Sensitivity analysis for set-valued models. Application to pollutant concentration maps

Céline Helbert
joint work with N. Fellmann (ICJ), C. Blanchet (ICJ), D. Sinoquet (IFPEN), A. Spagnol (IFPEN)
Centrale Lyon & ICJ
ENBIS 2026

Initial motivation

Excursion set estimation

- f a black box numerical model
- $\mathbf{x} \in \mathcal{X}$ are the design variables

=> we look for the feasible solutions of the constrained problem

$$\Gamma = \{\mathbf{x} \in \mathcal{X}, f(\mathbf{x}) \leq q\},$$

Initial motivation

Excursion set estimation

- f a black box numerical model
- $\mathbf{x} \in \mathcal{X}$ are the design variables

=> we look for the feasible solutions of the constrained problem

$$\Gamma = \{\mathbf{x} \in \mathcal{X}, f(\mathbf{x}) \leq q\},$$

- Presence of uncertain variables : manufacturing uncertainties or environmental uncertainties

=> the excursion set becomes a random set

$$\Gamma_U = \{\mathbf{x} \in \mathcal{X}, f(\mathbf{x}, \mathbf{U}) \leq q\}$$

Initial motivation

Excursion set estimation

- f a black box numerical model
- $\mathbf{x} \in \mathcal{X}$ are the design variables

=> we look for the feasible solutions of the constrained problem

$$\Gamma = \{\mathbf{x} \in \mathcal{X}, f(\mathbf{x}) \leq q\},$$

- Presence of uncertain variables : manufacturing uncertainties or environmental uncertainties

=> the excursion set becomes a random set

$$\Gamma_U = \{\mathbf{x} \in \mathcal{X}, f(\mathbf{x}, \mathbf{U}) \leq q\}$$

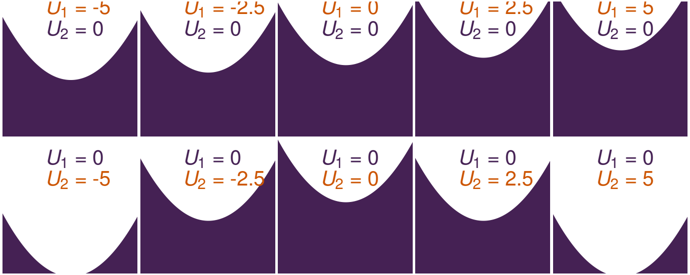
Goal : quantify the impact of U on excursion sets

A toy excursion set

Toy function from (El-Amri et al. 2021)

$$\forall(\mathbf{x},\mathbf{u}) \in [-5,5]^4, f(\mathbf{x},\mathbf{u}) = -x_1^2 + 5x_2 - u_1 + u_2^2 - 1.$$

Excursion set $\Gamma_{\mathbf{U}} = \{\mathbf{x} \in \mathcal{X}, f(\mathbf{x},\mathbf{U}) \leq 0\}$



How can we do sensitivity analysis on excursion sets ?

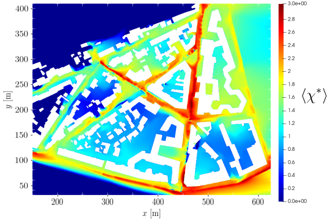
Another example

The case of pollutant concentration maps [**M. Pasquier**]

We consider a black box model which takes into account $\mathbf{u}$ a set of random parameters and provides a map, i.e. a function $\Phi_{\mathbf{u}} : (x_1, x_2) \mapsto \Phi_{\mathbf{u}}(x_1, x_2)$

