Application of HSIC-ANOVA indices to an industrial thermal-hydraulic case

SAMOURAI Workshop 2024 @ Institut Henri Poincaré

Simulation **A**nalytics and **M**eta-model-based solutions for **O**ptimization, **U**ncertainty and **R**eliability **A**nalys**I**s (ANR-20-CE46-0013)

Vincent Chabridon[1], Antoine Ajenjo[1], Gabriel Sarazin[2], Amandine Marrel[3], Sébastien Da Veiga[4]

December, 10th 2024

[1]EDF R&D, 6 quai Watier, 78401 Chatou, France
[2]CEA, DES, ISAS, DM2S, SGLS, LIAD, 91191 Gif-sur-Yvette, France
[3]CEA, DES, IRESNE, DER, Cadarache, 13108 Saint-Paul-Lez-Durance, France
[4]ENSAI, CREST - UMR 9194, F-35000 Rennes, France

Foreword

◆ **Acknowledgment:**

- ❑ To Delphine S. (for her patience during the 4-year-long project) and the co-organizers! ✌
- ❑ To the scientific partners of the ANR SAMOURAI Project!
  ➮ Special thanks to Gabriel S., Amandine M., Sébastien D.V. (WP #1) and Romain A.A.L., Julien B., Emmanuel V. (WP #2)!
- ❑ To the audience!

◆ **About the present work:**

- ❑ "Applicative" part of the work carried out in WP #1 by G. Sarazin during his postdoctoral work
  ➮ cea-04320711 ✌✌✌
- ❑ Mainly based on the use of the `sensitivity` R package
  ➮ Special thanks to B. Iooss (EDF R&D) as a maintainer and {S. Da Veiga, G. Sarazin} for HSIC-ANOVA codes!

1. Introduction

◆ A road trip through uncertainties!

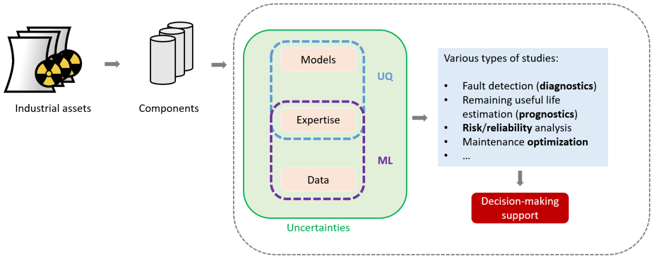

Figure 1: Dealing with uncertainties in an industrial process (©EDF).

☞ About UQ in industrial practice: [DRDT08, DR12]

Uncertainty Quantification in a nutshell!

◆ Verification, Validation & Uncertainty Quantification

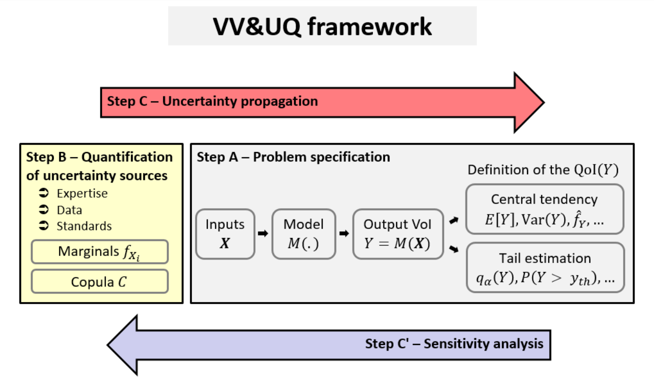

Figure 2: VV&UQ framework (©EDF).

☞ To go further into UQ in general: [Smi13, Sul15]
☞ To go further into Uncertainty Propagation: [MB15, DK22]
☞ To go further into Surrogate Modeling: [Bou18]
☞ To go further into SA: [SRA+08, DGIP21]

Main objectives of this talk

◆ **In this talk , you probably will . . .**

- ✔ Have a (very) short overview about a few sensitivity indices (Sobol', HSIC, HSIC-ANOVA) ☞ Wonderful textbook by [DGIP21]
- ✔ Have a first glimpse about applying these tools to **real-world** / **industrial** use cases!
- ✔ Benefit from a few insights about their practical advantages and drawbacks!

◆ **With this talk, you probably won't . . .**

- ✘ Become an expert in kernel-based sensitivity methods
  ☞ Go back to Gabriel's talk!

Main question

Do the HSIC-ANOVA indices bring a **significant improvement**, from an industrial viewpoint, in terms of **ranking**?

Table of contents

2. Background about GSA

GSA using Sobol' indices

◆ **Main notations and (⚠ strong) assumptions**

❏ G scalar-valued, black-box, deterministic computer model such that:

$$G : \left| \begin{array}{rcl} \mathcal{X} \subseteq \mathbb{R}^d & \to & \mathcal{Y} \subseteq \mathbb{R} \\ \mathbf{X} & \mapsto & Y = G(\mathbf{X}) \end{array} \right. \tag{1}$$

Assumption $\mathcal{A}_0$

Let $G \in \mathbb{L}^2(\mathbb{P}_{\mathbf{X}})$ where $\mathbb{L}^2(\mathbb{P}_{\mathbf{X}})$ is ($\approx$) "the set of all measurable functions g s.t. $\mathbb{E}\left[g^2(\mathbf{X})\right] < +\infty$".

Assumption $\mathcal{A}_1$

$\mathbf{X}$ is a second-order random vector of independent components, i.e.:

$$\mathbf{X} = (X_1, X_2, \ldots, X_d)^\top \sim \mathbb{P}_{\mathbf{X}} \quad \text{over} \quad \mathcal{X} = \bigtimes_{i=1}^{d} \mathcal{X}_i \tag{2}$$

◆ **Main notations and (⚠ strong) assumptions**

Assumption $\mathcal{A}_3$

Given-data context ➪ a **single** n-size i.i.d. learning sample of the couple $(\mathbf{X}, Y)$ is **available**!

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

with $\mathbb{P}_{\mathbf{X}^{(j)}} = \mathbb{P}_{\mathbf{X}}$
and $Y^{(j)} = G\left(X_1^{(j)}, X_2^{(j)}, \ldots, X_d^{(j)}\right), \forall j \in \{1, \ldots, n\}$

◆ A brief reminder about Sobol' indices

☆ Theorem (Hoeffding decomposition)

Under $\mathcal{A}_0$ and $\mathcal{A}_1$, $\exists!$ decomposition of G in $\mathbb{L}^2(\mathbb{P}_{\mathbf{X}})$ as follows:

$$G(\mathbf{x}) = \sum_{A \in \mathcal{P}_d} G_A(\mathbf{x}_A), \quad \mathbb{P}_{\mathbf{X}} - \text{a.s.}\,, \tag{4}$$

such that the following two properties hold:

(i) $G_\emptyset$ constant a.s.

(ii) $\forall\, A \in \mathcal{P}_d, A \neq \emptyset, \forall\, i \in A, \int_{E_i} G_A(\mathbf{x}_A)\mathbb{P}_{X_i}(\mathrm{d}x_i) = 0.$

The unique solution is given by, $\forall\, A \in \mathcal{P}_d$,

$$G_A(\mathbf{x}_A) = \sum_{B \subset A} (-1)^{|A|-|B|} \mathbb{E}\left[G(\mathbf{X}) \mid \mathbf{X}_B = \mathbf{x}_B\right] \quad \text{a.s.}\,. \tag{5}$$

◆ A brief reminder about Sobol' indices

☆ Corollary (FANOVA)

Assume $\mathcal{A}_0$. For any $A \in \mathcal{P}_d$, $A \neq \emptyset$, let $V_A = \mathrm{Var}\left(G_A(\mathbf{X}_A)\right)$. Then, under $\mathcal{A}_1$, one has:

$$V = \mathrm{Var}\left(G(\mathbf{X})\right) = \sum_{A \in \mathcal{P}_d, A \neq \emptyset} V_A. \tag{6}$$

Furthermore, for any $A \in \mathcal{P}_d$, $A \neq \emptyset$,

$$V_A = \sum_{B \subset A} (-1)^{|A|-|B|} \mathrm{Var}\left(\mathbb{E}\left[G(\mathbf{X}) \mid \mathbf{X}_B\right]\right). \tag{7}$$

GSA using Sobol' indices

◆ A brief reminder about Sobol' indices

✎ Definition (Sobol' indices)

