Sensitivity analysis via Sobol' indices, rank-based estimation and beyond

Agnès Lagnoux
Institut de Mathématiques de Toulouse
TOULOUSE - FRANCE

CIROQUO, July 2th 2021

OUTLINE OF THE LECTURES

Part I : From Sobol' indices to universal indices

Part II : Stochastic computer codes and an introduction to second-level sensitivity analysis

Part I
From Sobol' indices to universal indices

Outline of the talk

Framework

We consider a complicated regression function f defined on $E = E_1 \times E_2 \times \cdots \times E_p$ and valued in $\mathbb{R}^k$ depending on several variables :

$$y = f(x_1, \ldots, x_p), \tag{1}$$

where

1. the inputs x_i pour $i = 1, \ldots p$ are objects ;
2. f is deterministic and unknown. It is called a black-box.

Probabilistic frame

In order to quantify the influence of a variable, it is common to assume that the inputs are random :

$$X := (X_1, \ldots, X_p) \in E = E_1 \times \ldots \times E_p.$$

Then $f : E \to \mathbb{R}^k$ is a measurable function that can be evaluated on runs and the output code Y becomes random too :

$$Y = f(X_1, \ldots, X_p).$$

In this presentation, the inputs X_i are assumed to be mutually independent.

The so-called Sobol' indices

Classically to quantify the amount of randomness that a variable or a group of variables bring to Y, one computes the so-called Sobol' indices.

For instance, the first order Sobol' index with respect to $X_{\mathbf{u}} = (X_i, i \in \mathbf{u})$ is given by

$$S^{\mathbf{u}} = \frac{\mathrm{Var}(\mathbb{E}[Y|X_{\mathbf{u}}])}{\mathrm{Var}(Y)}$$

(assuming Y is scalar).

Such indices stem from the Hoeffding decomposition of the variance of f (or equivalently Y) that is assumed to lie in L^2.

Pick-Freeze estimation of Sobol' indices (I)

To fix ideas assume for example $p = 5$, $\mathbf{u} = \{1, 2\}$ so that $\sim \mathbf{u} = \{3, 4, 5\}$.

We consider the Pick-Freeze variable $Y_{\mathbf{u}}$ defined as follows :

- draw $X = (X_1, X_2, X_3, X_4, X_5)$,
- build $X_{\mathbf{u}} = (X_1, X_2, X_3', X_4', X_5')$.

Then, we compute

- $Y = f(X)$,
- $Y_{\mathbf{u}} = f(X_{\mathbf{u}})$.

A small miracle

$$\mathrm{Var}(\mathbb{E}[Y|X_{\mathbf{u}}]) = \mathrm{Cov}(Y, Y_{\mathbf{u}}). \text{ So that } S^{\mathbf{u}} = \frac{\mathrm{Cov}\,(Y, Y_{\mathbf{u}})}{\mathrm{Var}(Y)}.$$

Pick-Freeze estimation of Sobol' indices (II)

In practice Generate two N-samples.

- One N-sample of X : $\left(X^i\right)_{i=1,\dots,N}$.
- One N-sample of $X_{\mathbf{u}}$: $\left(X_{\mathbf{u}}^i\right)_{i=1,\dots,N}$.

Compute the code on both samples :

- $Y^i = f(X^i)_{i=1,\dots,N}$
- $Y_{\mathbf{u}}^i = f(X_{\mathbf{u}}^i)_{i=1,\dots,N}$.

Then estimate $S^{\mathbf{u}}$ by

$$S_{N,PF}^{\mathbf{u}} = \frac{\frac{1}{N}\sum Y^i Y_{\mathbf{u}}^i - \left(\frac{1}{N}\sum Y^i\right)\left(\frac{1}{N}\sum Y_{\mathbf{u}}^i\right)}{\frac{1}{N}\sum (Y^i)^2 - \left(\frac{1}{N}\sum Y^i\right)^2}$$

Pick-Freeze scheme (III) : some statistical questions

Is the Pick-Freeze estimator a "good" estimator of the Sobol' index ?

- Is it consistent ? Response : YES SLLN.
- If yes, at which rate of convergence ? Res. : YES CLT (cv in $\sqrt{N}$).
- Is it asymptotically efficient ? Resp. : YES.
- Is it possible to measure its performance for a fixed N ? Response : YES Berry-Esseen and/or concentration inequalities.

Ref. : A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. " Asymptotic normality et efficiency of a Sobol' index estimator", *ESAIM P&S*, 2013.

F. Gamboa, A. Janon, T. Klein, A. Lagnoux, and C. Prieur. " Statistical Inference for Sobol' Pick Freeze Monte Carlo method", *Statistics*, 2015.

Pick-Freeze scheme (IV) : consistency and CLT

$$S^{\mathbf{u}}_{N,PF} = \frac{\frac{1}{N}\sum Y^i Y^i_{\mathbf{u}} - \left(\frac{1}{N}\sum Y^i\right)\left(\frac{1}{N}\sum Y^i_{\mathbf{u}}\right)}{\frac{1}{N}\sum (Y^i)^2 - \left(\frac{1}{N}\sum Y^i\right)^2}, \quad S^{\mathbf{u}} = \frac{\mathrm{Var}\left(\mathbb{E}\left[Y|X_{\mathbf{u}}\right]\right)}{\mathrm{Var}(Y)}.$$

Theorem (Janon, Klein, Lagnoux, Nodet, Prieur (2015))

1. *One has* $S^{\mathbf{u}}_{N,PF} \xrightarrow[N\to\infty]{a.s.} S^{\mathbf{u}}$.
2. *If* $\mathbb{E}[Y^4] < \infty$, *then*

$$\sqrt{N}\left(S^{\mathbf{u}}_{N,PF} - S^{\mathbf{u}}\right) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}_1\left(0, \sigma^2_S\right)$$

where $\sigma^2_S = \frac{\mathrm{Var}((Y-\mathbb{E}[Y])[(Y^u-\mathbb{E}[Y])-S^{\mathbf{u}}(Y-\mathbb{E}[Y])])}{(\mathrm{Var}(Y))^2}$.

Pick-Freeze scheme (V) : concentration inequality

The Central Limit Theorem is a limit result. In real life, the number of experiments is finite. Concentration inequalities allow to quantify the error between the estimate and the index true value for a fixed value of N.