Input $\mathbf{U} = (\theta, U_\infty, q, \beta, \nu_{max})$ with,

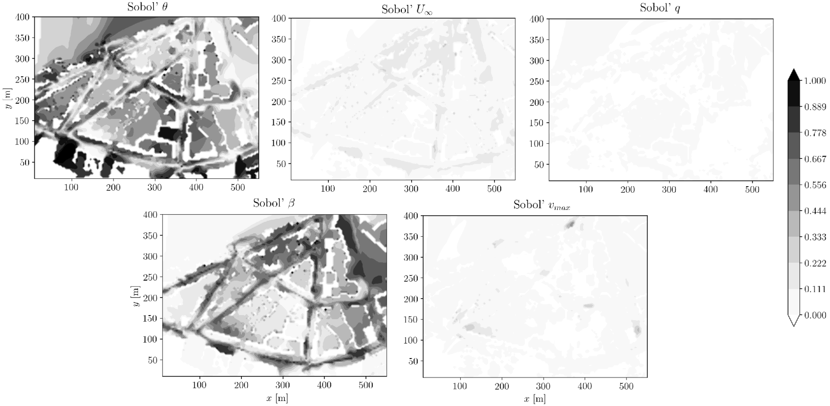
- Wind direction $\theta \sim \mathcal{N}_{[5,10]}(0.7, 0.5^2)$ [rad]
- Wind speed $U_\infty \sim \mathcal{N}_{[-0.35;1.75]}(8, 2)$ [m/s]
- Traffic volume $q \sim \mathcal{SN}_{[100;500]}(450, 100, -3)$ [vehicle/h]
- Proportion of diesel and petrol engine $\beta \sim \mathcal{U}([0, 1])$ [-]
- Speed limit $\nu_{\text{max}} \sim \mathcal{U}([30; 50])$ [km/h]



How can we do sensitivity analysis on maps ?

Sobol indices maps [**M. Pasquier**]

$$S_i(x_1, x_2) = \frac{\mathrm{Var}[\mathbb{E}[\Phi_{\mathbf{U}}(x_1, x_2) | U_i]]}{\mathrm{Var}[\Phi_{\mathbf{U}}(x_1, x_2)]}.$$



Generalized Sobol indices [M. Pasquier]

Generalized Sobol' indices (**vectorial_sobol**):

$$S_i^{\text{gen}} := \sum_{j=1}^{m} w_j S_i(x_1^{(j)}, x_2^{(j)}) \quad \text{with} \qquad w_j = \frac{\mathrm{Var}[\Phi_{\mathbf{U}}(x_1^{(j)}, x_2^{(j)})]}{\sum_{k=1}^{m} \mathrm{Var}[\Phi_{\mathbf{U}}(x_1^{(k)}, x_2^{(k)})]}.$$

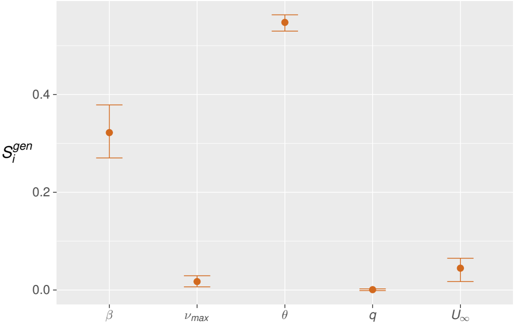
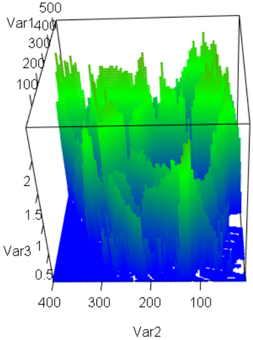


Figure: Estimated generalised first-order Sobol' indices.

From map-valued model to set-valued model

$$\Psi : \begin{array}{ccc} \mathcal{U} & \longrightarrow & \mathcal{L}(\mathcal{X}) \\ \mathbf{u} & \mapsto & \Gamma_{\mathbf{u}} = \{(x_1, x_2, c) \in \mathcal{D} \times [0, C_{max}], c \leq \Phi_{\mathbf{u}}(x_1, x_2)\} \end{array}$$



Goal : quantify the impact of U on random sets (excursion sets, maps, ...)

