [2017] and Marrel et al. [2015] explored spatiotemporal outputs. These methods generate sensitivity index maps that can be interpreted by connecting them to the underlying physical phenomena. However, there is a strong motivation, whether due to interpretation challenges or synthesis concerns, to find a single scalar index per input that captures its influence on the variability of the entire complex output. This is explored and presented in Gamboa et al. [2014], which defines aggregated Sobol' indices for vectorial and functional outputs. In Gamboa et al. [2021] the authors also propose an universal index that can be used in any metric space. Kernel-based indices, in particular the HSIC, can also be used with complex outputs, as proposed in Da Veiga [2015] and in El Amri and Marrel [2024] with functional outputs. Adaptation of HSIC to complex output is in fact straightforward, kernel methods being known to be flexible with respect to the type of data (see for example Shawe-Taylor and Cristianini [2004], where kernel methods are used with several types of data such as vectors, texts, trees...). The flexibility of the latter lies in the fact that they mostly rely only on defining a kernel in the relevant space (where random elements can be defined), making them particularly permissive in accommodating different input and output types. For this flexibility, low cost, ability to quantify changes across the entire output distribution, and their utility for both screening and ranking purposes, we propose to use HSIC-ANOVA indices to conduct a SA of set-valued models.

To perform kernel-based SA on a set-valued model, there are two main requirements. First, a framework of random sets is required. Indeed, a space of sets is not an easy space to work in, and describing randomness in such a space can be challenging. For example, Molchanov [2005] and Nguyen [2006] propose a complete theory of random sets. The second key to performing kernel-based SA on set-valued outputs is to have a kernel defined on sets. Given these two conditions, we can apply the kernel-based SA methodology to define sensitivity indices for set-valued outputs. This is the focus of this paper.

To this end, we first recall in Section 2 the general methodology for defining kernel-based sensitivity indices, and in particular the Hilbert-Schmidt Independence Criterion (HSIC), in a generic measurable output space. We then present our contributions in Section 3. First, we define the probabilistic framework for random sets. Then, we introduce a new kernel defined on a space of sets and study it in detail. In particular, we show that our kernel is *characteristic*, a crucial property for screening purposes. Finally, HSIC-ANOVA indices for set-valued outputs are derived, but their estimation is complicated by the presence of sets. Therefore, a nested Monte Carlo estimator is proposed to estimate the indices and its statistical properties are studied. Finally, numerical results obtained from two test cases related to excursion sets and from an industrial test case for electrical machine design are given in Section 4. Proofs of the results of section 3 and some additional numerical results are given in the appendix.

2 Kernel-based SA in a generic space $\mathcal{Z}$

In this section, we explain how HSIC-ANOVA indices can be derived from kernel theory in a generic space. For a complete kernel theory, the reader can refer to the book *Support Vector Machine* of Steinwart and Christmann [2008].

2.1 Kernel Embedding, MMD and HSIC

Let $(\mathcal{Z}, \mathcal{A})$ be a measurable space on which $k_\mathcal{Z}$ is a kernel, i.e., a symmetric and positive definite function $k_\mathcal{Z} : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$. As stated in Theorem 4.21 of Steinwart and Christmann [2008], there exists a unique Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}_{k_\mathcal{Z}}$ of reproducing kernel $k_\mathcal{Z}$, i.e a kernel such that

- $\forall z \in \mathcal{Z}, \quad k_\mathcal{Z}(\cdot, z) \in \mathcal{H}_{k_\mathcal{Z}}$
- $\forall z \in \mathcal{Z}, \quad \forall f \in \mathcal{H}_{k_\mathcal{Z}}, \quad \langle f, k_\mathcal{Z}(\cdot, z)\rangle_{\mathcal{H}_{k_\mathcal{Z}}} = f(z).$

Let $\mathcal{M}_+^1(\mathcal{Z})$ be the space of probability measures on $\mathcal{Z}$. By Lemma 3.1 of Muandet et al. [2017], a *bounded* and *measurable* kernel $k_\mathcal{Z}$ is a sufficient condition for $\mathcal{M}_+^1(\mathcal{Z})$ to be embedded in $\mathcal{H}_{k_\mathcal{Z}}$ by the kernel mean embedding defined in the following definition.

Definition 2.1. *The kernel mean embedding of $\mathcal{M}_+^1(\mathcal{Z})$ in $\mathcal{H}_{k_\mathcal{Z}}$ is defined as*

$$\begin{array}{rccc} & \mathcal{M}_+^1(\mathcal{Z}) & \to & \mathcal{H}_{k_\mathcal{Z}} \\ \mu_{k_\mathcal{Z}} \ : & \mathbb{P} & \mapsto & \mu_{k_\mathcal{Z}}(\mathbb{P}) = \int_\mathcal{Z} k_\mathcal{Z}(z, \cdot) d\mathbb{P}(z). \end{array}$$

The distance between the kernel mean embedding of two distributions $\mathbb{P}$ and $\mathbb{Q}$ is called the Maximum Mean Discrepancy (MMD) and is denoted by $\text{MMD}_{k_\mathcal{Z}}(\mathbb{P}, \mathbb{Q}) = ||\mu_{k_\mathcal{Z}}(\mathbb{P}) - \mu_{k_\mathcal{Z}}(\mathbb{Q})||_{\mathcal{H}_{k_\mathcal{Z}}}$ (see Figure 1). To be a distance between the distributions $\mathbb{P}$ and $\mathbb{Q}$, the kernel mean embedding $\mu_{k_\mathcal{Z}}$ must be injective. In this case, the kernel $\mu_{k_\mathcal{Z}}$ is said to be *characteristic*. The MMD can then be used to define HSIC for the sensitivity analysis

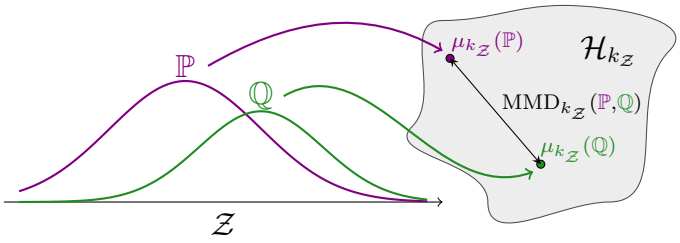


Figure 1 – Kernel mean embedding

context, as proposed in Da Veiga [2015].

Let $\eta : \mathcal{U} \to \mathcal{Z}$ be a measurable model with respect to the Borel σ algebras $\mathcal{B}_\mathcal{U}$ and $\mathcal{B}_\mathcal{Z}$, where $\mathcal{U} = \mathcal{U}_1 \times ... \times \mathcal{U}_p \subset \mathbb{R}^p$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which each input U_i is a random variable of probability distribution $\mathbb{P}_{U_i}$. For any subset of indices $A \subset \{1, ..., p\}$, $\boldsymbol{U}_A$ will denote the random vector $(U_i)_{i \in A}$. We are interested in knowing how the uncertainty of $Z = \eta(\boldsymbol{U})$ can be attributed to the different inputs (or group of inputs). To do this, kernel-based sensitivity analysis relies on measuring the dependence between $\boldsymbol{U}_A$ and Z by computing the MMD between the joint distribution $\mathbb{P}_{\boldsymbol{U}_A, Z}$ and the product of the marginal distributions $\mathbb{P}_{\boldsymbol{U}_A} \otimes \mathbb{P}_Z$. If this distance is zero, then $\boldsymbol{U}_A$ and Z

are independent, i.e. $\boldsymbol{U}_A$ has no effect on Z. This distance is called the Hilbert-Schmidt Independence Criterion (HSIC), first defined in Gretton et al. [2005a].

Definition 2.2. *For $A \subset \{1, ..., p\}$, let $k = k_A \otimes k_{\mathcal{Z}}$ be a kernel inducing the RKHS $\mathcal{H}_k$, with k_A a kernel on the space $\mathcal{U}_A = \otimes_{i \in A} \mathcal{U}_i$. The Hilbert-Schmidt Independence Criterion between $\boldsymbol{U}_A$ and Z, denoted by* $\mathrm{HSIC}(\boldsymbol{U}_A, Z)$*, is defined as*

$$\mathrm{HSIC}(\boldsymbol{U}_A, Z) := \mathrm{MMD}(\mathbb{P}_{\boldsymbol{U}_A, Z}, \mathbb{P}_{\boldsymbol{U}_A} \otimes \mathbb{P}_Z)^2 = ||\mu_k(\boldsymbol{U}_A, Z) - \mu_{k_A}(\boldsymbol{U}_A) \otimes \mu_{k_{\mathcal{Z}}}(Z)||^2_{\mathcal{H}_k}.$$

HSIC can then be used to test the independence between $\boldsymbol{U}_A$ and Z. For the test to be consistent, i.e. $\mathrm{HSIC}(\boldsymbol{U}_A, Z) = 0$ if and only if $\boldsymbol{U}_A$ and Z are independent, it is sufficient that both k_A and $k_{\mathcal{Z}}$ are characteristic (which doesn't necessarily mean that k is characteristic). This result can be obtained by combining Theorem 3.11 in Lyons [2013] and Proposition 29 in Sejdinovic et al. [2013] as explained in the introduction of Szabó and B. K. Sriperumbudur [2018]. To conduct screening, an independence test is performed for each input U_i : $(\mathcal{H}_0^i) : \ \mathrm{HSIC}(U_i, Z) = 0 \quad$ versus $\quad (\mathcal{H}_1^i) : \ \mathrm{HSIC}(U_i, Z) > 0$.

One of the reasons HSIC has become popular is that it can be expressed very simply in terms of kernel functions. From Lemma 1 of Gretton et al. [2005a] we have

$$\begin{aligned} \mathrm{HSIC}(\boldsymbol{U}_A, Z) = & \ \mathbb{E}[k_A(\boldsymbol{U}_A, \boldsymbol{U}_A')k_{\mathcal{Z}}(Z, Z')] \\ & + \mathbb{E}[k_A(\boldsymbol{U}_A, \boldsymbol{U}_A')]\mathbb{E}[k_{\mathcal{Z}}(Z, Z')] \\ & - 2\mathbb{E}[\mathbb{E}[k_A(\boldsymbol{U}_A, \boldsymbol{U}_A')|\boldsymbol{U}_A]\mathbb{E}[k_{\mathcal{Z}}(Z, Z')|Z]], \end{aligned} \tag{1}$$

where $(\boldsymbol{U}_A', Z')$ is an independent copy of $(\boldsymbol{U}_A, Z)$. This expression makes estimating the HSIC very easy, since it can be estimated with biased or unbiased estimators based on U- and V-statistics as introduced in Gretton et al. [2005a] and Song et al. [2007]. If the sample size is large enough, asymptotic estimation of the p-value associated to the independence test can be used, as suggested in Gretton et al. [2007]. For smaller samples, a permutation-based technique can be used (see De Lozzo and Marrel [2016]).

However, HSICs are not sufficient for ranking. The values $\mathrm{HSIC}(U_i, Z)$ must first be normalized in order to compare them and rank the inputs. In the context of sensitivity analysis, several normalizations have been studied, such as normalization by $\sum_{i=1}^p \mathrm{HSIC}(U_i, Z)$ (see Spagnol et al. [2019]) or by $\sqrt{\mathrm{HSIC}(U_i, U_i)\,\mathrm{HSIC}(Z, Z)}$ (Da Veiga [2015]). These indices have been used for ranking purposes, but they are not as satisfactory as the ANOVA decomposition of Sobol' indices. However, a recent study (Da Veiga [2021]) provides an ANOVA-like decomposition that justifies the use of HSIC for ranking purposes.

2.2 HSIC-ANOVA indices

In Da Veiga [2021] an ANOVA-like decomposition of the HSIC is proposed, which makes it usable to rank the inputs by influence. This decomposition requires strong assumptions, particularly the independence between the inputs and the ANOVA property of the input kernel.

Definition 2.3 (Orthogonal and ANOVA kernel)**.** *Let $\mathcal{U}$ be a measurable space. A kernel $k : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ is said to be orthogonal with respect to a probability measure $\nu \in \mathcal{M}_1^+(\mathcal{U})$ if*

$$\forall u \in \mathcal{U}, \int_{\mathcal{U}} k(u, z) d\nu(z) = 0.$$

A kernel $K : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ is said to be ANOVA w.r.t. ν if it can be decomposed as $K = 1 + k$ where k is orthogonal w.r.t. ν.

Theorem 2.1 (ANOVA decomposition of HSIC)**.** *Assuming that:*

i. The inputs $U_1, ..., U_p$ are mutually independent;

ii. Each input has an ANOVA kernel K_i w.r.t. the input distribution $\mathbb{P}_{U_i}$. For any group of inputs $\boldsymbol{U}_A$ with $A \subset \{1, ..., p\}$, the associated kernel is defined by

$$K_A = \bigotimes_{i \in A} K_i;$$

iii. For any $A \subset \{1, ..., p\}$, $\mathbb{E}[K_A(\boldsymbol{U}_A, \boldsymbol{U}_A)] < +\infty$ and $\mathbb{E}[k_{\mathcal{Z}}(Z, Z)] < +\infty$.

Then the ANOVA decomposition of the HSIC *is given by*

$$\mathrm{HSIC}(\boldsymbol{U}, Z) = \sum_{A \subseteq \{1,...,p\}} \sum_{B \subseteq A} (-1)^{|A|-|B|} \mathrm{HSIC}\left(\boldsymbol{U}_B, Z\right).$$

In addition to providing a decomposition of the HSIC, an ANOVA input kernel also simplifies the HSIC expression of the Equation 1, which becomes

$$\mathrm{HSIC}(\boldsymbol{U}_A, Z) = \mathbb{E}\left[(K_A(\boldsymbol{U}_A, \boldsymbol{U}_A{}') - 1) k_{\mathcal{Z}}(Z, Z')\right].$$

Given an independent and identically distributed (iid) sample $(\boldsymbol{U}_A^{(1)}, Z^{(1)}), ..., (\boldsymbol{U}_A^{(n)}, Z^{(n)})$ of $(\boldsymbol{U}_A, Z)$, the previous expression is easily estimated by a U-statistic

$$\widehat{\mathrm{HSIC}}\left(\boldsymbol{U}_A, Z\right) = \frac{2}{n(n-1)} \sum_{i<j}^{n} \left(K_A\left(U_A^{(i)}, U_A^{(j)}\right) - 1\right) k_{\mathcal{Z}}\left(Z^{(i)}, Z^{(j)}\right). \tag{2}$$

The decomposition given in Theorem 2.1 allows SA indices to be defined in a similar way to Sobol' indices.