Under $\mathcal{A}_0$ and $\mathcal{A}_1$ and $A \in \mathcal{P}_d$, one can define:

- The Sobol' index associated to A:

$$S_A = \frac{V_A}{V} = \frac{\sum_{B \subset A} (-1)^{|A|-|B|} \mathrm{Var}\left(\mathbb{E}\left[G(\mathbf{X}) \mid \mathbf{X}_B\right]\right)}{\mathrm{Var}\left(G(\mathbf{X})\right)}.$$

- The first-order index for the variable X_j: $S_j = S_{\{j\}}$.
- The closed index associated to A ($\equiv$ 1st-order of $\mathbf{X}_A$):

$$S_A^{\text{clos}} = \sum_{A' \subset A} S_{A'} = \frac{\mathrm{Var}\left(\mathbb{E}\left[G(\mathbf{X}) \mid \mathbf{X}_A\right]\right)}{\mathrm{Var}\left(G(\mathbf{X})\right)}.$$

- The total index associated to $\mathbf{X}_A$:

$$S_A^{\text{Tot}} = 1 - S_A^{\text{clos}}.$$

◆ An illustration of Sobol' indices

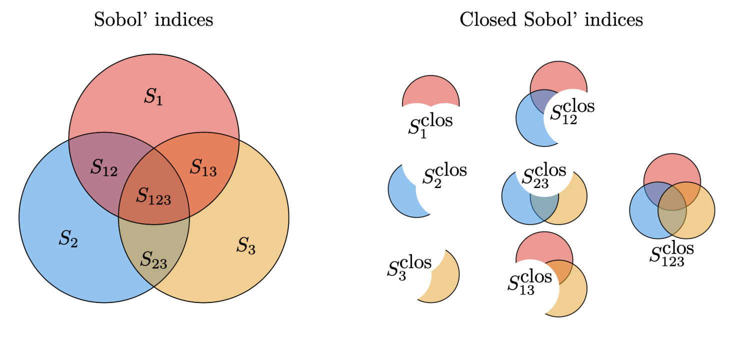

Figure 3: Illustration of the Sobol' and closed Sobol' indices for a three-input model (source: [II24]).

GSA using Sobol' indices

◆ **Estimation of Sobol' indices**

- ❑ A large panel of estimators do exist!
- ❑ Given-data contex ➪ k-nearest-neighbor estimator will be used in the following! ☞ [DVGIP21]
- ❑ In practice ➪ function `shapleysobol_knn()` in `sensitivity`
- ❑ ⚠ These estimators are known to show bias (in practice).

★ Take-home message #1 ★

✌ To keep in mind ✌

GSA requirements	S_j	S_j^{Tot}
Ranking	✓	✓
Screening	✗	✓
Given-data	✓	✗
Small data	✓	✗
Input dependency	✗	✗
Invariance U.M.T.	✓	✓

GSA using HSIC indices

◆ HSIC in a nutshell

Assumption $\mathcal{A}_4$

☞ Ingredients:

- ❑ let $\mathcal{H}_{\mathcal{X}}$ an RKHS of functions $\mathcal{X} \to \mathbb{R}$ with kernel $k_{\mathcal{X}}$;
- ❑ let $\mathcal{H}_{\mathcal{Y}}$ an RKHS of functions $\mathcal{Y} \to \mathbb{R}$ with kernel $k_{\mathcal{Y}}$;
- ❑ a couple of random vectors $(U, V) \sim \mathbb{P}_{(U,V)}$ on $\mathcal{X} \times \mathcal{Y}$ of marginal distributions $\mathbb{P}_U$ and $\mathbb{P}_V$, resp.

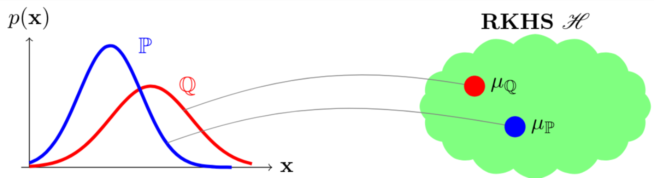

Figure 4: Illustration of the "kernel mean embedding" principle (source: [MFSS17])

GSA using HSIC indices

◆ HSIC in a nutshell

N.B.: HSIC ➪ Hilbert-Schmidt Independence Criterion

✎ Definition (HSIC indices)

☞ First formulation:

$$\text{HSIC}(U,V) = ||\mu_{\mathbb{P}_{(U,V)}} - \mu_{\mathbb{P}_U \times \mathbb{P}_V}||^2_{\mathcal{H}_\mathcal{X} \times \mathcal{H}_\mathcal{Y}} \tag{8}$$

where $\mu_{\mathbb{P}_{(U,V)}}$ and $\mu_{\mathbb{P}_U \times \mathbb{P}_V}$ are the **kernel mean embeddings** of the joint / product of marginal distributions, defined as:

$$\mu_{\mathbb{P}_{(U,V)}} = \int_{\mathcal{X}\times\mathcal{Y}} k_\mathcal{X}(\cdot,u)k_\mathcal{Y}(\cdot,v)\mathrm{d}\mathbb{P}_{(U,V)}(u,v) \tag{9}$$

$$\mu_{\mathbb{P}_U \times \mathbb{P}_V} = \int_{\mathcal{X}\times\mathcal{Y}} k_\mathcal{X}(\cdot,u)k_\mathcal{Y}(\cdot,v)\mathrm{d}\mathbb{P}_U(u)\mathrm{d}\mathbb{P}_V(v) \tag{10}$$

GSA using HSIC indices

◆ **HSIC in a nutshell**

✎ **Definition (HSIC indices)**

☞ Second formulation:

$$\begin{aligned}\mathrm{HSIC}(U,V) &= \mathbb{E}_{U,U',V,V'}\left[k_{\mathcal{X}}(U,U')k_{\mathcal{Y}}(V,V')\right] \\ &+ \mathbb{E}_{U,U'}\left[k_{\mathcal{X}}(U,U')\right]\mathbb{E}_{V,V'}\left[k_{\mathcal{Y}}(V,V'\right] \\ &- 2\mathbb{E}_{U,V}\left[\mathbb{E}_{U'}\left[k_{\mathcal{X}}(U,U')\right]\mathbb{E}_{V'}\left[k_{\mathcal{Y}}(V,V'\right]\right]\end{aligned} \tag{11}$$

where (U,V) and (U',V') are independent copies $\sim \mathbb{P}_{(U,V)}$.

☆ **Fundamental Property (HSIC)**

Assume that $k_{\mathcal{X}}$ and $k_{\mathcal{Y}}$ are **characteristic kernels** (✌), one has:

$$\mathrm{HSIC}(U,V) = 0 \Leftrightarrow U \perp V.$$

◆ Estimation of HSIC indices

❏ Two kinds of estimators are available for HSIC(X_i, Y)
➪ **U-stat.** vs. V-stat. ☞ [DVGIP21]

❏ One often uses a normalized sensitivity index (plug-in estimator):

$$\widehat{R^2_{\mathrm{HSIC},i}} = \frac{\widehat{\mathrm{HSIC}}(X_i, Y)}{\sqrt{\widehat{\mathrm{HSIC}}(X_i, X_i)\ \widehat{\mathrm{HSIC}}(Y, Y)}} \tag{12}$$

❏ Kernel choice (type + hyperparameters):

➤ mainstream approach (tabular data) ➪ **well-guided**!

$$k_{\mathcal{Z}}(z, z') = \exp\{-\theta ||z - z'||^2\} \quad \text{(Gaussian / RBF kernel)} \tag{13}$$

with $\theta = 1/\sigma^2$ (σ^2 the variance of Z, to be estimated empirically)

➤ specific task ➪ **user-defined** ⚠ Characteristic kernels!