Using soundly Bennett inequality, one gets

Proposition (Gamboa, Janon, Klein, Lagnoux, Prieur (2015))

Let $\mathbf{u}$ *be a subset of* $\{1, \dots, p\}$. *Then,*

$$\mathbb{P}\left(|S_N^{\mathbf{u}} - S^{\mathbf{u}}| \geqslant t\right) \leqslant 2\exp\left(-\frac{N\mathrm{Var}(Y)^2}{128}\left(1-\frac{1}{N}\right)^2\left(\frac{t}{3+2t}\right)^2\right).$$

Extensions

Multidimensional and functional outputs

F. Gamboa, A. Janon, T. Klein, and A. Lagnoux. “Sensitivity analysis for multidimensional and functional outputs”. *Electron. J. Stat*, (2014). Volume 8, no. 1, pp 575–603.

Drawbacks of the Pick-Freeze estimation

- The cost (=number of evaluations of the function f) of the estimation of the p first-order Sobol' indices is quite expensive : $(p+1)N$.
- This methodology is based on a particular design of experiment that may not be available in practice. For instance, when the practitioner only has access to real data.

$\Rightarrow$ *We are then interested in an estimator based on a N-sample only.*

Mighty estimation based on ranks (I)

Here we assume that the inputs X_i for $i = 1, \ldots, p$ are scalar and we want to estimate the Sobol' index S^1 with respect to X_1 :

$$S^1 = \frac{\mathrm{Var}\left(\mathbb{E}[Y|X_1]\right)}{\mathrm{Var}(Y)}$$

To do so, we consider a N-sample of the input/output pair (X_1, Y) given by

$$(X_1^1, Y_1), (X_1^2, Y_2), \ldots, (X_1^N, Y_N).$$

Mighty estimation based on ranks (II)

The pairs $(X_1^{(1)}, Y_{(1)}), (X_1^{(2)}, Y_{(2)}), \ldots, (X_1^{(N)}, Y_{(N)})$ are rearranged in such a way that

$$X_1^{(1)} < \ldots < X_1^{(N)}.$$

Example

- $N = 6$
- Original sample $(1, 5), (2, 9), (-2, 3), (6, -4), (0, 8)$
- Rearranged sample $(-2, 3), (0, 8), (1, 5), (2, 9), (6, -4)$.

Ref. : S. Chatterjee. "A new coefficient of Correlation", *JASA*, 2020.

F. Gamboa, P. Gremaud, T. Klein, and A. Lagnoux. " Global Sensitivity Analysis : a new generation of mighty estimators based on rank statistics", *Preprint Arxiv.* 2021.

Mighty estimation based on ranks (III)

We introduce

$$S^1_{N,Rank} = \frac{\frac{1}{N}\sum_{i=1}^{N-1} Y_{(i)} Y_{(i+1)} - \left(\frac{1}{N}\sum_{i=1}^{N} Y_i\right)^2}{\frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} Y_i\right)^2}.$$

Theorem (Gamboa, Gremaud, Klein, Lagnoux, 2021)

1. *One has* $S^1_{N,Rank} \xrightarrow[N\to\infty]{a.s.} S^1$.
2. *If the* X_i*'s are uniformly distributed and under some mild assumptions on* f, *then*

$$\sqrt{N}\left(S^1_{N,Rank} - S^1\right) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}_1\left(0, \sigma_R^2\right).$$

A real data example (I)

When designing a future aircraft, the manufacturer needs to satisfy the so-called TLAR="top level aircraft requirements" that summarize the expected performance of the future aircraft.

One important task is to identify the TLARS that influence the most the operating cost of an aircraft.

This example is borrowed from

- Peteilh, N., Klein, T., Druot, T. Y., Bartoli, N., & Liem, R. P. (2020). Challenging Top Level Aircraft Requirements based on operations analysis and data-driven models, application to takeoff performance design requirements. In AIAA AVIATION 2020 FORUM (p. 3171).
- Marouane Felloussi's project for the computation of the estimators based on the rank's method

A real data example (II)

We restrict our study to 3 TLARS (input variables) :

1. TOFL="take of length" $\in [1500, 5000]$ in m.,
2. altp="altitude of the airport" $\in [0, 2500]$ in m.,
3. ΔT_{ISA}="delta of temperature" $\in [-30, 30]$ in °C,

and study their influence on 5 different costs (output variables) :

1. the block fuel,
2. the block time,
3. the cash operating cost,
4. the direct operating cost,
5. the total fuel=block fuel+reserves.

A real data example (III)

		Estimation method	
		P&F	Rank
Output names	Input names	Index values	
Block fuel	TOFL	70.70%	73.5%
	altp	13.48%	8.8%
	ΔT_{ISA}	14.75%	8.2%
Block time	TOFL	67.54%	69.4%
	altp	12.44%	6.5%
	ΔT_{ISA}	18.30%	19.6%
Cash operating cost	TOFL	70.70%	73.5%
	altp	13.48%	8.8%
	ΔT_{ISA}	14.75%	8.2%
Direct operating cost	TOFL	70.73%	73.5%
	altp	13.49%	8.8%
	ΔT_{ISA}	14.79%	8.2%
Total fuel	TOFL	70.76%	73.6%
	altp	13.41%	8.8%
	ΔT_{ISA}	14.55%	7.8%

Red-thread example : a non-linear model (I)

Let us consider the following non-linear model

$$Y = \exp\{X_1 + 2X_2\},$$

where X_1 and X_2 are independent standard Gaussian random variables. Then tedious computations lead to the Sobol' indices S^1 and S^2 :

$$S^1 = (e-1)/(e^5-1) \approx 0.0117$$
$$S^2 = (e^4-1)/(e^5-1) \approx 0.3636$$

Red-thread example : a non-linear model (II)

Code jupyter TP_Sob.ipynb

Motivation for a new index

Sobol' indices are based on a variance decomposition.

- They only quantify the influence around the mean.
- In practice, one may be interested in the median or even in a quantile rather than the mean.
- It may also occur (eg. symmetric function variables with identical two first moments) that the Sobol' indices are not suitable to discriminate the role of the inputs.

Toy example (I)

Let X_1 and X_2 be two independent random variables with distinct distributions sharing the four first moments. Consider

$$Y = X_1 + X_2 + X_1^2 X_2^2.$$

Then

$$\begin{aligned} \mathrm{Var}\left(\mathbb{E}\left[Y|X_1\right]\right) &= \mathrm{Var}(X_1 + X_1^2 \mathbb{E}\left[X_2^2\right]) \\ &= \mathrm{Var}(X_2 + X_2^2 \mathbb{E}\left[X_1^2\right]) = \mathrm{Var}\left(\mathbb{E}\left[Y|X_2\right]\right). \end{aligned}$$

Y is a symmetrical function of X_1, X_2 but if X_1 and X_2 have different distributions, X_1 and X_2 should act differently.

It seems important to consider sensitivity indices that take into account not only the two first moments but the whole distribution.