Definition 2.4 (HSIC-ANOVA indices)**.** *The first-order and total-order HSIC-ANOVA indices can be defined by*

$$\forall i \in \{1, ..., p\}\ S_i^{\mathrm{HSIC}} := \frac{\mathrm{HSIC}(U_i, Z)}{\mathrm{HSIC}(\boldsymbol{U}, Z)} \text{ and } S_{T_i}^{\mathrm{HSIC}} := 1 - \frac{\mathrm{HSIC}(\boldsymbol{U}_{-i}, Z)}{\mathrm{HSIC}(\boldsymbol{U}, Z)},$$

with $\boldsymbol{U}_{-i} = (U_1, ..., U_{i-1}, U_{i+1}, ..., U_d)$. These indices can be generalized to groups of inputs.

These indices can be used to rank the inputs by influence by ranking their indices. It is also possible to use either first-order or total-order indices for screening, as we still have

$$S_i^{\mathrm{HSIC}} = 0 \iff S_{T_i}^{\mathrm{HSIC}} = 0 \iff U_i \perp \Gamma,$$

as shown in Sarazin et al. [2022]. The authors also propose associated independence tests that have better statistical power than the usual tests performed with HSIC indices.

It is important to note that the three assumptions of Theorem 2.1, and especially ii., restrict the choice of the input kernel to be ANOVA. Few kernels are known to be ANOVA. The best known are the Sobolev kernels (see Theorem 3.3 of Sarazin et al. [2023]), which correspond to the reproducing kernel of some Sobolev spaces. These kernels are ANOVA with respect to the uniform law on $[0, 1]$ and are also characteristic (Remark 3.4 and

Proposition 3.6 of Sarazin et al. [2023]). It is also possible to derive ANOVA kernels from classic kernels. One of the possible transformations is suggested in Ginsbourger et al. [2016]: for any given kernel k, its ANOVA counterpart K w.r.t. ν is defined by

$$K(x,y) = 1 + k(x,y) - \int k(x,z)d\nu(z) - \int k(z,y)d\nu(z) + \int\int k(z,z')d\nu(z)d\nu(z'). \tag{3}$$

In the case of input variables uniformly distributed on $[0,1]$, the previous transformation is known analytically for some classic kernels given in the appendix of Ginsbourger et al. [2016].

Given ANOVA and characteristic input kernels K_i and a characteristic output kernel $k_{\mathcal{Z}}$, first-order and total-order HSIC-ANOVA indices can be computed and used for both screening and ranking. This approach has the strength of being very permissive about the nature of the output. Indeed, once one has a measurable and bounded characteristic kernel on an output space on which a random element can be defined, one can derive HSIC-ANOVA sensitivity indices to perform a sensitivity analysis of the complex model. We propose to do so in the next section dealing with an output space of sets.

3 Kernel-based Sensitivity Analysis for sets

In this section, we apply the methodology of the previous section to the case of set-valued outputs, i.e., where $\mathcal{Z}$ is a space of sets. We introduce a probabilistic framework for random sets on which we propose a customized kernel and study its property so that it can be used to define HSIC-ANOVA indices for sets. We also introduce and study the estimation of these new indices. Basic concepts of functional analysis and topology useful in this section can be found, for example, in Rudin [1987].

3.1 Probabilistic framework of random sets

A general theory of random sets can be found in Molchanov [2005]. In our case, however, we propose to define a random set as $\Gamma = \eta(\boldsymbol{U})$, where η is a measurable function from $\mathcal{U}$ to a space of sets to be defined. First, we define the space of sets we are working in, and a corresponding σ-algebra.

Let $\mathscr{L}(\mathcal{X})$ be the space of all Lebesgue-measurable subsets of a compact space $\mathcal{X} \subset \mathbb{R}^d$. Let $\delta : \mathscr{L}(\mathcal{X}) \times \mathscr{L}(\mathcal{X}) \to \mathbb{R}$ be the Lebesgue measure of the symmetric difference defined by $\delta(\gamma_1, \gamma_2) = \lambda(\gamma_1 \Delta \gamma_2)$ where Δ is the symmetric difference defined by $\gamma_1 \Delta \gamma_2 = (\gamma_1 \cup \gamma_2) \backslash (\gamma_1 \cap \gamma_2)$ and λ is the Lebesgue measure on $\mathcal{X}$. $L^2(\mathcal{X})$ denotes the space of squared Lebesgue integrable functions on $\mathcal{X}$, and the L^2 norm is denoted by $||\cdot||_2$. δ can also be seen in terms of the L^2 norm of the indicator functions of γ_1 and γ_2, as given in the following lemma.

Lemma 3.1. *Let $\gamma_1, \gamma_2 \in \mathscr{L}(\mathcal{X})$, we have*

$$\delta(\gamma_1, \gamma_2) = ||\mathbb{1}_{\gamma_1} - \mathbb{1}_{\gamma_2}||_2^2,$$

and

$$\delta(\gamma_1, \gamma_2) = 0 \Leftrightarrow \gamma_1 = \gamma_2 \ \lambda\text{-almost everywhere.}$$

Thus δ is a pseudo-metric in $\mathscr{L}(X)$, since $\delta(\gamma_1, \gamma_2) = 0$ only implies that $\gamma_1 = \gamma_2$ λ-almost everywhere. To work in a metric space, we quotient $\mathscr{L}(\mathcal{X})$ by the equivalence relation $\delta(\gamma_1, \gamma_2) = 0$. The resulting quotient space is called $\mathscr{L}^*(\mathcal{X})$. δ remains well defined on $\mathscr{L}_2^\star(\mathcal{X}) = \mathscr{L}^\star(\mathcal{X}) \times \mathscr{L}^\star(\mathcal{X})$ and becomes a distance sometimes called the Fréchet-Nikodym-Aronszajn metric (see Marczewski and Steinhaus [1958]). With this distance, $\mathscr{L}^*(\mathcal{X})$ can be provided with the Borel σ-algebra $\mathcal{B}_\Gamma := \mathcal{B}(\mathscr{L}^*(\mathcal{X}), \delta)$ to make it a measurable space. A random set is then simply defined as $\Gamma = \eta(\boldsymbol{U})$ where $\eta : \mathcal{U} \to \mathscr{L}^*(\mathcal{X})$ is measurable with respect to $\mathcal{B}_{\boldsymbol{U}}$ and $\mathcal{B}_\Gamma$. The probability distribution of a random set Γ is then defined as the push-forward probability measure of $\mathbb{P}_{\mathcal{U}}$ through η.

Examples of random set are numerous in industrial applications, especially in the context of optimization or inversion, where determining the feasible set of solutions is crucial. In the presence of uncertainties, this feasible set is random and can be called an excursion set as given in the following definition.

Definition 3.1. *Let $g : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$ be a measurable function. The excursion set Γ_g associated with the constraint $g \leq 0$ is defined by*

$$\Gamma_g = \{x \in X, \quad g(x, \boldsymbol{U}) \leq 0\}.$$

3.2 A kernel between sets

Now that we have a framework of random sets, we want to follow the methodology of Section 2 but with $\mathcal{Z} = \mathscr{L}^*(\mathcal{X})$ to define HSIC indices for sets. To do this, we first need a kernel on sets, i.e. a symmetric and positive definite function $k : \mathscr{L}^*(\mathcal{X}) \times \mathscr{L}^*(\mathcal{X}) \to \mathbb{R}$. We propose the function k_{set} defined by

$$\forall \gamma_1, \gamma_2 \in \mathscr{L}^*(\mathcal{X}), \quad k_{set}(\gamma_1, \gamma_2) := e^{-\frac{\lambda(\gamma_1 \Delta \gamma_2)}{2\sigma^2}},$$

where σ is a positive scalar. k_{set} is inspired by the classic Gaussian kernel, since the Lebesgue measure of the symmetric difference is equal to the L^2 norm of the difference between two indicator functions, as shown in Lemma 3.1. This function was shown to be positive definite in Balança and Herbin [2012] when the space of sets is an indexing collection, but the proof is similar in our case. Moreover, k_{set} is bounded and measurable.

Proposition 3.1. *The function $k_{set} : \mathscr{L}^*(\mathcal{X}) \times \mathscr{L}^*(\mathcal{X}) \to \mathbb{R}$ defined by*

$$\forall \gamma_1, \gamma_2 \in \mathscr{L}^*(\mathcal{X}), \quad k_{set}(\gamma_1, \gamma_2) = e^{-\frac{\lambda(\gamma_1 \Delta \gamma_2)}{2\sigma^2}},$$

is symmetric and positive definite for any positive scalar σ which means that k_{set} is a kernel.

Proposition 3.2. *The kernel k_{set} is bounded and measurable from $\mathscr{L}_2^\star(\mathcal{X}) = \mathscr{L}^\star(\mathcal{X}) \times \mathscr{L}^\star(\mathcal{X})$ to $\mathbb{R}$.*

As in Definition 2.1, this allows to define the mean embedding of random set distributions $\mu_{k_{set}}$. We then show that it is injective, i.e. k_{set} is characteristic.

Proposition 3.3. *The kernel k_{set} is characteristic.*

We now have all the ingredients to use HSIC-ANOVA indices for set-valued outputs. We can now screen the inputs using independence tests and rank them by their influence on the random set Γ using either first-order or total-order indices, that we denote $S_i^{\mathrm{H}_{set}}$ and $S_{T_i}^{\mathrm{H}_{set}}$. HSIC-ANOVA indices applied to sets provide a solution for the screening and ranking of the inputs of a set-valued model. However, to achieve these purposes, they must first be estimated, but the presence of sets raises some difficulties.

3.3 Estimation of HSIC-ANOVA indices for sets

As given in Equation 2, $\mathrm{HSIC}\left(\boldsymbol{U}_A, \Gamma\right)$ can be estimated by

$$\widehat{\mathrm{HSIC}}\left(\boldsymbol{U}_A, \Gamma\right) = \frac{2}{n(n-1)} \sum_{i<j}^{n} \left(K_A\left(\boldsymbol{U}_A^{(i)}, \boldsymbol{U}_A^{(j)}\right) - 1 \right) k_{set}\left(\Gamma^{(i)}, \Gamma^{(j)}\right),$$

where $(\boldsymbol{U}_A^{(i)}, \Gamma^{(i)}), i = 1, ..., n$ is an iid sample of $(\boldsymbol{U}_A, \Gamma)$.

For set-valued models, it is common to have access only to the knowledge of whether a given point $x \in \mathcal{X}$ is in a set output Γ or not. In terms of estimation, this means that the Lebesgue measure of the symmetric difference and thus k_{set} must be estimated.

For two random sets $(\Gamma^{(i)}, \Gamma^{(j)})$, $k_{set}(\Gamma^{(i)}, \Gamma^{(j)}) = e^{-\frac{\lambda(\Gamma^{(i)} \Delta \Gamma^{(j)})}{2\sigma^2}}$ can be written as

$$k_{set}(\Gamma^{(i)}, \Gamma^{(j)}) = \exp\left(-\frac{\lambda(\mathcal{X})}{2\sigma^2} \mathbb{E}[\mathbb{1}_{\Gamma^{(i)} \Delta \Gamma^{(j)}}(\boldsymbol{X}) | (\Gamma^{(i)}, \Gamma^{(j)})] \right),$$

where $\boldsymbol{X} \sim \mathcal{U}(\mathcal{X})$ using that $\lambda(\Gamma) = \lambda(\mathcal{X}) \mathbb{E}_{\boldsymbol{X} \sim \mathcal{U}(\mathcal{X})}[\mathbb{1}_{\Gamma}(\boldsymbol{X}) | \Gamma]$. Then, given an iid sample $(\boldsymbol{X}^{(1)}, ..., \boldsymbol{X}^{(m)})$ of $\boldsymbol{X}$, we can estimate $k_{set}(\Gamma^{(i)}, \Gamma^{(j)})$ by

$$\widehat{k_{set}}(\Gamma^{(i)}, \Gamma^{(j)}) = \exp\left(-\frac{\lambda(\mathcal{X})}{2\sigma^2} \frac{1}{m} \sum_{k=1}^{m} \mathbb{1}_{\Gamma^{(i)} \Delta \Gamma^{(j)}}(\boldsymbol{X}^{(k)}) \right).$$

We now need to plug this estimator into our previous one, resulting in a Nested Monte Carlo (NMC) estimator. To be compatible with the framework of NMC estimator, it is necessary for samples of $\boldsymbol{X}$ to be drawn independently for every pair of indices (i, j) within the outer loop. Let $(\boldsymbol{X}_{i,j}^{(k)})$ be an iid sample of $\boldsymbol{X} \sim \mathcal{U}(\mathcal{X})$, with $k \in \{1, ..., m\}$ and $(i, j) \in \{1, ..., n\}^2$ s.t. $i < j$. Including the previous estimation of the kernel k_{set}, the NMC estimator of $\mathrm{HSIC}(\boldsymbol{U}_A, \Gamma)$ is given by

$$\widehat{\mathrm{HSIC}}^{nest}\left(\boldsymbol{U}_A, \Gamma\right) = \frac{2}{n(n-1)} \sum_{i<j}^{n} \left(K_A\left(\boldsymbol{U}_A^{(i)}, \boldsymbol{U}_A^{(j)}\right) - 1 \right) e^{-\frac{\lambda(\mathcal{X})}{2\sigma^2} \frac{1}{m} \sum_{k=1}^{m} \mathbb{1}_{\Gamma^{(i)} \Delta \Gamma^{(j)}}(\boldsymbol{X}_{i,j}^{(k)})}. \quad (4)$$

Note that this estimator is biased like any NMC estimator (see Rainforth et al. [2016]). For each pair $(\Gamma^{(i)}, \Gamma^{(j)})$, the previous estimator requires checking whether $\boldsymbol{X}_{i,j}^{(k)} \in \Gamma^{(i)} \Delta \Gamma^{(j)}$ for each $k \in \{1, ..., m\}$. Each of these requires checking whether $\boldsymbol{X}_{i,j}^{(k)} \in \Gamma^{(i)}$ and $\boldsymbol{X}_{i,j}^{(k)} \in \Gamma^{(j)}$, corresponding in the example of excursion sets, to compute $g(\boldsymbol{X}_{i,j}^{(k)}, \boldsymbol{U}_A^{(i)})$. This means $n(n-1)m$ evaluations to estimate the index, which is not affordable. To solve this problem, we propose to reuse the same $\boldsymbol{X}^{(k)}$ for each (i, j). By doing so, we only need to test whether

$\boldsymbol{X}^{(k)} \in \Gamma^{(i)}$ for each k and i, reducing the number of evaluations to nm. The estimator is then given by

$$\widehat{\widehat{\mathrm{HSIC}}}(\boldsymbol{U}_A, \Gamma) = \frac{2}{n(n-1)} \sum_{i<j}^{n} \left(K_A\left(\boldsymbol{U}_A^{(i)}, \boldsymbol{U}_A^{(j)}\right) - 1 \right) e^{-\frac{\lambda(\mathcal{X})}{2\sigma^2} \frac{1}{m} \sum_{k=1}^{m} \mathbb{1}_{\Gamma^{(i)} \Delta \Gamma^{(j)}}(\boldsymbol{X}^{(k)})}.$$

By simulating a single m-sample of $\boldsymbol{X}$, $\widehat{\widehat{\mathrm{HSIC}}}(\boldsymbol{U}_A, \Gamma)$ is no longer a classic NMC estimator. Nevertheless, we show that its quadratic risk goes to 0 and we give an upper bound.