Table of Contents

Table of Contents

Vorob'ev-based indices

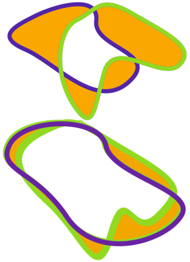
Sobol indices :

$$S_i = \frac{\mathrm{Var}\,\mathbb{E}[Y|U_i]}{\mathrm{Var}\,Y} = 1 - \frac{\mathbb{E}\,\mathrm{Var}[Y|U_i]}{\mathrm{Var}\,Y}$$

Goal

Propose a similar decomposition through the theory of random sets (Molchanov 2005).

$$\begin{array}{lcl}
\mathbb{E} & \longleftrightarrow & Q_{0.5} = \{x \in \mathcal{X},\ \mathbb{P}(x \in \Gamma) \geq 0.5\} \\
\mathrm{Var} & \longleftrightarrow & \mathbb{E}(\lambda(\Gamma \Delta Q_{0.5})) \\
\mathrm{Var}(\Gamma|U_i) & \longleftrightarrow & \mathbb{E}(\lambda(\Gamma \Delta Q^i_{0.5})|U_i) \\
\mathbb{E}\mathrm{Var}(\Gamma|U_i) & \longleftrightarrow & \mathbb{E}(\lambda(\Gamma \Delta Q^i_{0.5}))
\end{array}$$



where $Q^i_{0.5} = \{\mathbf{x} \in \mathcal{X}, \mathbb{P}(\mathbf{x} \in \Gamma|U_i) \geq 0.5\}$.

Vorob'ev-based indices

Definition

"Sobol indices" on random sets:

$$S_i^V = 1 - \frac{\mathbb{E}[\lambda(\Gamma \Delta Q_{0.5}^i)]}{\mathbb{E}[\lambda(\Gamma \Delta Q_{0.5})]}$$

Proposition

$0 \leq S_i^V \leq 1$

Estimation

Double loop ...

Results

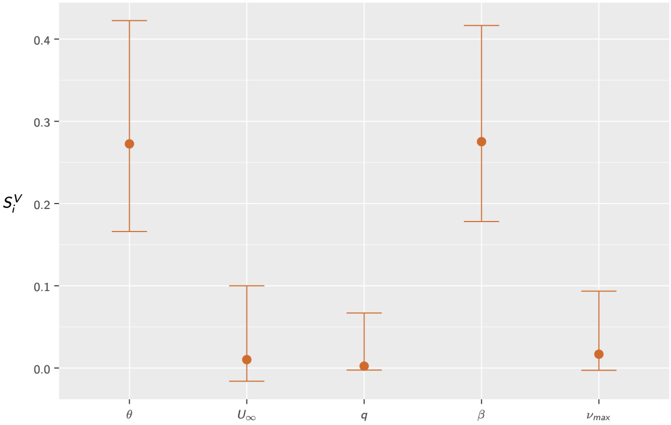


Figure: Estimation of S_V^i for each input with 1024 model evaluations and a confidence interval estimated with 100 bootstrap resamples.

Table of Contents

Universal indices

CENTRALE LYON

Definition (Fort, Klein, and Lagnoux 2021)

$$S_{\mathbf{u}}^{\text{Univ}}(T_a, \mathbb{Q}) = \frac{\int_\Omega Var\left(\mathbb{E}[T_a(Z)|X_{\mathbf{u}}]\right) d\mathbb{Q}(a)}{\int_\Omega Var(T_a(Z)) d\mathbb{Q}(a)}. \tag{1}$$

where $T_a : (a, x) \in \mathcal{X}^m \times \mathcal{X} \mapsto T_a(x) \in \mathbb{R}$ is a scalar which sums up object x.
Notice that this index lies in $[0, 1]$ and shares the same properties as its Sobol counterparts.

CENTRALE LYON

Definition (in the case of random sets)

$$S_i^{\text{Univ}}(\gamma, \mathbb{Q}) = \frac{\int_0^1 \operatorname{Var} \mathbb{E}\left[\lambda(\Gamma \Delta \gamma_a) \mid U_i\right] d\mathbb{Q}(a)}{\int_0^1 \operatorname{Var}\left(\lambda(\Gamma \Delta \gamma_a)\right) d\mathbb{Q}(a)}.$$

where $(\gamma_a)_{a \in [0,1]} \subset \mathcal{X}$ is a collection of test sets parameterised by $a \in [0, 1]$ and $\mathbb{Q}$ is a probability distribution on $[0, 1]$ to choose.