Outline of the talk

Construction of the Cramér-von Mises indices (I)

Let $Z = f(X_1, \dots, X_p) \in \mathbb{R}^k$ be the code output and F be its cumulative distribution function defined for $t = (t_1, \dots, t_k) \in \mathbb{R}^k$ by

$$F(t) = \mathbb{P}(Z \leqslant t) = \mathbb{E}\left[\mathbb{1}_{\{Z \leqslant t\}}\right] =: \mathbb{E}[Y(t)].$$

Let $F^{\mathbf{u}}(t)$ be the conditional cumulative distribution function (conditionally Z knowing $X_{\mathbf{u}}$) :

$$F^{\mathbf{u}}(t) = \mathbb{P}(Z \leqslant t | X_{\mathbf{u}}) = \mathbb{E}\left[\mathbb{1}_{\{Z \leqslant t\}} | X_{\mathbf{u}}\right] = \mathbb{E}[Y(t) | X_{\mathbf{u}}].$$

Construction of the Cramér-von Mises indices (II)

First, we perform the Hoeffding decomposition of $Y(t)$:

$$
\begin{aligned}
Y(t) = \mathbb{1}_{\{Z\leqslant t\}} = & \underbrace{\mathbb{E}[Y(t)]}_{\text{Mean effect}} \\
& + \underbrace{(\mathbb{E}[Y(t)|X_{\mathbf{u}}] - \mathbb{E}[Y(t)]) + (\mathbb{E}[Y(t)|X_{\sim\mathbf{u}}] - \mathbb{E}[Y(t)])}_{\text{First order effects}} \\
& + \underbrace{R(t,\mathbf{u})}_{\text{Second order effects or interaction}:=IA} .
\end{aligned}
$$

Construction of the Cramér-von Mises indices (III)

Second, we compute the variance of both sides of the previous equation :

$$\begin{aligned}\mathrm{Var}(Y(t)) =& \mathbb{E}\left[(F^{\mathbf{u}}(t) - F(t))^2\right] + \mathbb{E}\left[(F^{\sim\mathbf{u}}(t) - F(t))^2\right] \\ &+ \mathrm{Var}(R(t,\mathbf{u}))\end{aligned}$$

by the decorrelation of the different terms involved in the Hoeffding decomposition.

Construction of the Cramér-von Mises indices (IV)

Finally, it remains to integrate in $t \in \mathbb{R}^k$ with respect to the distribution of Z and to normalize to get :

$$S_{2,CVM}^{\mathbf{u}} := \frac{\int_{\mathbb{R}^k} \mathbb{E}\left[(F(t) - F^{\mathbf{u}}(t))^2\right] dF(t)}{\int_{\mathbb{R}^k} F(t)(1 - F(t))dF(t)},$$

involving the Cramér-von Mises distance between $\mathcal{L}(Z)$ and $\mathcal{L}(Z|X_{\mathbf{u}})$

Properties and remarks

These new indices share the same properties as the classical Sobol' indices, namely,

1. the different contributions sum to 1 ;
2. the indices are invariant by any translation, by any isometry, and by any nondegenerated scaling of the components of Y.

Despite the fact that the Cramér-von Mises indices have no clear dual formulation, our method represents at least three advantages :

1. the index always exists whatever the output distribution ;
2. such an integration weights the support of the output distribution ;
3. the index can be easily estimated using a Pick-Freeze scheme.

Estimation of the Cramér-von-Mises indices

1. First approach - Pick-Freeze estimation
2. Second approach - Pick-Freeze and U-stats
3. Third approach - Ranks

Second approach - Pick-Freeze estimation (I)

Principle :

- Multiple Monte-Carlo estimation procedure (one to handle the integration part, one to handle the Pick-Freeze part).
- Cost to estimate all first-order indices : $N(1+p+1)$.
- CLT OK.

First approach - Pick-Freeze estimation (II)

To fix ideas assume for example $p = 5$, $\mathbf{u} = \{1, 2\}$ so that $\sim \mathbf{u} = \{3, 4, 5\}$.
We consider the Pick-Freeze variable $Z^{\mathbf{u}}$ defined as follows :

- draw $X = (X_1, X_2, X_3, X_4, X_5)$,
- build $X_{\mathbf{u}} = (X_1, X_2, X_3', X_4', X_5')$.

Then, we compute

- $Z = f(X)$,
- $Z_{\mathbf{u}} = f(X_{\mathbf{u}})$.

First approach - Pick-Freeze estimation (III)

The estimation of the numerator $N^{\mathbf{u}}_{2,CVM}$ of $S^{\mathbf{u}}_{2,CVM}$ is based on

$$\begin{aligned} N^{\mathbf{u}}_{2,CVM} &= \int_{\mathbb{R}^k} \mathbb{E}\left[(F(t) - F^{\mathbf{u}}(t))^2\right] dF(t) \\ &= \mathbb{E}\left[\mathbb{E}\left[(F(W) - F^{\mathbf{u}}(W))^2\right]\right] \\ &= \mathbb{E}\left[\mathrm{Var}\left(\mathbb{E}\left[Y(W)|X_{\mathbf{u}}\right]\right)\right] \\ &= \mathbb{E}\left[\mathrm{Cov}\left(Y(W), Y_{\mathbf{u}}(W)\right)\right] \\ &= \mathbb{E}\left[\mathrm{Cov}\left(\mathbb{1}_{Z\leqslant W}, \mathbb{1}_{Z_{\mathbf{u}}\leqslant W}\right)\right] \end{aligned}$$

where W is an independent copy of Z.

First approach - Pick-Freeze estimation (IV)

Then the estimation stands on a double Monte Carlo : we generate

1. two N-samples of Z : $(Z_j^{\mathbf{u},1}, Z_j^{\mathbf{u},2})$, $1 \leqslant j \leqslant N$; (Pick-Freeze)
2. a third independent N-sample of Z : W_k, $1 \leqslant k \leqslant N$

resulting in

$$N_{2,CVM,PF}^{\mathbf{u}} = \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Z_j^{\mathbf{u},1}\leqslant W_k\}}\mathbb{1}_{\{Z_j^{\mathbf{u},2}\leqslant W_k\}} - \left[\frac{1}{2N}\sum_{j=1}^{N}\left(\mathbb{1}_{\{Z_j^{\mathbf{u},1}\leqslant W_k\}}+\mathbb{1}_{\{Z_j^{\mathbf{u},2}\leqslant W_k\}}\right)\right]^2\right\}.$$

First approach - Pick-Freeze estimation (V)

- Low estimation cost.
- Whatever the dimension of the output.
- The sample required for their estimation also provides a Sobol' indices estimation.