❏ In practice ➪ functions `sensiHSIC()` (and `testHSIC()`) in `sensitivity`

★ Take-home message #2 ★

N.B.: ⚠ notation ➪ $H_j := \mathrm{HSIC}(X_j, Y)$

✌ To keep in mind ✌

GSA requirements	S_j	S_j^{Tot}	H_j
Ranking	✓	✓	✗
Screening	✗	✓	✓
Given-data	✓	✗	✓
Small data	✓	✗	✓
Input dependency	✗	✗	✓
Invariance U.M.T.	✓	✓	✗

◆ HSIC-ANOVA in a nutshell

☞ **See Gabriel's talk for further details!**

Assumption $\mathcal{A}_5$

The reproducing kernel $k_{\mathcal{X}}$ is of the form:

$$k_{\mathcal{X}}(\mathbf{x}, \mathbf{x}') = \prod_{j=1}^{d} (1 + k_j(x_j, x_j')) \tag{14}$$

where, $\forall\, j \in \{1, \dots, d\}$, $k_j(\cdot, \cdot)$ is the reproducing kernel of a RKHS $\mathcal{H}_j$ of real functions depending only on variable x_j and such that $1 \notin \mathcal{H}_j$. Furthermore, $\forall\, j \in \{1, \dots, d\}$, and $\forall\, x_j \in \mathcal{X}_j$, one has:

$$\int_{\mathcal{X}_j} k_j(x_j, x_j') \mathrm{d}\mathbb{P}_{X_j}(x_j') = 0.$$

<u>N.B.:</u> ✌ using univariate kernels associated to RKHSs which do not include constant functions.

◆ HSIC-ANOVA in a nutshell

☆ Theorem (ANOVA for HSIC)

Under the same assumptions as for Hoeffding decomposition and $\mathcal{A}_5$ (+ Mercer thm. holds):

$$\text{HSIC}(\mathbf{X}, Y) = \sum_{A \subseteq \mathcal{P}_d} \text{HSIC}_A \,, \tag{15}$$

where each term is given by:

$$\text{HSIC}_A = \sum_{B \subset A} (-1)^{|A|-|B|} \text{HSIC}(\mathbf{X}_B, Y) \,, \tag{16}$$

and $\text{HSIC}(\mathbf{X}_B, Y)$ is defined with kernel

$$k_B(\mathbf{x}_B, \mathbf{x}'_B) = \prod_{j \in B} (1 + k_j(x_j, x'_j))$$

on the inputs.

GSA using HSIC-ANOVA indices

◆ **HSIC-ANOVA in a nutshell**

✎ **Definition (HSIC-ANOVA indices)**

Under similar assumptions mentioned previously, and with $A \in \mathcal{P}_d$, one can define:

- The normalized HSIC-ANOVA index associated to A:

$$S_A^{\text{HSIC}} = \frac{\text{HSIC}_A}{\text{HSIC}(\mathbf{X}, Y)}.$$

- The total HSIC-ANOVA index associated to A:

$$S_A^{\text{Tot,HSIC}} = \sum_{B \subseteq \mathcal{P}_d, B \cap A \neq \emptyset} S_B^{\text{HSIC}} = 1 - \frac{\text{HSIC}(\mathbf{X}_{-A}, Y)}{\text{HSIC}(\mathbf{X}, Y)}.$$

◆ **HSIC-ANOVA in a nutshell**

☆ Fundamental Property (HSIC-ANOVA)

From previous theorem, one has:

$$\sum_{A \subseteq \mathcal{P}_d} S_A^{\text{HSIC}} = 1.$$

◆ Estimation of HSIC-ANOVA indices

- ❑ Two kinds of estimators are available for HSIC-ANOVA
  ➪ U-stat. vs. **V-stat.** ⚠ ☞ [DVGIP21]
- ❑ Kernel choice ➪ **Sobolev** $r = 1$ (for both inputs and output)
- ❑ In practice ➪ functions `sensiHSIC()` (and `testHSIC()`) in `sensitivity`

★ Take-home message #3 ★

N.B.: ⚠ notation ➪ $H_j := \text{HSIC}(X_j, Y)$

✌ To keep in mind ✌

GSA requirements	S_j	S_j^{Tot}	H_j	S_j^{H}	$S_j^{\text{Tot,H}}$
Ranking	✓	✓	✗	✓	✓
Screening	✗	✓	✓	✓	✓
Given-data	✓	✗	✓	✓	✓
Small data	✓	✗	✓	✓	✓
Input dependency	✗	✗	✓	✗	✗
Invariance U.M.T.	✓	✓	✗	✗	✗

3. Appli. #1 – Mystery Case

The "Mystery Case"

◆ **A (very) few words**

❏ Computer model: $G : \mathcal{X} \subseteq \mathbb{R}^{d=5} \to \mathbb{R}^+$

❏ Inputs: $\mathbf{X} = (X_\beta, X_\gamma, X_\eta, X_\rho, X_\pi)$ ➩ some quantities of materials

❏ Probabilistic model:

- ➤ $\forall\, j \in \{1, \ldots, d\},\ X_j \sim \mathcal{U}([a_j, b_j])$
- ➤ The components of $\mathbf{X}$ are mutually independent!

❏ Output variable of interest: $Y = G(\mathbf{X})$ ➩ a distance of propagation of the output phenomenon

❏ Given-data context: i.i.d. sample of size $n = 2 \times 10^3$ for $\left(\mathbf{X}^{(j)}, Y^{(j)}\right)_{(1 \leq j \leq n)}$

Goal of the study

☞ Analyse the relative importance of the inputs and the identify the interactions.

The "Mystery Case"

◆ Scenarios

- ❑ The idea: the problem does involve either all, or a part of the inputs ➪ several scenarios have to be tested!
- ❑ 4 analyses are conducted:
    1. $d = 5$, $\mathbf{X} = (X_\beta, X_\gamma, X_\eta, X_\rho, X_\pi)$
    2. $d = 3$, $\mathbf{X} = (X_\beta, X_\gamma, X_\rho)$
    3. $d = 3$, $\mathbf{X} = (X_\beta, X_\gamma, X_\eta)$
    4. $d = 3$, $\mathbf{X} = (X_\beta, X_\eta, X_\rho)$

Specific question

☞ Are there some special combinations between materials that influence Y globally?

◆ **Uncertainty analysis**

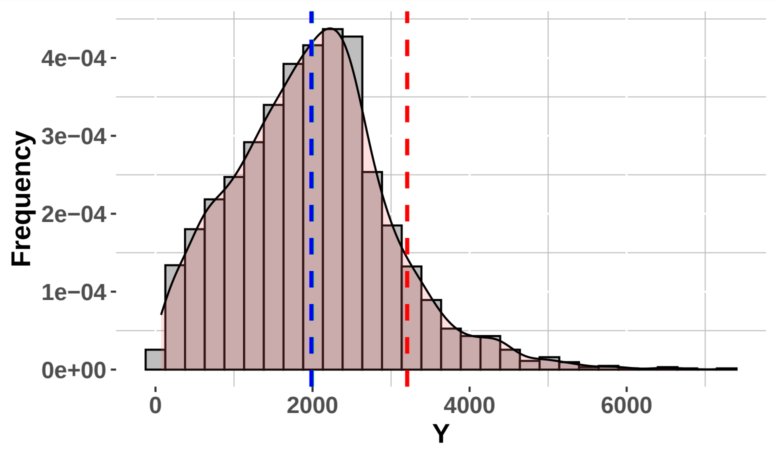

Figure 5: Output distribution (histogram and KDE) – Case #1.

The "Mystery Case" – Case #1 ($d = 5$, $\mathbf{X} = (X_\beta, X_\gamma, X_\eta, X_\rho, X_\pi)$)

◆ Input-output visualization

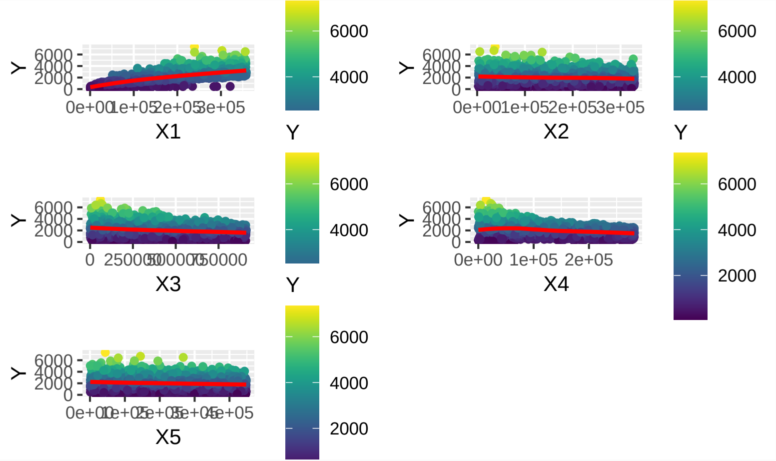

Figure 6: Scatter plots – Case #1.