- Centered balls: $\gamma_a = B(x^0, a)$ with $\mathbb{Q} \sim \mathcal{U}([0, \frac{1}{2}])$
- Centered squares: $\gamma_a = \{x \in [0, 1]^3, ||x - x^0||_\infty \leq a\}$ with $\mathbb{Q} \sim \mathcal{U}([0, \frac{1}{2}])$
- Slices along the i-th dimension: $\gamma_a^i = \{x \in [0, 1]^3, x^i \leq a\}$ with $\mathbb{Q} \sim \mathcal{U}([0, 1])$
- Vorob'ev quantiles: $\gamma_a = \{x \in [0, 1]^3, \mathbb{P}(x \in \Gamma) \geq a\}$ with $\mathbb{Q} \sim \mathcal{N}(\frac{1}{2}, 0.05^2)$

Results

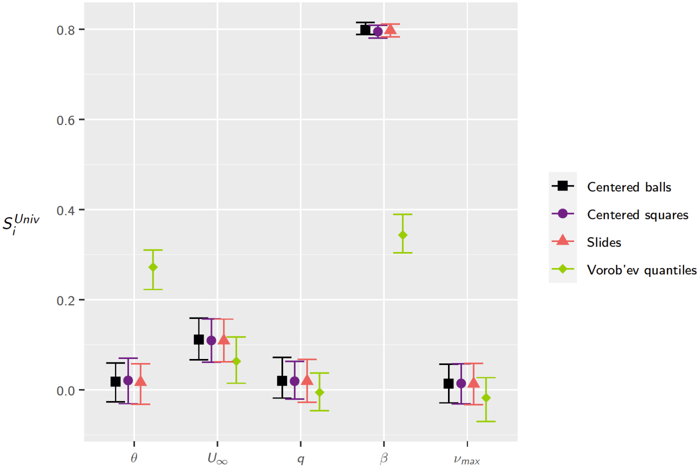


Figure: Estimation of the universal indices S_i^{Univ} for each input and for four different test sets and $N_a = 100$, with 1000 model evaluations. Confidence intervals are obtained with 100 bootstrap samples

Table of Contents

From Maximum Mean Discrepancy to HSIC

Let $k_{\mathcal{Z}} : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ be a kernel defined on a measurable space $\mathcal{Z}$.

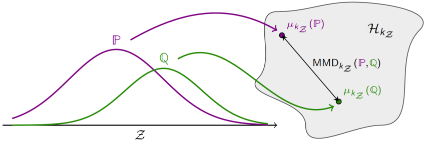


Figure: Kernel mean embedding

$$\mu_{k_{\mathcal{Z}}}(\mathbb{P}) = \mathbb{E}_{X\sim\mathbb{P}}[k_{\mathcal{Z}}(X,\cdot)]$$

From Maximum Mean Discrepancy to HSIC

Let $k_{\mathcal{Z}} : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ be a kernel defined on a measurable space $\mathcal{Z}$.

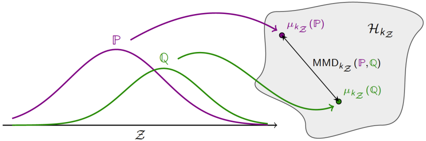


Figure: Kernel mean embedding

$$\mu_{k_{\mathcal{Z}}}(\mathbb{P}) = \mathbb{E}_{X\sim\mathbb{P}}[k_{\mathcal{Z}}(X,\cdot)]$$

$$\mathrm{MMD}^2_{k_{\mathcal{Z}}}(\mathbb{P},\mathbb{Q}) = \|\mu_{k_{\mathcal{Z}}}(\mathbb{P}) - \mu_{k_{\mathcal{Z}}}(\mathbb{Q})\|^2_{\mathcal{H}_{k_{\mathcal{Z}}}} = \mathbb{E}[k_{\mathcal{Z}}(X,X')] + \mathbb{E}[k_{\mathcal{Z}}(Y,Y')] - 2\,\mathbb{E}[k_{\mathcal{Z}}(X,Y)]$$

where $X \perp X' \sim \mathbb{P}$, $Y \perp Y' \sim \mathbb{Q}$

From Maximum Mean Discrepancy to HSIC

Let $k_{\mathcal{Z}} : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ be a kernel defined on a measurable space $\mathcal{Z}$.