Theorem (Gamboa, Klein, Lagnoux (2018))

$S^{\mathbf{u}}_{2,CVM,PF}$ is strongly convergent as N goes to infinity.
If $\mathbb{E}[\|Z\|^4] < +\infty$, the sequence $S^{\mathbf{u}}_{2,CVM,PF}$ is asymptotically Gaussian. More precisely, $\sqrt{N}\left(S^{\mathbf{u}}_{2,CVM,PF} - S^{\mathbf{u}}_{2,CVM}\right)$ converge in law to a centered Gaussian variable with explicit variance.

Ref. : F. Gamboa, T. Klein, and A. Lagnoux. "Sensitivity analysis based on Cramér-von Mises distance ", *SIAM UQ*, 2018.

Second approach - U-statistics estimation (I)

Principle :

- Dealing simultaneously with the Sobol' part and the integration part to get rid of the additional N-sample $(W_k)_{1\leqslant k\leqslant N}$.
- Cost to estimate all first-order indices : $N(p+1)$.
- Elementary proof of the CLT using a CLT for U-stats (Hoeffding 1948) and the classical delta method.

Second approach - U-statistics estimation (II)

It suffices to rewrite $S^{\mathbf{u}}_{2,\mathrm{CVM}}$ as

$$S^{\mathbf{u}}_{2,\mathrm{CVM}} = \frac{I(\Phi_1) - I(\Phi_2)}{I(\Phi_3) - I(\Phi_4)},$$

where $m(1) = m(3) = 2$, $m(2) = m(4) = 3$,

$$\begin{aligned}
\Phi_1(\mathbf{z}_1, \mathbf{z}_2) &= \mathbb{1}_{\{z_2 \leqslant z_1\}} \mathbb{1}_{\{z_2^{\mathbf{u}} \leqslant z_1\}} \\
\Phi_2(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3) &= \mathbb{1}_{\{z_2 \leqslant z_1\}} \mathbb{1}_{\{z_3^{\mathbf{u}} \leqslant z_1\}} \\
\Phi_3(\mathbf{z}_1, \mathbf{z}_2) &= \mathbb{1}_{\{z_2 \leqslant z_1\}} \\
\Phi_4(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3) &= \mathbb{1}_{\{z_2 \leqslant z_1\}} \mathbb{1}_{\{z_3 \leqslant z_1\}}
\end{aligned}$$

denoting by $\mathbf{z}_i$ the pair $(z_i, z_i^{\mathbf{u}})$ and, for $j = 1, \dots, 4$,

$$I(\Phi_j) = \int_{\mathbb{R}^k} \Phi_j(\mathbf{z}_1, \dots, \mathbf{z}_{m(j)}) d\mathbb{P}_2^{u, \otimes m(j)}(\mathbf{z}_1 \dots, \mathbf{z}_{m(j)}).$$

Second approach - U-statistics estimation (III)

Finally, one considers the empirical version of $S^{\mathbf{u}}_{2,\mathrm{CVM}}$:

$$S^{\mathbf{u}}_{2,CVM,Ustat} = \frac{U_{1,N} - U_{2,N}}{U_{3,N} - U_{4,N}},$$

where, for $j = 1, \ldots, 4$,

$$U_{j,N} = \binom{N}{m(j)}^{-1} \sum_{1 \leqslant i_1 < \cdots < i_{m(j)} \leqslant N} \Phi^s_j \left(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_{m(j)}} \right)$$

and the function :

$$\Phi^s_j(\mathbf{z}_1, \ldots, \mathbf{z}_{m(j)}) = \frac{1}{(m(j))!} \sum_{\tau \in \mathcal{S}_{m(j)}} \Phi_j(\mathbf{z}_{\tau(1)}, \ldots, \mathbf{z}_{\tau(m(j))})$$

is the symmetrized version of Φ_j.

Second approach - U-statistics estimation (IV)

The estimator $S^{\mathbf{u}}_{2,\mathrm{CVM},\mathrm{Ustat}}$ has been proved to be consistent and asymptotically Gaussian.

Theorem (Gamboa, Klein, Lagnoux, Moreno (2021))

If for $j = 1, \ldots, 4$, $\mathbb{E}\left[\Phi^s_j\left(\mathbf{Z}_1, \ldots, \mathbf{Z}_{m(j)}\right)^2\right] < \infty$ then

$$\sqrt{N}\left(S^{\mathbf{u}}_{2,CVM,Ustat} - S^{\mathbf{u}}_{2,CVM}\right) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}_1(0, \sigma^2)$$

where the asymptotic variance σ^2 is explicitly known.

Ref. : F. Gamboa, T. Klein, A. Lagnoux, and L. Moreno. "Sensitivity analysis in general metric spaces", *RESS*, 2021.

Third approach - Rank-based estimation (I)

Principle :

- Only when the inputs are scalar and to estimate the first-order indices.
- Cost to estimate all first-order indices : N.
- CLT in progress.

Let $\pi_i(j)$ be the rank of X_i^j in the sample $(X_i^1, \ldots, X_i^N)$ of X_i and define

$$N_i(j) = \begin{cases} \pi_i^{-1}(\pi_i(j)+1) & \text{if } \pi_i(j)+1 \leqslant N, \\ \pi_i^{-1}(1) & \text{if } \pi_i(j) = N. \end{cases}$$

Third approach - Rank-based estimation (II)

Then the empirical estimator $S^i_{2,\mathrm{CVM,Rank}}$ of $S^i_{2,\mathrm{CVM}}$ is given by the ratio between

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\left[\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Z_j\leqslant Z_i\}}\mathbb{1}_{\{Z_{N(j)}\leqslant Z_i\}}\right]-\left[\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Z_j\leqslant Z_i\}}\right]^2\right\}$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\left[\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Z_j\leqslant Z_i\}}\right]-\left[\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Z_j\leqslant Z_i\}}\right]^2\right\}.$$

<ins>Ref.</ins> : F. Gamboa, P. Gremaud, T. Klein, and A. Lagnoux. " Global Sensitivity Analysis : a new generation of mighty estimators based on rank statistics", *Preprint Arxiv.* 2021.

A real data example : giant cell arthritis (I)

"*giant cell arthritis (GCA) is a vasculitis of unknown etiology that affects large and medium sized vessels and occurs almost exclusively in patients 50 years or older*".

- This disease may lead to severe side effects (loss of visual accuity, fever, headache,...). The risks of not treating it include the threat of blindness and major vessels occlusion.
- A patient with suspected GCA can receive a therapy based on Prednisone. Unfortunately, a treatment with high Prednisone doses may cause severe complications.