The "Mystery Case" – Case #1 ($d = 5$, $\mathbf{X} = (X_\beta, X_\gamma, X_\eta, X_\rho, X_\pi)$)

◆ **Sobol' indices**

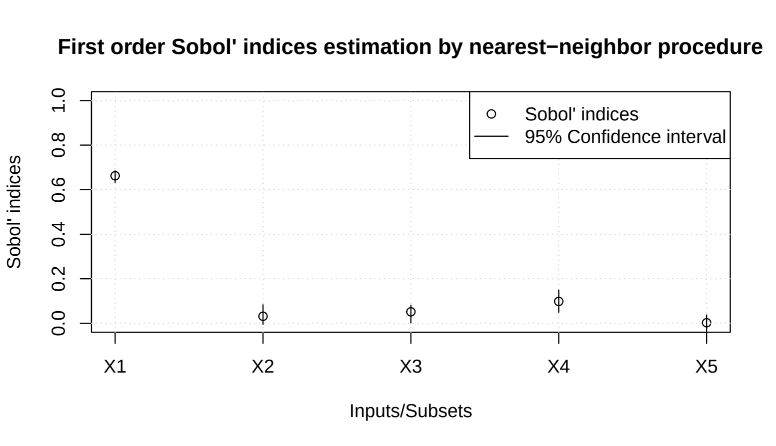

Figure 7: First-order Sobol' indices – Case #1.

The "Mystery Case" – Case #1 ($d = 5$, $\mathbf{X} = (X_\beta, X_\gamma, X_\eta, X_\rho, X_\pi)$)

◆ **Sobol' indices**

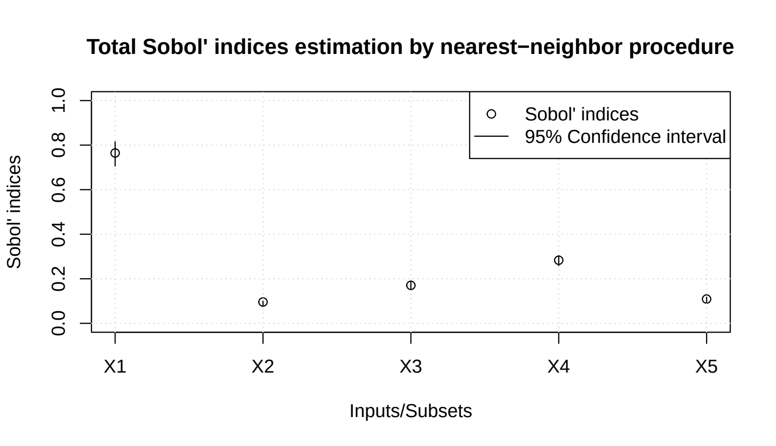

Figure 8: Total Sobol' indices – Case #1.

The "Mystery Case" – Case #1 ($d = 5$, $\mathbf{X} = (X_\beta, X_\gamma, X_\eta, X_\rho, X_\pi)$)

◆ **Sobol' indices**

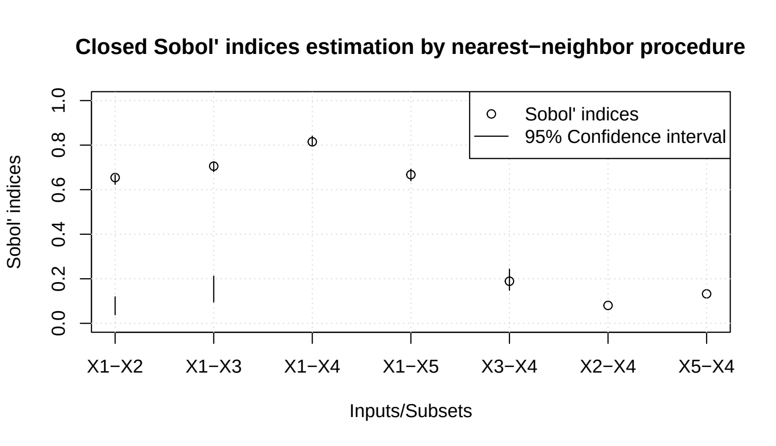

Figure 9: Closed Sobol' indices – Case #1.

◆ R2-HSIC indices

N.B.: ⚠ HSIC indices with Gaussian / RBF kernels.

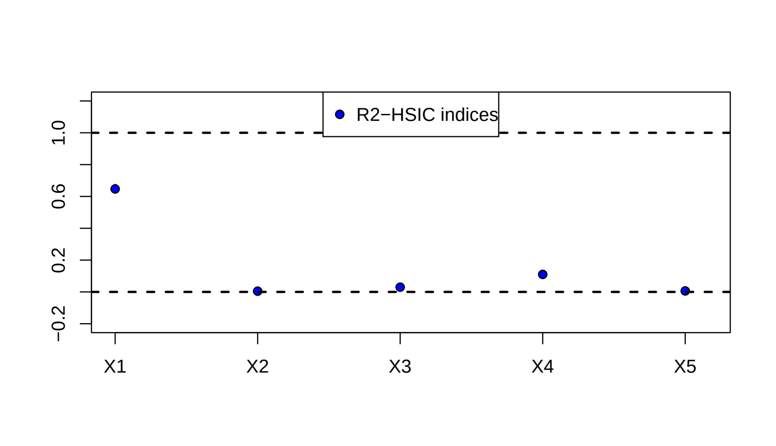

Figure 10: R2-HSIC indices – Case #1.

◆ HSIC-ANOVA indices

N.B.: ⚠ HSIC indices with Sobolev ($r = 1$) kernels.

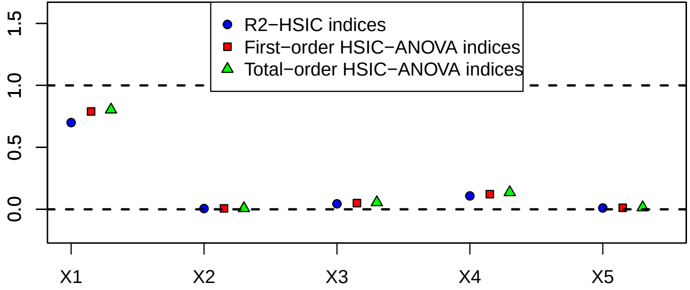

Figure 11: HSIC-ANOVA indices – Case #1.

◆ **Uncertainty analysis**

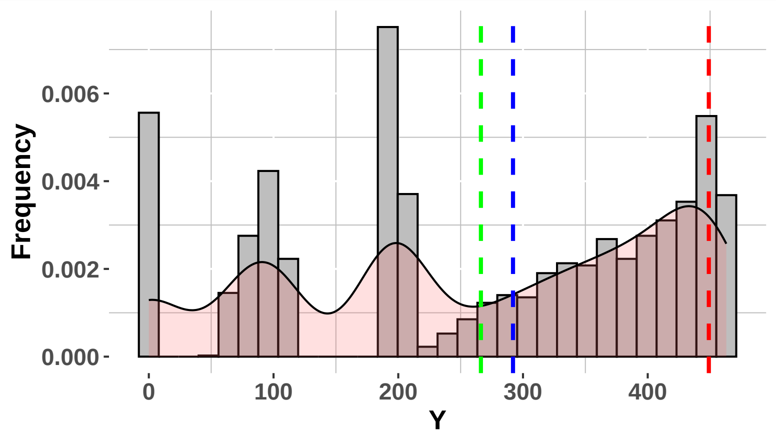

Figure 12: Output distribution (histogram and KDE) – Case #2.

◆ Input-output visualization

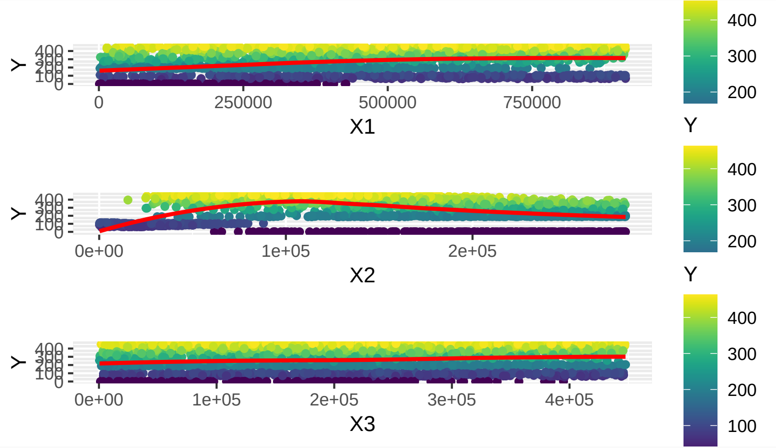

Figure 13: Scatter plots – Case #2.