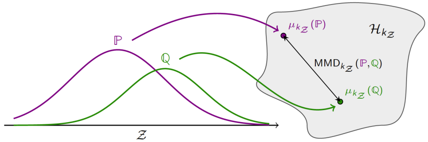


Figure: Kernel mean embedding

$$\mu_{k_{\mathcal{Z}}}(\mathbb{P}) = \mathbb{E}_{X\sim\mathbb{P}}[k_{\mathcal{Z}}(X,\cdot)]$$

$$\mathrm{MMD}^2_{k_{\mathcal{Z}}}(\mathbb{P},\mathbb{Q}) = \|\mu_{k_{\mathcal{Z}}}(\mathbb{P}) - \mu_{k_{\mathcal{Z}}}(\mathbb{Q})\|^2_{\mathcal{H}_{k_{\mathcal{Z}}}} = \mathbb{E}[k_{\mathcal{Z}}(X,X')] + \mathbb{E}[k_{\mathcal{Z}}(Y,Y')] - 2\,\mathbb{E}[k_{\mathcal{Z}}(X,Y)]$$

where $X \perp X' \sim \mathbb{P}$, $Y \perp Y' \sim \mathbb{Q}$

$\mathrm{MMD}_{k_{\mathcal{Z}}}(\mathbb{P},\mathbb{Q})$ distance $\iff$ $k_{\mathcal{Z}}$ caracteristic (mean embedding injective)

From Maximum Mean Discrepancy to HSIC

CENTRALE LYON

Hilbert Schmidt Independence Criterion (HSIC), Gretton et al. 2006

$$\begin{aligned}\mathrm{HSIC}_{K_i}(U_i, Y) &= \mathrm{MMD}^2_{K_i}\left(\mathbb{P}_{(U_i,Y)}, \mathbb{P}_{U_i} \otimes \mathbb{P}_Y\right) \\ &= \left\|\mu_{K_i}(\mathbb{P}_{(U_i,Y)}) - \mu_{k_{U_i}}(\mathbb{P}_{U_i}) \otimes \mu_{k_Y}(\mathbb{P}_Y)\right\|^2_{\mathcal{H}_{K_i}}\end{aligned}$$

où $K_i = k_{U_i} \otimes k_Y$.

Proposition (case of caracteristic kernels)

$$\mathrm{HSIC}_{K_i}(U_i, Y) = 0 \iff U_i \perp Y$$

HSIC for random sets

Issue

How to define a kernel for sets noted k_{set} ?

$$HSIC_{K_i \otimes k_{set}}(U_i, \Gamma) = ||\mu_{K_i \otimes k_{set}}(\mathbb{P}_{(U_i,\Gamma)}) - \mu_{K_i \otimes k_{set}}(\mathbb{P}_{U_i} \otimes \mathbb{P}_{\Gamma})||^2_{\mathcal{H}_{K_i \otimes k_{set}}}$$

HSIC for random sets

Issue

How to define a kernel for sets noted k_{set} ?

$$HSIC_{K_i \otimes k_{set}}(U_i, \Gamma) = ||\mu_{K_i \otimes k_{set}}(\mathbb{P}_{(U_i,\Gamma)}) - \mu_{K_i \otimes k_{set}}(\mathbb{P}_{U_i} \otimes \mathbb{P}_{\Gamma})||^2_{\mathcal{H}_{K_i \otimes k_{set}}}$$

Proposition (Fellmann et al. 2023)

The kernel k_{set} defined by:

$$k_{set}(\gamma_1, \gamma_2) = \exp\left(-\frac{\lambda(\gamma_1 \Delta \gamma_2)}{2\sigma^2}\right)$$

is characteristic.

HSIC for random sets

Issue

How to define a kernel for sets noted k_{set} ?

$$HSIC_{K_i \otimes k_{set}}(U_i, \Gamma) = ||\mu_{K_i \otimes k_{set}}(\mathbb{P}_{(U_i,\Gamma)}) - \mu_{K_i \otimes k_{set}}(\mathbb{P}_{U_i} \otimes \mathbb{P}_\Gamma)||^2_{\mathcal{H}_{K_i \otimes k_{set}}}$$

Proposition (Fellmann et al. 2023)

The kernel k_{set} defined by:

$$k_{set}(\gamma_1, \gamma_2) = \exp\left(-\frac{\lambda(\gamma_1 \Delta \gamma_2)}{2\sigma^2}\right)$$

is characteristic.