A real data example : giant cell arthritis (II)

When confronted to a patient with suspected GCA, the clinician must adopt a strategy among :

- A : Treat none of the patients ;
- B : Proceed to the biopsy and treat all the positive patients ;
- C : Proceed to the biopsy and treat all the patients whatever their result ;
- D : Treat all the patients.

optimizing the patient outcomes measured in terms of utility. The basic idea is that a patient with perfect health is assigned a utility of 1 and the expected utility of the other patients (not perfectly healthy) is calculated subtracting some “disutilities” from this perfect score of 1.

A real data example : giant cell arthritis (III)

The base value of some input parameters are reliable while the others are really uncertain that leads us to consider them as random.

As a consequence, if Y_A, Y_B, Y_C and Y_D represent the outcomes corresponding to the four different strategies A to D, the clinician aims to determine

$$\max\{\mathbb{E}[Y_A], \mathbb{E}[Y_B], \mathbb{E}[Y_C], \mathbb{E}[Y_D]\}.$$

A real data example : giant cell arthritis (IV)

	Sensitivity meas.	Ranking	CPU time
$N = 10^2$	Multivariate	1 6 2 3 5 7 4	0.0624
	Borgonovo *et al.*	1 3 6 2 5 7 4	1.5132
	Cramér-von Mises	1 6 2 3 7 5 4	0.9048
$N = 10^3$	Multivariate	1 6 2 3 7 5 4	0.0156
	Borgonovo *et al.*	1 6 2 5 7 3 4	57.8452
	Cramér-von Mises	1 6 2 3 7 5 4	10.1089
$N = 10^4$	Multivariate	1 6 2 3 7 5 4	0.0312
	Borgonovo *et al.*	1 6 2 7 3 5 4	$5.1988\ 10^3$
	Cramér-von Mises	1 6 2 3 7 5 4	436.8028

Red-thread example : a non-linear model (I)

Let us consider the following non linear model

$$Y = \exp\{X_1 + 2X_2\},$$

where X_1 and X_2 are independent standard Gaussian random variables. Then tedious computations lead to the Cramér-von-Mises indices $S^1_{2,CVM}$ and $S^2_{2,CVM}$:

$$S^1_{2,CVM} = \frac{6}{\pi} \arctan 2 - 2 \approx 0.1145$$
$$S^2_{2,CVM} = \frac{6}{\pi} \arctan \sqrt{19} - 2 \approx 0.5693.$$

Red-thread example : a non-linear model (II)

Code TP_CVM.ipynb

Outline of the talk

Framework

We consider a family of test functions parametrized by m elements of $\mathcal{X}$ with $m \in \mathbb{N}^*$. For any $a = (a_i)_{i=1,\ldots,m} \in \mathcal{X}^m$, we consider the test functions

$$\begin{array}{ccc} \mathcal{X}^m \times \mathcal{X} & \to & \mathbb{R} \\ (a, x) & \mapsto & T_a(x). \end{array}$$

We assume that $T_a(\cdot) \in L^2(\mathbb{P}^{\otimes m} \otimes \mathbb{P})$ where $\mathbb{P}$ denotes the distribution of Z.

Definition of the general metric space index

Recall the expression of the Cramér-von-Mises index

$$S^{\mathbf{u}}_{2,CVM} = \frac{\int_{\mathbb{R}^k} \mathbb{E}_{X^{\mathbf{u}}}\left[\left(\mathbb{E}_Z[\mathbb{1}_{Z\leqslant t}] - \mathbb{E}_Z[\mathbb{1}_{Z_{\mathbf{u}}\leqslant t}]\right)^2\right] dF(t)}{\int_{\mathbb{R}^k} F(t)(1-F(t))dF(t)},$$

where $\mathbb{E}_U$ stands for the expectation with respect to the r.v. U.

The general metric space sensitivity index with respect to $\mathbf{u}$ is defined by

$$S^{\mathbf{u}}_{2,GMS} := \frac{\int_{\mathcal{X}^m} \mathbb{E}_{X_{\mathbf{u}}}\left[\left(\mathbb{E}_Z[T_a(Z)] - \mathbb{E}_Z[T_a(Z)|X_{\mathbf{u}}]\right)^2\right] d\mathbb{P}^{\otimes m}(a)}{\int_{\mathcal{X}^m} \mathrm{Var}(T_a(Z))d\mathbb{P}^{\otimes m}(a)}.$$

Definition of the general metric space index

By construction, $S^{\mathbf{u}}_{2,GMS} \in [0,1]$ and

1. the different contributions sum to 1 ;
2. the indices are invariant by any translation, by any isometry and by any non-degenerated scaling of the components of Z.

Particular examples

1. For $\mathcal{X} = \mathbb{R}$, $m = 0$ and T_a given by $T_a(x) = x$, one recovers the classical Sobol' indices.
2. For $\mathcal{X} = \mathbb{R}^k$ and $m = 0$, one can recover the Sobol' indices for vectorial outputs in Gamboa *et al*. and Lamboni *et al*.
3. For $\mathcal{X} = \mathbb{R}^k$, $m = 1$ and T_a given by $T_a(x) = \mathbb{1}_{\{x \leqslant a\}}$, one recovers the index based on the Cramér-von-Mises distance.
4. Consider that $\mathcal{X} = \mathcal{M}$ is a manifold, $m = 2$ and T_a is given by

$$T_a(x) = \mathbb{1}_{\{x \in \widetilde{B}(a_1, a_2)\}} = \mathbb{1}_{\{\|x-(a_1+a_2)/2\| \leqslant \|a_1-a_2\|/2\}},$$

where $\widetilde{B}(a_1, a_2)$ will stand for the ball of diameter $\overline{a_1 a_2}$. One recovers the indices defined in Fraiman *et al*.

Estimation of the general metric space indices

1. First approach - Pick-Freeze estimation
2. Second approach - Pick-Freeze and U-stats
3. Third approach - Ranks

First approach - Pick-Freeze estimation (I)

Principle :

- Multiple Monte-Carlo estimation procedure (one to handle the integration part, one to handle the Pick-Freeze part).
- Cost to estimate all first-order indices : $N(m+p+1)$.
- Non trivial proof of the CLT using Donsker theorem and the functional delta method.

Design of experiment :

- a classical Pick-Freeze N-sample, that is two N-samples of Z : $(Z_j, Z_j^{\mathbf{u}})$, $1 \leqslant j \leqslant N$;
- m other N-samples of Z independent of $(Z_j, Z_j^{\mathbf{u}})_{1 \leqslant j \leqslant N}$: $W_{l,k}$, $1 \leqslant l \leqslant m$, $1 \leqslant k \leqslant N$.

