



**HAL**  
open science

## Kernel-based sensitivity analysis on excursion sets

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

► **To cite this version:**

Christophette Blanchet-Scalliet, Noé Fellmann, Céline Helbert, Adrien Spagnol, Delphine Sinoquet. Kernel-based sensitivity analysis on excursion sets. MASCOT-NUM2023, Apr 2023, Le Croisic, France. hal-04098720

**HAL Id: hal-04098720**

**<https://hal.science/hal-04098720v1>**

Submitted on 17 May 2023

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

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



Distributed under a Creative Commons Attribution 4.0 International License

## Introduction

**Goal:** Quantifying the influence of **uncertain inputs** on **feasible sets** associated to a constrained robust optimization problem.

$$x^* = \arg \min_{x \in \mathcal{X}} \mathbb{E}_U[f(x, U)] \text{ s.t. } \mathbb{P}_U[g(x, U) \leq 0] \geq \alpha.$$

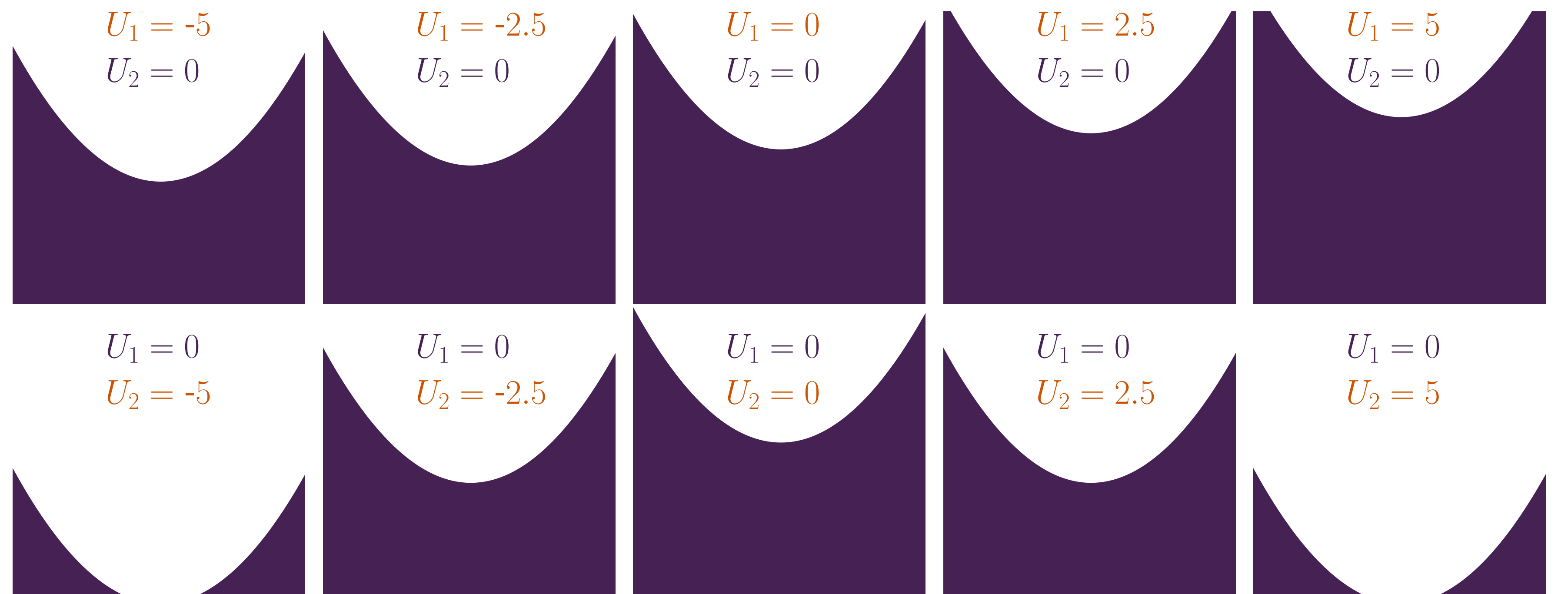
**Answer:** Performing **Sensitivity analysis** on **excursion sets**  $\Gamma_U$  using **kernel-based** methods. We propose a kernel  $k_{set}$  with which we compute **HSIC-ANOVA** indices.

$$(U_1, \dots, U_p) \mapsto \Gamma = \{x \in \mathcal{X}, g(x, U) \leq 0\}$$

**Example:**

$$\Gamma = \{x \in [-5, 5]^2, x_1^2 + 5x_2 - U_1 + U_2^2 - 1 \leq 0\}$$

→  $U_2$  seems more influential than  $U_1$  on the feasible sets.