HSIC decomposition

$$S_i^{H_{set}} := \frac{HSIC^{ANOVA}(U_i, \Gamma)}{HSIC^{ANOVA}(\mathbf{U}, \Gamma)} = \frac{\mathbb{E}\left[(K_i(U_i, U_i') - 1)k_{set}(\Gamma, \Gamma')\right]}{\mathbb{E}\left[(K(\mathbf{U}, \mathbf{U}') - 1)k_{set}(\Gamma, \Gamma')\right]}$$

Results

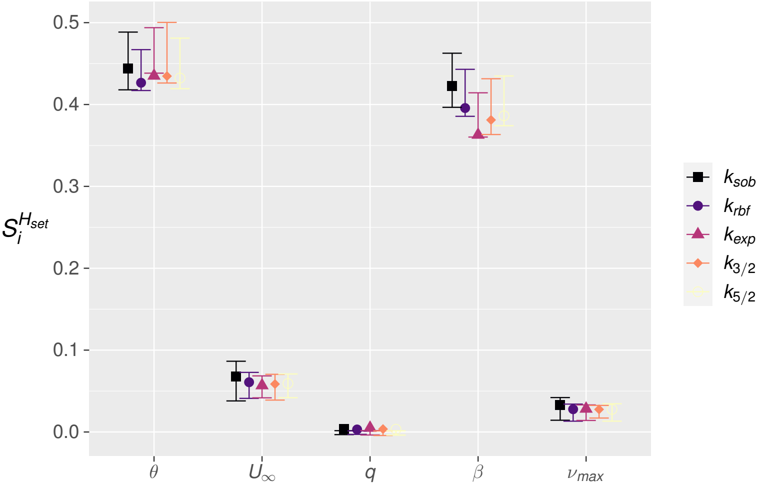


Figure: Estimation of $S_i^{\mathrm{H}_{set}}$ for five input kernels, 1000 model evaluations. Confidence intervals are obtained by bootstrap with 100 resamples

Table of Contents

Comparison

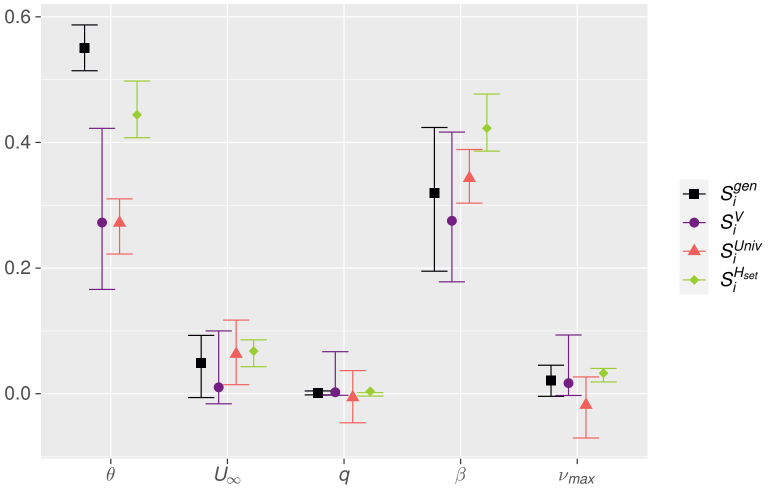


Figure: Comparison of the four indices with a total budget of n = 1000 model evaluations. 100 bootstrap sample are used to estimate confidence intervals