First approach - Pick-Freeze estimation (II)

The estimator of the numerator of $S^{\mathbf{u}}_{2,\mathrm{GMS}}$ is then given by

$$\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Bigg\{\Bigg[\frac{1}{N}\sum_{j=1}^{N} T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j)\,T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j^{\mathbf{u}})\Bigg] - \Bigg[\frac{1}{2N}\sum_{j=1}^{N}\left(T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j)+T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j^{\mathbf{u}})\right)\Bigg]^2\Bigg\}$$

while the one of the denominator is

$$\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Bigg\{\Bigg[\frac{1}{2N}\sum_{j=1}^{N}\left(T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j)^2+T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j^{\mathbf{u}})^2\right)\Bigg] - \Bigg[\frac{1}{2N}\sum_{j=1}^{N}\left(T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j)+T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j^{\mathbf{u}})\right)\Bigg]^2\Bigg\}.$$

Second approach - U-statistics estimation (I)

Principle :

- Dealing simultaneously with the Sobol' part and the integration part with respect to $d\mathbb{P}^{\otimes m}(a)$ to get rid of the additional N-samples $(W_{k,l})_{1\leqslant k\leqslant N, 1\leqslant l\leqslant m}$.
- Cost to estimate all first-order indices : $N(p+1)$.
- Elementary proof of the CLT using a CLT for U-stats (Hoeffding 1948) and the classical delta method.

Second approach - U-statistics estimation (II)

It suffices to rewrite $S_{2,\mathrm{GMS}}^{\mathbf{u}}$ as

$$S_{2,\mathrm{GMS}}^{\mathbf{u}} = \frac{I(\Phi_1) - I(\Phi_2)}{I(\Phi_3) - I(\Phi_4)},$$

where,

$$\begin{aligned}
\Phi_1(\mathbf{z}_1, \ldots, \mathbf{z}_{m+1}) &= T_{z_1,\ldots,z_m}(z_{m+1}) T_{z_1,\ldots,z_m}(z_{m+1}^{\mathbf{u}}) \\
\Phi_2(\mathbf{z}_1, \ldots, \mathbf{z}_{m+2}) &= T_{z_1,\ldots,z_m}(z_{m+1}) T_{z_1,\ldots,z_m}(z_{m+2}^{\mathbf{u}}) \\
\Phi_3(\mathbf{z}_1, \ldots, \mathbf{z}_{m+1}) &= T_{z_1,\ldots,z_m}(z_{m+1})^2 \\
\Phi_4(\mathbf{z}_1, \ldots, \mathbf{z}_{m+2}) &= T_{z_1,\ldots,z_m}(z_{m+1}) T_{z_1,\ldots,z_m}(z_{m+2})
\end{aligned}$$

denoting by $\mathbf{z}_i$ the pair $(z_i, z_i^{\mathbf{u}})$ and, for $j = 1, \ldots, 4$,

$$I(\Phi_j) = \int_{\mathcal{X}^{m(j)}} \Phi_j(\mathbf{z}_1, \ldots, \mathbf{z}_{m(j)}) d\mathbb{P}_2^{u,\otimes m(j)}(\mathbf{z}_1 \ldots, \mathbf{z}_{m(j)}).$$

Second approach - U-statistics estimation (III)

Finally, one considers the empirical version of $S^{\mathbf{u}}_{2,\mathrm{GMS}}$:

$$S^{\mathbf{u}}_{2,GMS,Ustat} = \frac{U_{1,N} - U_{2,N}}{U_{3,N} - U_{4,N}},$$

where, for $j = 1, \ldots, 4$,

$$U_{j,N} = \binom{N}{m(j)}^{-1} \sum_{1 \leqslant i_1 < \cdots < i_{m(j)} \leqslant N} \Phi^s_j \left(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_{m(j)}} \right)$$

and the function :

$$\Phi^s_j(\mathbf{z}_1, \ldots, \mathbf{z}_{m(j)}) = \frac{1}{(m(j))!} \sum_{\tau \in \mathcal{S}_{m(j)}} \Phi_j(\mathbf{z}_{\tau(1)}, \ldots, \mathbf{z}_{\tau(m(j))})$$

is the symmetrized version of Φ_j.

Second approach - U-statistics estimation (IV)

The estimator $S^{\mathbf{u}}_{2,\mathrm{GMS,Ustat}}$ has been proved to be consistent and asymptotically Gaussian.

Theorem (Gamboa, Klein, Lagnoux, Moreno (2021))

If for $j = 1, \dots, 4$, $\mathbb{E}\left[\Phi_j^s\left(\mathbf{Z}_1, \dots, \mathbf{Z}_{m(j)}\right)^2\right] < \infty$ *then*

$$\sqrt{N}\left(S^{\mathbf{u}}_{2,GMS,Ustat} - S^{\mathbf{u}}_{2,GMS}\right) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}_1(0,\sigma^2)$$

where the asymptotic variance σ^2 *is explicitly known.*

Third approach - Rank-based estimation (I)

Principle :

- Only when the inputs are scalar and to estimate the first-order indices.
- Cost to estimate all first-order indices : N.
- CLT in progress.

Let $\pi_i(j)$ be the rank of X_i^j in the sample $(X_i^1, \ldots, X_i^N)$ of X_i and define

$$N_i(j) = \begin{cases} \pi_i^{-1}(\pi_i(j)+1) & \text{if } \pi_i(j)+1 \leqslant N, \\ \pi_i^{-1}(1) & \text{if } \pi_i(j) = N. \end{cases}$$

Third approach - Rank-based estimation (II)

Then the empirical estimator $\widehat{S}^i_{2,\mathrm{GMS,Rank}}$ of $S^i_{2,\mathrm{GMS}}$ is given by the ratio between

$$\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\left\{\left[\frac{1}{N}\sum_{j=1}^{N}T_{Z_{i_1},\cdots,Z_{i_m}}(Z_j)\,T_{Z_{i_1},\cdots,Z_{i_m}}(Z_{N_i(j)})\right]\right.$$
$$\left.-\left[\frac{1}{N}\sum_{j=1}^{N}T_{Z_{i_1},\cdots,Z_{i_m}}(Z_j)\right]^2\right\}$$

and

$$\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\left\{\left[\frac{1}{N}\sum_{j=1}^{N}T_{Z_{i_1},\cdots,Z_{i_m}}(Z_j)^2\right]-\left[\frac{1}{N}\sum_{j=1}^{N}T_{Z_{i_1},\cdots,Z_{i_m}}(Z_j)\right]^2\right\}.$$

Definition of the universal index

We have defined

$$S^{\mathbf{u}}_{2,GMS} := \frac{\int_{\mathcal{X}^m} \mathbb{E}\left[(\mathbb{E}[T_a(Z)] - \mathbb{E}[T_a(Z)|X_{\mathbf{u}}])^2\right] d\mathbb{P}^{\otimes m}(a)}{\int_{\mathcal{X}^m} \mathrm{Var}(T_a(Z)) d\mathbb{P}^{\otimes m}(a)}.$$

One may extend this definition allowing a to live in another space and integrating with respect to a different probability measure $\mathbb{Q}$ than $\mathbb{P}$.