Proposition 3.4. *With the previous notations, we have*

$$\mathbb{E}\left(\widehat{\widehat{\mathrm{HSIC}}}(\boldsymbol{U}_A, \Gamma) - \mathrm{HSIC}(\boldsymbol{U}_A, \Gamma) \right)^2 \leq 2 \left(\frac{2\sigma_1^2}{n(n-1)} + \frac{4(n-2)\sigma_2^2}{n(n-1)} + \frac{L^2 \sigma_3^2}{m} \right),$$

where

- $\sigma_1^2 = Var\left(\left(K_A\left(\boldsymbol{U}_A, \boldsymbol{U}_A{}'\right) - 1 \right) k_{set}(\Gamma, \Gamma') \right)$,
- $\sigma_2^2 \;\; = Var\left(\mathbb{E}\left[\left(K_A\left(\boldsymbol{U}_A, \boldsymbol{U}_A{}'\right) - 1 \right) k_{set}(\Gamma, \Gamma') | (\Gamma, \boldsymbol{U}_A) \right] \right)$,
- $\sigma_3^2 = \mathbb{E}\left[\left(K_A\left(\boldsymbol{U}_A, \boldsymbol{U}_A{}'\right) - 1 \right)^2 \mathrm{Var}\left(\mathbb{1}_{\Gamma \Delta \Gamma'}(\boldsymbol{X}) | (\boldsymbol{U}_A, \boldsymbol{U}_A{}', \Gamma, \Gamma') \right) \right]$,
- $L = \frac{\lambda(\mathcal{X})}{2\sigma^2}$,

and where $(\boldsymbol{U}_A{}', \Gamma')$ *is an independent copy of* $(\boldsymbol{U}_A, \Gamma)$.

The quadratic risk has a rate of $\mathcal{O}(\frac{1}{n} + \frac{1}{m})$, which tends to say that we should use $n = m$. In the case of the classic NMC estimator (see Rainforth et al. [2018]), here without reusing the same samples of $\boldsymbol{X}$, the rate is $\mathcal{O}(\frac{1}{n} + \frac{1}{m^2})$ as shown in the Appendix 7.1.6. However, even if the convergence rate is better, the number of evaluations required is $(n-1)$ times larger, which makes our choice to use the same sample of $\boldsymbol{X}$ more efficient. Since the previous result is only an upper bound, we can also hope to get closer to a $\mathcal{O}(\frac{1}{n} + \frac{1}{m^2})$ convergence rate in the application tests. That is, we can expect to have achieved convergence of our estimator without having to take high n and m. This will be highlighted in the numerical results presented in the next section. By plugging the estimator $\widehat{\widehat{\mathrm{HSIC}}}$, we denote $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ and $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ the estimators of the first-order and total-order HSIC ANOVA indices on sets.

4 Numerical Experiments

In this part, our goal is to quantify the influence of the inputs $\boldsymbol{U}$ on different excursion sets Γ_g (defined in Definition 3.1). To do this, we will consider three test cases. We first consider an analytically known function g defined on $\mathcal{X} \subset \mathbb{R}^2$ and $\mathcal{U} \subset \mathbb{R}^2$, borrowed from El Amri et al. [2023]. In this first toy case, we will also study numerically the quadratic risk bound given in Proposition 3.4. The second test case is related to an optimization problem with a stationary harmonic oscillator, from Cousin et al. [2022], on which we want to quantify the impact of some uncertain inputs on the feasible sets. In the last case, a sensitivity analysis is performed in the context of a bi-objective robust optimization of a permanent magnet-assisted synchronous reluctance machine for electrical vehicle application. For each example, our sensitivity analysis is performed in two steps:

- Screening: We compute the p-values associated with the test $\mathrm{HSIC}(U_i, \Gamma_g) = 0$ versus $\mathrm{HSIC}(U_i, \Gamma_g) > 0$. We use permutation based estimation. If the p-value is

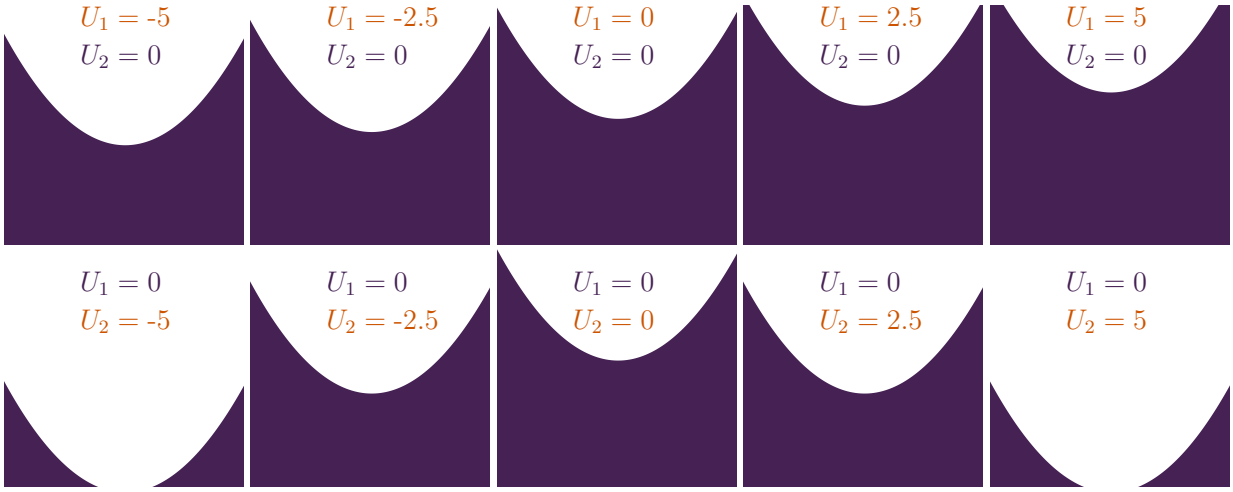


Figure 2 – Excursion set of the constraint $g \leq 0$ for $U_1 \in \{-5, -2.5, 0, 2.5, 5\}$ and $U_2 = 0$ (first row) and for $U_1 = 0$ and $U_2 \in \{-5, -2.5, 0, 2.5, 5\}$. (second row)

greater than a risk α, the input is negligible, and influential otherwise. We use $\alpha = 0.05$ but this value can be changed depending on the application.

- Ranking: We compute the first-order indices $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ and the total-order indices $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ of all inputs. The first-order indices $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ are used to rank the inputs. With the total-order indices $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ we can quantify the HSIC interaction effects.

The numerical implementation of the indices is done in the R language, using the `sensitivity` package and in particular the `sensiHSIC` and `testHSIC` functions, which allows to compute HSIC-ANOVA indices and p-values.

4.1 Excursion sets of a toy function

In this part, we will estimate the previous HSIC-ANOVA indices on the excursion set Γ_g defined by the following function g from El Amri et al. [2023],

$$\forall \boldsymbol{x}, \boldsymbol{u} \in [-5, 5]^2 \times [-5, 5]^2 \quad g(\boldsymbol{x}, \boldsymbol{u}) = -x_1^2 + 5x_2 - u_1 + u_2^2 - 1.$$

Visual conjecture (Figure 2) shows that u_2 seems to induce more changes in Γ_g than u_1. To verify this assumption, we compute the p-values and the indices $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$, $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ for each input $U_i \sim \mathcal{U}([-5, 5])$ for $i \in \{1, 2, 3\}$, where U_3 is an additional dummy input that does not appear in the function g. We first use $n = m = 100$ and then $n = m = 1000$. The hyperparameter σ^2 of the kernel k_{set} is chosen to be equal to the empirical mean of $\lambda(\Gamma_g^{(i)} \Delta \Gamma_g^{(j)})$ with $i > j$. We compute the indices for five characteristic ANOVA kernels. The first is the Sobolev kernel K_{Sob} of order 1, defined by:

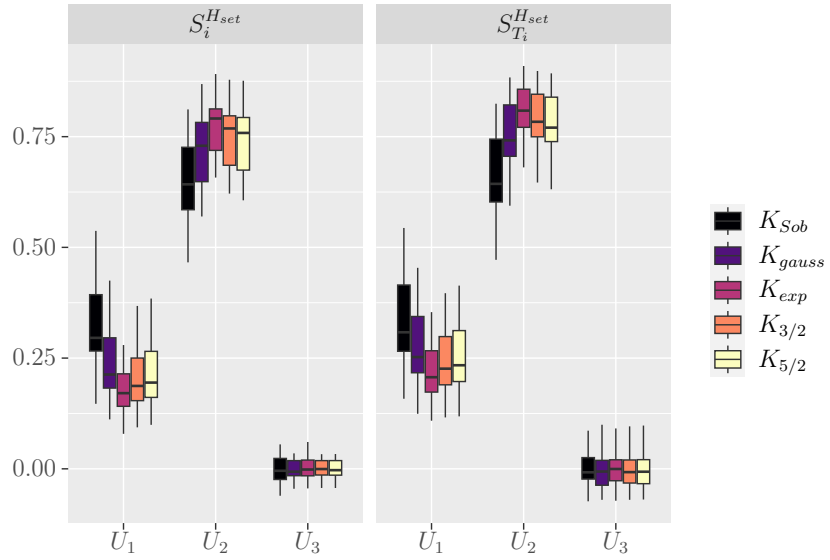
$$K_{Sob}(x, y) = 1 + (x - \frac{1}{2})(y - \frac{1}{2}) + \frac{1}{2}[(x - y)^2 - |x - y| + \frac{1}{6}].$$

The four others are obtained using the transformation given in (3) based on four classic kernels: the Gaussian kernel, the Laplace kernel and the Matérn 3/2 and 5/2. For these kernels, the previous transformation is known analytically and can be found in the Appendix

of Ginsbourger et al. [2016]). The hyper-parameter is taken to be equal to the empirical standard deviation of the inputs. Since the inputs are not uniformly distributed on $[0, 1]$, we apply the inverse of the cumulative distribution function, as suggested in Da Veiga [2021]. We repeat the estimation 20 times over 20 independent samples. The results are shown in Figure 3 for $n = m = 100$ and in Figure 4 for $n = m = 1000$. The acceptance rates of the independence hypothesis are given in the table on the left side of the figure, and the variability of the indices is visualized in the boxplots on the right side.

Kernel	U_1	U_2	U_3
K_{Sob}	0	0	95
K_{gauss}	0	0	95
K_{exp}	0	0	95
$K_{3/2}$	0	0	95
$K_{5/2}$	0	0	90

(a) Acceptance rates (%) over 20 independence tests with a risk of 5%



(b) Estimations of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ and $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$

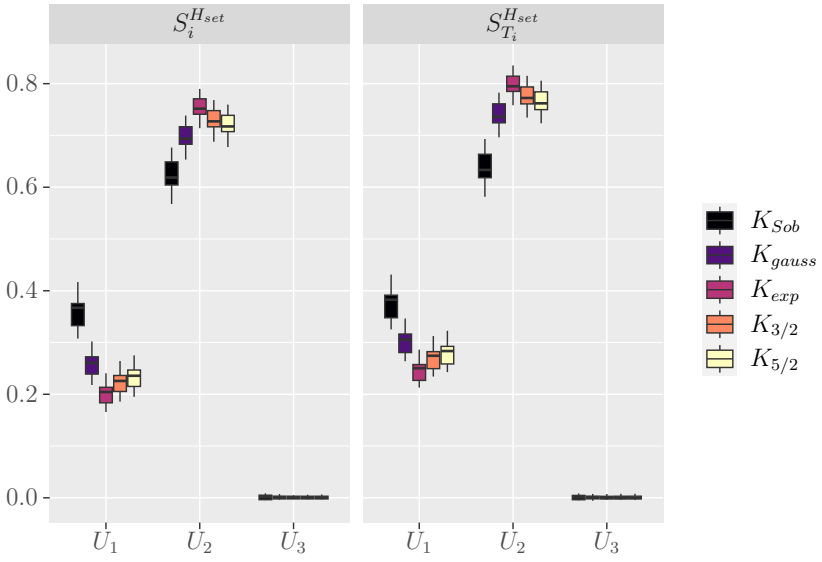
Figure 3 – Acceptance rates (a) and estimation of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ and $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ (b) for the excursion set Γ_g computed for 5 kernels with $n = 100$, $m = 100$ and repeated 20 times

With these results, we classify the inputs as influential or negligible (screening) and rank them by influence (ranking). 100% of the p-values of U_1 and U_2 are below the threshold of 0.05, so they are classified as influential. U_3 has a p-value greater than 0.05 around 95% of the time, so it is correctly identified as negligible by the independence test. The first-order and total-order indices give the expected results that U_2 has a greater influence than U_1 on the excursion sets. More precisely, it shows that U_2 alone explains about 70% of $\mathrm{HSIC}(\boldsymbol{U}, \Gamma_g)$ (depending on the input kernel), while U_1 explains about 25%. The artificial input U_3 is responsible for 0%, as expected. The rest is some kind of interaction between U_1 and U_2. This is confirmed by looking at the total-order indices, whose values are a bit greater than the first-order. However, this part of the interaction is not to be interpreted as in the case of the Sobol' indices. It is actually an open question to interpret what kind of interactions are detected by the HSIC-ANOVA indices. This is studied in Sarazin et al. [2023] in the case of Sobolev kernels. The effect of the input kernel on the HSIC-ANOVA indices is also an open question, and in this toy case we can say that the input kernel has a non-negligible effect on the estimates of the p-values and the indices. However, in this case it does not change the results of the screening and the ranking. It also seems that the Sobolev kernel behaves differently from the other four kernels.