Comparison

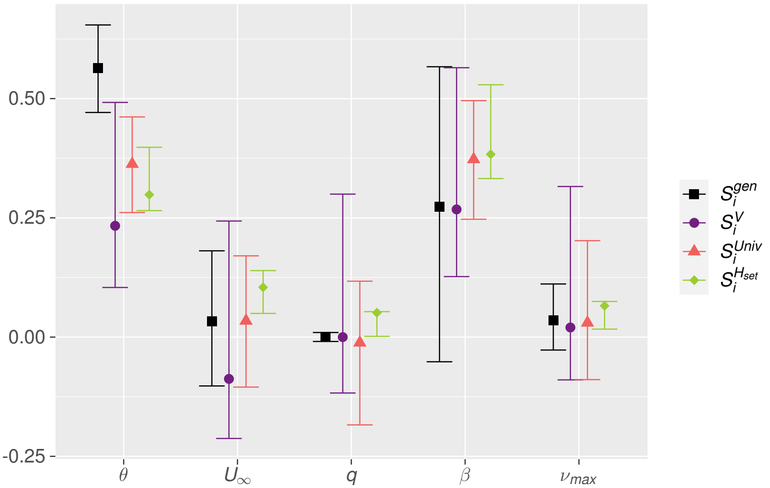


Figure: Comparison of the four indices with a total budget of $n = 100$ model evaluations. 100 bootstrap sample are used to estimate confidence intervals

Comparison and conclusion

	Ranking	Screening	Evaluations	Limitations
S_i^{gen}	✓ ANOVA	$\sim$ threshold	$\sim (p+1)n$	✗ pointwise influence
S_i^{V}	✗ no decomposition	✗ no screening method	✗ n^2 (double loop)	✓ no choice to be made
S_i^{Univ}	✓ ANOVA	$\sim$ threshold	$\sim$ n but big confidence intervals	✗ choice of the test sets
$S_i^{H_{set}}$	✓ ANOVA	✓ independence test	✓ n and small confidence intervals	$\sim$ choice of kernel ✗ Interactions interpretations

- HSIC-based indices seem to be the most efficient except if we are interested in the interactions between the inputs
- Methodology presented for map-valued outputs but can be used for any set-valued outputs

CENTRALE LYON

Thank you for your attention !

References I

- Borgonovo, E. (2007). “A new uncertainty importance measure”. In: *Reliability Engineering and System Safety* 92.6, pp. 771–784. ISSN: 0951-8320. DOI: https://doi.org/10.1016/j.ress.2006.04.015. URL: https://www.sciencedirect.com/science/article/pii/S0951832006000883.
- El-Amri, Reda et al. (2021). *A sampling criterion for constrained Bayesian optimization with uncertainties*. arXiv: 2103.05706 [stat.ML].
- Fellmann, Noé et al. (2023). *Kernel-based sensitivity analysis for (excursion) sets*. arXiv: 2305.09268 [math.ST].
- Fort, Jean-Claude, Thierry Klein, and Agnès Lagnoux (2021). “Global Sensitivity Analysis and Wasserstein Spaces”. In: *SIAM/ASA Journal on Uncertainty Quantification* 9.2, pp. 880–921. DOI: 10.1137/20M1354957. eprint: https://doi.org/10.1137/20M1354957. URL: https://doi.org/10.1137/20M1354957.

References II

- Gamboa, Fabrice, Thierry Klein, and Agnès Lagnoux (2018). "Sensitivity Analysis Based on Cramér–von Mises Distance". In: *SIAM/ASA Journal on Uncertainty Quantification* 6.2, pp. 522–548. DOI: 10.1137/15M1025621. eprint: https://doi.org/10.1137/15M1025621. URL: https://doi.org/10.1137/15M1025621.
- Gamboa, Fabrice et al. (2022). "Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics". In: *Bernoulli* 28.4, pp. 2345–2374. DOI: 10.3150/21-BEJ1421. URL: https://doi.org/10.3150/21-BEJ1421.
- Gretton, Arthur et al. (2005). "Measuring Statistical Dependence with Hilbert-Schmidt Norms". In: *Algorithmic Learning Theory*. Ed. by Sanjay Jain, Hans Ulrich Simon, and Etsuji Tomita. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 63–77. ISBN: 978-3-540-31696-1.
- Molchanov, Ilya (2005). *Theory of random sets*. Vol. 19. 2. Springer.

References III

Spagnol, Adrien (July 2020). "Kernel-based sensitivity indices for high-dimensional optimization problems". Theses. Université de Lyon. URL: https://tel.archives-ouvertes.fr/tel-03173192.