Definition (Fort, Klein, and Lagnoux (2021))

$$S^{\mathbf{u}}_{2,\mathrm{Univ}}(T_a, \mathbb{Q}) := \frac{\int_{\Omega} \mathbb{E}\left[(\mathbb{E}[T_a(Z)] - \mathbb{E}[T_a(Z)|X_{\mathbf{u}}])^2\right] d\mathbb{Q}(a)}{\int_{\Omega} \mathrm{Var}(T_a(Z)) d\mathbb{Q}(a)}.$$

Ref. : F. Gamboa, T. Klein, A. Lagnoux, and L. Moreno. "Sensitivity analysis in general metric spaces ", *RESS*, 2021.

J.-C. Fort, T. Klein, and A. Lagnoux. "Global sensitivity analysis and Wasserstein spaces", *SIAM UQ*, 2021.

Numerical application (I)

Consider F, F_1, and F_2 three elements of $\mathcal{M}_2(\mathbb{R})$ and, for $a = (F_1, F_2)$, the family of test functions

$$T_a(F) = T_{(F_1,F_2)}(F) = \mathbb{1}_{W_2(F_1,F) \leqslant W_2(F_1,F_2)}.$$

Then, for all $\mathbf{u} \subset \{1, \cdots, p\}$, the index is

$$S^{\mathbf{u}}_{2,W_2} = S^{\mathbf{u}}_{2,\mathrm{Univ}}((F_1, F_2, F) \mapsto T_{F_1,F_2}(F), \mathbb{P}^{\otimes 2})$$

$$= \frac{\int_{\mathcal{W}_2(\mathbb{R}) \times \mathcal{W}_2(\mathbb{R})} \mathbb{E}\left[\left(\mathbb{E}[\mathbb{1}_{W_2(F_1,\mathbb{F}) \leqslant W_2(F_1,F_2)}] - \mathbb{E}[\mathbb{1}_{W_2(F_1,\mathbb{F}) \leqslant W_2(F_1,F_2)} | X_{\mathbf{u}}]\right)^2\right] d\mathbb{P}^{\otimes 2}(F_1, F_2)}{\int_{\mathcal{W}_2(\mathbb{R}) \times \mathcal{W}_2(\mathbb{R})} \mathrm{Var}(\mathbb{1}_{W_2(F_1,\mathbb{F}) \leqslant W_2(F_1,F_2)}) d\mathbb{P}^{\otimes 2}(F_1, F_2)}$$

$$= \frac{\int_{\mathcal{W}_2(\mathbb{R}) \times \mathcal{W}_2(\mathbb{R})} \mathrm{Var}\left(\mathbb{E}[\mathbb{1}_{W_2(F_1,\mathbb{F}) \leqslant W_2(F_1,F_2)} | X_{\mathbf{u}}]\right) d\mathbb{P}^{\otimes 2}(F_1, F_2)}{\int_{\mathcal{W}_2(\mathbb{R}) \times \mathcal{W}_2(\mathbb{R})} \mathrm{Var}(\mathbb{1}_{W_2(F_1,\mathbb{F}) \leqslant W_2(F_1,F_2)}) d\mathbb{P}^{\otimes 2}(F_1, F_2)}.$$

Numerical application (II)

Let X_1, X_2, X_3 be 3 independent random variables Bernoulli distributed with parameter p_1, p_2, and p_3 respectively. We consider the c.d.f.-valued code f, the output of which is given by

$$\mathbb{F}(t) = \frac{t}{1 + X_1 + X_2 + X_1 X_3} \mathbb{1}_{0 \leqslant t \leqslant 1 + X_1 + X_2 + X_1 X_3} + \mathbb{1}_{1 + X_1 + X_2 + X_1 X_3 < t}.$$

Numerical application (III)

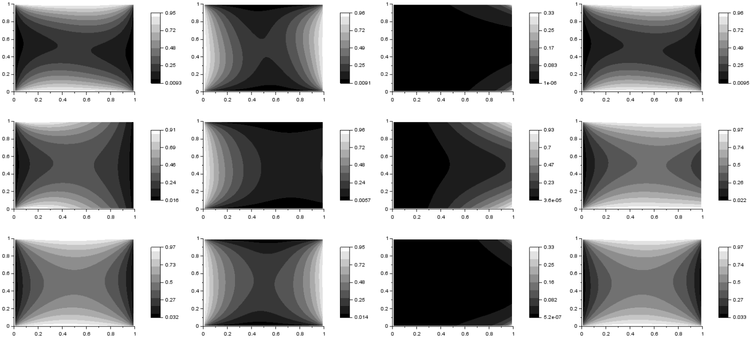

Figure – Values of the indices S^{1}_{2,W_2}, S^{2}_{2,W_2}, S^{3}_{2,W_2}, and $S^{1,3}_{2,W_2}$ (from left to right) with respect to the values of p_1 and p_2 (varying from 0 to 1). In the first row (resp. second and third), p_3 is fixed to $p_3 = 0.01$ (resp. 0.5 and 0.99).

Numerical application (IV)

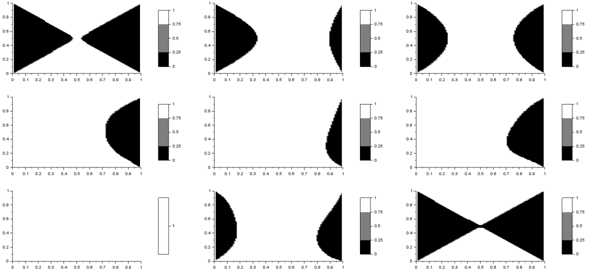

Figure – In the first row of the figure, regions where $S^1_{2,W_2} \geqslant S^2_{2,W_2}$ (black), $S^1_{2,W_2} \leqslant S^2_{2,W_2}$ (white), and $S^1_{2,W_2} = S^2_{2,W_2}$ (gray) with respect to p_1 and p_2 varying from 0 to 1 and, from left to right, $p_3 = 0.01$, 0.5, and 0.99. Analogously, the second (resp. last) row considers the regions with S^1_{2,W_2} and S^3_{2,W_2} (resp. S^2_{2,W_2} and S^3_{2,W_2}) with respect to p_1 and p_3 (resp. p_2 and p_3) varying from 0 to 1 and, from left to right, $p_2 = 0.01$, 0.5, and 0.99 (resp. $p_1 = 0.01$, 0.5, and 0.99).