Taking $n = m = 1000$ in Figure 4 reduces the variance of the indices, but was not

Kernel	U_1	U_2	U_3
K_{Sob}	0	0	95
K_{gauss}	0	0	95
K_{exp}	0	0	100
$K_{3/2}$	0	0	95
$K_{5/2}$	0	0	95

(a) Acceptance rates (%) over 20 independence tests with a risk of 5%



(b) Estimations of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ and $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$

Figure 4 – Acceptance rates (a) and estimation of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ and $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ (b) for the excursion set Γ_g computed for 5 kernels with $n = 1000$, $m = 1000$ and repeated 20 times

necessary to screen and rank the inputs. To support this comment, we compute the relative quadratic risk of the estimator $\widehat{\widehat{\mathrm{HSIC}}}(U_1, \Gamma_g)$ defined by:

$$\mathcal{R}(\widehat{\widehat{\mathrm{HSIC}}}(U_1, \Gamma_g)) = \mathbb{E}\left(\frac{\widehat{\widehat{\mathrm{HSIC}}}(U_1, \Gamma_g) - \mathrm{HSIC}(U_1, \Gamma_g)}{\mathrm{HSIC}(U_1, \Gamma_g)} \right)^2 .$$

We also compute the associated upper bound given in Proposition 3.4 and the upper bound for a classic NMC estimator (with independent $\boldsymbol{X}_{i,j}^{(k)}$ for all $1 \leq i, j \leq n$, see equation 4). The "true" value of HSIC, used to compute the quadratic risk, and the constants σ_1, σ_2, and σ_3 are computed for $n = m = 3000$. The Sobolev kernel is used as the input kernel. We plot the risk and the two bounds in the Figure 5 from $n = m = 30$ to $n = m = 500$. We observe, as expected, that the quadratic risk of the estimator is below the upper bound of Proposition 3.4, but also below the NMC upper bound we would have by taking independent $\boldsymbol{X}_{i,j}^{(k)}$. This example confirms that we have not lost too much in terms of variance of the estimator by reusing the same $\boldsymbol{X}^{(k)}$. It also seems that we do not necessarily need to use high values of $n = m$. Therefore, in the next example, we will only compute the indices for $n = m = 100$.

4.2 Excursion sets of the optimization of an oscillator

In Cousin et al. [2022], an optimization is performed with three probabilistic constraints. We consider the first two constraints, which are defined by the following functions :

$$g_1(x_1, x_2, u_1, u_2, u_p, u_{r_1}, u_{r_2}) = u_{r_1} - \max_{t \in [0,T]} \mathcal{Y}'(x_1 + u_1, x_2 + u_2, u_p; t),$$

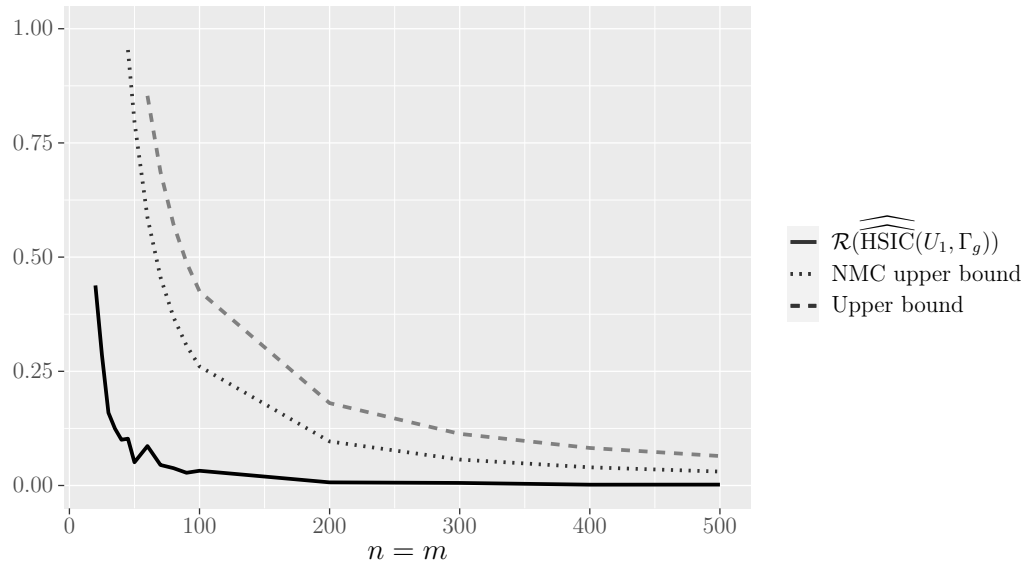


Figure 5 – Evolution of $\widehat{\mathcal{R}(\widehat{\mathrm{HSIC}}(U_1, \Gamma_g))}$ and of the associated upper bounds for the excursion set Γ_g

Table 1 – Definition of the uncertain inputs

Uncertainty	Distribution	Uncertainty	Distribution
U_1	$\mathcal{U}[-0.3, 0.3]$	U_{r_1}	$\mathcal{N}(1, 0.1^2)$
U_2	$\mathcal{U}[-1, 1]$	U_{r_2}	$\mathcal{N}(2.5, 0.25^2)$
U_p	$\mathcal{U}[0.5, 1.5]$	U_{r_3}	$\mathcal{N}(15, 3^2)$

$$g_2(x_1, x_2, u_1, u_2, u_p, u_{r_1}, u_{r_2}) = u_{r_2} - \max_{t \in [0,T]} \mathcal{Y}''(x_1 + u_1, x_2 + u_2, u_p; t),$$

where $\mathcal{Y}(x_1 + u_1, x_2 + u_2, u_p; t)$ is the solution of the harmonic oscillator defined by:

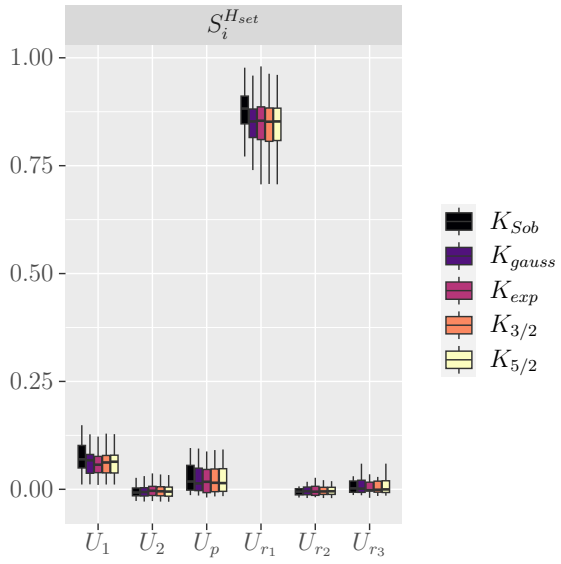
$$(x_1 + u_1)\mathcal{Y}''(t) + u_p\mathcal{Y}'(t) + (x_2 + u_2)\mathcal{Y}(t) = \eta(t).$$

The deterministic input domain is $\mathcal{X} = [1, 5] \times [20, 50]$. The uncertain input probability distributions are given in Table 1. U_{r_3} is initially a random input associated with a third constraint. In our case it will play the role of a dummy input to check if it is recognized as negligible. We study the impact of the uncertain inputs on the excursion sets Γ_{g_1} associated with the constraint $g_1 \leq 0$ and on the excursion sets Γ_{g_2} associated with $g_2 \leq 0$. Since kernel-based methods are appropriate for vectorial outputs, we also consider the case where an output is the pair of the two excursion sets $(\Gamma_{g_1}, \Gamma_{g_2})$, each associated with a constraint. For each case, we compute $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$, $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ and the associated p-values for each uncertain input with $n = 100$ and $m = 100$. We again use the output kernel k_{set} and $k_{set} \otimes k_{set}$ for the pairs of excursion sets. Note that the kernel $k_{set} \otimes k_{set}$ may not be characteristic. The same 5 ANOVA kernels are used as in the previous example. We repeat the estimation 20 times to again obtain the acceptance rates and boxplots of the first-order indices given in Figures 6 to 8. The total-order indices are given in Figure 12 in the Appendix 7.2.

Acceptance rates tables can be used for screening and the boxplots for ranking. For the random set associated with the constraint $g_1 \leq 0$ (Figure 6), only U_{r_1} is always detected

Kernel	U_1	U_2	U_p	U_{r_1}	U_{r_2}	U_{r_3}
K_{Sob}	15	95	55	0	100	90
K_{gauss}	15	100	60	0	100	95
K_{exp}	25	90	60	0	100	95
$K_{3/2}$	15	90	60	0	100	95
$K_{5/2}$	15	100	60	0	100	95

(a) Acceptance rates (%) over 20 independence tests with a risk of 5%

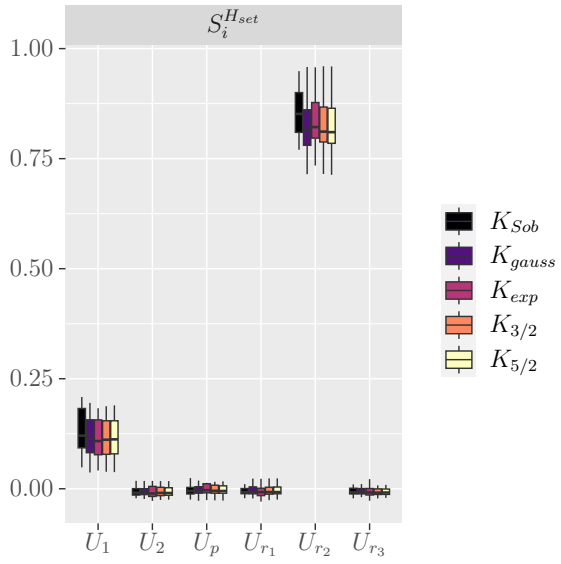


(b) Estimations of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$

Figure 6 – Acceptance rates (a) and estimation of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ (b) for the excursion set Γ_{g_1} computed for 5 kernels with $n = 100$, $m = 100$ and repeated 20 times

Kernel	U_1	U_2	U_p	U_{r_1}	U_{r_2}	U_{r_3}
K_{Sob}	0	100	95	95	0	95
K_{gauss}	0	100	95	95	0	95
K_{exp}	0	100	95	100	0	95
$K_{3/2}$	0	100	95	100	0	95
$K_{5/2}$	0	100	95	95	0	95

(a) Acceptance rates (%) over 20 independence tests with a risk of 5%

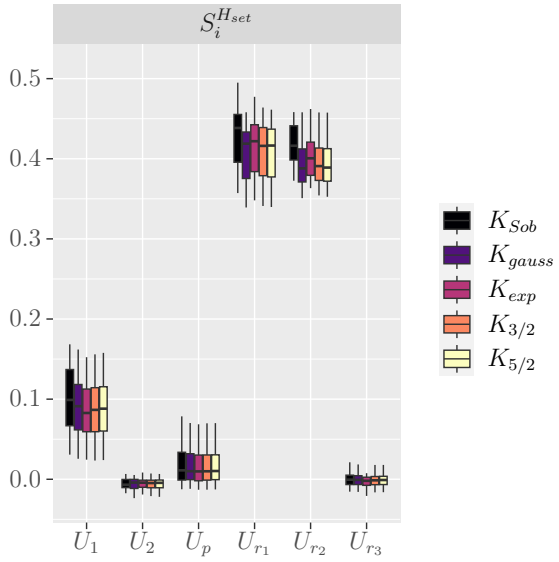


(b) Estimations of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$

Figure 7 – Acceptance rates (a) and estimation of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ (b) for the excursion set Γ_{g_2} computed for 5 kernels with $n = 100$, $m = 100$ and repeated 20 times

Kernel	U_1	U_2	U_p	U_{r_1}	U_{r_2}	U_{r_3}
K_{Sob}	0	95	60	0	0	90
K_{gauss}	0	100	60	0	0	90
K_{exp}	0	95	60	0	0	90
$K_{3/2}$	0	100	60	0	0	90
$K_{5/2}$	0	100	60	0	0	90

