Global sensitivity analysis and dimension reduction

Clémentine PRIEUR
Grenoble Alpes University

Summer school on GSA & Poincaré inequalities
6-8 July 2022, Toulouse

Part I

Variance-based sensitivity analysis: I will present the general framework of variance-based sensitivity analysis, from the definition of sensitivity indices to the presentation of different inference strategies.

See, e.g., Saltelli et al. [2000], Faivre et al. [2013], Da Veiga et al. [2021].

$$\boxed{\text{inputs}} \longrightarrow \boxed{\text{model}} \longrightarrow \boxed{\text{output}}$$

$$\boxed{\left.\begin{matrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_d \end{matrix}\right\}} \longrightarrow \boxed{\mathcal{M}} \longrightarrow \boxed{Y = \mathcal{M}(X_1, \ldots, X_d)}$$

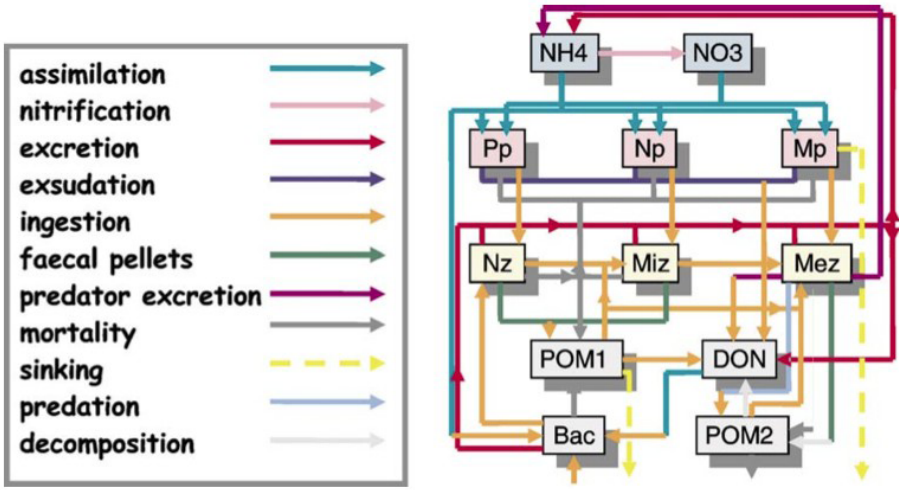
One wishes to quantify the sensitivity of the output Y to the inputs $X_1, \ldots, X_d$.

The model $\mathcal{M}$ is most of the time complex, expensive to evaluate. Each input factor can be a scalar, a vector, or even a function. The output (or QoI Quantity of Interest) can also be scalar, vectorial or functional.

Application to a biogeochemical model: ecosystem model (MODECOGeL) of the Ligurian Sea

Joint work with IGE Lab (Grenoble, FRANCE)

MODECOGeL is a one-dimensional coupled hydrodynamical-biological model.



• hydrodynamic model: 1-D vertical simplification of primitive equations for the ocean, 5 state variables;
• ecosystem model: marine biogeochemistry, 12 biological state variables.

Inputs/Outputs:

- ▷ 74 scalar input parameters;
- ▷ spatio-temporal outputs.

Main issue: calibration of the model.

Sensitivity Analysis is a preliminary step to this calibration task.

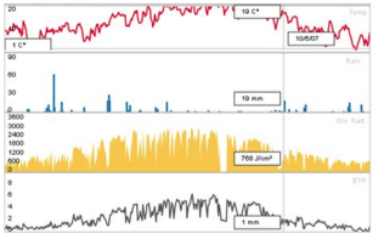
Agro-climatic model for the water status management of vineyard

Joint work with INRA and iTK (Montpellier, FRANCE)

Project objective: control of grape/wine quality. SA as decision support.

INPUTS

- **Vine-plot**
  - ➡ **22 scalar parameters:** soil texture, rooting depth, vegetation size, row orientation, ...
- **Weather data during the growing season**
  - ➡ **4 correlated (daily) functional inputs**

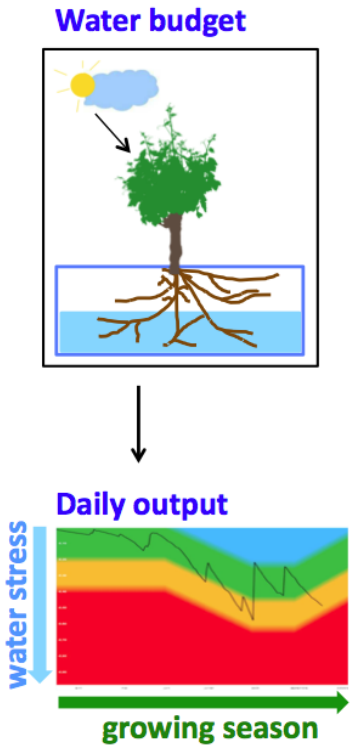


Temperature

Rain

Solar radiation

PET



Background :

$$\mathcal{M} : \begin{cases} \mathbb{R}^d & \rightarrow & \mathbb{R} \\ \mathbf{x} & \mapsto & y = \mathcal{M}(x_1, \ldots, x_d) \end{cases}$$

Goal : find how model outputs vary with inputs changes.

Different strategies :

- Qualitative analysis : non-linear behaviors? possible interactions?
  ex. : screening .
- Quantitative analysis : factorial hierarchisation, statistical tests H_0 "negligible input"
  ex. : sensitivity Sobol' indices

Sensitivity analysis may help identifying inappropriate models.

Background :

$$\mathcal{M} : \begin{cases} \mathbb{R}^d & \rightarrow & \mathbb{R} \\ \mathbf{x} & \mapsto & y = \mathcal{M}(x_1, \ldots, x_d) \end{cases}$$

Goal : find how model outputs vary with inputs changes.

Different strategies :

- Qualitative analysis : non-linear behaviors? possible interactions?
  ex. : screening .
- Quantitative analysis : factorial hierarchisation, statistical tests H_0 "negligible input"
  ex. : sensitivity Sobol' indices

Sensitivity analysis may help identifying inappropriate models.

Background :

$$\mathcal{M} : \begin{cases} \mathbb{R}^d & \rightarrow & \mathbb{R} \\ \mathbf{x} & \mapsto & y = \mathcal{M}(x_1, \ldots, x_d) \end{cases}$$

Goal : find how model outputs vary with inputs changes.

Different strategies :

- Qualitative analysis : non-linear behaviors? possible interactions?
  ex. : screening .
- Quantitative analysis : factorial hierarchisation, statistical tests H_0 "negligible input"
  ex. : sensitivity Sobol' indices

Sensitivity analysis may help identifying inappropriate models.

Various approaches for quantitative sensitivity :

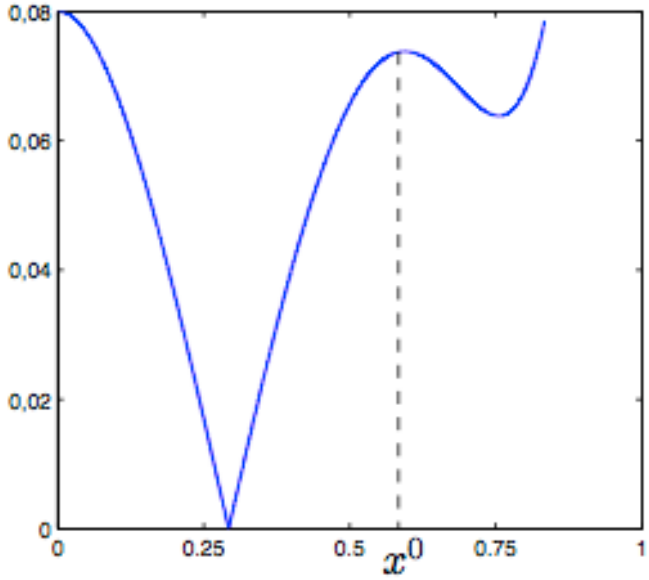
Local approaches :

$\mathcal{M}(\mathbf{x}) \approx \mathcal{M}(\mathbf{x^0}) + \sum_{i=1}^{d} \left(\frac{\partial \mathcal{M}}{\partial x_i}\right)_{\mathbf{x^0}} (x_i - x_i^0)$ (Taylor approximation).

First order sensitivity index for input i : $\left(\frac{\partial \mathcal{M}}{\partial x_i}\right)_{\mathbf{x^0}}$.

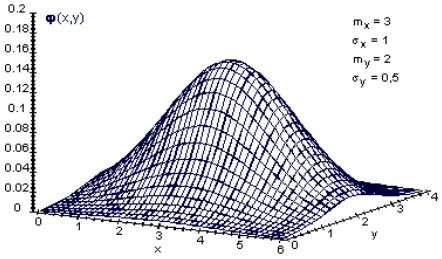
Pros : Low computational cost even for large d

Cons : local approaches, not well-suited for highly nonlinear models

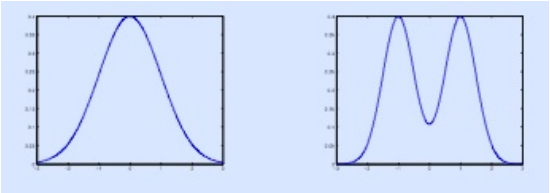


Global uncertainty quantification framework :

Uncertain input parameters are modeled by a probability distribution μ on $\mathcal{X}$, from experts' knowledge or from observations.



E.g., if the inputs are independent, this probability distribution is characterized by its marginals: $\mu(d\mathbf{x}) = \prod_{i=1}^{d} \mu_i(dx_i)$.



General setup : (Hoeffding, 1948; Sobol', 1993)

$Y = \mathcal{M}(X_1, \ldots, X_d)$, $(X_1, \ldots, X_d) \sim \mu$. In the following, we assume :

i) the X_i are independent: $\mu(d\mathbf{x}) = \prod_{i=1}^{d} \mu_i(dx_i)$;
ii) $\forall\, i = 1, \ldots, d$, $X_i \sim \mathcal{U}([0, 1])$.

Assumption ii) is not necessary but lightens the presentation.

The case of correlated inputs raises several issues and will be discussed later.

General setup : (Hoeffding, 1948; Sobol', 1993)

$Y = \mathcal{M}(X_1, \ldots, X_d)$, $(X_1, \ldots, X_d) \sim \mu$. In the following, we assume :

i) the X_i are independent: $\mu(d\mathbf{x}) = \prod_{i=1}^{d} \mu_i(dx_i)$;
ii) $\forall\, i = 1, \ldots, d$, $X_i \sim \mathcal{U}([0, 1])$.

Assumption ii) is not necessary but lightens the presentation.

The case of correlated inputs raises several issues and will be discussed later.

General setup : (Hoeffding, 1948; Sobol', 1993)

$Y = \mathcal{M}(X_1, \ldots, X_d)$, $(X_1, \ldots, X_d) \sim \mu$. In the following, we assume :

i) the X_i are independent: $\mu(d\mathbf{x}) = \prod_{i=1}^{d} \mu_i(dx_i)$;
ii) $\forall\, i = 1, \ldots, d$, $X_i \sim \mathcal{U}([0, 1])$.

Assumption ii) is not necessary but lightens the presentation.

The case of correlated inputs raises several issues and will be discussed later.

Towards Sobol sensitivity indices

Is the output Y more or less variable when input are fixed?

$\mathrm{Var}\left(Y|X_i = x_i\right)$, how to choose x_i? $\Rightarrow E\left[V\left(Y|X_i\right)\right]$

the smaller this quantity, (i.e. fixing X_i), the smaller is the variance of Y when fixing the ith input: variable X_i has a strong impact.

Theorem (Total variance)

$\mathrm{Var}(Y) = \mathrm{Var}\left[\mathbb{E}\left(Y|X_i\right)\right] + \mathbb{E}\left[\mathrm{Var}\left(Y|X_i\right)\right].$

Definition (First order Sobol' Index)

$i = 1, \ldots, d$

$$0 \leq S_i = \frac{V\left[E\left(Y|X_i\right)\right]}{\mathrm{Var}(Y)} \leq 1$$

ex. : linear output $Y = \sum_{i=1}^{d} \beta_i X_i$, we get $S_i = \frac{\beta_i^2 \mathrm{Var}(X_i)}{\mathrm{Var}(Y)} = \rho_i^2$, with ρ_i linear correlation coefficient.