## Kernel-based Sensitivity Analysis

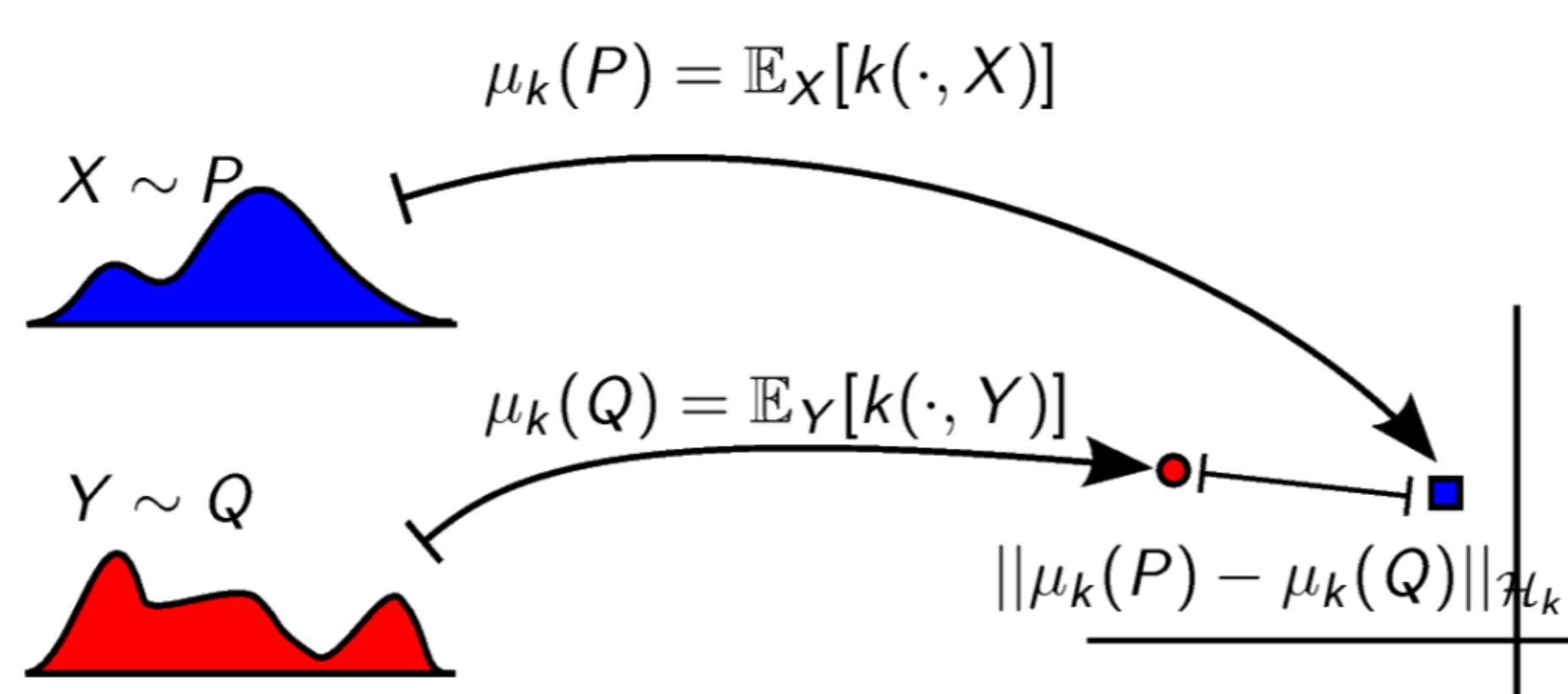


Fig. 1: Kernel mean embedding [6]