(a) Acceptance rates (%) over 20 independence tests with a risk of 5%



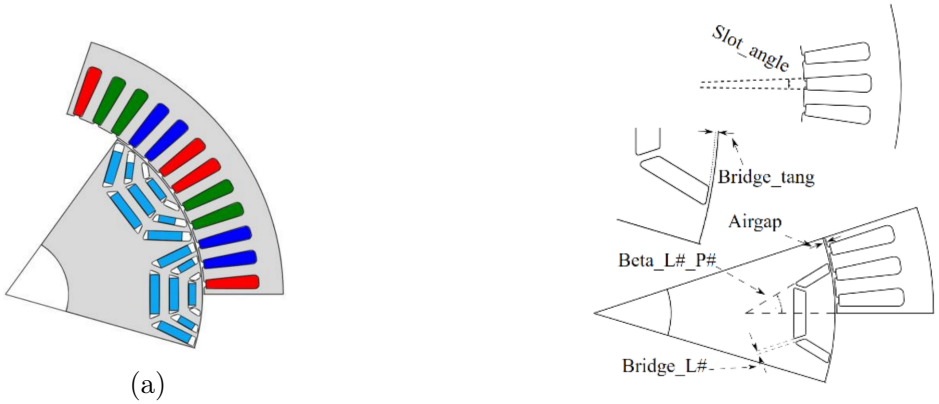
(b) Estimations of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$

Figure 8 – Acceptance rates (a) and estimation of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ (b) for the pair of excursion sets $(\Gamma_{g_1}, \Gamma_{g_2})$ computed for 5 kernels with $n = 100$, $m = 100$ and repeated 20 times

as influential. U_1 is also detected as influential most of the time (with an acceptance rate of about 20%), and U_p is detected as influential only 40% of the time. Between these three inputs, U_{r_1} is much more influential, since $\hat{\hat{S}}_{r_1}^{\mathrm{H}_{set}} \approx 80\%$. Then U_1 and U_p are responsible for about 10% and 3% of $\mathrm{HSIC}(\boldsymbol{U}, \Gamma_{g_1})$. For the second constraint $g_2 \leq 0$ (Figure 7), only U_1 and U_{r_2} are influential with their first-order index of about 10% and 80%, respectively. If we consider the pair of the two random sets (Figure 8), we get a kind of compromise between the two previous cases: U_1, U_{r_1} and U_{r_2} are always recognized as influential. In terms of ranking, U_{r_1} and U_{r_2} have almost the same influence, with their first-order index around 40% and U_1 remaining around 10%. It is important to note that considering pairs of random sets is completely different from considering the random set associated with the pair of constraints, denoted $\Gamma_{(g_1,g_2)}$, i.e., the intersection of the two sets. The results of the latter are shown in Figure 11 in Appendix 7.2 and are indeed very different from the results in Figure 8. In this test case, we observe that the choice of the input kernel has a limited impact on the indices. Regardless of the stage of screening or ranking, the conclusions remain the same for the five kernels.

4.3 Sensitivity analysis for robust conception of an electrical machine

In [Reyes Reyes et al. 2024], a sensitivity analysis is performed in the context of robust conception of an electrical machine. The studied machine is a permanent magnet assisted synchronous reluctance motor which is one of the most used machines nowadays in electrical vehicles. The purpose of the authors is to take into account the uncertainties (manufactoring and assembly tolerances) on geometric and magnetic parameters of the machine components in the optimization of the mean torque (to be maximized) and the



(a) (b)

Figure 9 – (a) Geometry of the permanent magnet assisted synchronous reluctance motor - (b) Design parameters for one layer (# is the number of layer)

Input parameters	**Lower bound** $x_{\min}$	**Upper bound** $x_{\max}$	**Manufacturing Tolerance** U
Slot angle	2.47°	3.27°	±0.1°
β_{L1P1}	27.03°	29.66°	±0.33°
β_{L1P2}	37.03°	39.66°	±0.33°
β_{L2P1}	31.03°	33.66°	±0.33°
β_{L2P2}	47.03°	49.66°	±0.33°
β_{L3P1}	33.7°	37°	±0.33°
β_{L3P2}	59.7°	63°	±0.33°
Airgap	0.55 mm	0.65 mm	±0.03 mm
Bridge_{L1}	2.6 mm	2.98 mm	±0.05 mm
Bridge_{L2}	0.9 mm	1.18 mm	±0.05 mm
Bridge_{L3}	0.5 mm	0.62 mm	±0.03 mm
Bridge_{tang}	0.4 mm	0.6 mm	±0.05 mm

Table 2 – Geometrical variables (see [Reyes Reyes et al. 2024] for more details)

torque ripples (to be minimized). The resulting optimization problem is the following

$$\min_{x\in\mathcal{X}} \left(\mathbb{E}\left[f_1(x,\boldsymbol{U})\right], \mathbb{E}\left[f_2(x,\boldsymbol{U})\right]\right)$$

where f_1 and f_2 are two real-valued objective functions defined on $\mathcal{X} \subset \mathbb{R}^{12}$ (respectively the opposite of the mean torque and the torque ripple) and $\mathcal{U} \subset \mathbb{R}^{14}$. 12 of the 14 uncertain inputs are manufacturing tolerances on each x, summarized in Table 2, where, for example, $\pm 0.1°$ means that the slot angle uncertainty U_1 follows a uniform law on $[-0.1, 0.1]$. The other two, U_{13} and U_{14}, describe the magnetic material properties and follow uniform distributions on $[-1, 1]$.

The purpose of our study is to screen and rank the uncertain parameters according to their impact on the quantities of interest in this optimization. Since we want to minimize an objective function, we are interested in quantifying the effect of $\boldsymbol{U}$ on sets where the objective function takes low values. This can be done by looking at the random set of the form $\Gamma_{f_1} = \{x \in \mathcal{X}, f_1(x,\boldsymbol{U}) \leq q_1\}$, where q_1 is a threshold to be selected. It quantifies the

	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}	U_{11}	U_{12}	U_{13}	U_{14}
Γ_{F_1}	0.40	0.90	0.75	0.90	0.75	1.00	1.00	0.40	0.90	0.90	0.85	1.00	0.00	0.00
Γ_{F_2}	0.70	0.00	0.00	0.75	0.00	1.00	0.65	0.15	0.95	0.70	0.90	0.60	0.00	0.40
$(\Gamma_{F_1}, \Gamma_{F_2})$	0.25	0.00	0.05	0.90	0.05	1.00	0.95	0.10	0.85	0.75	0.80	0.75	0.00	0.00

Table 3 – Acceptance rates (%) over 20 independence tests with a risk of 5% for the excursion sets Γ_{F_1}, Γ_{F_2} and the pair $(\Gamma_{F_1}, \Gamma_{F_2})$ with $-q_1 = 420$ N.m and $q_2 = 7\%$ computed with the Sobolev input kernel and with $n = 100$, $m = 100$.

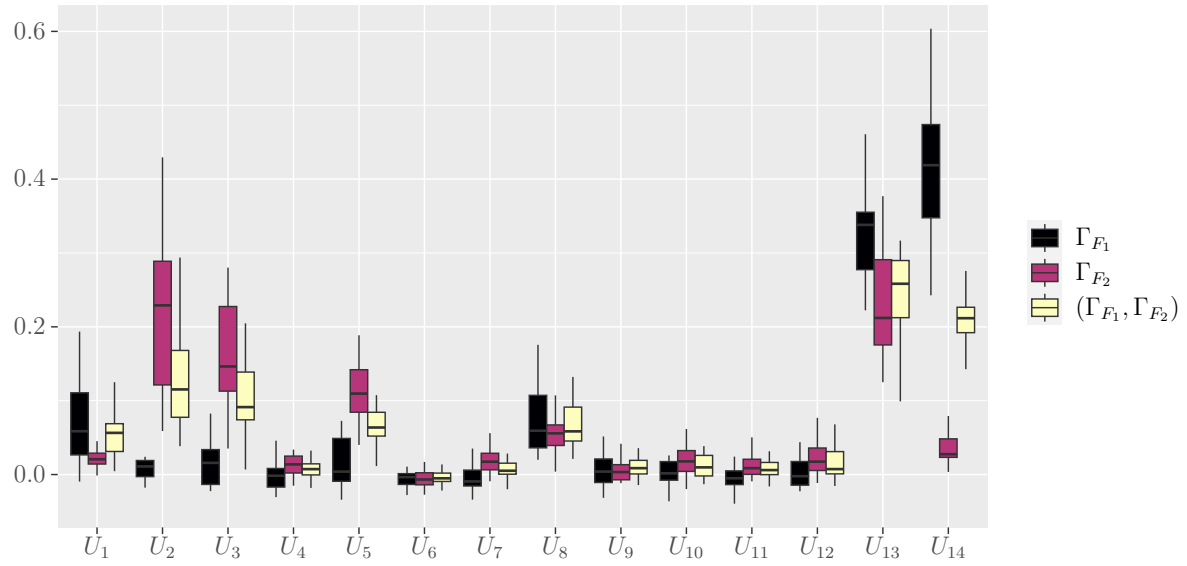


Figure 10 – Estimation of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ for the excursion sets Γ_{F_1}, Γ_{F_2} and the pair $(\Gamma_{F_1}, \Gamma_{F_2})$ with $-q_1 = 420$ N.m and $q_2 = 7\%$ computed with the Sobolev input kernel and with $n = 100$, $m = 100$. 20 replicates.

effect of $\boldsymbol{U}$ on the sets where f_1 is below q_1, but does not take into account the variations of f_1 within this low-valued region. For this reason, we propose to look at the effect of the uncertain inputs on the set Γ_{F_1} defined by $\Gamma_{F_1} = \{(\boldsymbol{x}, x_{13}) \in \mathcal{X} \times [f_1^{min}, q_1], x_{13} \leq f_1(\boldsymbol{x}, \boldsymbol{U})\}$ where f_1^{min} is a lower bound to only consider sets of low values of f_1. This also corresponds to the excursion set associated with $F_1 : \mathcal{X} \times [f_1^{min}, q_1] \times \mathcal{U} \to \mathbb{R}$ defined by $F_1(\boldsymbol{x}, \boldsymbol{u}) = x_{13} - f_1(\boldsymbol{x}_{-13}, \boldsymbol{u})$. Γ_{F_2} is defined similarly.

We study three different cases : sensitivity analysis on Γ_{F_1}, on Γ_{F_2}, and on the pair of excursion sets $(\Gamma_{F_1}, \Gamma_{F_2})$. We use the thresholds $-q_1 = 420$ N.m and $q_2 = 7\%$ where lies the Pareto front in Figure 9 of [Reyes Reyes et al. 2024]. For these three cases, we take $n = m = 100$ and use only the Sobolev kernel. With 20 replicates, the acceptance rates with a risk of 0.05% are given in Table 3 and the boxplots of the first-order indices are plotted in Figure 10.

Table 3 shows that the inputs influencing the excursion sets Γ_{F_1} and Γ_{F_2} are different. Only U_{13} and U_{14} are always tested as influential for Γ_{F_1} and U_2, U_3, U_5 and U_{13} for Γ_{F_2}. They correspond to the most influential inputs in Figure 10 and can be ranked. For example, U_2 has the most influence on Γ_{F_2}, followed by U_{13}, U_3 and U_5. Detected about 95% of the time as independent of the output, U_2, U_4, U_6, U_7, U_9, U_{10}, U_{12} for Γ_{F_1}, and U_6,

U_9, U_{11} for Γ_{F_2} could be removed to simplify each model. The remaining inputs have an acceptance rate between 15% and 85%, which means that they are sometimes classified as independent and sometimes not. They correspond to inputs that have a small effect on the output, as can be observed in Figure 10. By taking the pair of the two sets, the influential inputs are the one that were influential on at least one of the sets. This results in five dominant inputs that can be ranked by their first order index: $U_{13} \succeq U_{14} \succeq U_2 \succeq U_3 \succeq U_5$. The results associated to the excursion sets of the forms $\Gamma_{f_i} = \{x \in \mathcal{X}, f_i(x, \boldsymbol{U}) \leq q_i\}$ are similar in this case and given in Appendix 7.2.

5 Conclusion

In this paper, we propose a method to perform sensitivity analysis on set-valued outputs through kernel-based sensitivity analysis, which relies on the choice of a kernel between sets. We introduce the kernel k_{set}, which is based on the symmetric difference between two sets. We show that it is characteristic, which is an essential property for performing screening. We then adapt the recent HSIC-ANOVA index to set-valued outputs and introduce an efficient estimator. Finally, we compute the indices on three test cases including a real application for robust design of electrical motor. The proposed method allows to screen and rank the uncertain inputs according to their impact on the excursion sets. For future research, it could be interesting to find and study other set-valued output kernels. Other types of approaches to perform sensitivity analysis for sets could also be investigated, such as using universal indices from Fort et al. [2021], or by using classic random set theory from Molchanov [2005].

In the context of robust optimization, the presented method and especially the screening results could be used to reduce the dimension of the uncertain space by quantifying the impact of uncertain inputs on the optimization constraints. Reducing the dimension of the uncertain space can then be useful to reduce the computational cost of a joint space metamodel that could be used within a Bayesian optimization.

More generally, sensitivity analysis for sets can also be used when dealing with numerical codes with set-valued outputs. This occurs in several areas: in the field of viability, where the outputs are sets called viability kernels, or, for example, in flood risk, where the output is the map of flooded areas.

6 Aknowledgment

The authors thank Gabriel Sarazin for his numerous fruitful discussions and comments. We are also grateful to the reviewers and the associate editor for their relevant and helpful comments. This research was conducted with the support of the consortium in Applied Mathematics CIROQUO, gathering partners in technological research and academia in the development of advanced methods for Computer Experiments.

References

Balança, P. and Herbin, E. (2012), « A set-indexed Ornstein-Uhlenbeck process », *Electronic Communications in Probability* 17, 1–14.

Borgonovo, E. (2007), « A new uncertainty importance measure », *Reliability Engineering and System Safety* 92.6, 771–784.

Cousin, A., Garnier, J., Guiton, M., and Munoz Zuniga, M. (2022), « A two-step procedure for time-dependent reliability-based design optimization involving piece-wise stationary Gaussian processes », *Structural and Multidisciplinary Optimization* 65.4, 120.

Da Veiga, S. (2021), « Kernel-based ANOVA decomposition and Shapley effects–Application to global sensitivity analysis », *Preprint*.

Da Veiga, S. (2015), « Global sensitivity analysis with dependence measures », *Journal of Statistical Computation and Simulation* 85.7, 1283–1305.