Towards Sobol sensitivity indices

Is the output Y more or less variable when input are fixed?

$\mathrm{Var}\left(Y|X_i = x_i\right)$, how to choose x_i? $\Rightarrow E\left[V\left(Y|X_i\right)\right]$

the smaller this quantity, (i.e. fixing X_i), the smaller is the variance of Y when fixing the ith input: variable X_i has a strong impact.

Theorem (Total variance)

$\mathrm{Var}(Y) = \mathrm{Var}\left[\mathbb{E}\left(Y|X_i\right)\right] + \mathbb{E}\left[\mathrm{Var}\left(Y|X_i\right)\right].$

Definition (First order Sobol' Index)

$i = 1, \ldots, d$

$$0 \leq S_i = \frac{V\left[E\left(Y|X_i\right)\right]}{\mathrm{Var}(Y)} \leq 1$$

ex. : linear output $Y = \sum_{i=1}^{d} \beta_i X_i$, we get $S_i = \frac{\beta_i^2 \mathrm{Var}(X_i)}{\mathrm{Var}(Y)} = \rho_i^2$, with ρ_i linear correlation coefficient.

Towards Sobol sensitivity indices

Is the output Y more or less variable when input are fixed?

$\mathrm{Var}\left(Y|X_i = x_i\right)$, how to choose x_i? $\Rightarrow E\left[V\left(Y|X_i\right)\right]$

the smaller this quantity, (i.e. fixing X_i), the smaller is the variance of Y when fixing the ith input: variable X_i has a strong impact.

Theorem (Total variance)

$\mathrm{Var}(Y) = \mathrm{Var}\left[\mathbb{E}\left(Y|X_i\right)\right] + \mathbb{E}\left[\mathrm{Var}\left(Y|X_i\right)\right].$

Definition (First order Sobol' Index)

$i = 1, \ldots, d$

$$0 \leq S_i = \frac{V\left[E\left(Y|X_i\right)\right]}{\mathrm{Var}(Y)} \leq 1$$

ex. : linear output $Y = \sum_{i=1}^{d} \beta_i X_i$, we get $S_i = \frac{\beta_i^2 \mathrm{Var}(X_i)}{\mathrm{Var}(Y)} = \rho_i^2$, with ρ_i linear correlation coefficient.

Towards Sobol sensitivity indices

Is the output Y more or less variable when input are fixed?

$\mathrm{Var}\left(Y|X_i = x_i\right)$, how to choose x_i? $\Rightarrow E\left[V\left(Y|X_i\right)\right]$

the smaller this quantity, (i.e. fixing X_i), the smaller is the variance of Y when fixing the ith input: variable X_i has a strong impact.

Theorem (Total variance)

$\mathrm{Var}(Y) = \mathrm{Var}\left[\mathbb{E}\left(Y|X_i\right)\right] + \mathbb{E}\left[\mathrm{Var}\left(Y|X_i\right)\right]$.

Definition (First order Sobol' Index)

$i = 1, \ldots, d$

$$0 \leq S_i = \frac{V\left[E\left(Y|X_i\right)\right]}{\mathrm{Var}(Y)} \leq 1$$

ex. : linear output $Y = \sum_{i=1}^{d} \beta_i X_i$, we get $S_i = \frac{\beta_i^2 \mathrm{Var}(X_i)}{\mathrm{Var}(Y)} = \rho_i^2$, with ρ_i linear correlation coefficient.

Towards Sobol sensitivity indices

Is the output Y more or less variable when input are fixed?

$\mathrm{Var}\left(Y|X_i = x_i\right)$, how to choose x_i? $\Rightarrow E\left[V\left(Y|X_i\right)\right]$

the smaller this quantity, (i.e. fixing X_i), the smaller is the variance of Y when fixing the ith input: variable X_i has a strong impact.

Theorem (Total variance)

$\mathrm{Var}(Y) = \mathrm{Var}\left[\mathbb{E}\left(Y|X_i\right)\right] + \mathbb{E}\left[\mathrm{Var}\left(Y|X_i\right)\right].$

Definition (First order Sobol' Index)

$i = 1, \ldots, d$

$$0 \leq S_i = \frac{V\left[E\left(Y|X_i\right)\right]}{\mathrm{Var}(Y)} \leq 1$$

ex. : linear output $Y = \sum_{i=1}^{d} \beta_i X_i$, we get $S_i = \frac{\beta_i^2 \mathrm{Var}(X_i)}{\mathrm{Var}(Y)} = \rho_i^2$, with ρ_i linear correlation coefficient.

Towards Sobol sensitivity indices

Is the output Y more or less variable when input are fixed?

$\mathrm{Var}\left(Y|X_i = x_i\right)$, how to choose x_i? $\Rightarrow E\left[V\left(Y|X_i\right)\right]$

the smaller this quantity, (i.e. fixing X_i), the smaller is the variance of Y when fixing the ith input: variable X_i has a strong impact.

Theorem (Total variance)

$\mathrm{Var}(Y) = \mathrm{Var}\left[\mathbb{E}\left(Y|X_i\right)\right] + \mathbb{E}\left[\mathrm{Var}\left(Y|X_i\right)\right]$.

Definition (First order Sobol' Index)

$i = 1, \ldots, d$

$$0 \leq S_i = \frac{V\left[E\left(Y|X_i\right)\right]}{\mathrm{Var}(Y)} \leq 1$$

ex. : linear output $Y = \sum_{i=1}^{d} \beta_i X_i$, we get $S_i = \frac{\beta_i^2 \mathrm{Var}(X_i)}{\mathrm{Var}(Y)} = \rho_i^2$, with ρ_i linear correlation coefficient.

Toy case:

$$Y = X_1^2 + X_2 \qquad X_i \sim \mathcal{U}([0,1]) \quad X_1 \perp\!\!\!\perp X_2$$

$\mathbb{E}(Y|X_1) = X_1^2 + \mathbb{E}(X_2) \Rightarrow \mathrm{Var}[\mathbb{E}(Y|X_1)] = \mathrm{Var}(X_1^2) = \frac{4}{45}$

$\mathbb{E}(Y|X_2) = \mathbb{E}(X_1^2) + X_2 \Rightarrow \mathrm{Var}[\mathbb{E}(Y|X_2)] = \mathrm{Var}(X_2) = \frac{1}{12}$

$\mathrm{Var}(Y) = \mathrm{Var}(X_1^2) + \mathrm{Var}(X_2) = \frac{31}{180}$

$$S_1 = \frac{16}{31} \approx 0,516\,, \; S_2 = \frac{15}{31} \approx 0,484$$

$S_1 + S_2 = 1$, additive model

Toy case:

$$Y = X_1^2 + X_2 \qquad X_i \sim \mathcal{U}([0,1]) \quad X_1 \perp\!\!\!\perp X_2$$

$$\mathbb{E}(Y|X_1) = X_1^2 + \mathbb{E}(X_2) \Rightarrow \mathrm{Var}\left[\mathbb{E}(Y|X_1)\right] = \mathrm{Var}(X_1^2) = \frac{4}{45}$$

$$\mathbb{E}(Y|X_2) = \mathbb{E}(X_1^2) + X_2 \Rightarrow \mathrm{Var}\left[\mathbb{E}(Y|X_2)\right] = \mathrm{Var}(X_2) = \frac{1}{12}$$

$$\mathrm{Var}(Y) = \mathrm{Var}(X_1^2) + \mathrm{Var}(X_2) = \frac{31}{180}$$

$$S_1 = \frac{16}{31} \approx 0,516\,, \;\; S_2 = \frac{15}{31} \approx 0,484$$

$S_1 + S_2 = 1$, additive model

Toy case:

$$Y = X_1^2 + X_2 \qquad X_i \sim \mathcal{U}([0,1]) \quad X_1 \perp\!\!\!\perp X_2$$

$$\mathbb{E}(Y|X_1) = X_1^2 + \mathbb{E}(X_2) \Rightarrow \mathrm{Var}[\mathbb{E}(Y|X_1)] = \mathrm{Var}(X_1^2) = \frac{4}{45}$$

$$\mathbb{E}(Y|X_2) = \mathbb{E}(X_1^2) + X_2 \Rightarrow \mathrm{Var}[\mathbb{E}(Y|X_2)] = \mathrm{Var}(X_2) = \frac{1}{12}$$

$$\mathrm{Var}(Y) = \mathrm{Var}(X_1^2) + \mathrm{Var}(X_2) = \frac{31}{180}$$

$$S_1 = \frac{16}{31} \approx 0,516\,, \; S_2 = \frac{15}{31} \approx 0,484$$

$S_1 + S_2 = 1$, additive model

Toy case:

$$Y = X_1^2 + X_2 \qquad X_i \sim \mathcal{U}([0,1]) \quad X_1 \perp\!\!\!\perp X_2$$

$$\mathbb{E}(Y|X_1) = X_1^2 + \mathbb{E}(X_2) \Rightarrow \mathrm{Var}\left[\mathbb{E}(Y|X_1)\right] = \mathrm{Var}(X_1^2) = \frac{4}{45}$$

$$\mathbb{E}(Y|X_2) = \mathbb{E}(X_1^2) + X_2 \Rightarrow \mathrm{Var}\left[\mathbb{E}(Y|X_2)\right] = \mathrm{Var}(X_2) = \frac{1}{12}$$

$$\mathrm{Var}(Y) = \mathrm{Var}(X_1^2) + \mathrm{Var}(X_2) = \frac{31}{180}$$

$$S_1 = \frac{16}{31} \approx 0,516\,, \; S_2 = \frac{15}{31} \approx 0,484$$

$S_1 + S_2 = 1$, additive model

Toy case:

$$Y = X_1^2 + X_2 \qquad X_i \sim \mathcal{U}([0,1]) \quad X_1 \perp\!\!\!\perp X_2$$

$$\mathbb{E}(Y|X_1) = X_1^2 + \mathbb{E}(X_2) \Rightarrow \mathrm{Var}[\mathbb{E}(Y|X_1)] = \mathrm{Var}(X_1^2) = \tfrac{4}{45}$$

$$\mathbb{E}(Y|X_2) = \mathbb{E}(X_1^2) + X_2 \Rightarrow \mathrm{Var}[\mathbb{E}(Y|X_2)] = \mathrm{Var}(X_2) = \tfrac{1}{12}$$

$$\mathrm{Var}(Y) = \mathrm{Var}(X_1^2) + \mathrm{Var}(X_2) = \tfrac{31}{180}$$

$$S_1 = \frac{16}{31} \approx 0,516\,, \; S_2 = \frac{15}{31} \approx 0,484$$

$S_1 + S_2 = 1$, additive model

Toy case:

$$Y = X_1^2 + X_2 \quad X_i \sim \mathcal{U}([0,1]) \quad X_1 \perp\!\!\!\perp X_2$$

$$\mathbb{E}(Y|X_1) = X_1^2 + \mathbb{E}(X_2) \Rightarrow \mathrm{Var}[\mathbb{E}(Y|X_1)] = \mathrm{Var}(X_1^2) = \frac{4}{45}$$

$$\mathbb{E}(Y|X_2) = \mathbb{E}(X_1^2) + X_2 \Rightarrow \mathrm{Var}[\mathbb{E}(Y|X_2)] = \mathrm{Var}(X_2) = \frac{1}{12}$$

$$\mathrm{Var}(Y) = \mathrm{Var}(X_1^2) + \mathrm{Var}(X_2) = \frac{31}{180}$$

$$S_1 = \frac{16}{31} \approx 0,516\,,\ S_2 = \frac{15}{31} \approx 0,484$$

$S_1 + S_2 = 1$, additive model

More generally,

Theorem (Hoeffding decomposition)

$\mathcal{M} : [0,1]^d \to \mathbb{R}$, $\int_{[0,1]^d} \mathcal{M}^2(x)dx < \infty$

$\mathcal{M}$ has an unique decomposition

$$\mathcal{M}_0 + \sum_{i=1}^{d} \mathcal{M}_i(x_i) + \sum_{1 \leq i < j \leq d} \mathcal{M}_{i,j}(x_i, x_j) + \ldots + \mathcal{M}_{1,\ldots,d}(x_1, \ldots, x_d)$$

under the constraint

- *$\mathcal{M}_0$ constant,*
- $\forall 1 \leq s \leq d, \forall 1 \leq i_1 < \ldots < i_s \leq d, \forall 1 \leq p \leq s$

$$\int_0^1 \mathcal{M}_{i_1,\ldots,i_s}(x_{i_1}, \ldots, x_{i_s})dx_{i_p} = 0$$

Consequences : $\mathcal{M}_0 = \int_{[0,1]^d} \mathcal{M}(x)dx$ and the terms of the decomposition are orthogonal.