With  $K = k_X \otimes k_Y$ , the Hilbert Schmidt Independence Criterion (HSIC) is given by:

$$\begin{aligned} \text{HSIC}_{k_X, k_Y}(X, Y) &= \|\mu_K(X, Y) - \mu_{k_X}(X) \otimes \mu_{k_Y}(Y)\|_{\mathcal{H}_K}^2 \\ &= \mathbb{E}[k_X(X, X')k_Y(Y, Y')] \\ &\quad + \mathbb{E}[k_X(X, X')]\mathbb{E}[k_Y(Y, Y')] \\ &\quad - 2\mathbb{E}[\mathbb{E}[k_X(X, X')|X]\mathbb{E}[k_Y(Y, Y')|Y]]. \end{aligned}$$

When  $K$  is **characteristic** (injectivity of the mean embedding),

$$\text{HSIC}_{k_X, k_Y}(X, Y) = 0 \text{ iff } X \perp Y.$$

→ **screening**

Assuming that the inputs are **independent** and that the input kernels are **ANOVA**,

$$\text{HSIC}(\mathbf{X}, Y) = \sum_{A \subseteq \{1, \dots, d\}} \sum_{B \subseteq A} (-1)^{|A|-|B|} \text{HSIC}(\mathbf{X}_B, Y).$$

HSIC-ANOVA indices [3] are then defined as:

$$S_i^{\text{HSIC}} := \frac{\text{HSIC}(X_i, Y)}{\text{HSIC}(\mathbf{X}, Y)},$$

$$S_{T_i}^{\text{HSIC}} := 1 - \frac{\text{HSIC}(\mathbf{X}_{-i}, Y)}{\text{HSIC}(\mathbf{X}, Y)}$$

and are suited for **ranking** (and screening [5]).

## HSIC for set-valued outputs

With  $A \Delta B = A \cup B - B \cap A$  and  $\mu$  the Lebesgue measure, we define a kernel on closed sets by:

$$\forall \Gamma_1, \Gamma_2 \in \mathcal{F}(\mathcal{X}), k_{set}(\Gamma_1, \Gamma_2) = \exp\left(-\frac{\mu(\Gamma_1 \Delta \Gamma_2)}{2\sigma^2}\right).$$

### Proposition 1

$k_{set}$  is a kernel [1] and is characteristic.

Given an input  $U_i$  and an associated ANOVA kernel  $k_i$ , the HSIC on sets that we call  $H_{set}$  is defined by:

$$\begin{aligned} H_{set}(U_i, \Gamma) &:= \text{HSIC}_{k_i, k_{set}}(U_i, \Gamma) \\ &= \mathbb{E}[(k_i(U_i, U_i') - 1)k_{set}(\Gamma, \Gamma')]. \end{aligned}$$

It can be estimated with its U-statistic:

$$\widehat{H}_{set}(U_i, \Gamma) = \frac{2}{n(n-1)} \sum_{i < j} (k_i(U_i^{(i)}, U_i^{(j)}) - 1) k_{set}(\Gamma^{(i)}, \Gamma^{(j)}).$$

$k_{set}$  is rarely exactly known and requires to be estimated. Given  $X^{(1)}, \dots, X^{(m)}$  an iid sample of  $X \sim \mathcal{U}(\mathcal{X})$ , we propose the estimator:

$$\widehat{k}_{set}(\Gamma^{(i)}, \Gamma^{(j)}) = e^{-\frac{\mu(\mathcal{X})}{2\sigma^2} \frac{1}{m} \sum_{k=1}^m \mathbb{1}_{\Gamma^{(i)} \Delta \Gamma^{(j)}}(X^{(k)})},$$

leading finally to the **nested**  $H_{set}$  estimator:

$$\widehat{\widehat{H}}_{set}(U_i, \Gamma) = \frac{2}{n(n-1)} \sum_{i < j} (k_i(U_i^{(i)}, U_i^{(j)}) - 1) e^{-\frac{\mu(\mathcal{X})}{2\sigma^2} \frac{1}{m} \sum_{k=1}^m \mathbb{1}_{\Gamma^{(i)} \Delta \Gamma^{(j)}}(X^{(k)})}.$$

### Proposition 2

The quadratic risk of the nested estimator  $\widehat{\widehat{H}}_{set}$  verifies:

$$\mathbb{E} \left( \widehat{\widehat{H}}_{set}(U_i, \Gamma) - H_{set}(U_i, \Gamma) \right)^2 \leq 2 \left( \frac{2\sigma_1^2}{n(n-1)} + \frac{4(n-2)\sigma_2^2}{n(n-1)} + \frac{K^2\sigma_3^2}{m} \right).$$

HSIC-ANOVA indices on sets are then denoted  $S_i^{\widehat{\widehat{H}}_{set}}$  and  $S_{T_i}^{\widehat{\widehat{H}}_{set}}$ .