Da Veiga, S., Gamboa, F., Iooss, B., and Prieur, C. (2021), *Basics and Trends in Sensitivity Analysis*, Society for Industrial and Applied Mathematics.

De Lozzo, M. and Marrel, A. (2016), « New improvements in the use of dependence measures for sensitivity analysis and screening », *Journal of Statistical Computation and Simulation* 86.15, 3038–3058.

— (2017), « Sensitivity analysis with dependence and variance-based measures for spatio-temporal numerical simulators », *Stochastic environmental research and risk assessment* 31, 1437–1453.

El Amri, M. and Marrel, A. (2024), « More powerful HSIC-based independence tests, extension to space-filling designs and functional data », *International Journal for Uncertainty Quantification* 14.2.

El Amri, R., Le Riche, R., Helbert, C., Blanchet-Scalliet, C., and Da Veiga, S. (2023), « A Sampling Criterion for Constrained Bayesian Optimization with Uncertainties », *The SMAI Journal of computational mathematics* 9, 285–309.

Fort, J.-C., Klein, T., and Lagnoux, A. (2021), « Global Sensitivity Analysis and Wasserstein Spaces », *SIAM/ASA Journal on Uncertainty Quantification* 9.2, 880–921.

Gamboa, F., Gremaud, P., Klein, T., and Lagnoux, A. (2022), « Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics », *Bernoulli* 28.4, 2345–2374.

Gamboa, F., Janon, A., Klein, T., and Lagnoux, A. (2014), « Sensitivity analysis for multidimensional and functional outputs », *Electronic Journal of Statistics* 8.1, 575–603.

Gamboa, F., Klein, T., and Lagnoux, A. (2018), « Sensitivity analysis based on Cramér–von Mises distance », *SIAM/ASA Journal on Uncertainty Quantification* 6.2, 522–548.

Gamboa, F., Klein, T., Lagnoux, A., and Moreno, L. (2021), « Sensitivity analysis in general metric spaces », *Reliability Engineering and System Safety* 212.

Ginsbourger, D., Roustant, O., Schuhmacher, D., Durrande, N., and Lenz, N. (2016), « On ANOVA decompositions of kernels and Gaussian random field paths », *Monte Carlo and Quasi-Monte Carlo Methods*, Springer, 315–330.

Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005a), « Measuring statistical dependence with Hilbert-Schmidt norms », *International conference on algorithmic learning theory*, 63–77.

Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B., and Smola, A. (2007), « A kernel statistical test of independence », *Advances in neural information processing systems* 20.

Gretton, A., Herbrich, R., Smola, A., Bousquet, O., and Schölkopf, B. (2005b), « Kernel Methods for Measuring Independence », *Journal of Machine Learning Research* 6.70, 2075–2129.

Hida, T., Kuo, H.-H., Potthoff, J., and Streit, L. (2013), *White noise: an infinite dimensional calculus*, vol. 253, Springer Science & Business Media.

Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008), « Computer model calibration using high-dimensional output », *Journal of the American Statistical Association* 103.482, 570–583.

Iooss, B. and Lemaître, P. (2015), « A review on global sensitivity analysis methods », *Uncertainty management in simulation-optimization of complex systems: algorithms and applications*, 101–122.

Lyons, R. (2013), « Distance covariance in metric spaces », *The Annals of Probability* 41.5.

Marczewski, E. and Steinhaus, H. (1958), « On a certain distance of sets and the corresponding distance of functions », *Colloquium Mathematicum* 6, 319–327.

Marrel, A., Iooss, B., Jullien, M., Laurent, B., and Volkova, E. (2011), « Global sensitivity analysis for models with spatially dependent outputs », *Environmetrics* 22.3, 383–397.

Marrel, A., Perot, N., and Mottet, C. (2015), « Development of a surrogate model and sensitivity analysis for spatio-temporal numerical simulators », *Stochastic environmental research and risk assessment* 29, 959–974.

Molchanov, I. (2005), *Theory of random sets*, vol. 19, 2, Springer.

Morris, M. D. (1991), « Factorial Sampling Plans for Preliminary Computational Experiments », *Technometrics* 33.2, 161–174.

Muandet, K., Fukumizu, K., Sriperumbudur, B., Schölkopf, B., et al. (2017), « Kernel mean embedding of distributions: A review and beyond », *Foundations and Trends® in Machine Learning* 10.1-2, 1–141.

Nguyen, H. T. (2006), *An introduction to random sets*, CRC press.

Perrin, T., Roustant, O., Rohmer, J., Alata, O., Naulin, J., Idier, D., Pedreros, R., Moncoulon, D., and Tinard, P. (2021), « Functional principal component analysis for global sensitivity analysis of model with spatial output », *Reliability Engineering & System Safety* 211, 107522.

Rainforth, T., Cornish, R., Yang, H., Warrington, A., and Wood, F. (2018), « On nesting monte carlo estimators », *International Conference on Machine Learning*, 4267–4276.

Rainforth, T., Cornish, R., Yang, H., and Wood, F. (2016), « On the pitfalls of nested Monte Carlo », *preprint*.

Reyes Reyes, A., Nasr, A., Sinoquet, D., and Hlioui, S. (2024), « Study on the impact of uncertain design parameters on the perfomances of a permanent magnet assisted synchronous reluctance motor », *Sci. Tech. Energ. Transition*.

Rudin, W. (1987), *Real and complex analysis*, McGraw-Hill.

Sarazin, G., Marrel, A., Da Veiga, S., and Chabridon, V. (2022), « Test d'indépendance basé sur les indices HSIC-ANOVA d'ordre total », *53èmes Journées de Statistique de la SFdS*.

Sarazin, G., Marrel, A., Veiga, S. da, and Chabridon, V. (2023), « New insights into the feature maps of Sobolev kernels: application in global sensitivity analysis ».

Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013), « Equivalence of distance-based and RKHS-based statistics in hypothesis testing », *The Annals of Statistics* 41.5.

Shawe-Taylor, J. and Cristianini, N. (2004), *Kernel Methods for Pattern Analysis*, Cambridge University Press, 291–326.

Sobol', I. M. (1993), « Sensitivity estimates for nonlinear mathematical models », *Math. Model. Comput. Exp.* 1, 407.

Sobol', I. M. (2001), « Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates », *Mathematics and Computers in Simulation* 55.1, 271–280.

Song, L., Smola, A., Gretton, A., Borgwardt, K., and Bedo, J. (2007), « Supervised Feature Selection via Dependence Estimation », *Proceedings of the 24th international conference on Machine learning*, 823–830.

Spagnol, A., Riche, R. L., and Veiga, S. D. (2019), « Global Sensitivity Analysis for Optimization with Variable Selection », *SIAM/ASA Journal on Uncertainty Quantification* 7.2, 417–443.

Sriperumbudur, B. K., Fukumizu, K., and Lanckriet, G. R. (2011), « Universality, Characteristic Kernels and RKHS Embedding of Measures », *Journal of Machine Learning Research* 12.70, 2389–2410.

Steinwart, I. and Christmann, A. (2008), « Kernels and reproducing kernel hilbert spaces », *Support Vector Machines*, Springer, 110–163.

Szabó, Z. and Sriperumbudur, B. K. (2018), « Characteristic and Universal Tensor Product Kernels », *Journal of Machine Learning Research* 18.233, 1–29.

Ziegel, J., Ginsbourger, D., and Dümbgen, L. (2022), « Characteristic kernels on Hilbert spaces, Banach spaces, and on sets of measures », *preprint*.

7 Appendix

7.1 Proofs

7.1.1 Proof of Lemma 3.1

Lemma 3.1. *Let $\gamma_1, \gamma_2 \in \mathscr{L}(\mathcal{X})$, we have*

$$\delta(\gamma_1, \gamma_2) = ||\mathbb{1}_{\gamma_1} - \mathbb{1}_{\gamma_2}||_2^2,$$

and

$$\delta(\gamma_1, \gamma_2) = 0 \Leftrightarrow \gamma_1 = \gamma_2 \ \lambda\text{-almost everywhere.}$$

Proof.

$$\begin{aligned}
\delta(\gamma_1, \gamma_2) &= \lambda(\gamma_1 \Delta \gamma_2) \\
&= \int_{\mathcal{X}} \mathbb{1}_{\gamma_1 \Delta \gamma_2} d\lambda \\
&= \int_{\mathcal{X}} (\mathbb{1}_{\gamma_1 \cup \gamma_2} - \mathbb{1}_{\gamma_1 \cap \gamma_2}) d\lambda \\
&= \int_{\mathcal{X}} (\mathbb{1}_{\gamma_1} + \mathbb{1}_{\gamma_2} - 2\mathbb{1}_{\gamma_1 \cap \gamma_2}) d\lambda \\
&= \int_{\mathcal{X}} (\mathbb{1}_{\gamma_1} - \mathbb{1}_{\gamma_2})^2 d\lambda \\
&= ||\mathbb{1}_{\gamma_1} - \mathbb{1}_{\gamma_2}||_2^2
\end{aligned}$$

$$\begin{aligned}
\delta(\gamma_1, \gamma_2) = 0 &\Leftrightarrow ||\mathbb{1}_{\gamma_1} - \mathbb{1}_{\gamma_2}||_2^2 = 0 \\
&\Leftrightarrow \mathbb{1}_{\gamma_1}(x) = \mathbb{1}_{\gamma_2}(x) \text{ for } \lambda\text{-almost every } x \in \mathcal{X} \\
&\Leftrightarrow \gamma_1 = \gamma_2 \ \lambda\text{-almost everywhere.}
\end{aligned}$$

□

7.1.2 Proof of Proposition 3.1

Proposition 3.1. *The function $k_{set} : \mathscr{L}^*(\mathcal{X}) \times \mathscr{L}^*(\mathcal{X}) \to \mathbb{R}$ defined by*

$$\forall \gamma_1, \gamma_2 \in \mathscr{L}^*(\mathcal{X}), \quad k_{set}(\gamma_1, \gamma_2) = e^{-\frac{\lambda(\gamma_1 \Delta \gamma_2)}{2\sigma^2}},$$

is symmetric and positive definite for any positive scalar σ which means that k_{set} is a kernel.

Proof. The proof is similar as the one of Lemma 2.1. in Balança and Herbin [2012] but is recalled and adapted here.

Let $\alpha_1, ..., \alpha_n \in \mathbb{R}$ and $\gamma_1, ..., \gamma_n \in \mathscr{L}^*(\mathcal{X})$. Let's show that

$$\sum_{kl}^{n} e^{-\frac{\lambda(\gamma_k \Delta \gamma_l)}{2\sigma^2}} \alpha_k \alpha_l \geq 0.$$

First let's use Lemma 3.1 to write:

$$\sum_{kl}^{n} e^{-\frac{\lambda(\gamma_k \Delta \gamma_l)}{2\sigma^2}} \alpha_k \alpha_l = \sum_{kl}^{n} e^{-\frac{\left|\left|\mathbb{1}_{\gamma_k} - \mathbb{1}_{\gamma_l}\right|\right|^2_{L^2(\mathcal{X})}}{2\sigma^2}} \alpha_k \alpha_l.$$

Then we use the Bochner-Milos theorem (Theorem 1.1 of Hida et al. [2013]), which says that there exists a random variable $G \in L^2(\mathcal{X})$ such that

$$\forall f \in L^2(\mathcal{X}),\ \mathbb{E}(e^{i\langle f, G\rangle_{L^2(\mathcal{X})}}) = e^{-\frac{1}{2}||f||^2_{L^2(\mathcal{X})}}.$$

This allows to derive :

$$\begin{aligned}
\sum_{kl}^{n} e^{-\frac{\lambda(\gamma_k \Delta \gamma_l)}{2\sigma^2}} \alpha_k \alpha_l &= \sum_{kl}^{n} e^{-\frac{\left|\left|\mathbb{1}_{\gamma_k} - \mathbb{1}_{\gamma_l}\right|\right|^2_{L^2(\mathcal{X})}}{2\sigma^2}} \alpha_k \alpha_l \\
&= \sum_{kl}^{n} \mathbb{E} e^{i\langle \frac{\mathbb{1}_{\gamma_k} - \mathbb{1}_{\gamma_l}}{\sqrt{2}\sigma}, G\rangle_{L^2(\mathcal{X})}} \alpha_k \alpha_l \\
&= \mathbb{E} \sum_{kl}^{n} e^{\frac{i}{\sqrt{2}\sigma}\langle \mathbb{1}_{\gamma_k}, G\rangle_{L^2(\mathcal{X})}} \alpha_k e^{-\frac{i}{\sqrt{2}\sigma}\langle \mathbb{1}_{\gamma_l}, G\rangle_{L^2(\mathcal{X})}} \alpha_l \\
&= \mathbb{E} \left| \sum_{k}^{n} e^{\frac{i}{\sqrt{2}\sigma}\langle \mathbb{1}_{\gamma_k}, G\rangle_{L^2(\mathcal{X})}} \alpha_k \right|^2 \geq 0.
\end{aligned}$$

□

7.1.3 Proof of Proposition 3.2

Proposition 3.2. *The kernel k_{set} is bounded and measurable from $\mathscr{L}_2^\star(\mathcal{X}) = \mathscr{L}^\star(\mathcal{X}) \times \mathscr{L}^\star(\mathcal{X})$ to $\mathbb{R}$.*

Proof. k_{set} is clearly bounded by 1. In term of measurability, we are studying the measurability of k_{set} with respect to the two Borels σ-algebras $\mathcal{B}(\mathscr{L}_2^\star(\mathcal{X}), \delta_2)$ and $\mathcal{B}(\mathbb{R}, |.|)$ with $\delta_2((\gamma_1^A, \gamma_1^A), (\gamma_2^B, \gamma_2^B)) = \delta(\gamma_1^A, \gamma_1^B) + \delta(\gamma_2^A, \gamma_2^B)$. As $\delta : \mathscr{L}_2^\star(\mathcal{X}) \to \mathbb{R}^+$ is a distance on $\mathscr{L}^\star(\mathcal{X})$, it is a continuous function from $\mathscr{L}_2^\star(\mathcal{X})$ to $\mathbb{R}$. k_{set} is then continuous as exp is continuous. Finally the continuity of k_{set} implies measurability as we are working in two Borel σ-algebras. □

7.1.4 Proof of Proposition 3.3

Proposition 3.3. *The kernel k_{set} is characteristic.*

Proof. The proof is based on the Proposition 5.2. of Ziegel et al. [2022] which is recalled here with our notations.