Estimation of Vorob'ev indices

$$\tilde{S}_i^V = 1 - \frac{\sum_{j=1}^n \hat{\lambda}_m(\Gamma^{(j)} \Delta \hat{Q}^i_{0.5}(U_i^{(j)}))}{\sum_{j=1}^n \hat{\lambda}_m(\Gamma^{(j)} \Delta \hat{Q}_{0.5})}.$$

where

- $\Gamma^{(j)} = \Psi(U^{(j)})$
- $\Gamma_l(U_i^{(j)}) = \Psi(U_i^{(j)}, U_{-i}^{(j,l)})$
- $\hat{Q}^i_{0.5}(U_i^{(j)}) := \{x \in \mathcal{X}, \frac{1}{n} \sum_{l=1}^n \mathbf{1}_{\Gamma_l(U_i^{(j)})}(x) \geq 0.5\}$
- $\hat{\lambda}_m(A) = \frac{1}{m} \sum_{k=1}^m \mathbf{1}_A(\mathbf{x}^{(l)})$.

Estimation of universal indices (rank-based estimation)

The estimator $\widehat{S}_i^{\text{Univ}}$ of S_i^{Univ} is given by the ratio between

$$\widehat{S}_{num} = \frac{1}{N_a}\sum_{l=1}^{N_a}\left[\frac{1}{n}\sum_{j=1}^{n}\left(\hat{\lambda}_m(\Gamma^{(j)}\Delta\gamma_{A_l})\right)\left(\hat{\lambda}_m(\Gamma^{(N_i(j))}\Delta\gamma_{a_l})\right)\right] - \frac{1}{N_a}\sum_{l=1}^{N_a}\left[\frac{1}{n}\sum_{j=1}^{n}\left(\hat{\lambda}_m(\Gamma^{(j)}\Delta\gamma_{a_l})\right)\right]^2$$

and,

$$\widehat{S}_{den} = \frac{1}{N_a}\sum_{l=1}^{N_a}\left[\frac{1}{n}\sum_{j=1}^{n}\left(\hat{\lambda}_m(\Gamma^{(j)}\Delta\gamma_{a_l})\right)^2\right] - \frac{1}{N_a}\sum_{l=1}^{N_a}\left[\frac{1}{n}\sum_{j=1}^{n}\left(\hat{\lambda}_m(\Gamma^{(j)}\Delta\gamma_{a_l})\right)\right]^2,$$

where (a_l) is an iid sample of the law $\mathbb{Q}$ and $N_i(j)$ is the index in the sample (U_i^l) that comes after U_i^j when (U_i^l) is sorted in ascending order (see Gamboa et al. 2022 for details).

Estimation of HSIC

The indices can be estimated using:

$$\widehat{\mathrm{H}_{set}}(U_i, \Gamma) = \frac{2}{n(n-1)} \sum_{j<l}^{n} \left(K_i \left(U_i^{(j)}, U_i^{(l)} \right) - 1 \right) \exp\left(-\frac{\lambda(\mathcal{X})}{2\sigma^2} \hat{\lambda}_m(\Gamma^{(j)} \Delta \Gamma^{(l)})\right).$$

Input kernels:

- the Sobolev kernel of order 1, $k_{sob}(x, y) = 1 + (x - \frac{1}{2})(y - \frac{1}{2}) + \frac{1}{2}[(x - y)^2 - |x - y| + \frac{1}{6}]$
- the Gaussian kernel, $k_{rbf}(x, y) = e^{-\frac{1}{2}\left(\frac{x-y}{\sigma}\right)^2}$ with $\sigma > 0$,
- the Laplace kernel, $k_{exp}(x, y) = e^{-\frac{|x-y|}{h}}$ with $h > 0$,
- the Matérn 3/2, $k_{3/2}(x, y) = \left(1 + \sqrt{3}\frac{|x-y|}{h}\right) e^{-\sqrt{3}\frac{|x-y|}{h}}$ with $h > 0$,
- the Matérn 5/2, $k_{5/2}(x, y) = \left(1 + \sqrt{5}\frac{|x-y|}{h} + \frac{5}{3}\frac{|x-y|}{h^2}\right) e^{-\sqrt{5}\frac{|x-y|}{h}}$ with $h > 0$.