The computation of each term in the decomposition writes:

- $\mathcal{M}_i(x_i) = \int_{[0,1]^{d-1}} \mathcal{M}(x) \Pi_{p \neq i} dx_p - \mathcal{M}_0$
- $i \neq j$
  $\mathcal{M}_{i,j}(x_i, x_j) = \int_{[0,1]^{d-2}} \mathcal{M}(x) \Pi_{p \neq i,j} dx_p - \mathcal{M}_0 - \mathcal{M}_i(x_i) - \mathcal{M}_j(x_j)$
- ...

$\Rightarrow$ computation of multiple integrals.

Consequences : $\mathcal{M}_0 = \int_{[0,1]^d} \mathcal{M}(x)dx$ and the terms of the decomposition are orthogonal.

The computation of each term in the decomposition writes:

- $\mathcal{M}_i(x_i) = \int_{[0,1]^{d-1}} \mathcal{M}(x) \Pi_{p \neq i} dx_p - \mathcal{M}_0$
- $i \neq j$
  $\mathcal{M}_{i,j}(x_i, x_j) = \int_{[0,1]^{d-2}} \mathcal{M}(x) \Pi_{p \neq i,j} dx_p - \mathcal{M}_0 - \mathcal{M}_i(x_i) - \mathcal{M}_j(x_j)$
- ...

$\Rightarrow$ computation of multiple integrals.

<u>Consequences :</u> $\mathcal{M}_0 = \int_{[0,1]^d} \mathcal{M}(x)dx$ and the terms of the decomposition are orthogonal.

The computation of each term in the decomposition writes:

- $\mathcal{M}_i(x_i) = \int_{[0,1]^{d-1}} \mathcal{M}(x) \Pi_{p \neq i} dx_p - \mathcal{M}_0$
- $i \neq j$
  $\mathcal{M}_{i,j}(x_i, x_j) = \int_{[0,1]^{d-2}} \mathcal{M}(x) \Pi_{p \neq i,j} dx_p - \mathcal{M}_0 - \mathcal{M}_i(x_i) - \mathcal{M}_j(x_j)$
- ...

$\Rightarrow$ computation of multiple integrals.

<u>Consequences :</u> $\mathcal{M}_0 = \int_{[0,1]^d} \mathcal{M}(x)dx$ and the terms of the decomposition are orthogonal.

The computation of each term in the decomposition writes:

- $\mathcal{M}_i(x_i) = \int_{[0,1]^{d-1}} \mathcal{M}(x) \Pi_{p \neq i} dx_p - \mathcal{M}_0$
- $i \neq j$
  $\mathcal{M}_{i,j}(x_i, x_j) = \int_{[0,1]^{d-2}} \mathcal{M}(x) \Pi_{p \neq i,j} dx_p - \mathcal{M}_0 - \mathcal{M}_i(x_i) - \mathcal{M}_j(x_j)$
- $\ldots$

$\Rightarrow$ computation of multiple integrals.

Variance decomposition : $X_1, \ldots, X_d$ i.i.d. $\sim \mathcal{U}([0,1])$

$$Y = \mathcal{M}(X) = \mathcal{M}_0 + \sum_{i=1}^{d} \mathcal{M}_i(X_i) + \ldots + \mathcal{M}_{1,\ldots,d}(X_1, \ldots, X_d)$$

- $\mathcal{M}_0 = \mathbb{E}(Y)$,
- $\mathcal{M}_i(X_i) = \mathbb{E}(Y|X_i) - \mathbb{E}(Y)$,
- $i \neq j$ $\mathcal{M}_{i,j}(X_i, X_j) = \mathbb{E}(Y|X_i, X_j) - \mathbb{E}(Y|X_i) - \mathbb{E}(Y|X_j) + \mathbb{E}(Y)$,
- $\ldots$

$$\mathrm{Var}(Y) = \sum_{i=1}^{d} \mathrm{Var}(\mathcal{M}_i(X_i)) + \ldots + \mathrm{Var}(\mathcal{M}_{1,\ldots,d}(X_1, \ldots, X_d))$$

Variance decomposition : $X_1, \ldots, X_d$ i.i.d. $\sim \mathcal{U}([0,1])$

$$Y = \mathcal{M}(X) = \mathcal{M}_0 + \sum_{i=1}^{d} \mathcal{M}_i(X_i) + \ldots + \mathcal{M}_{1,\ldots,d}(X_1, \ldots, X_d)$$

- $\mathcal{M}_0 = \mathbb{E}(Y)$,
- $\mathcal{M}_i(X_i) = \mathbb{E}(Y|X_i) - \mathbb{E}(Y)$,
- $i \neq j$ $\mathcal{M}_{i,j}(X_i, X_j) = \mathbb{E}(Y|X_i, X_j) - \mathbb{E}(Y|X_i) - \mathbb{E}(Y|X_j) + \mathbb{E}(Y)$,
- $\ldots$

$$\mathrm{Var}(Y) = \sum_{i=1}^{d} \mathrm{Var}(\mathcal{M}_i(X_i)) + \ldots + \mathrm{Var}(\mathcal{M}_{1,\ldots,d}(X_1, \ldots, X_d))$$

Definition (Sobol' indices)

$$\forall\, i = 1, \ldots, d \;\; S_i = \frac{\mathrm{Var}\left(\mathcal{M}_i(X_i)\right)}{\mathrm{Var}(Y)} = \frac{\mathrm{Var}\left[\mathbb{E}\left(Y|X_i\right)\right]}{\mathrm{Var}(Y)}$$

$$\forall\, i \neq j \;\; S_{i,j} = \frac{\mathrm{Var}\left(\mathcal{M}_{i,j}(X_i, X_j)\right)}{\mathrm{Var}(Y)} =$$

$$\frac{\mathrm{Var}\left[\mathbb{E}\left(Y|X_i, X_j\right)\right] - \mathrm{Var}\left[\mathbb{E}\left(Y|X_i\right)\right] - \mathrm{Var}\left[\mathbb{E}\left(Y|X_j\right)\right]}{\mathrm{Var}(Y)}$$

$$\ldots$$

$$1 = \sum_{i=1}^{d} S_i + \sum_{i \neq j} S_{i,j} + \ldots + S_{1,\ldots,d}$$

Sobol' indices :

Definition (Total indices)

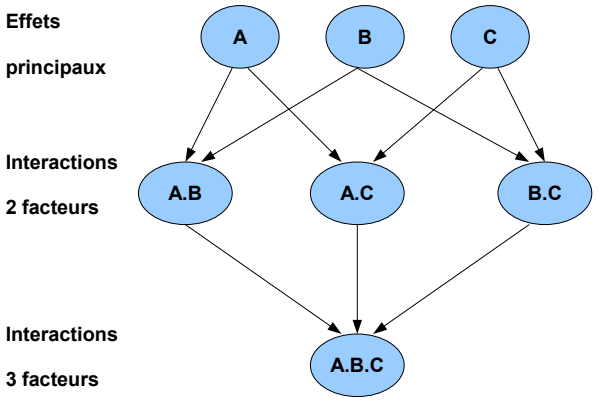
$$i = 1, \ldots, d \qquad S_i^{\text{tot}} = \sum_{\mathbf{v} \subseteq \{1,\ldots,d\}\,,\; i \in \mathbf{v}} S_{\mathbf{v}} \,.$$

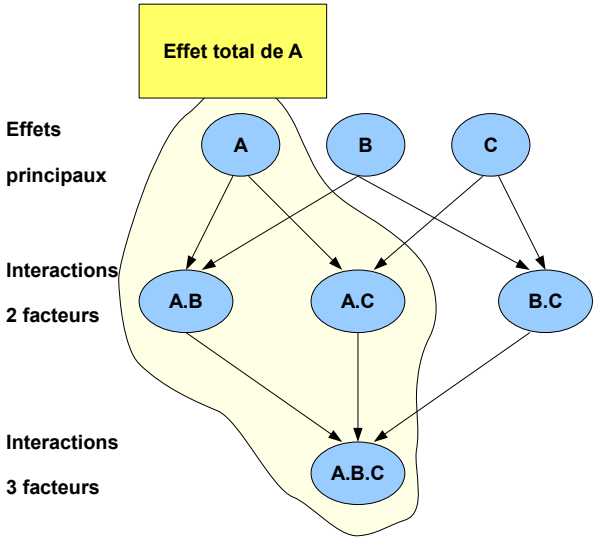
$X_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d)$

Using orthogonality of the decomposition combined with the total variance theorem, we prove

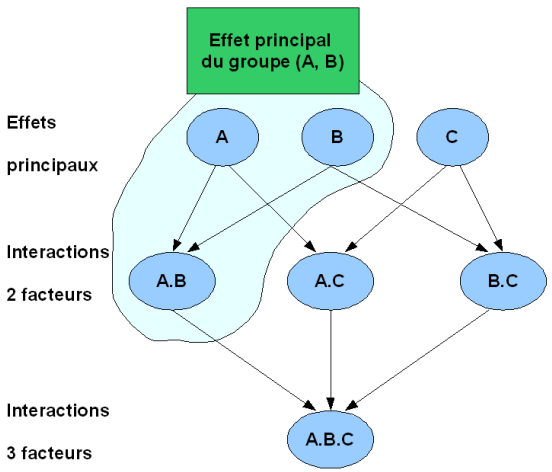
$$S_i^{\text{tot}} = \frac{\mathbb{E}\left[\mathrm{Var}\left(Y | X_{-i}\right)\right]}{\mathrm{Var}(Y)} = 1 - \frac{\mathrm{Var}\left[\mathbb{E}\left(Y | X_{-i}\right)\right]}{\mathrm{Var}(Y)} \,.$$

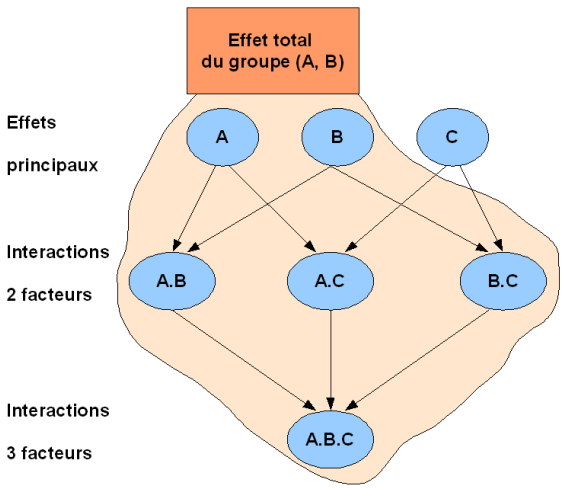
Indices with factor:





Indices with groupe of factors:





Fact : Analytical expressions of Sobol' indices, with integrals in high dimensional spaces, are rarely available.

We present different approaches for inference:

1. Pick-freeze estimators (hypothesis $\mathbb{L}^2$ with the model);
2. Given data estimators (under mild regularity assumptions on the model);
3. Spectral estimators (additional hypotheses of regularity).

If the model is too costly to assess, we fit a metamodel before applying these techniques.

ex.: parametric and non-parametric regressions, Gaussian metamodel...

Fact : Analytical expressions of Sobol' indices, with integrals in high dimensional spaces, are rarely available.

We present different approaches for inference:

1. Pick-freeze estimators (hypothesis $\mathbb{L}^2$ with the model);
2. Given data estimators (under mild regularity assumptions on the model);
3. Spectral estimators (additional hypotheses of regularity).

If the model is too costly to assess, we fit a metamodel before applying these techniques.

ex.: parametric and non-parametric regressions, Gaussian metamodel...

Fact : Analytical expressions of Sobol' indices, with integrals in high dimensional spaces, are rarely available.

We present different approaches for inference:

1. Pick-freeze estimators (hypothesis $\mathbb{L}^2$ with the model);
2. Given data estimators (under mild regularity assumptions on the model);
3. Spectral estimators (additional hypotheses of regularity).

If the model is too costly to assess, we fit a metamodel before applying these techniques.

ex.: parametric and non-parametric regressions, Gaussian metamodel...