Proposition 7.1. *Let $\mathcal{P}$ be a Polish space, H a separable Hilbert space, T a measurable and injective mapping from $\mathcal{P}$ to H, and φ the Laplace transform of a finite Borel measure ν on $[0, +\infty)$ such that $\nu \neq 0$ and supp$\nu \neq \{0\}$. Then, the kernel k on $\mathcal{P} \times \mathcal{P}$ defined by*

$$k(x, x') := \varphi\left(\|T(x) - T(x')\|_H^2\right), \quad (x, x') \in \mathcal{X}^2$$

is integrally strictly positive definite with respect to $\mathcal{M}(\mathcal{X})$.

Using the previous Proposition, we will show that k_{set} is integrally strictly positive which is a sufficient condition to be characteristic (see Theorem 7 in B. K. Sriperumbudur et al. [2011]).

To apply the Proposition, we first introduce the notation,

$$\mathscr{F}_{\text{bin}}^{\star} = \{f \in L^2(\mathcal{X}) : \exists \gamma \in \mathscr{L}(\mathcal{X}) \text{ such that } f = \mathbb{1}_\gamma \ \lambda\text{-almost everywhere}\}.$$

We want to apply the proposition with $\mathcal{P} = \mathscr{L}^*(\mathcal{X})$, $H = L^2(\mathcal{X})$, $\varphi(\cdot) = \exp(-\frac{\cdot}{2\sigma^2})$ and T defined by

$$\begin{array}{rccc} T: & \mathscr{L}^\star(\mathcal{X}) & \longrightarrow & \mathscr{F}_{\text{bin}}^{\star} \subset H \\ & [\gamma]_\delta & \longmapsto & [\mathbb{1}_\gamma]_2 \,, \end{array}$$

where $[\cdot]_\delta$ and $[\cdot]_2$ denote equivalence classes in $\mathscr{L}(\mathcal{X})$ and $\mathscr{F}_{\text{bin}}$ respectively. We need to show that $\mathscr{L}^*(\mathcal{X})$ is Polish, that $L^2(\mathcal{X})$ is a separable Hilbert space, and that T is measurable and injective.

- Let's first show that is T a well-defined measurable and injective mapping from $\mathcal{P}$ to H:
  - T is well-defined and in the same time injective as for any $\gamma_1, \gamma_2 \in \mathscr{L}(\mathcal{X})$,

$$\begin{aligned} [\gamma_1]_\delta = [\gamma_2]_\delta &\Leftrightarrow \delta(\gamma_1, \gamma_2) = 0 \\ &\Leftrightarrow \lambda(\gamma_1 \Delta \gamma_2) = 0 \\ &\Leftrightarrow \left\|\mathbb{1}_{\gamma_1} - \mathbb{1}_{\gamma_2}\right\|_2 = 0 \\ &\Leftrightarrow [\mathbb{1}_{\gamma_1}]_2 = [\mathbb{1}_{\gamma_2}]_2 \,. \end{aligned}$$

  - T is measurable (with respect to the Borel σ-algebras $\mathcal{B}\left(\mathscr{L}^\star(\mathcal{X}), \delta\right)$ and $\mathcal{B}\left(\mathscr{F}_{\text{bin}}^{\star}, ||\cdot||_2\right)$) because it is continuous as it is an isometry from $\left(\mathscr{L}^\star(\mathcal{X}), \sqrt{\delta}\right)$ to $(\mathscr{F}_{\text{bin}}^{\star}, ||\cdot||_2)$.
  - Let us show that T is also surjective which will be useful later on. Let $[f]_2 \in \mathscr{F}_{\text{bin}}^{\star}$. There exist $\gamma \in \mathscr{L}(\mathcal{X})$ such that $f = \mathbb{1}_\gamma$ λ-almost everywhere which implies that $||f - \mathbb{1}_\gamma||_2 = 0$ i.e. $[f]_2 = [\mathbb{1}_\gamma]_2 = T([\gamma]_\delta)$.
- $L^2(\mathcal{X})$ is a separable Hilbert space as $\mathcal{X}$ is compact so separable.
- Let us now show that $\mathscr{L}^*(\mathcal{X})$ is Polish i.e. a topological space homeomorphic to a separable complete metric space. Luckily, as T is a surjective isometry, it is an homeorphism from $\mathscr{L}^\star(\mathcal{X})$ to $\mathscr{F}_{\text{bin}}^{\star}$ which is a metric space. Then, it only remains to prove that $\mathscr{F}_{\text{bin}}^{\star}$ is both complete and separable.
  - $\mathscr{F}_{\text{bin}}^{\star} \subset L^2(\mathcal{X})$ which is separable so $\mathscr{F}_{\text{bin}}^{\star}$ is separable.
  - As $L^2(\mathcal{X})$ is complete, having $\mathscr{F}_{\text{bin}}^{\star}$ closed is a sufficient condition for $\mathscr{F}_{\text{bin}}^{\star}$ to be complete. Let us show then that it is closed. Let $\mathbb{1}_{\gamma_n} \underset{n\to+\infty}{\overset{L_2}{\to}} f$ with $\gamma_n \in \mathscr{L}(\mathcal{X})$ $\forall n$. The L_2 convergence implies that there is a sub-sequence $(\mathbb{1}_{\gamma_{\phi(n)}})_n$ that converges almost everywhere pointwise to f (as stated in Theorem 3.13 of Rudin [1987]). It means that there exists a λ-null set $\mathcal{N}$ s.t. $\forall x \notin \mathcal{N}, \mathbb{1}_{\gamma_{\phi(n)}}(x) \underset{n\to+\infty}{\to} f(x)$. As $\mathbb{1}_{\gamma_{\phi(n)}}(x)$ is a sequence of 0 and 1, we have that its limit, $f(x)$, belong to $\{0, 1\}$. So, $f = \mathbb{1}_{f^{-1}(\{1\})}$ λ-almost everywhere and $f^{-1}(\{1\}) \in \mathscr{L}(\mathcal{X})$ as f is measurable. So $f \in \mathscr{F}_{\text{bin}}$. $\mathscr{F}_{\text{bin}}^{\star}$ is thus closed.

Thus k_{set} is integrally strictly positive definite with respect to $\mathcal{M}(\mathscr{L}^*(\mathcal{X}))$, the set of signed measure on $\mathscr{L}^*(\mathcal{X})$, which implies that it is characteristic. □

7.1.5 Proof of Proposition 3.4

Proposition 3.4. *With the previous notations, we have*

$$\mathbb{E}\left(\widehat{\mathrm{HSIC}}\left(\boldsymbol{U}_A,\Gamma\right)-\mathrm{HSIC}(\boldsymbol{U}_A,\Gamma)\right)^2\leq 2\left(\frac{2\sigma_1^2}{n(n-1)}+\frac{4(n-2)\sigma_2^2}{n(n-1)}+\frac{L^2\sigma_3^2}{m}\right),$$

where

- $\sigma_1^2 = Var\left(\left(K_A\left(\boldsymbol{U}_A,\boldsymbol{U}_A{}'\right)-1\right)k_{set}(\Gamma,\Gamma')\right)$,
- $\sigma_2^2 \;= Var\left(\mathbb{E}\left[\left(K_A\left(\boldsymbol{U}_A,\boldsymbol{U}_A{}'\right)-1\right)k_{set}(\Gamma,\Gamma')|(\Gamma,\boldsymbol{U}_A)\right]\right)$,
- $\sigma_3^2 = \mathbb{E}\left[\left(K_A\left(\boldsymbol{U}_A,\boldsymbol{U}_A{}'\right)-1\right)^2\mathrm{Var}\left(\mathbb{1}_{\Gamma\Delta\Gamma'}(\boldsymbol{X})|(\boldsymbol{U}_A,\boldsymbol{U}_A{}',\Gamma,\Gamma')\right)\right]$,
- $L=\frac{\lambda(\mathcal{X})}{2\sigma^2}$,

and where $(\boldsymbol{U}_A{}',\Gamma')$ *is an independent copy of* $(\boldsymbol{U}_A,\Gamma)$.

Proof. Let $f(\boldsymbol{u},z)=g(\boldsymbol{u})h(z)$ with $g(\boldsymbol{u})=\left(K_A\left(\boldsymbol{u}_1,\boldsymbol{u}_2\right)-1\right)$, $h(z)=e^{-\frac{\lambda(\mathcal{X})}{2\sigma^2}\boldsymbol{z}}$, and let $\Phi(\boldsymbol{x},\gamma_1,\gamma_2)=\mathbb{1}_{\gamma_1\Delta\gamma_2}(x)$. Let us have iid samples $(\boldsymbol{U}_A^{(i)},\Gamma^{(i)})$, $i=1,...,n$ of $(\boldsymbol{U}_A,\Gamma)$ and $(\boldsymbol{X}^{(1)},...,\boldsymbol{X}^{(m)})$ of $\boldsymbol{X}\sim\mathcal{U}(\mathcal{X})$. We denote

$$\mathrm{H}=\mathrm{HSIC}(\boldsymbol{U}_A,\Gamma)=\mathbb{E}[f(\boldsymbol{U},\boldsymbol{U}',\mathbb{E}[\Phi(\boldsymbol{X},\Gamma,\Gamma')|(\Gamma,\Gamma')]],$$

$$\mathrm{H}_{n,m}=\widehat{\mathrm{HSIC}_u}\left(\boldsymbol{U}_A,\Gamma\right)=\frac{2}{n(n-1)}\sum_{i<j}^{n}f(\boldsymbol{U}_A^{(i)},\boldsymbol{U}_A^{(j)},\frac{1}{m}\sum_{k=1}^{m}\Phi(\boldsymbol{X}^{(k)},\Gamma^{(i)},\Gamma^{(j)}),$$

and

$$\mathrm{H}_n=\frac{2}{n(n-1)}\sum_{i<j}^{n}f(\boldsymbol{U}_A^{(i)},\boldsymbol{U}_A^{(j)},\mathbb{E}[\Phi(\boldsymbol{X},\Gamma^{(i)},\Gamma^{(j)})|(\Gamma^{(i)},\Gamma^{(j)})]).$$

First we split the risk into two terms:

$$\mathbb{E}|\,\mathrm{H}_{n,m}-\mathrm{H}\,|^2\leq 2\mathbb{E}|\,\mathrm{H}_n-\mathrm{H}\,|^2+2\mathbb{E}|\,\mathrm{H}_n-\mathrm{H}_{n,m}\,|^2$$

The first term is the variance of a classic U-statistic of order 2:

$$\mathbb{E}|\,\mathrm{H}_n-\mathrm{H}\,|^2=\frac{2\sigma_1^2}{n(n-1)}+\frac{4(n-2)\sigma_2^2}{n(n-1)}.$$

The second term can be developed:

$$\mathbb{E}|\,\mathrm{H}_n-\mathrm{H}_{nm}\,|^2=\frac{4}{n^2(n-1)^2}\left(\sum_{i<j}^{n}\sum_{p<l}^{n}\mathbb{E}\left(E_{ij}-E_{ij,m}\right)\left(E_{pl}-E_{pl,m}\right)\right),$$

with

$$E_{ij}=f(\boldsymbol{U}_A^{(i)},\boldsymbol{U}_A^{(j)},\mathbb{E}[\Phi(\boldsymbol{X},\Gamma^{(i)},\Gamma^{(j)})|(\Gamma^{(i)},\Gamma^{(j)})])$$

and

$$E_{ij,m}=f(\boldsymbol{U}_A^{(i)},\boldsymbol{U}_A^{(j)},\frac{1}{m}\sum_{k=1}^{m}\Phi(\boldsymbol{X}^{(k)},\Gamma^{(i)},\Gamma^{(j)}).$$

As $\boldsymbol{X}^{(1)}, ..., \boldsymbol{X}^{(m)}$ are common to each $E_{ij,m}$, the terms $E_{ij,m}$ and $E_{pl,m}$ are not independent even if i, j, p, l are pairwise distinct. We can still bound them but we will lose one order of convergence in m. We first have

$$\begin{aligned}
|\mathbb{E}\left(E_{ij} - E_{ij,m}\right)\left(E_{pl} - E_{pl,m}\right)| &\leq \left(\mathbb{E}\left|E_{ij} - E_{ij,m}\right|^2 \mathbb{E}\left|E_{pl} - E_{pl,m}\right|^2\right)^{\frac{1}{2}} &(5)\\
&= \left(\mathbb{E}\left|E_{12} - E_{12,m}\right|^2\right) &(6)\\
&\leq \frac{L^2}{m}\mathbb{E}\left[\left(K_A\left(\boldsymbol{U}_A, \boldsymbol{U}_A{}'\right) - 1\right)^2 \mathrm{Var}\left(\Phi(\boldsymbol{X}, \Gamma, \Gamma')|(\boldsymbol{U}_A, \boldsymbol{U}_A{}', \Gamma, \Gamma')\right)\right],
\end{aligned}$$

using Cauchy-Schwarz inequality in (5) and that h is L-lipschitz. Summing each term, we obtain

$$\mathbb{E}|\,\mathrm{H}_n - \mathrm{H}_{nm}\,|^2 \leq \frac{L^2}{m}\mathbb{E}\left[\left(K_A\left(\boldsymbol{U}_A, \boldsymbol{U}_A{}'\right) - 1\right)^2 \mathrm{Var}\left(\Phi(\boldsymbol{X}, \Gamma, \Gamma')|(\boldsymbol{U}_A, \boldsymbol{U}_A{}', \Gamma, \Gamma')\right)\right].$$