## Toy case 1 [4]

$$\forall x, u \in [-5, 5]^2 \times [-5, 5]^3 \quad g_1(x_1, x_2, u_1, u_2, u_3) = -x_1^2 + 5x_2 - u_1 + u_2^2 - 1$$

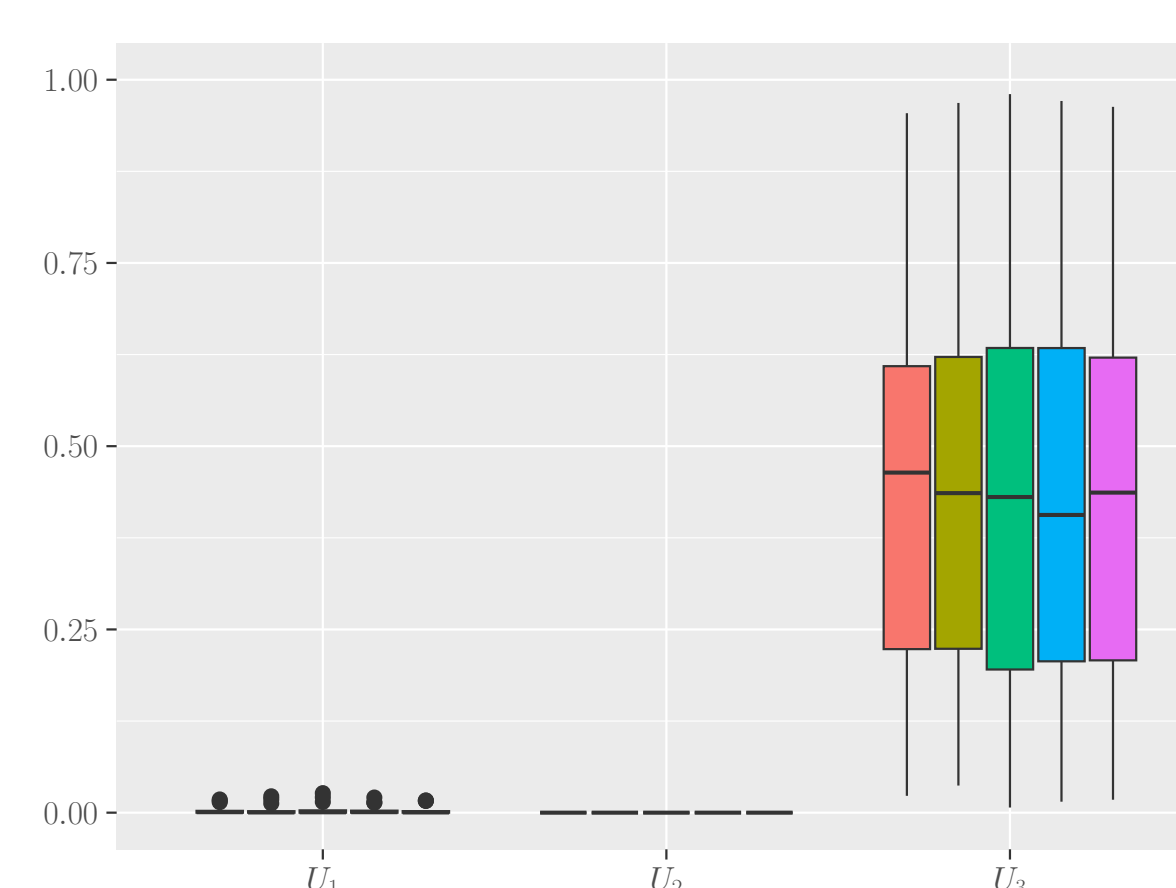


Fig. 2: p-value of  $\widehat{\widehat{H}}_{set}$  for  $n = m = 100$

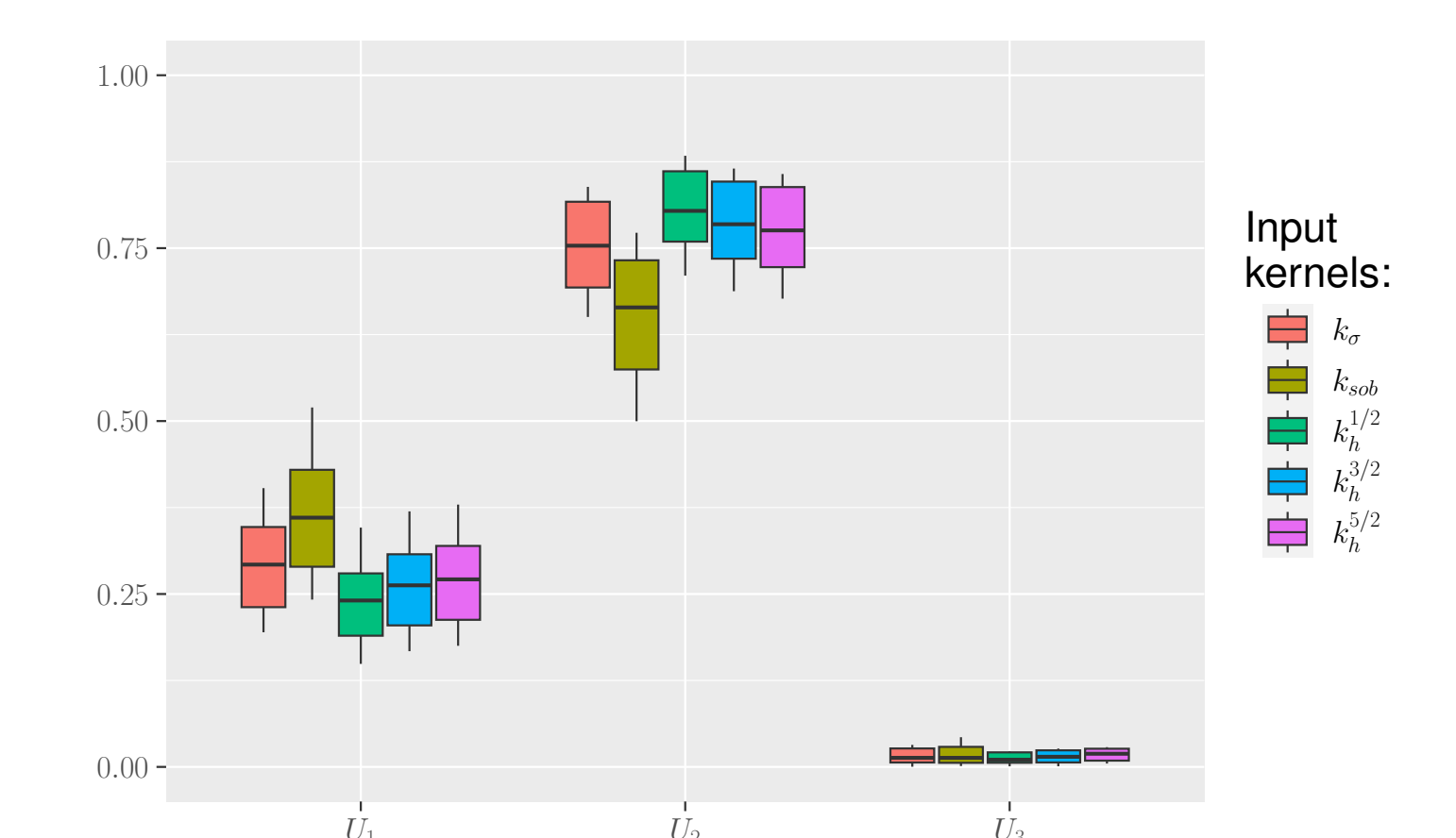


Fig. 3:  $S_{T_i}^{\widehat{\widehat{H}}_{set}}$  for  $n = m = 100$

## Toy case 2 [2]

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