The "Mystery Case" – Case #2 ($d = 3$, $\mathbf{X} = (X_\beta, X_\gamma, X_\rho)$)

◆ **Sobol' indices**

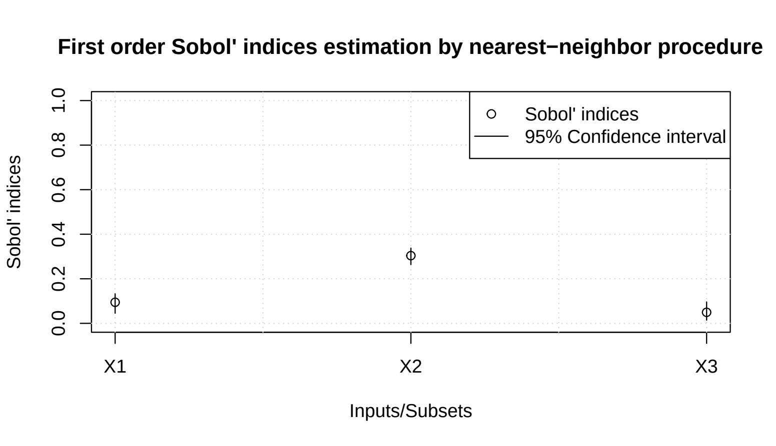

Figure 14: First-order Sobol' indices – Case #2.

The "Mystery Case" – Case #2 ($d = 3$, $\mathbf{X} = (X_\beta, X_\gamma, X_\rho)$)

◆ **Sobol' indices**

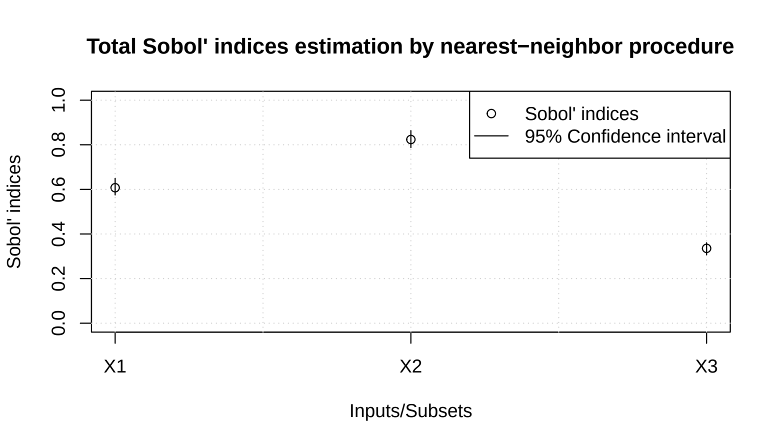

Figure 15: Total Sobol' indices – Case #2.

The "Mystery Case" – Case #2 ($d = 3$, $\mathbf{X} = (X_\beta, X_\gamma, X_\rho)$)

◆ **Sobol' indices**

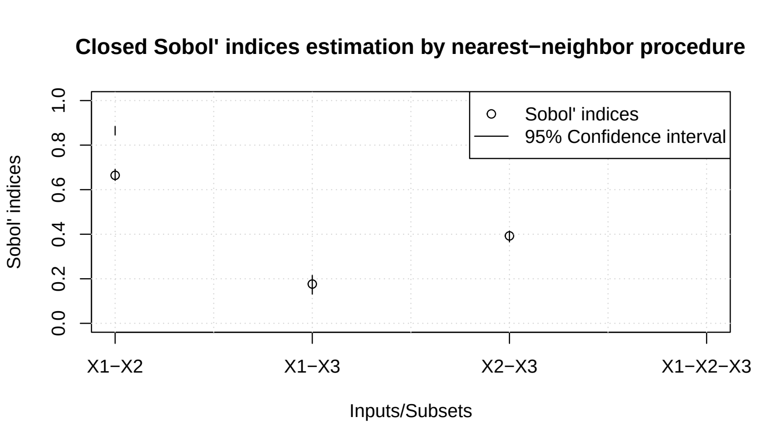

Figure 16: Closed Sobol' indices – Case #2.

The “Mystery Case” – Case #2 ($d = 3$, $\mathbf{X} = (X_\beta, X_\gamma, X_\rho)$)

◆ R2-HSIC indices

N.B.: ⚠ HSIC indices with Gaussian / RBF kernels.

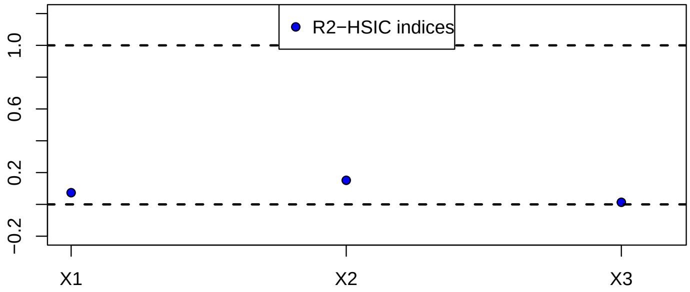

Figure 17: R2-HSIC indices – Case #2.

◆ R2-HSIC and HSIC-ANOVA indices

N.B.: ⚠ HSIC indices with Sobolev ($r = 1$) kernels.

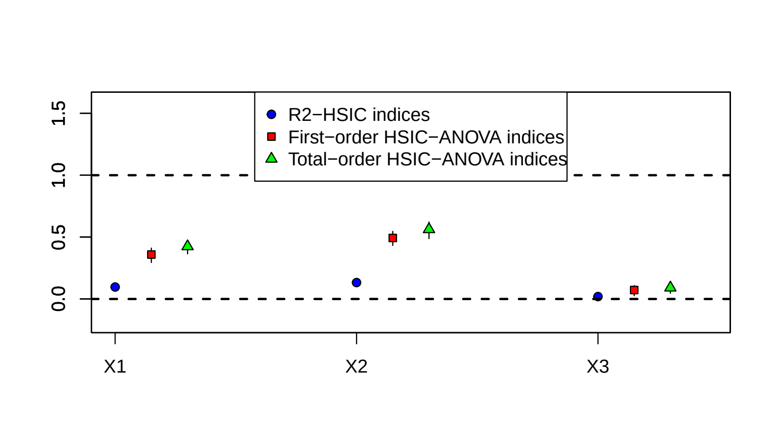

Figure 18: HSIC-ANOVA indices – Case #2.

The "Mystery Case" – Case #3 ($d = 3$, $\mathbf{X} = (X_\beta, X_\gamma, X_\eta)$)

◆ **Uncertainty analysis**

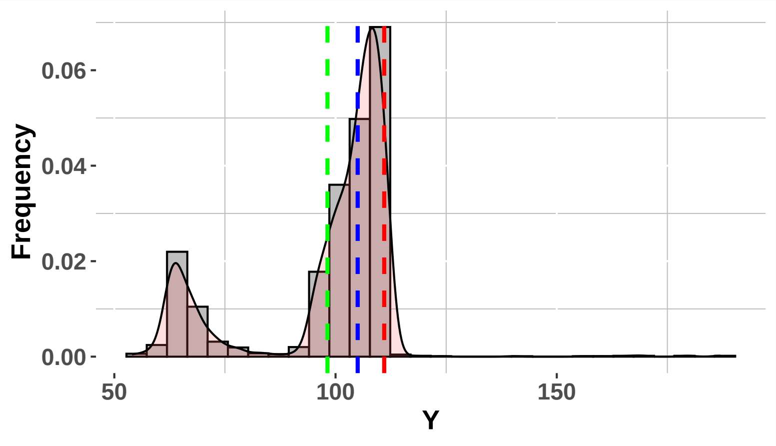

Figure 19: Output distribution (histogram and KDE) – Case #3.