Monte-Carlo type Approaches : (Sobol' 93, Saltelli 02, Mauntz, ...)

Idea : X'_{-i} indep. copy of X_{-i}, $Y = \mathcal{M}(X_i, X_{-i})$, $Y^i = \mathcal{M}(X_i, X'_{-i})$

We have $S_i = \frac{\mathrm{Cov}(Y, Y^i)}{\mathrm{Var}(Y)}$, the idea is based on empirical formulas.

Two independent samples A and B (Monte-Carlo, LHS)

$$A = \begin{pmatrix} x^A_{1,1} & \dots & x^A_{d,1} \\ \vdots & & \vdots \\ x^A_{1,n} & \dots & x^A_{d,n} \end{pmatrix} \qquad B = \begin{pmatrix} x^B_{1,1} & \dots & x^B_{d,1} \\ \vdots & & \vdots \\ x^B_{1,n} & \dots & x^B_{d,n} \end{pmatrix}$$

From A and of B, we create d sampling matrices C_i, $i = 1, \dots, d$:

$$C_i = \begin{pmatrix} x^A_{1,1} & \dots & x^B_{i,1} & \dots & x^A_{d,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x^A_{1,n} & \dots & x^B_{i,n} & \dots & x^A_{d,n} \end{pmatrix}$$

We compute $(1+d) \times n$ the model $\mathcal{M}$:

$$y^B = \begin{pmatrix} y_1^B \\ \cdot \\ \cdot \\ \cdot \\ y_n^B \end{pmatrix} \quad \text{and} \quad \forall\, 1 \leq i \leq d \quad y^{C_i} = \begin{pmatrix} y_1^{C_i} \\ \cdot \\ \cdot \\ \cdot \\ y_n^{C_i} \end{pmatrix}$$

`sobolEff()` (Janon *et al.*, 2014 & 2016)

- $\hat{V}_i = \frac{1}{n}\sum_{k=1}^{n} y_k^B y_k^{C_i} - \left(\frac{1}{n}\sum_{k=1}^{n} \frac{y_k^B + y_k^{C_i}}{2}\right)^2$ numerator of the first-order index
- $\hat{V} = \frac{1}{n}\sum_{k=1}^{n} \frac{(y_k^B)^2 + (y_k^{C_i})^2}{2} - \left(\frac{1}{n}\sum_{k=1}^{n} \frac{y_k^B + y_k^{C_i}}{2}\right)^2$ denominator

This type of estimators is known as pick-freeze estimators.

<u>Remarks:</u>

Pick-freeze estimators can be defined for any subset $\mathbf{u} \subseteq \{1, \ldots, d\}$.

In practice, we can replace MC or LHS samplings by QMC (hyp. of regular variations).

What about the statistical properties of pick-freeze estimators?

- Is it consistent? yes, proof by using the Strong Law of Large Numbers.
- If yes, at which rate of convergence? yes, CLT (cv in $\sqrt{n}$).
- Is it asymptotically efficient? yes.
- Is it possible to measure its performance for a fixed n? yes, Berry-Esseen and/or concentration inequalities.

see, Janon *et al.* (2014,2016) or Gamboa *et al.* (2014)

As an example, let us state in the next slide a central limit theorem. From such a CLT, one can also deduce asymptotic confidence intervals or hypothesis testing, e.g., on the nullity of Sobol' index associated to $\mathbf{u} \subseteq \{1, \ldots, d\}$.

$$S_{\mathbf{u}}^{\text{clo}} = \frac{\text{Var}\left[\mathbb{E}\left(Y|\mathbf{X}_{\mathbf{u}}\right)\right]}{\text{Var}[Y]}$$

$$\widehat{S_{\mathbf{u}}^{\text{clo}}} = \frac{\frac{1}{n}\sum_{k=1}^{n} Y_k^B Y_k^{C_{\mathbf{u}}} - \left(\frac{1}{n}\sum_{k=1}^{n} \frac{Y_k^B + Y_k^{C_{\mathbf{u}}}}{2}\right)^2}{\frac{1}{n}\sum_{k=1}^{n} \frac{(Y_k^B)^2 + (Y_k^{C_{\mathbf{u}}})^2}{2} - \left(\frac{1}{n}\sum_{k=1}^{n} \frac{Y_k^B + Y_k^{C_{\mathbf{u}}}}{2}\right)^2}.$$

Theorem (Janon *et al.*, 2014)

1. *One has* $\widehat{S_{\mathbf{u}}^{\text{clo}}} \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} S_{\mathbf{u}}^{\text{clo}}$.
2. *If* $\mathbb{E}\left(Y^4\right) < \infty$, *then* $\sqrt{n}\left(\widehat{S}_{\mathbf{u}}^{\text{clo}} - S_{\mathbf{u}}^{\text{clo}}\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}\left(0, \sigma_{\mathbf{u}}^2\right)$

   with $\sigma_{\mathbf{u}}^2 = \frac{\text{Var}\left[(Y-\mathbb{E}(Y))(Y_{\mathbf{u}}-\mathbb{E}(Y)) - \frac{S_{\mathbf{u}}^{\text{clo}}}{2}\left((Y-\mathbb{E}(Y))^2 + (Y_{\mathbf{u}}-\mathbb{E}(Y))^2\right]\right]}{(\text{Var}[Y])^2}$.

Using Bennett's concentration inequality, one gets for fixed sample size n:

Proposition (Janon *et al.*, 2016; Gamboa *et al.*, 2014)

Let $\mathbf{u}$ *be a subset of* $\{1, \ldots, d\}$*. Let* $b > 0$ *and* $t > 0$*. Let* $Y \in [-b, b]$*. Then,*

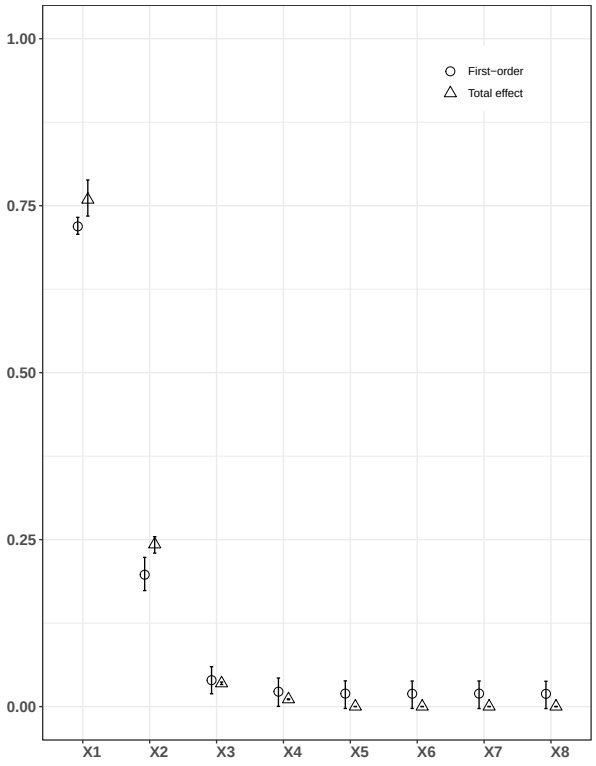
$$\mathbb{P}\left(\widehat{S_{\mathbf{u}}^{\text{clo}}} \geqslant S_{\mathbf{u}}^{\text{clo}} + t\right) \leqslant \exp\left(-\frac{n\mathrm{Var}[Y]^2}{128}\left(1 - \frac{1}{n}\right)^2\left(\frac{t}{1+t}\right)^2\right).$$

Assume further that $\frac{9}{8n} \leq t \leq 1$*, then*

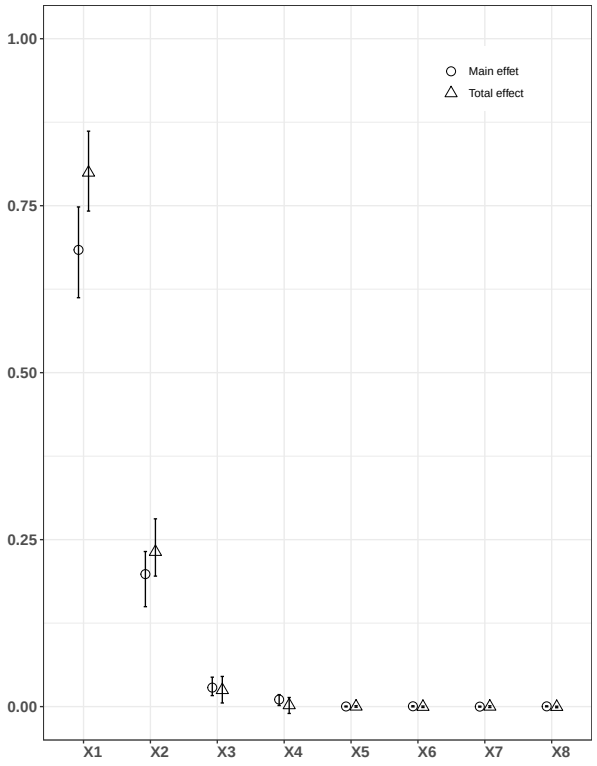
$$\mathbb{P}\left(\widehat{S_{\mathbf{u}}^{\text{clo}}} \leqslant S_{\mathbf{u}}^{\text{clo}} - t\right) \leqslant \exp\left(-\frac{n\mathrm{Var}[Y]^2}{128}\left(t - \frac{9}{8n}\right)^2\right).$$

The Sobol' g-function: $f(x) = \prod_{i=1}^{d} f_i(x_i)$ with $f_i(x_i) = \frac{|4x_i - 2| + a_i}{1 + a_i}$,

- $d = 8$,
- $a_1 = 0$, $a_2 = 1$, $a_3 = 4.5$, $a_4 = 9$, $a_i = 99$ for $5 \leq i \leq 8$,
- $n = 5000$, $b = 100$, $IC(0.95)$.

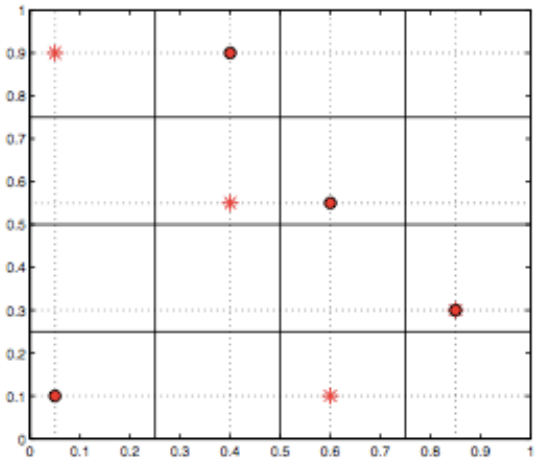


`sobolEff`



`sobol2007`

Replicated latin hypercubes: (Tissot *et al.*)



Definition (Replicated Latin Hypercube Sampling)

$k = 1, \ldots, n$

$$\mathbf{x}_k = \left(\frac{\pi_1(k) - U_{1,\pi_1(k)}}{n}, \ldots, \frac{\pi_d(k) - U_{d,\pi_d(k)}}{n} \right)$$

$$\tilde{\mathbf{x}}_k = \left(\frac{\tilde{\pi}_1(k) - U_{1,\tilde{\pi}_1(k)}}{n}, \ldots, \frac{\tilde{\pi}_d(k) - U_{d,\tilde{\pi}_d(k)}}{n} \right)$$

We have two matrices B and $\widetilde{B}$ at our disposal

$$B = \begin{pmatrix} x_{1,1} & \dots & x_{d,1} \\ \vdots & & \vdots \\ x_{1,n} & \dots & x_{d,n} \end{pmatrix} \qquad \widetilde{B} = \begin{pmatrix} \tilde{x}_{1,1} & \dots & \tilde{x}_{d,1} \\ \vdots & & \vdots \\ \tilde{x}_{1,n} & \dots & \tilde{x}_{d,n} \end{pmatrix}$$

We evaluate model $\mathcal{M}$ at 2n points (corresponding to the n lines of B, *resp.* $\widetilde{B}$). Then for $i_0 \in \{1, \dots, d\}$, we estimate S_{i_0} by permuting the lines of $\widetilde{B}$ as explained below.