Numerical application (V)

- Only 450 calls of the computer code are allowed to estimate the indices $S^{\mathbf{u}}(\mathbb{F})$ and $S^{\mathbf{u}}_{2,W_2}$ for $\mathbf{u} = \{1\}$, $\{2\}$, and $\{3\}$. Hence, the sample size allowed in the rank-based procedure is $N = 450$. In the Pick-Freeze methodology, the estimation of the Wasserstein indices $S^{\mathbf{u}}_{2,W_2}$ requires one initial output sample, three extra output samples to get the Pick-Freeze versions (one for each index) and two extra samples to handle the integration leading to an allowed sample size $N = \lfloor 450/6 \rfloor = 75$ for the indices.
- We only focus on the first-order indices since, as explained previously, the rank-based procedure has not been developed yet for higher-order indices.
- We repeat the estimation procedure $n_r = 200$ times.

Numerical application (VI)

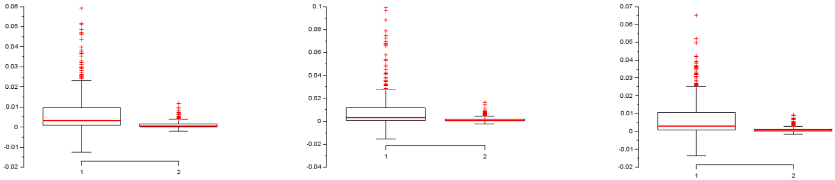

Figure – Here $p_1 = 1/3$, $p_2 = 2/3$, and $p_3 = 3/4$. Boxplots of the mean square errors of the estimation of the Wasserstein indices $S^{\mathbf{u}}_{2,W_2}$ with a fixed sample size N and $n_r = 200$ replications. The indices with respect to $\mathbf{u} = \{1\}$, $\{2\}$, and $\{3\}$ are displayed from left to right. The results of the Pick-Freeze estimation procedure with $N = 75$ for the Wasserstein indices $S^{\mathbf{u}}_{2,W_2}$ are provided in the left side of each graphic. The results of the rank-based methodology with $N = 450$ are provided in the right side of each graphic.

The Gaussian Plume Model (I)

We consider a point source that emits contaminant into a uni-directional wind in an infinite domain. Such a model is also applied, for instance, to volcanic eruptions, pollen and insect dispersals, and is called the Gaussian plume model (GPM). The contaminant concentration at location $(x, y, 0)$ rewrites as :

$$C(x, y, 0) = \frac{Q}{2\pi K x} e^{\frac{-u(y^2+H^2)}{4Kx}}, \tag{2}$$

where Q is the emission rate, u the wind speed, K the diffusion, and H the effective height.

The Gaussian Plume Model (II)

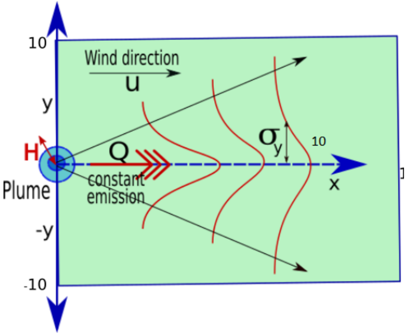

Figure – Plume model (2). Cross section at $z = 0$ of a contaminant plume emitted from a continuous point source, with wind direction aligned with the x-axis.

The Gaussian Plume Model (III)

In this setting, the function f that defines the output of interest is then given by :

$$\begin{array}{rccc} f: & \mathbb{R}^3 & \to & L^2(R^2) \\ & (Q,K,u) & \mapsto & f(Q,K,u) = \left(C(x,y,0)\right)_{(x,y)\in\mathbb{R}^2}, \end{array}$$

where Q, K, and u are assumed to be all independent with uniform distribution $\mathcal{U}(0,10)$.

The Gaussian Plume Model (IV)

A first step consists in performing a GSA for spatial data, namely an ubiquous sensitivity analysis. In other words, the sensitivity indices are computed location after location leading to a sensitivity map.

The Gaussian Plume Model (V)

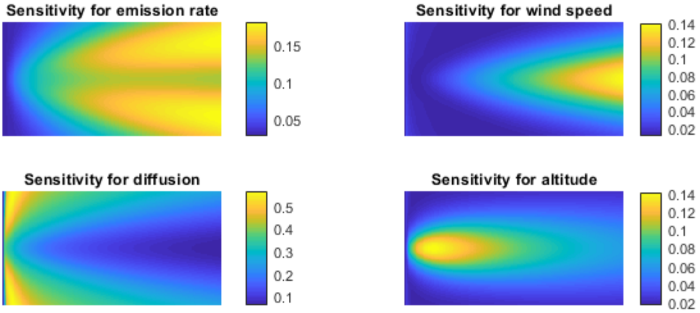

Figure – Plume model (2). Ubiquous sensitivity analysis with respect to the emission rate Q (top left), the wind speed u (top right), the diffusion K (bottom left), and the effective height H (bottom right).

The Gaussian Plume Model (VI)

For two pollution concentrations C_1 and C_2 with domain in the ground level (in $\mathbb{R}^2$), the distance used is the classical L^2 distance

$$d(C_1, C_2) = \sqrt{\iint (C_1(x,y,0) - C_2(x,y,0))^2 dxdy}.$$

To quantify the sensitivity on the contaminant concentration with respect to Q, K, and u, we consider the family of functions T_a given by $T_{(a_1,a_2)}(b) = \mathbb{1}_{b \in B_{(a_1,a_2)}}$, where a_1, a_2, and b square-integrable are applications from $\mathbb{R}^2$ to $\mathbb{R}$ and $B_{(a_1,a_2)}$ stands for the L^2-ball centered at a_1 with radius $\overline{a_1 a_2}$ (whence $m = 2$).

The Gaussian Plume Model (VII)

	N=1000			N=2000			N=5000		
H	K	Q	u	K	Q	u	K	Q	u
1	0.1365	0.1216	0.1330	0.1124	0.1419	0.1453	0.1425	0.1431	0.1562
2	0.1028	0.1197	0.1212	0.1291	0.1317	0.1171	0.1222	0.1627	0.1143
10	0.0813	0.0891	0.1010	0.1081	0.1077	0.1256	0.0893	0.0831	0.1001
20	0.1027	0.0246	0.1041	0.0620	0.0942	0.1030	0.0913	0.0091	0.0329

Table – Sensitivity indices for the plume model (2).

Thanks for your attention !

Questions ?