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

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

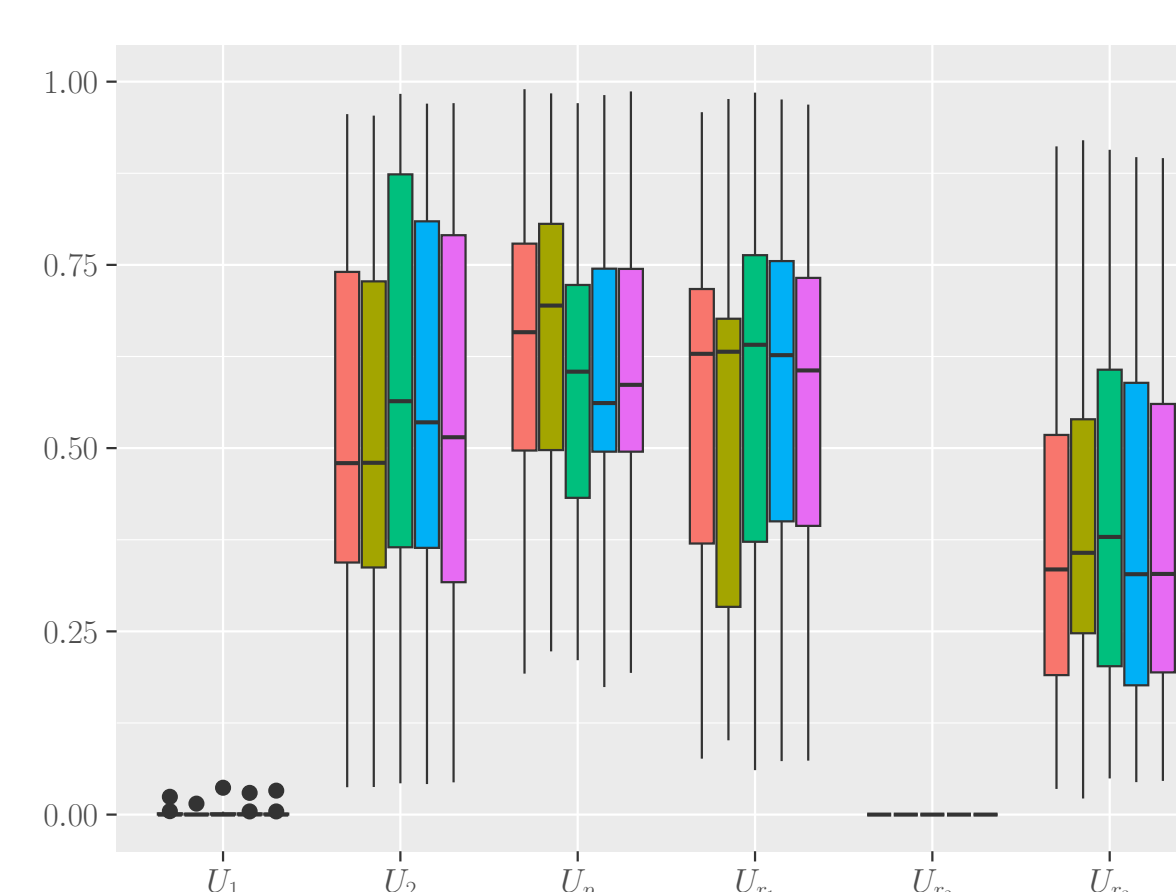


Fig. 4: p-value of  $\widehat{\widehat{H}}_{set}$  for  $n = m = 100$

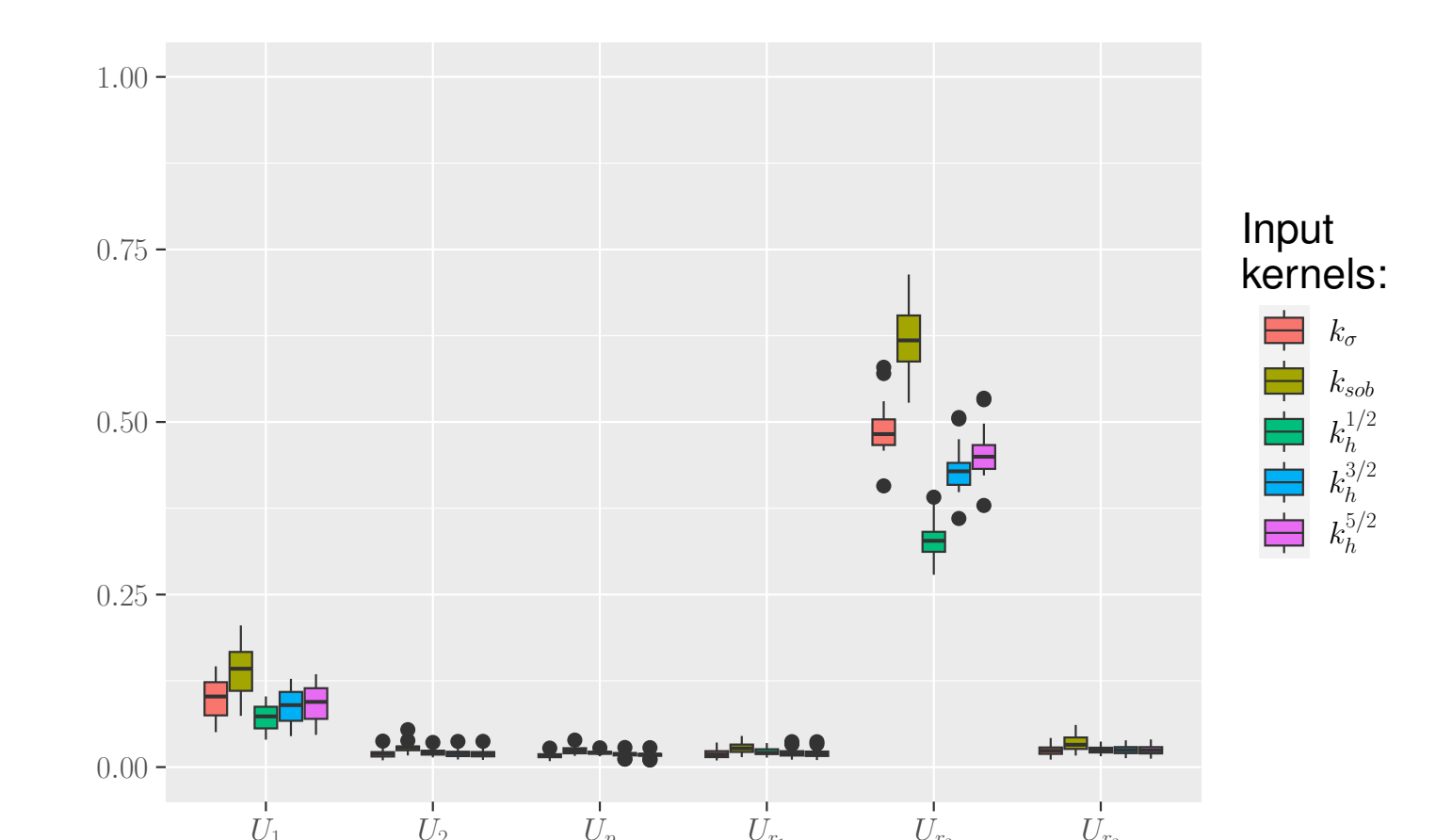


Fig. 5:  $S_{T_i}^{\widehat{\widehat{H}}_{set}}$  for  $n = m = 100$

## Outlook

- ▶ Testing the indices on **industrial applications** (as viability kernels, or air pollutant concentration maps) and compared to other indices defined for set-valued output models.
- ▶ Incorporating this method inside a **robust optimization** methodology: reducing the uncertain input space dimension to get cheaper meta-models.

## References

- [1] P. Balança, E. Herbin. A set-indexed Ornstein-Uhlenbeck process. 2012.
- [2] A. Cousin et al. A two-step procedure for time-dependent reliability-based design optimization involving piece-wise stationary Gaussian processes. 2022.
- [3] S. da Veiga. Kernel-based ANOVA decomposition and Shapley effects - Application to global sensitivity analysis. 2021.
- [4] R. El Amri et al. A sampling criterion for constrained Bayesian optimization with uncertainties. 2021
- [5] G. Sarazin et al. Test d'indépendance basé sur les indices HSIC-ANOVA d'ordre total. 2022.
- [6] D. Sejdinovic. Learning with Approximate Kernel Embeddings. RegML Workshop. 2017.