Putting all results together, we get

$$\mathbb{E}|\,\mathrm{H}_{n,m} - \mathrm{H}\,|^2 \leq 2\left(\frac{2\sigma_1^2}{n(n-1)} + \frac{4(n-2)\sigma_2^2}{n(n-1)} + \frac{L^2\sigma_3^2}{m}\right).$$

□

7.1.6 Convergence rate in the case of Independent m sample

If an m sample $\boldsymbol{X}_{ij}^{(k)}$ is drawn independently for each (i, j), we can obtain an asymptotic rate of $\mathcal{O}(\frac{1}{n} + \frac{1}{m^2})$. Indeed in (5) we use the independence between $(\boldsymbol{X}_{ij}^{(k)})_k$ and $(\boldsymbol{X}_{pl}^{(k)})_k$ for i, j, p, l pairwise distinct:

$$\begin{aligned}
|\mathbb{E}\left(E_{ij} - E_{ij,m}\right)\left(E_{pl} - E_{pl,m}\right)| &= |\mathbb{E}\left(E_{ij} - E_{ij,m}\right)\mathbb{E}\left(E_{pl} - E_{pl,m}\right)|\\
&= |\mathbb{E}\left(E_{12} - E_{12,m}\right)|^2
\end{aligned}$$

Then by applying Taylor Lagrange's formula to h with $a = \mathbb{E}(\Phi(\boldsymbol{X}, \Gamma^{(1)}, \Gamma^{(2)})|(\Gamma^{(1)}, \Gamma^{(2)}))$ and $b = \frac{1}{m}\sum_{k=1}^m \Phi(\boldsymbol{X}^{(k)}, \Gamma^{(1)}, \Gamma^{(2)})$, we have the existence of $\theta \in \mathbb{R}^+$ such that,

$$\begin{aligned}
E_{12,m} - E_{12} &= \left(K_A\left(\boldsymbol{U}_A, \boldsymbol{U}_A{}'\right) - 1\right)\left(h(b) - h(a)\right)\\
&= \left(K_A\left(\boldsymbol{U}_A, \boldsymbol{U}_A{}'\right) - 1\right)\left[h'(a)(b-a) + \frac{h''(\theta)}{2}(b-a)^2\right].
\end{aligned}$$

Then we take the expectation and use the tower property and we use that Monte Carlo estimators are unbiased (i.e. $\mathbb{E}(b|(\Gamma^{(1)}, \Gamma^{(2)})) = a$) which leads to

$$\begin{aligned}
\mathbb{E}(E_{12,m} - E_{12}) &= \mathbb{E}\left[\mathbb{E}(E_{12,m} - E_{12}|\boldsymbol{U}_A, \boldsymbol{U}_A', \Gamma, \Gamma')\right]\\
&= \mathbb{E}\left[\left(K_A\left(\boldsymbol{U}_A, \boldsymbol{U}_A{}'\right) - 1\right)\mathbb{E}\left(\frac{h''(\theta)}{2}(b-a)^2|\boldsymbol{U}_A, \boldsymbol{U}_A{}', \Gamma, \Gamma'\right)\right].
\end{aligned}$$

Then, as h'' is bounded by L,

$$|\mathbb{E}(E_{12,m} - E_{12})| \leq \frac{L^2}{2m}\mathbb{E}\left[|K_A\left(\boldsymbol{U}_A, \boldsymbol{U}_A{}'\right) - 1|\,\mathrm{Var}\left(\Phi(\boldsymbol{X}, \Gamma, \Gamma')|(\boldsymbol{U}_A, \boldsymbol{U}_A{}', \Gamma, \Gamma')\right)\right],$$

which finally leads to

$$|\mathbb{E}\left(E_{ij} - E_{ij,m}\right)\left(E_{pl} - E_{pl,m}\right)| \leq \frac{L^4}{4m^2}\sigma_4^4,$$

where

$$\sigma_4^2 = \mathbb{E}\left[|K_A\left(\boldsymbol{U}_A, \boldsymbol{U}_A{}'\right) - 1|\,\mathrm{Var}\left(\Phi(\boldsymbol{X}, \Gamma, \Gamma')|(\boldsymbol{U}_A, \boldsymbol{U}_A{}', \Gamma, \Gamma')\right)\right].$$

If $i = p$ or $i = l$ or $j = p$ or $j = l$, we lost the independence so we use the previous result that

$$|\mathbb{E}\left(E_{ij} - E_{ij,m}\right)\left(E_{pl} - E_{pl,m}\right)| \leq \frac{L^2}{m}\sigma_3^2.$$

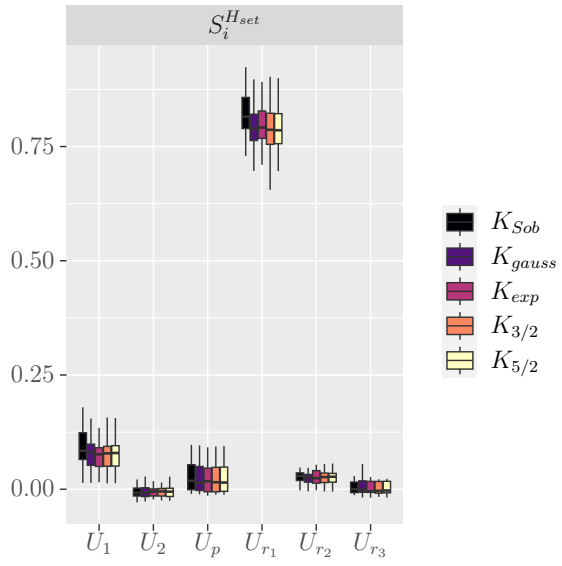
There are $\frac{n(n-1)(n-2)(n-3)}{4}$ i, j, p, l pairwise distinct with $i < j$ and $p < l$ so we finally obtain that,

$$\begin{aligned}\mathbb{E}|\,\mathrm{H}_{n,m} - \mathrm{H}\,|^2 &\leq 2\left(\frac{2\sigma_1^2}{n(n-1)} + \frac{4(n-2)\sigma_2^2}{n(n-1)} + \frac{L^2 2(2n-3)\sigma_3^2}{n(n-1)m} + \frac{L^4(n-2)(n-3)\sigma_4^4}{4n(n-1)m^2}\right) \\ &= \mathcal{O}(\frac{1}{n} + \frac{1}{m^2}).\end{aligned}$$

7.2 Figures

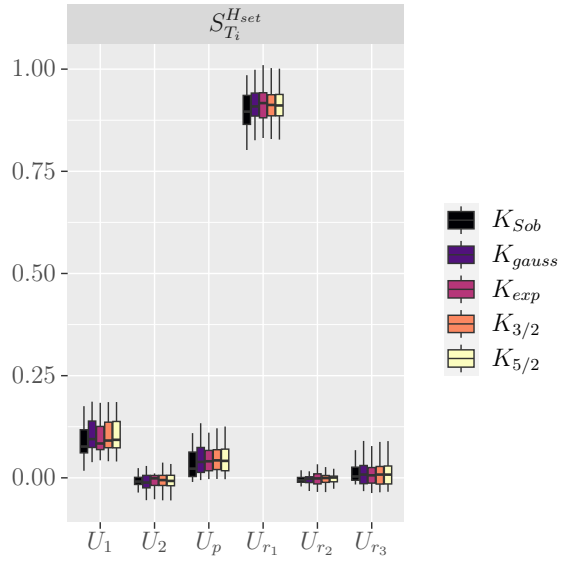
	U_1	U_2	U_p	U_{r_1}	U_{r_2}	U_{r_3}
K_{Sob}	5	95	55	0	75	90
K_{Sob}	5	100	60	0	75	95
K_{Sob}	15	95	55	0	65	95
K_{Sob}	10	95	55	0	65	95
K_{Sob}	5	100	55	0	70	95

(a) Acceptance rate (%) over 20 independence tests with a risk of 5%

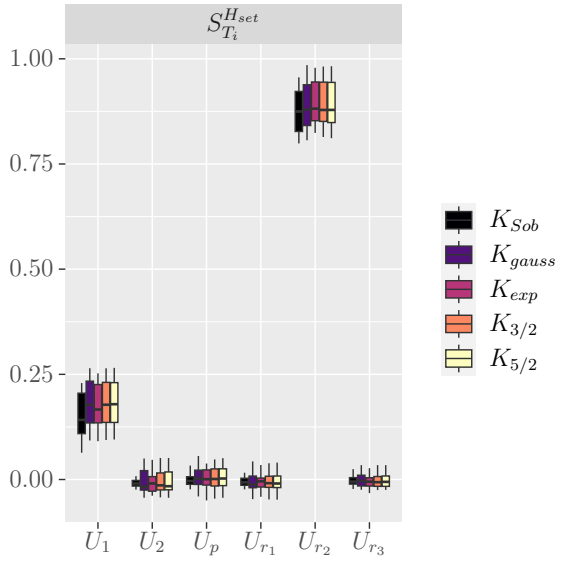


(b) Estimations of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$

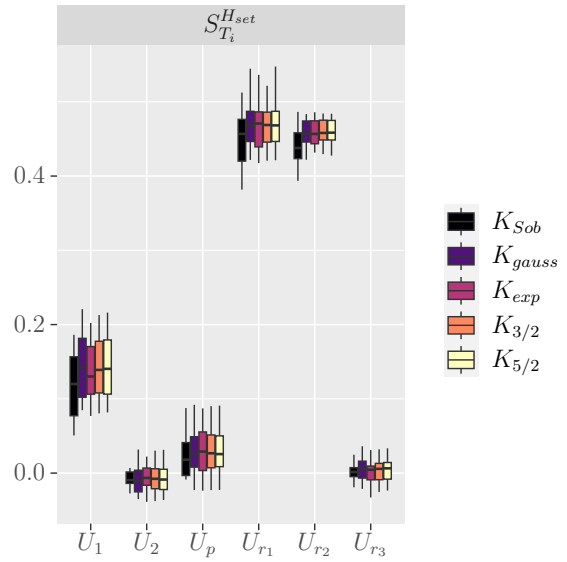
Figure 11 – Acceptance rate (a) and estimation of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ (b) for excursion set $\Gamma_{(g_1,g_2)}$ computed for 5 kernels with $n = 100$, $m = 100$ and repeated 20 times



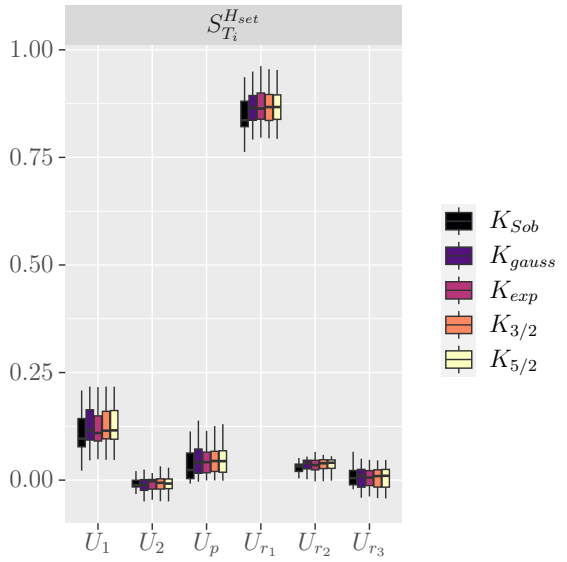
(a) Γ_{g_1}



(b) Γ_{g_2}



(c) $(\Gamma_{g_1}, \Gamma_{g_2})$



(d) $\Gamma_{(g_1,g_2)}$

Figure 12 – Estimation of the total-order indices $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ for the oscillator case computed for 5 kernels with $n = 100$, $m = 100$ and repeated 20 times

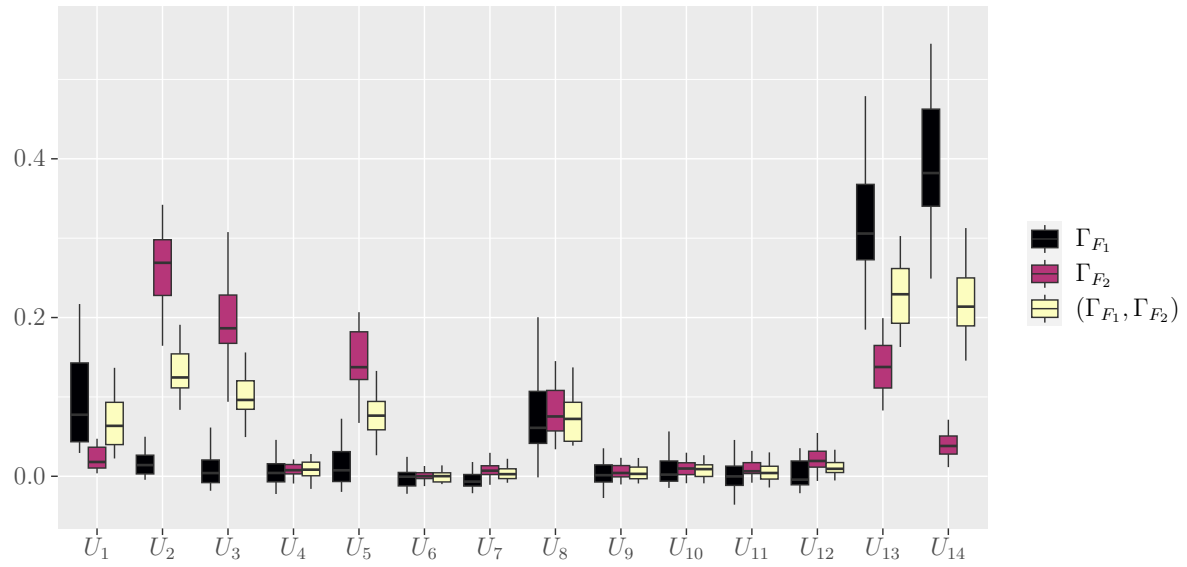


Figure 13 – Estimation of $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ for the excursion sets Γ_{f_1}, Γ_{f_2} and the pair $(\Gamma_{f_1}, \Gamma_{f_2})$ with $-q_1 = 420$ N.m and $q_2 = 7\%$ computed with the Sobolev input kernel and with $n = 100$, $m = 100$. 20 replicates.