The "Mystery Case" – Case #3 ($d = 3$, $\mathbf{X} = (X_\beta, X_\gamma, X_\eta)$)

◆ Sobol' indices

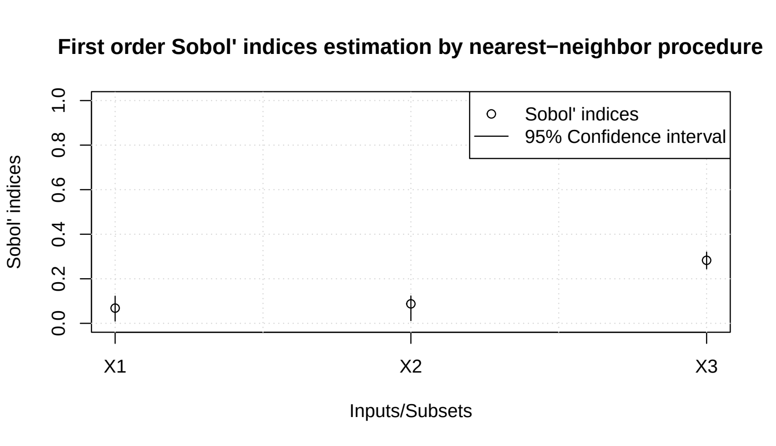

Figure 20: First-order Sobol' indices – Case #3.

The "Mystery Case" – Case #3 ($d = 3$, $\mathbf{X} = (X_\beta, X_\gamma, X_\eta)$)

◆ Sobol' indices

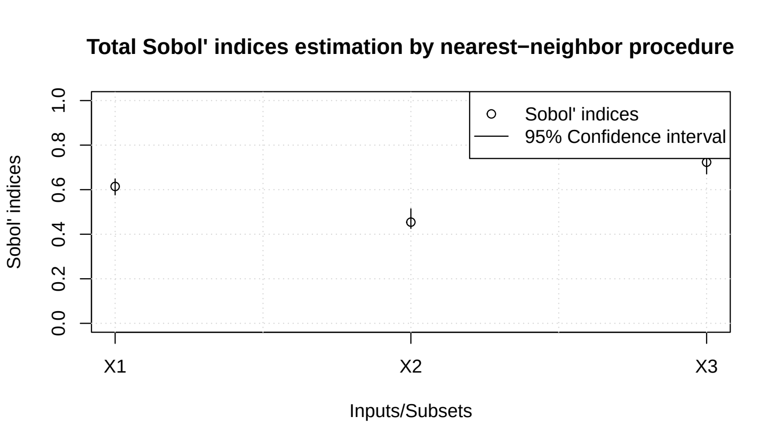

Figure 21: Total Sobol' indices – Case #3.

The "Mystery Case" – Case #3 ($d = 3, \mathbf{X} = (X_\beta, X_\gamma, X_\eta)$)

◆ Sobol' indices

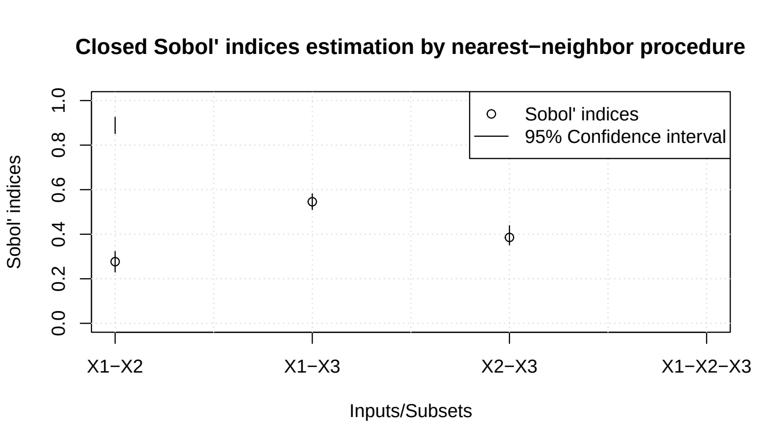

Figure 22: Closed Sobol' indices – Case #3.

◆ R2-HSIC indices

N.B.: ⚠ HSIC indices with Gaussian / RBF kernels.

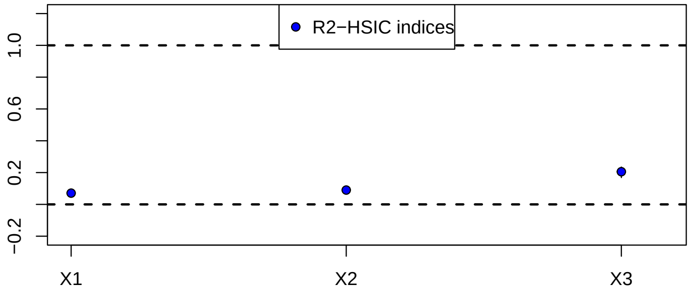

Figure 23: R2-HSIC indices – Case #3.

◆ R2-HSIC and HSIC-ANOVA indices

N.B.: ⚠ HSIC indices with Sobolev ($r = 1$) kernels.

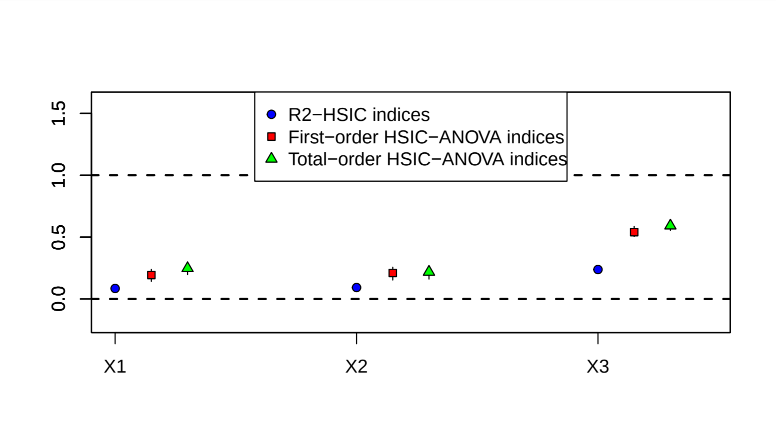

Figure 24: HSIC and HSIC-ANOVA indicices – Case #3.

The "Mystery Case" – Case #4 ($d = 3$, $\mathbf{X} = (X_\beta, X_\eta, X_\rho)$)

◆ **Uncertainty analysis**

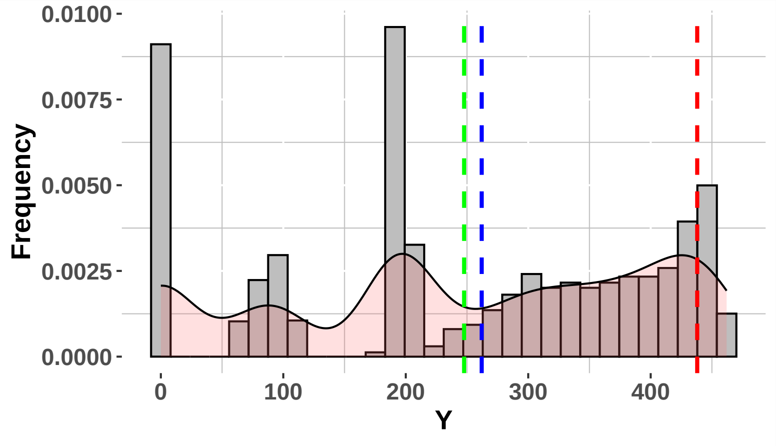

Figure 25: Output distribution (histogram and KDE) – Case #3.

The "Mystery Case" – Case #4 ($d = 3$, $\mathbf{X} = (X_\beta, X_\eta, X_\rho)$)

◆ Sobol' indices

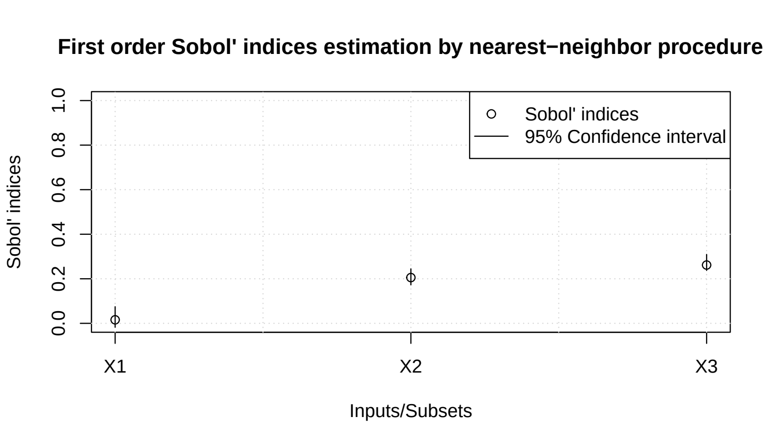

Figure 26: First-order Sobol' indices – Case #4.