Permutation of lines:

$$\begin{cases} \widetilde{B} = (\tilde{x}_{i,k})_{1 \le i \le d, 1 \le k \le n} & \mapsto & \widetilde{B}_{i_0} = (\tilde{x}^{i_0}_{i,k})_{1 \le i \le d, 1 \le k \le n} \\ (L_k,\ 1 \le k \le n) & \mapsto & (L_{\tilde{\pi}^{-1}_{i_0} \circ \pi_{i_0}(k)},\ 1 \le k \le n) \end{cases}$$

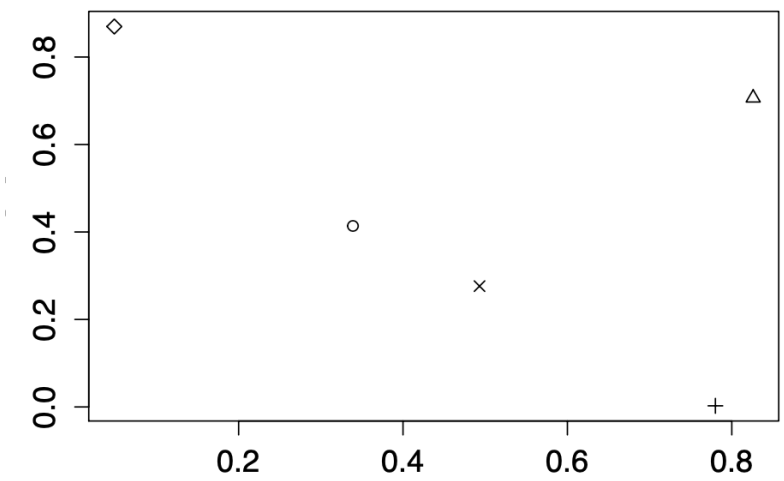
Then, $\tilde{x}^{i_0}_{i_0,k} = \tilde{x}_{i_0, \tilde{\pi}^{-1}_{i_0} \circ \pi_{i_0}(k)} = x_{i_0,k}$, $k = 1, \dots, n$. The column i_0 of $\widetilde{B}_{i_0}$ coincides with the one of B.

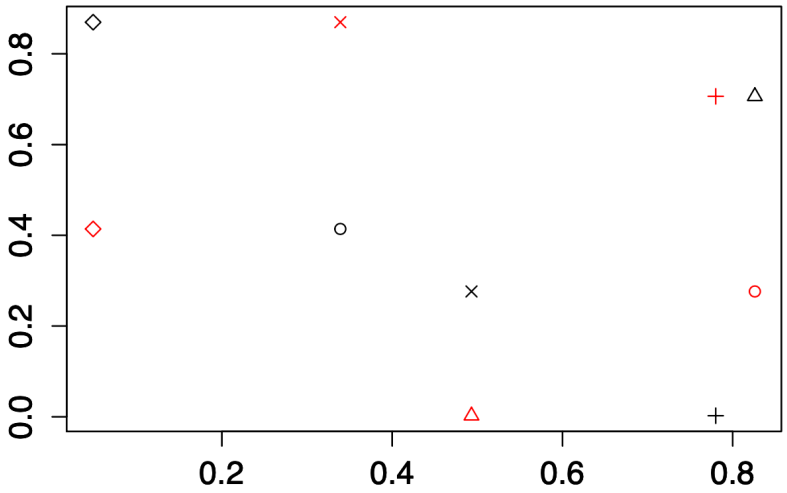
Thus to estimate S_{i_0}, we use B and $\widetilde{B}_{i_0}$ in the pick-freeze formula, instead of B and C_{i_0}.

II.1- Monte Carlo based Sobol' index inference

Caption:
point 1 $\circ$ point 2 $\triangle$ point 3 $+$ point 4 $\times$ point 5 $\diamond$

Design B (left), B and $\widetilde{B}$ (right)



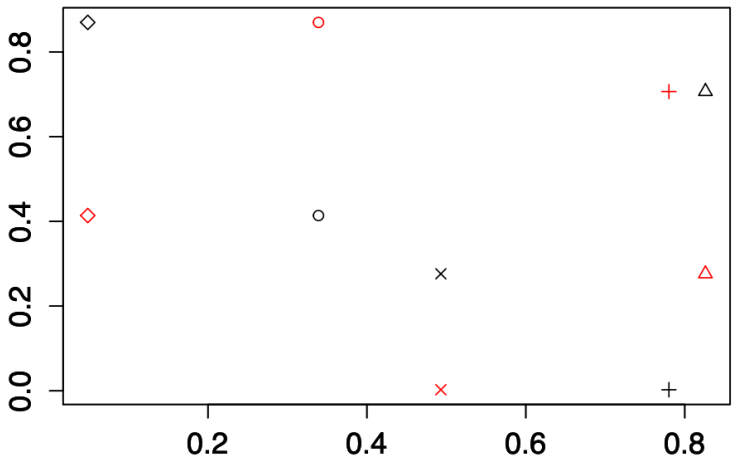


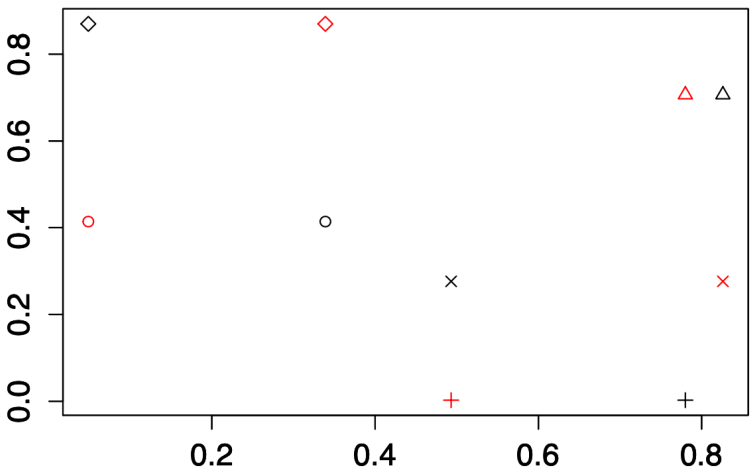
II- Monte Carlo based Sobol' index inference

Caption:

point 1 $\circ$ point 2 $\triangle$ point 3 $+$ point 4 $\times$ point 5 $\diamond$

Design B and $\widetilde{B}_1$ (left), B and $\widetilde{B}_2$ (right)





Asymptotic confidence intervals with variance smaller than for MC.

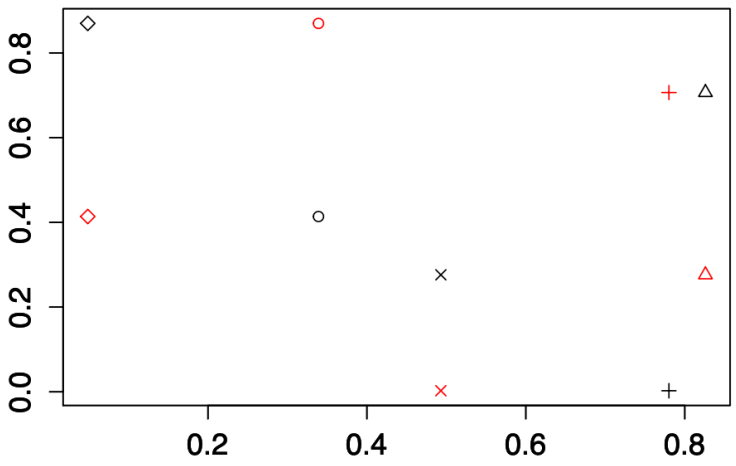
Possible extension to indices of order two (via orthogonal arrays of strength 2).

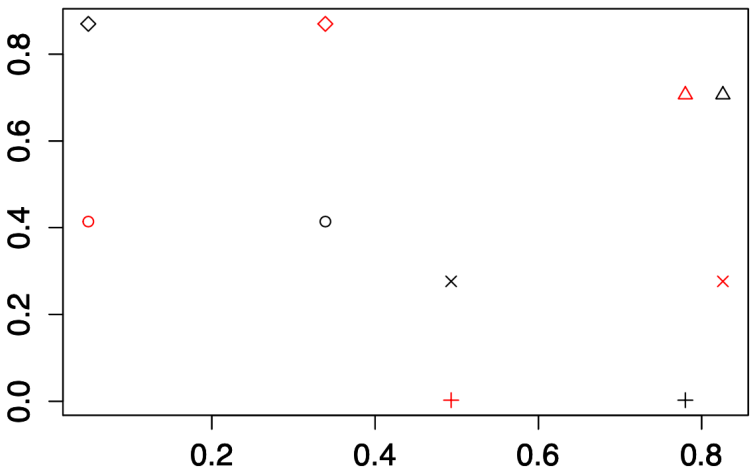
II- Monte Carlo based Sobol' index inference

Caption:

point 1 $\circ$ point 2 $\triangle$ point 3 $+$ point 4 $\times$ point 5 $\diamond$

Design B and $\widetilde{B}_1$ (left), B and $\widetilde{B}_2$ (right)





Asymptotic confidence intervals with variance smaller than for MC.

Possible extension to indices of order two (via orthogonal arrays of strength 2).

The trick cannot be used to estimate total Sobol' indices due to constraints inherent to the construction of OAs of strength $d-1$. If one wants to estimate total Sobol' indices, the best is to use Saltelli's trick (see Saltelli, 2002):

For $A \subseteq \{1, \ldots, d\}$, let us use the notation $U_A = \mathrm{Var}\left(\mathbb{E}(Y|\mathbf{X}_A)\right) + \mathbb{E}^2(Y)$ and $\mathbf{x}^{\sim A} = \left(\mathbf{x}_A, \mathbf{x}'_{-A}\right)$.

	$\mathbf{x}'$	$\mathbf{x}^{\sim 1}$	$\mathbf{x}^{\sim 2}$	$\mathbf{x}^{\sim 3}$	$\mathbf{x}^{\sim 4}$	$\mathbf{x}^{\sim \{2,3,4\}}$	$\mathbf{x}^{\sim \{1,3,4\}}$	$\mathbf{x}^{\sim \{1,2,4\}}$	$\mathbf{x}^{\sim \{1,2,3\}}$	$\mathbf{x}^{\sim \{1,2,3,4\}}$
$\mathbf{x}'$	V									
$\mathbf{x}^{\sim 1}$	U_{-1}	V								
$\mathbf{x}^{\sim 2}$	U_{-2}	U_{-12}	V							
$\mathbf{x}^{\sim 3}$	U_{-3}	U_{-13}	U_{-23}	V						
$\mathbf{x}^{\sim 4}$	U_{-4}	U_{-14}	U_{-24}	U_{-34}	V					
$\mathbf{x}^{\sim \{2,3,4\}}$	U_1		U_{12}	U_{13}	U_{14}	V				
$\mathbf{x}^{\sim \{1,3,4\}}$	U_2	U_{12}		U_{23}	U_{24}	U_{-12}	V			
$\mathbf{x}^{\sim \{1,2,4\}}$	U_3	U_{13}	U_{23}		U_{34}	U_{-13}	U_{-23}	V		
$\mathbf{x}^{\sim \{1,2,3\}}$	U_4	U_{14}	U_{24}	U_{34}		U_{-14}	U_{-24}	U_{-34}	V	
$\mathbf{x}^{\sim \{1,2,3,4\}}$	$\mathbb{E}^2\ Y$	U_1	U_2	U_3	U_4	U_{-1}	U_{-2}	U_{-3}	U_{-4}	V

Table: The table gives for each cell the term that can be estimated by evaluating the model on the corresponding input vectors, $d = 4$. For example, U_{-12} can be estimated from the evaluation of the model on two n-samples $\mathbf{x}^{\sim 2}$ and $\mathbf{x}^{\sim 1}$.

In conclusion, Saltelli's trick lead to the estimation of all first-order and total indices at a cost of $n(d+2)$ model evaluations and to the estimation of all first-order, second-order and total indices at a cost of $n(2d+2)$ model evaluations.

Conclusions about Monte Carlo type inference :

We recommend the following (see Gilquin *et al.*, 2019):

First and second order Sobol' indices: R package `sensitivity`, function `sobolrep` with `total=FALSE`.
The cost is $2n$ with $n = q^2$, q a prime number.

First, second order and total Sobol' indices: R package `sensitivity`, function `sobolrep` with `total=TRUE`.
The cost is $n(d+2)$ with $n = q^2$, q a prime number (see Gilquin *et al.*, 2019).

Pick-freeze estimator is based on a specific design of experiments 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, that is a given data estimator.