The "Mystery Case" – Case #4 ($d = 3$, $\mathbf{X} = (X_\beta, X_\eta, X_\rho)$)

◆ **Sobol' indices**

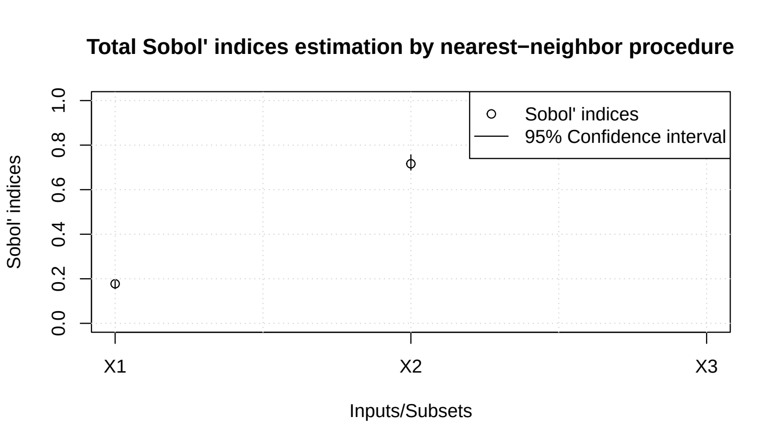

Figure 27: Total Sobol' indices – Case #4.

The "Mystery Case" – Case #4 ($d = 3, \mathbf{X} = (X_\beta, X_\eta, X_\rho)$)

◆ Sobol' indices

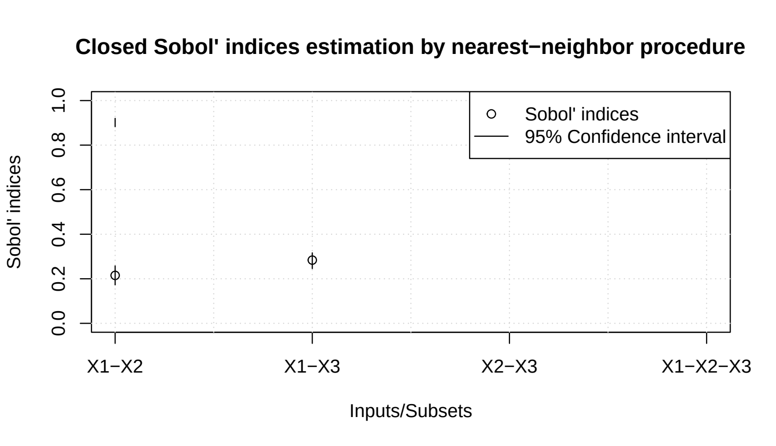

Figure 28: Closed Sobol' indices – Case #4.

The "Mystery Case" – Case #4 ($d = 3$, $\mathbf{X} = (X_\beta, X_\eta, X_\rho)$)

◆ R2-HSIC indices

N.B.: ⚠ HSIC indices with Gaussian / RBF kernels.

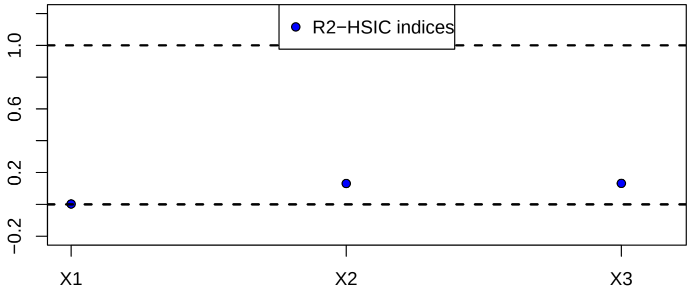

Figure 29: R2-HSIC indices – Case #4.

The "Mystery Case" – Case #4 ($d = 3$, $\mathbf{X} = (X_\beta, X_\eta, X_\rho)$)

◆ HSIC and HSIC-ANOVA indices

N.B.: ⚠ HSIC indices with Sobolev ($r = 1$) kernels.

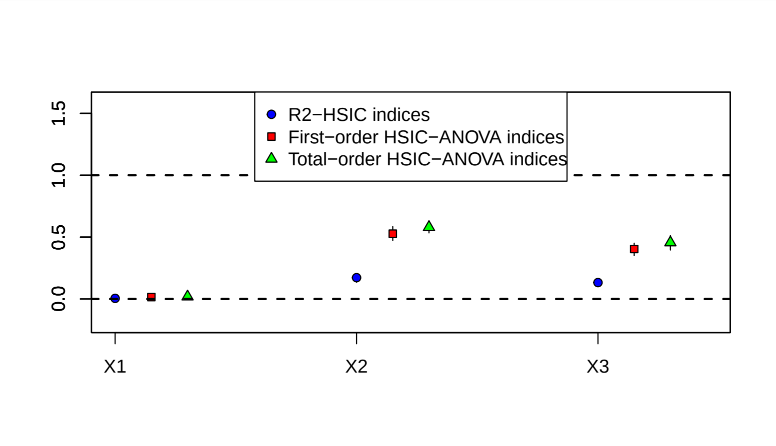

Figure 30: HSIC and HSIC-ANOVA indices – Case #4.

4. Appli. #2 – TH

Thermal-hydraulic accident scenario use case

◆ **A (very) few words**

- ❑ Reliability and risk assessment of critical nuclear systems/components:
  - ➤ Deterministic analyses (a.k.a. "conservative procedures")
  - ➤ "**Best-estimate plus uncertainty**" (**BEPU**) analyses
- ❑ Safety analyses using a set of accident scenarios, e.g., for thermal-hydraulic issues:
  - ➤ small-break loss-of-coolant accident
  - ➤ **intermediate-break loss-of-coolant accident** (**IBLOCA**)
  - ➤ large-break loss-of-coolant accident

Goal of the study

☞ Analyse the relative importance of the inputs and the identify the interactions.

◆ A (very) few words

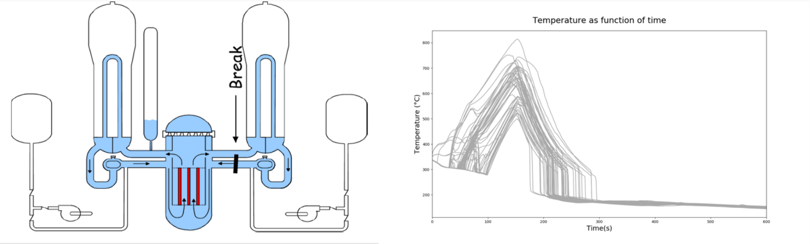

Figure 31: Illustrative scheme of an IBLOCA scenario (@CEA)

–

Simulation trajectories of the Peak Cladding Temperature (PCT) (@EDF).