Let us present rank estimator of S_1 from Gamboa *et al.* (2021) .

Let's consider a n-sample of the input/output pair (X_1, Y) given by $(X_{1,1}, Y_1), \ldots, (X_{1,n}, Y_n)$.

The pairs $(X_{1,(1)}, Y_{(1)}), \ldots, (X_{1,(n)}, Y_{(n)})$ 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: $(-2, 3), (0, 8), (1, 5), (2, 9), (6, -4)$

Pick-freeze estimator is based on a specific design of experiments 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, that is a given data estimator.

Let us present rank estimator of S_1 from Gamboa *et al.* (2021) .

Let's consider a n-sample of the input/output pair (X_1, Y) given by $(X_{1,1}, Y_1), \ldots, (X_{1,n}, Y_n)$.

The pairs $(X_{1,(1)}, Y_{(1)}), \ldots, (X_{1,(n)}, Y_{(n)})$ 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: $(-2, 3), (0, 8), (1, 5), (2, 9), (6, -4)$

We define $Y_{(n+1)} = Y_{(1)}$. We then introduce

$$\widehat{S_1}^{\text{rank}} = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_{(i)} Y_{(i+1)} - \left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2}.$$

Theorem (Gamboa *et al.*, 2021, see also Chatterjee, 2020)

1. *Assume that $X_i \sim \mathcal{U}[0,1])$, $i = 1, \ldots, n$ and that $\mathcal{M}$ is bounded. One has $\widehat{S_1}^{\text{rank}} \xrightarrow[n\to\infty]{a.s.} S_1$.*
2. *Assume that $X_i \sim \mathcal{U}[0,1])$, $i = 1, \ldots, n$ and that $\mathcal{M}$ is twice differentiable wrt its first coordinate with bounded first derivatives. Then*

$$\sqrt{n}\left(\widehat{S_1}^{\text{rank}} - S_1\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(0, \sigma^2_{\text{rank}}\right).$$

Rank estimator is limited to first-order Sobol' index estimation.

In Broto *et al.* (2020), the authors propose a given data estimator based on **nearest neighbors**. This estimator can be defined for any order of interaction.

Consistency is proven under regularity assumptions on the model. No CLT is proven.

Devroye *et al.* (2013,2020) propose an estimator of $\mathbb{E}[\mathbb{E}[Y|X]^2]$ based on two independent n-samples:

- the first one among which the first nearest neighbor of x is searched to estimate $x \mapsto \mathbb{E}[Y|X = x]$,
- the second one to build a plug-in estimator.

They prove a CLT ($\sqrt{n}$). Nevertheless the recentering factor is not $\mathbb{E}[\mathbb{E}[Y|X]^2]$ but the expected value of their estimator. In addition, the bias is negligible if and only if the dimension $d \leqslant 3$.

Ishigami toy model: $\mathcal{M}(x) = \sin(x_1) + 7\sin^2(x_2) + 0.1x_3^4\sin(x_1)$, $X_i \sim \mathcal{U}([-\pi, \pi])$, $i = 1, 2, 3$.

We compare `sobolrank` with `sobolrep` with $2 \times n = 2 \times 19^2 = 2 \times 361 = 722$ model evaluations, $n_{\text{rep}} = 100$. Root mean square errors are computed with 100 samples.

`sobolrank`	0.03635195	0.03440188	0.04715759
`sobolrep`	0.04199731	0.04436713	0.07468821

For the same number of model evaluations, `sobolrep` also provides second-order Sobol' indices. However it requires a pick-freeze design based on replicated OAs of strength 2.

For sake of clarity in the presentation, we consider the case $d = 2$.

$Y = \sum_{\mathbf{k}=(k_1,k_2)\in\mathbb{Z}^2} c_{\mathbf{k}}(\mathcal{M})\Phi_{1,k_1}(X_1)\Phi_{2,k_2}(X_2)$

with , for all $i = 1, 2$, $\left(\Phi_{i,k}\right)_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathbb{L}^2([0,1])$ and $\Phi_{i,0} \equiv 1$.

$\mathcal{M}_0 = c_0(\mathcal{M})$,

$\mathcal{M}_1(X_1) = \sum_{k_1\in\mathbb{Z}^*} c_{k_1,0}(\mathcal{M})\Phi_{1,k_1}(X_1)$,
$\mathcal{M}_2(X_2) = \sum_{k_2\in\mathbb{Z}^*} c_{0,k_2}(\mathcal{M})\Phi_{2,k_2}(X_2)$,

$\mathcal{M}_{1,2}(X_1, X_2) = \sum_{k_1\in\mathbb{Z}^*,k_2\in\mathbb{Z}^*} c_{k_1,k_2}(\mathcal{M})\Phi_{1,k_1}(X_1)\Phi_{2,k_2}(X_2)$.

We have with Parseval identity:

- $\mathrm{Var}\left(\mathcal{M}_1(X_1)\right) = \sigma_1^2 = \sum_{k_1\in\mathbb{Z}^*} |c_{k_1,0}(\mathcal{M})|^2$, (idem for σ_2^2),
- $\mathrm{Var}\left(\mathcal{M}_{1,2}(X_1, X_2)\right) = \sigma_{1,2}^2 = \sum_{k_1\in\mathbb{Z}^*,k_2\in\mathbb{Z}^*} |c_{k_1,k_2}(\mathcal{M})|^2$,
- $\mathrm{Var}\left(Y\right) = \sigma^2 = \sum_{(k_1,k_2)\in\mathbb{Z}\times\mathbb{Z},(k_1,k_2)\neq(0,0)} |c_{k_1,k_2}(\mathcal{M})|^2$.

ex. : orthogonal polynomials, wavelet basis, Fourier basis.

For sake of clarity in the presentation, we consider the case $d = 2$.

$Y = \sum_{\mathbf{k}=(k_1,k_2)\in\mathbb{Z}^2} c_{\mathbf{k}}(\mathcal{M})\Phi_{1,k_1}(X_1)\Phi_{2,k_2}(X_2)$

with , for all $i = 1, 2$, $\left(\Phi_{i,k}\right)_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathbb{L}^2([0,1])$ and $\Phi_{i,0} \equiv 1$.

$\mathcal{M}_0 = c_0(\mathcal{M})$,

$\mathcal{M}_1(X_1) = \sum_{k_1\in\mathbb{Z}^*} c_{k_1,0}(\mathcal{M})\Phi_{1,k_1}(X_1)$,
$\mathcal{M}_2(X_2) = \sum_{k_2\in\mathbb{Z}^*} c_{0,k_2}(\mathcal{M})\Phi_{2,k_2}(X_2)$,

$\mathcal{M}_{1,2}(X_1, X_2) = \sum_{k_1\in\mathbb{Z}^*,k_2\in\mathbb{Z}^*} c_{k_1,k_2}(\mathcal{M})\Phi_{1,k_1}(X_1)\Phi_{2,k_2}(X_2)$.

We have with Parseval identity:

- $\mathrm{Var}\left(\mathcal{M}_1(X_1)\right) = \sigma_1^2 = \sum_{k_1\in\mathbb{Z}^*} |c_{k_1,0}(\mathcal{M})|^2$, (idem for σ_2^2),
- $\mathrm{Var}\left(\mathcal{M}_{1,2}(X_1, X_2)\right) = \sigma_{1,2}^2 = \sum_{k_1\in\mathbb{Z}^*,k_2\in\mathbb{Z}^*} |c_{k_1,k_2}(\mathcal{M})|^2$,
- $\mathrm{Var}\left(Y\right) = \sigma^2 = \sum_{(k_1,k_2)\in\mathbb{Z}\times\mathbb{Z},(k_1,k_2)\neq(0,0)} |c_{k_1,k_2}(\mathcal{M})|^2$.

ex. : orthogonal polynomials, wavelet basis, Fourier basis.

For sake of clarity in the presentation, we consider the case $d = 2$.

$Y = \sum_{\mathbf{k}=(k_1,k_2)\in\mathbb{Z}^2} c_{\mathbf{k}}(\mathcal{M})\Phi_{1,k_1}(X_1)\Phi_{2,k_2}(X_2)$

with , for all $i = 1, 2$, $\left(\Phi_{i,k}\right)_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathbb{L}^2([0,1])$ and $\Phi_{i,0} \equiv 1$.

$\mathcal{M}_0 = c_0(\mathcal{M})$,

$\mathcal{M}_1(X_1) = \sum_{k_1\in\mathbb{Z}^*} c_{k_1,0}(\mathcal{M})\Phi_{1,k_1}(X_1)$,
$\mathcal{M}_2(X_2) = \sum_{k_2\in\mathbb{Z}^*} c_{0,k_2}(\mathcal{M})\Phi_{2,k_2}(X_2)$,

$\mathcal{M}_{1,2}(X_1, X_2) = \sum_{k_1\in\mathbb{Z}^*,k_2\in\mathbb{Z}^*} c_{k_1,k_2}(\mathcal{M})\Phi_{1,k_1}(X_1)\Phi_{2,k_2}(X_2)$.

We have with Parseval identity:

- $\mathrm{Var}\left(\mathcal{M}_1(X_1)\right) = \sigma_1^2 = \sum_{k_1\in\mathbb{Z}^*} |c_{k_1,0}(\mathcal{M})|^2$, (idem for σ_2^2),
- $\mathrm{Var}\left(\mathcal{M}_{1,2}(X_1, X_2)\right) = \sigma_{1,2}^2 = \sum_{k_1\in\mathbb{Z}^*,k_2\in\mathbb{Z}^*} |c_{k_1,k_2}(\mathcal{M})|^2$,
- $\mathrm{Var}\left(Y\right) = \sigma^2 = \sum_{(k_1,k_2)\in\mathbb{Z}\times\mathbb{Z},(k_1,k_2)\neq(0,0)} |c_{k_1,k_2}(\mathcal{M})|^2$.

ex. : orthogonal polynomials, wavelet basis, Fourier basis.

For sake of clarity in the presentation, we consider the case $d = 2$.

$Y = \sum_{\mathbf{k}=(k_1,k_2)\in\mathbb{Z}^2} c_{\mathbf{k}}(\mathcal{M})\Phi_{1,k_1}(X_1)\Phi_{2,k_2}(X_2)$

with , for all $i = 1, 2$, $\left(\Phi_{i,k}\right)_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathbb{L}^2([0,1])$ and $\Phi_{i,0} \equiv 1$.

$\mathcal{M}_0 = c_0(\mathcal{M})$,

$\mathcal{M}_1(X_1) = \sum_{k_1\in\mathbb{Z}^*} c_{k_1,0}(\mathcal{M})\Phi_{1,k_1}(X_1)$,
$\mathcal{M}_2(X_2) = \sum_{k_2\in\mathbb{Z}^*} c_{0,k_2}(\mathcal{M})\Phi_{2,k_2}(X_2)$,

$\mathcal{M}_{1,2}(X_1, X_2) = \sum_{k_1\in\mathbb{Z}^*,k_2\in\mathbb{Z}^*} c_{k_1,k_2}(\mathcal{M})\Phi_{1,k_1}(X_1)\Phi_{2,k_2}(X_2)$.

We have with Parseval identity:

- $\mathrm{Var}\left(\mathcal{M}_1(X_1)\right) = \sigma_1^2 = \sum_{k_1\in\mathbb{Z}^*} |c_{k_1,0}(\mathcal{M})|^2$, (idem for σ_2^2),
- $\mathrm{Var}\left(\mathcal{M}_{1,2}(X_1, X_2)\right) = \sigma_{1,2}^2 = \sum_{k_1\in\mathbb{Z}^*,k_2\in\mathbb{Z}^*} |c_{k_1,k_2}(\mathcal{M})|^2$,
- $\mathrm{Var}\left(Y\right) = \sigma^2 = \sum_{(k_1,k_2)\in\mathbb{Z}\times\mathbb{Z},(k_1,k_2)\neq(0,0)} |c_{k_1,k_2}(\mathcal{M})|^2$.

<u>ex. :</u> orthogonal polynomials, wavelet basis, Fourier basis.

Inference scheme:

If D is an experimental design with $[0,1]^2$, we propose the quadrature formula:

$$\hat{c}_{k_1,k_2}(\mathcal{M}, D) = \frac{1}{\text{card} D} \sum_{\mathbf{x}=(x_1,x_2)\in D} \mathcal{M}(x) e^{-2i\pi(k_1 x_1 + k_2 x_2)} .$$

We then infer each part of variance with a truncation:

- $\hat{\sigma}_1^2(\mathcal{M}, K_1, D) = \sum_{k_1 \in K_1} |\hat{c}_{k_1,0}(\mathcal{M}, D)|^2$, with $K_1 \subset \mathbb{Z}^*$ of finite cardinal, (idem for $\hat{\sigma}_2^2$),
- $\hat{\sigma}_{1,2}^2(\mathcal{M}, K_{1,2}, D) = \sum_{(k_1,k_2)\in K_{1,2}} |\hat{c}_{k_1,k_2}(\mathcal{M}, D)|^2$, with $K_{1,2} \subset \mathbb{Z}^* \times \mathbb{Z}^*$ of finite cardinal.

We infer the total variance with $\hat{\sigma}^2(\mathcal{M}, D) = \hat{c}_{0,0}(\mathcal{M}^2, D) - \hat{c}_{0,0}(\mathcal{M}, D)^2$.

The estimators of Sobol' indices can be written as:

$$\hat{S}_i = \frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} , \quad i = 1, 2 , \quad S_{1,2} = \frac{\hat{\sigma}_{1,2}^2}{\hat{\sigma}^2} .$$

Inference scheme:

If D is an experimental design with $[0,1]^2$, we propose the quadrature formula:

$$\hat{c}_{k_1,k_2}(\mathcal{M}, D) = \frac{1}{\text{card} D} \sum_{\mathbf{x}=(x_1,x_2)\in D} \mathcal{M}(x) e^{-2i\pi(k_1 x_1 + k_2 x_2)} .$$

We then infer each part of variance with a truncation:

- $\hat{\sigma}_1^2(\mathcal{M}, K_1, D) = \sum_{k_1 \in K_1} |\hat{c}_{k_1,0}(\mathcal{M}, D)|^2$, with $K_1 \subset \mathbb{Z}^*$ of finite cardinal, (idem for $\hat{\sigma}_2^2$),
- $\hat{\sigma}_{1,2}^2(\mathcal{M}, K_{1,2}, D) = \sum_{(k_1,k_2)\in K_{1,2}} |\hat{c}_{k_1,k_2}(\mathcal{M}, D)|^2$, with $K_{1,2} \subset \mathbb{Z}^* \times \mathbb{Z}^*$ of finite cardinal.

We infer the total variance with $\hat{\sigma}^2(\mathcal{M}, D) = \hat{c}_{0,0}(\mathcal{M}^2, D) - \hat{c}_{0,0}(\mathcal{M}, D)^2$.

The estimators of Sobol' indices can be written as:

$$\hat{S}_i = \frac{\hat{\sigma}_i^2}{\hat{\sigma}^2}, \quad i = 1, 2, \quad S_{1,2} = \frac{\hat{\sigma}_{1,2}^2}{\hat{\sigma}^2} .$$

Inference scheme:

If D is an experimental design with $[0,1]^2$, we propose the quadrature formula:

$$\hat{c}_{k_1,k_2}(\mathcal{M}, D) = \frac{1}{\text{card} D} \sum_{\mathbf{x}=(x_1,x_2)\in D} \mathcal{M}(x) e^{-2i\pi(k_1 x_1 + k_2 x_2)} .$$

We then infer each part of variance with a truncation:

- $\hat{\sigma}_1^2(\mathcal{M}, K_1, D) = \sum_{k_1 \in K_1} |\hat{c}_{k_1,0}(\mathcal{M}, D)|^2$, with $K_1 \subset \mathbb{Z}^*$ of finite cardinal, (idem for $\hat{\sigma}_2^2$),
- $\hat{\sigma}_{1,2}^2(\mathcal{M}, K_{1,2}, D) = \sum_{(k_1,k_2)\in K_{1,2}} |\hat{c}_{k_1,k_2}(\mathcal{M}, D)|^2$, with $K_{1,2} \subset \mathbb{Z}^* \times \mathbb{Z}^*$ of finite cardinal.

We infer the total variance with $\hat{\sigma}^2(\mathcal{M}, D) = \hat{c}_{0,0}(\mathcal{M}^2, D) - \hat{c}_{0,0}(\mathcal{M}, D)^2$.

The estimators of Sobol' indices can be written as:

$$\hat{S}_i = \frac{\hat{\sigma}_i^2}{\hat{\sigma}^2}, \quad i = 1, 2, \quad S_{1,2} = \frac{\hat{\sigma}_{1,2}^2}{\hat{\sigma}^2} .$$

Inference scheme:

If D is an experimental design with $[0,1]^2$, we propose the quadrature formula:

$$\hat{c}_{k_1,k_2}(\mathcal{M}, D) = \frac{1}{\text{card} D} \sum_{\mathbf{x}=(x_1,x_2)\in D} \mathcal{M}(x) e^{-2i\pi(k_1 x_1 + k_2 x_2)} .$$

We then infer each part of variance with a truncation:

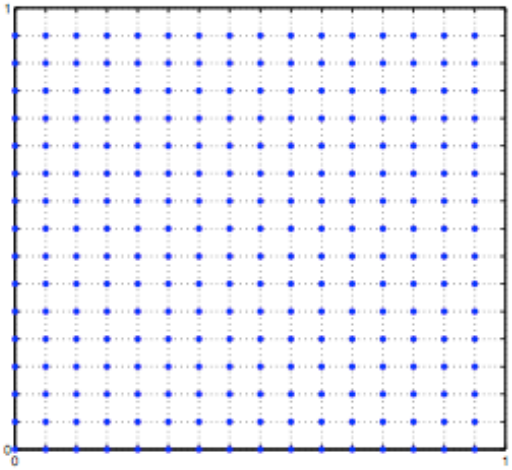
- $\hat{\sigma}_1^2(\mathcal{M}, K_1, D) = \sum_{k_1 \in K_1} |\hat{c}_{k_1,0}(\mathcal{M}, D)|^2$, with $K_1 \subset \mathbb{Z}^*$ of finite cardinal, (idem for $\hat{\sigma}_2^2$),
- $\hat{\sigma}_{1,2}^2(\mathcal{M}, K_{1,2}, D) = \sum_{(k_1,k_2)\in K_{1,2}} |\hat{c}_{k_1,k_2}(\mathcal{M}, D)|^2$, with $K_{1,2} \subset \mathbb{Z}^* \times \mathbb{Z}^*$ of finite cardinal.

We infer the total variance with $\hat{\sigma}^2(\mathcal{M}, D) = \hat{c}_{0,0}(\mathcal{M}^2, D) - \hat{c}_{0,0}(\mathcal{M}, D)^2$.

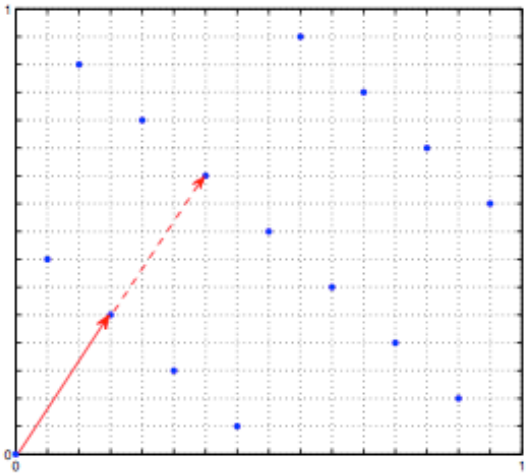
The estimators of Sobol' indices can be written as:

$$\hat{S}_i = \frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} , \ i = 1, 2 , \quad S_{1,2} = \frac{\hat{\sigma}_{1,2}^2}{\hat{\sigma}^2} .$$

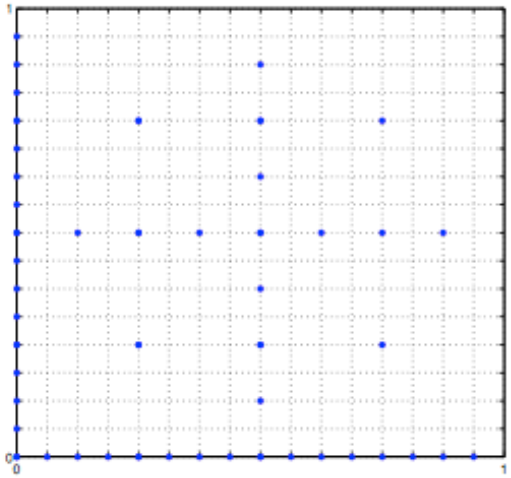
Classical designs:



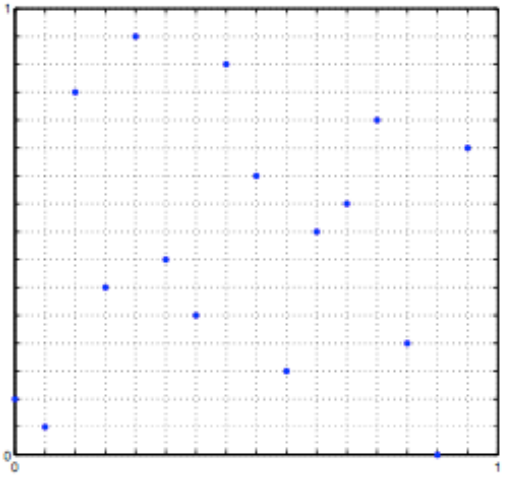
(a) grille régulière



(b) sous-groupe fini



(c) grille creuse



(d) tableau orthogonal

The performance of previous estimators is linked to the decreasing speed of Fourier spectrum (regularity) of $\mathcal{M}$. The techniques FAST and RBD are two particular cases of such approaches (after model regularisation). See Tissot & Prieur (2012) or Prieur & Tarantola (2017) for a review.

FAST: (Cukier *et al.*, 78) *Fourier Amplitude Sensitivity Test*

• we fix $K_{\mathbf{u}}$ an ensemble of a priori non negligible frequencies;

• we chose D cyclic group (design (b)) in order to control the quadrature error.

Remarks:

- if $\mathcal{M}$ regular, we can obtain a speed of convergence $>> \sqrt{n}$;
- for the total indices `fast99()` (no confidence intervals in the function) Saltelli *et al.*, 99.

The performance of previous estimators is linked to the decreasing speed of Fourier spectrum (regularity) of $\mathcal{M}$. The techniques FAST and RBD are two particular cases of such approaches (after model regularisation). See Tissot & Prieur (2012) or Prieur & Tarantola (2017) for a review.

FAST: (Cukier *et al.*, 78) *Fourier Amplitude Sensitivity Test*

• we fix $K_{\mathbf{u}}$ an ensemble of a priori non negligible frequencies;

• we chose D cyclic group (design (b)) in order to control the quadrature error.

Remarks:

- if $\mathcal{M}$ regular, we can obtain a speed of convergence $>> \sqrt{n}$;
- for the total indices `fast99()` (no confidence intervals in the function) Saltelli *et al.*, 99.

The performance of previous estimators is linked to the decreasing speed of Fourier spectrum (regularity) of $\mathcal{M}$. The techniques FAST and RBD are two particular cases of such approaches (after model regularisation). See Tissot & Prieur (2012) or Prieur & Tarantola (2017) for a review.

FAST: (Cukier *et al.*, 78) *Fourier Amplitude Sensitivity Test*

• we fix $K_{\mathbf{u}}$ an ensemble of a priori non negligible frequencies;

• we chose D cyclic group (design (b)) in order to control the quadrature error.

Remarks:

- if $\mathcal{M}$ regular, we can obtain a speed of convergence $>> \sqrt{n}$;
- for the total indices `fast99()` (no confidence intervals in the function) Saltelli *et al.*, 99.

The performance of previous estimators is linked to the decreasing speed of Fourier spectrum (regularity) of $\mathcal{M}$. The techniques FAST and RBD are two particular cases of such approaches (after model regularisation). See Tissot & Prieur (2012) or Prieur & Tarantola (2017) for a review.

FAST: (Cukier *et al.*, 78) *Fourier Amplitude Sensitivity Test*

• we fix $K_{\mathbf{u}}$ an ensemble of a priori non negligible frequencies;

• we chose D cyclic group (design (b)) in order to control the quadrature error.

Remarks:

- if $\mathcal{M}$ regular, we can obtain a speed of convergence $>> \sqrt{n}$;
- for the total indices `fast99()` (no confidence intervals in the function) Saltelli *et al.*, 99.