Thermal-hydraulic accident scenario use case

◆ A (very) few words

- ❏ Complex system: **primary circuit (cold leg) of a PWR**
- ❏ Scenario: **IBLOCA** (thermal-hydraulic issue)
- ❏ Sources of uncertainties:
  - ☞ Critical flowrates, Initial/boundary conditions, ...
- ❏ Probabilistic quantification of the input variables:
  - ☞ Marginal probability density functions (PDFs): $\mathcal{U}$, $\mathcal{LU}$, $\mathcal{N}$, $\mathcal{LN}$
- ❏ Goal of the study: **risk assessment**
  - ☞ Scalar model output $\rightarrow$ the 2^{nd} **peak of cladding temperature** (**PCT**)
  - ☞ Quantity of Interest (QoI) $\rightarrow$ a **high-order quantile** over the PCT
- ❏ Simulation computer model: **CATHARE2 (V2.5_3mod3.1) code**
  - ☞ Highly-nonlinear
  - ☞ Costly-to-evaluate (1 run $>$ 1 hour)
  - ☞ High-dimensional (≈ 100 inputs)
- ❏ Uncertainty propagation: Monte Carlo sample of $n = 1496$ simulations

◆ **Uncertainty analysis**

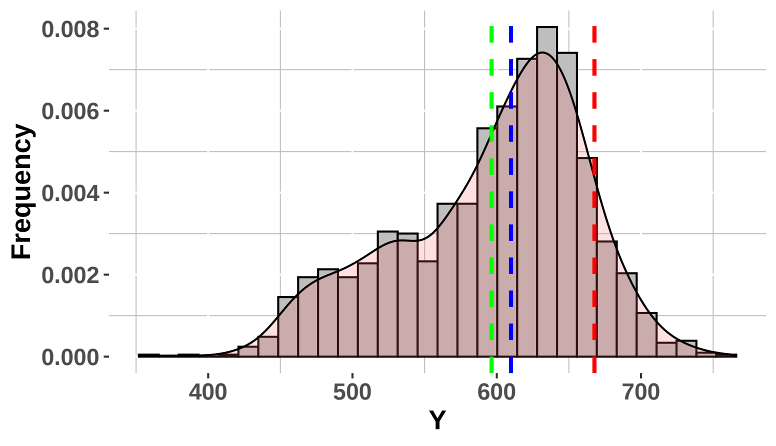

Figure 32: Output distribution (histogram and KDE).

◆ Sobol' indices

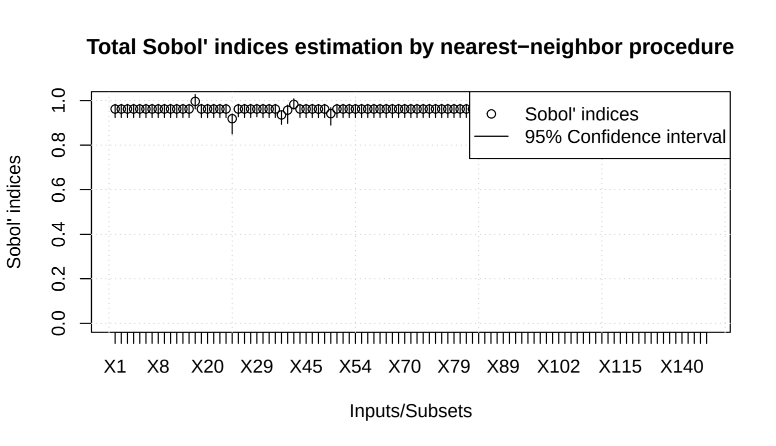

Figure 33: Total Sobol' indices.

◆ Sobol' indices

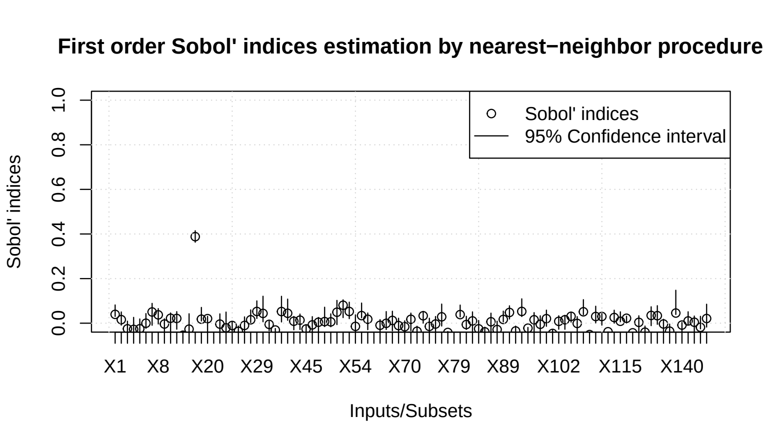

Figure 34: First-order Sobol' indices.

◆ HSIC indices

N.B.: ⚠ HSIC indices with Gaussian / RBF kernels.

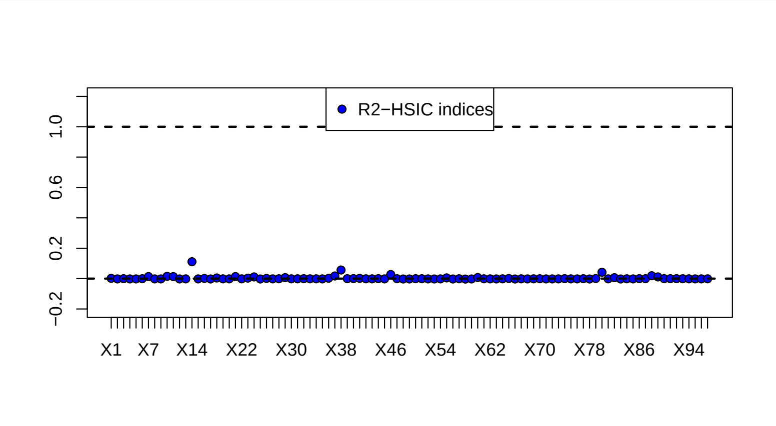

Figure 35: HSIC and HSIC-ANOVA indices.

◆ HSIC and HSIC-ANOVA indices

N.B.: ⚠ HSIC indices with Sobolev ($r = 1$) kernels.

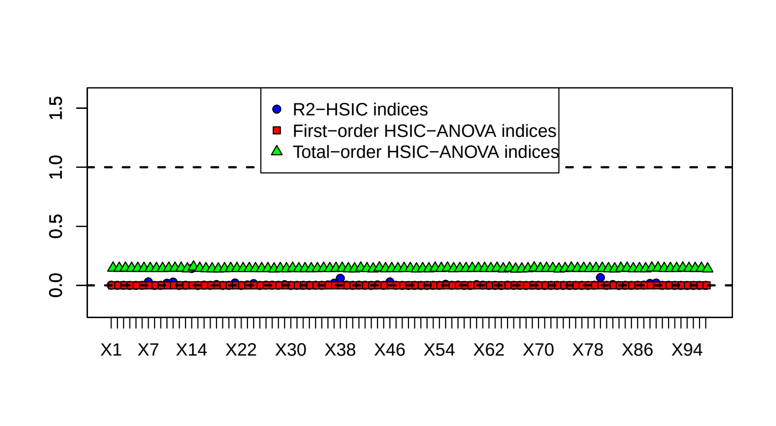

Figure 36: HSIC and HSIC-ANOVA indices.

Conclusion

Conclusion

✌ To REALLY keep in mind ✌

◆ **From an industrial viewpoint, HSIC-ANOVA . . .**

✔ Offer an **elegant** and sound theoretical framework!

✔ Benefit from **both** sides: Sobol' indices (ANOVA) and HSIC (beyond variance-based indices)!

✔ Show a great potential to detect **high-order** / **fine** interactions!

◆ **Some work still has to be done. . .**

✘ To better understand the **patterns** of interactions detected by the total index!

✘ To better understand the high-dimensional setting and the effect of the estimator type ➪ V-stat vs. U-stat?

✘ To implement these indices in **our open source platform** OpenTURNS (HSIC already in there!) ✌

Thank your for your attention!

Any question?

References i

J.-M. Bourinet.
Reliability analysis and optimal design under uncertainty – Focus on adaptive surrogate-based approaches.
HDR (French Accreditation to Supervise Research), Université Clermont Auvergne, 2018.
245 pages.

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

A. Der Kiureghian.
Structural and System Reliability.
Cambridge University Press, 2022.

E. De Rocquigny.
Modelling Under Risk and Uncertainty: An Introduction to Statistical, Phenomenological and Computational Methods.
Wiley series in Probability and Statistics. Wiley, 2012.

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

S. Da Veiga, F. Gamboa, B. Iooss, and C. Prieur.
Basics and Trends in Sensitivity Analysis: Theory and Practice in R.
Computational Science and Engineering. SIAM, 2021.

M. Il Idrissi.
Development of interpretability methods for certifying machine learning models applied to critical systems.
PhD thesis, Université de Toulouse, 2024.

References ii

J. Morio and M. Balesdent.
Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems: A Practical Approach.
Woodhead Publishing, Elsevier, 2015.

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

R. C. Smith.
Uncertainty Quantification: Theory, Implementation, and Applications.
Society for Industrial and Applied Mathematics, 2013.

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

T. J. Sullivan.
Introduction to Uncertainty Quantification, volume 63 of Texts in Applied Mathematics.
Springer International Publishing Switzerland, 2015.