RBD: (Tarantola *et al.*, 06) *Random Balance Designs*

• we choose D an orthogonal array of strength 1 (design (d)), randomized by a random permutation ($D(\pi)$));

• $K_{\mathbf{u}}$ choice of a priori non negligible frequencies.

Remarks:

- these estimators are known to be biased;
- we can correct a part of this biais (Tissot *et al.*, 2012);
- if the function is not regular enough, the bias remains important.

RBD: (Tarantola *et al.*, 06) *Random Balance Designs*

• we choose D an orthogonal array of strength 1 (design (d)), randomized by a random permutation ($D(\pi)$));

• $K_{\mathbf{u}}$ choice of a priori non negligible frequencies.

Remarks:

- these estimators are known to be biased;
- we can correct a part of this biais (Tissot *et al.*, 2012);
- if the function is not regular enough, the bias remains important.

RBD: (Tarantola *et al.*, 06) *Random Balance Designs*

• we choose D an orthogonal array of strength 1 (design (d)), randomized by a random permutation ($D(\pi)$));

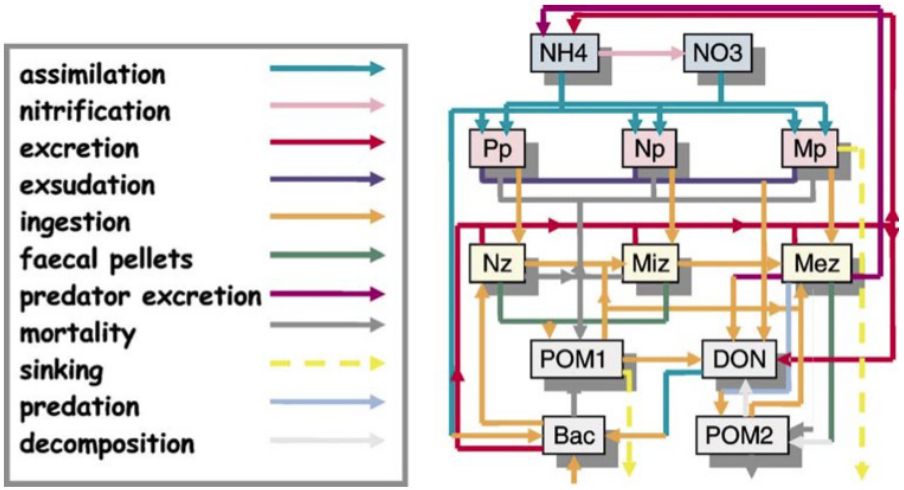
• $K_{\mathbf{u}}$ choice of a priori non negligible frequencies.

Remarks:

- these estimators are known to be biased;
- we can correct a part of this biais (Tissot *et al.*, 2012);
- if the function is not regular enough, the bias remains important.

see Prieur *et al*, 2019

MODECOGeL is a one-dimensional coupled hydrodynamical-biological model.



• hydrodynamic model: 1-D vertical simplification of primitive equations for the ocean, 5 state variables;
• ecosystem model: marine biogeochemistry, 12 biological state variables.

▷ 74 independent scalar parameters

Index	Name	Unit	Pdf	Mean	Std	Std/Mean
1	**PicP** max growth rate	t^{-1}	$\Gamma(25, 0.12)$	3.	0.6	20%
2	**NanP** max growth rate	t^{-1}	$\Gamma(25, 0.1)$	2.5	0.5	20%
3	**MicP** max growth rate	t^{-1}	$\Gamma(25, 0.08)$	2.	0.4	20%
4	dependence of **NO3** limitation to **NH4**	C^{-1}	$\Gamma(400, 0.00365)$	1.46	0.073	5%
5	**NO3** semisaturation for **PicP**	C	$\Gamma(4, 0.125)$	0.5	0.25	50%
6	**NO3** semisaturation for **NanP**	C	$\Gamma(4, 0.175)$	0.7	0.35	50%
7	**NO3** semisaturation for **MicP**	C	$\Gamma(4, 0.25)$	1.0	0.5	50%
8	**NH4** semisaturation for **PicP**	C	$\Gamma(4, 0.075)$	0.3	0.15	50%
9	**NH4** semisaturation for **NanP**	C	$\Gamma(4, 0.125)$	0.5	0.25	50%
10	**NH4** semisaturation for **MicP**	C	$\Gamma(4, 0.175)$	0.7	0.35	50%
11	optimal PAR for **PicP**	I	$\Gamma(25, 0.4)$	10.	2.	20%
12	optimal PAR for **NanP**	I	$\Gamma(25, 0.6)$	15.	3.	20%
13	optimal PAR for **MicP**	I	$\Gamma(25, 0.8)$	20.	4.	20%
14	variation of light limitation for **PicP**	–	$-\Gamma(4, 0.2)$	-0.8	0.4	50%
15	variation of light limitation for **NanP**	–	$-\Gamma(4, 0.175)$	-0.7	0.35	50%
16	variation of light limitation for **MicP**	–	$-\Gamma(4, 0.15)$	-0.6	0.3	50%
17	optimal temperature for **PicP**	T	$\mathcal{N}(15, 3^2)$	15.	3.	20%
18	optimal temperature for **NanP**	T	$\mathcal{N}(15, 3^2)$	15.	3.	20%
19	optimal temperature for **MicP**	T	$\mathcal{N}(16, 3.2^2)$	16.	3.2	20%
20	variation of temp. limitation for **PicP**	–	$-\Gamma(4, 0.125)$	-0.5	0.25	50%
21	variation of temp. limitation for **NanP**	–	$-\Gamma(4, 0.125)$	-0.5	0.25	50%
22	variation of temp. limitation for **MicP**	–	$-\Gamma(4, 0.1375)$	-0.55	0.275	50%
23	bacteria growth limitation	–	$\Gamma(4, 0.15)$	0.6	0.3	50%
24	semisaturation for **BAC** growth	C	$\Gamma(4, 0.125)$	0.5	0.25	50%
25	exudation ratio for **PicP**	–	$\Gamma(4, 0.015)$	0.06	0.03	50%
26	exudation ratio for **NanP**	–	$\Gamma(4, 0.0125)$	0.05	0.025	50%
27	exudation ratio for **MicP**	–	$\Gamma(4, 0.01)$	0.04	0.02	50%
28	max ingestion rate for **NanZ**	t^{-1}	$\Gamma(25, 0.12)$	3.	0.6	20%
29	max ingestion rate for **MicZ**	t^{-1}	$\Gamma(25, 0.08)$	2.	0.4	20%
30	max ingestion rate for **MesZ**	t^{-1}	$\Gamma(25, 0.06)$	1.5	0.3	20%
31	threshold ingestion for **NanZ**	C	$\Gamma(4, 0.0125)$	0.05	0.025	50%
32	threshold ingestion for **MicZ**	C	$\Gamma(4, 0.0075)$	0.03	0.015	50%
33	threshold ingestion for **MesZ**	C	$\Gamma(4, 0.0025)$	0.01	0.005	50%
34	semisaturation for ingestion by **NanZ**	C	$\Gamma(4, 0.125)$	0.5	0.25	50%
35	semisaturation for ingestion by **MicZ**	C	$\Gamma(4, 0.1875)$	0.75	0.375	50%
36	semisaturation for ingestion by **MesZ**	C	$\Gamma(4, 0.25)$	1.	0.5	50%
37	efficiency of **MesZ** on **MicP**	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
38	efficiency of **NanZ** on **BAC**	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
39	efficiency of **MicZ** on **NanZ**	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
40	efficiency of **MesZ** on **MicZ**	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
41	efficiency of **MicZ** on **MOP1**	–	$\beta(19.8, 79.2)$	0.2	0.04	20%
42	efficiency of **MesZ** on **MOP1**	–	$\beta(19.8, 79.2)$	0.2	0.04	20%
43	efficiency of **MesZ** on **MOP2**	–	$\beta(19.8, 79.2)$	0.2	0.04	20%
44	mortality rate for **PicP**	t^{-1}	$\Gamma(4, 0.015)$	0.06	0.03	50%
45	mortality rate for **NanP**	t^{-1}	$\Gamma(4, 0.0125)$	0.05	0.025	50%
46	mortality rate for **MicP**	t^{-1}	$\Gamma(4, 0.01)$	0.04	0.02	50%
47	mortality rate for **NanZ**	t^{-1}	$\Gamma(4, 0.015)$	0.06	0.03	50%
48	mortality rate for **MicZ**	t^{-1}	$\Gamma(4, 0.0125)$	0.05	0.025	50%
49	mortality rate for **MesZ**	t^{-1}	$\Gamma(4, 0.0075)$	0.03	0.015	50%
50	mortality rate for **BAC**	t^{-1}	$\Gamma(4, 0.015)$	0.06	0.03	50%
51	threshold for predation	C	$\Gamma(4, 0.005)$	0.02	0.01	50%
52	maximum predation rate on **MesZ**	t^{-1}	$\Gamma(4, 0.25)$	1.	0.5	50%
53	semisaturation for predation on **MesZ**	C	$\Gamma(4, 0.25)$	1.	0.5	50%
54	excreted fraction of predation on **MesZ**	–	$\beta(2.33, 4.67)$	0.333	0.167	50%
55	fraction of grazing used for growth of **NanZ**	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
56	fraction of grazing used for growth of **MicZ**	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
57	fraction of grazing used for growth of **MesZ**	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
58	fraction of POM used for growth of **MicZ**	–	$\beta(12, 12)$	0.5	0.1	20%
59	fraction of POM used for growth of **MesZ**	–	$\beta(12, 12)$	0.5	0.1	20%
60	excretion rate for **NanZ**	t^{-1}	$\Gamma(4, 0.0375)$	0.15	0.075	50%
61	excretion rate for **MicZ**	t^{-1}	$\Gamma(4, 0.025)$	0.1	0.05	50%
62	excretion rate for **MesZ**	t^{-1}	$\Gamma(4, 0.0125)$	0.05	0.025	50%
63	excretion rate for **BAC**	t^{-1}	$\Gamma(4, 0.0375)$	0.15	0.075	50%
64	temperature variation of excretion for **NanZ**	–	LogGamma	1.05	0.0525	5%
65	temperature variation of excretion for **MicZ**	–	LogGamma	1.05	0.0525	5%
66	temperature variation of excretion for **MesZ**	–	LogGamma	1.02	0.051	5%
67	temperature variation of excretion for **BAC**	–	LogGamma	1.04	0.052	5%
68	fraction of excretion as **DOM**	–	$\beta(2.75, 8.25)$	0.25	0.125	50%
69	POM1 decomposition rate	t^{-1}	$\Gamma(4, 0.01625)$	0.065	0.0325	50%
70	POM2 decomposition rate	t^{-1}	$\Gamma(4, 0.015)$	0.06	0.03	50%
71	sedimentation velocity for **MicP**	V	$\Gamma(4, 0.25)$	1.	0.5	50%
72	nitrification rate	t^{-1}	$\Gamma(4, 0.0075)$	0.03	0.015	50%
73	light attenuation coefficient in sea water	–	$\Gamma(25, 0.0016)$	0.04	0.008	20%
74	fraction of photosynthetically active radiation	–	$\Gamma(25, 0.02)$	0.5	0.1	20%

State variables

The ecosystem model provides a 12-component description of the ecosystem of the Ligurian Sea.

Variable	Acronym	Name
C_1	**NO3**	Nitrate
C_2	**NH4**	Ammonium
C_3	**PicP**	Picophytoplankton
C_4	**NanP**	Nanophytoplankton
C_5	**MicP**	Microphytoplankton
C_6	**NanZ**	Nanozooplankton
C_7	**MicZ**	Microzooplankton
C_8	**MesZ**	Mesozooplankton
C_9	**BAC**	Bacteria
C_{10}	**DON**	Dissolved organic nitrogen
C_{11}	**POM1**	Particulate organic matter (size 1)
C_{12}	**POM2**	Particulate organic matter (size 2)

The time evolution of each state variable is governed by the equation:

$$\frac{\partial C_i}{\partial t} = \mathtt{ADV}_i + \mathtt{DIFF}_i + \mathtt{SMS}_i \quad \text{with} \quad \mathtt{SMS}_i = \sum_{j \neq i} \mathtt{FLUX}(C_j \to C_i)$$

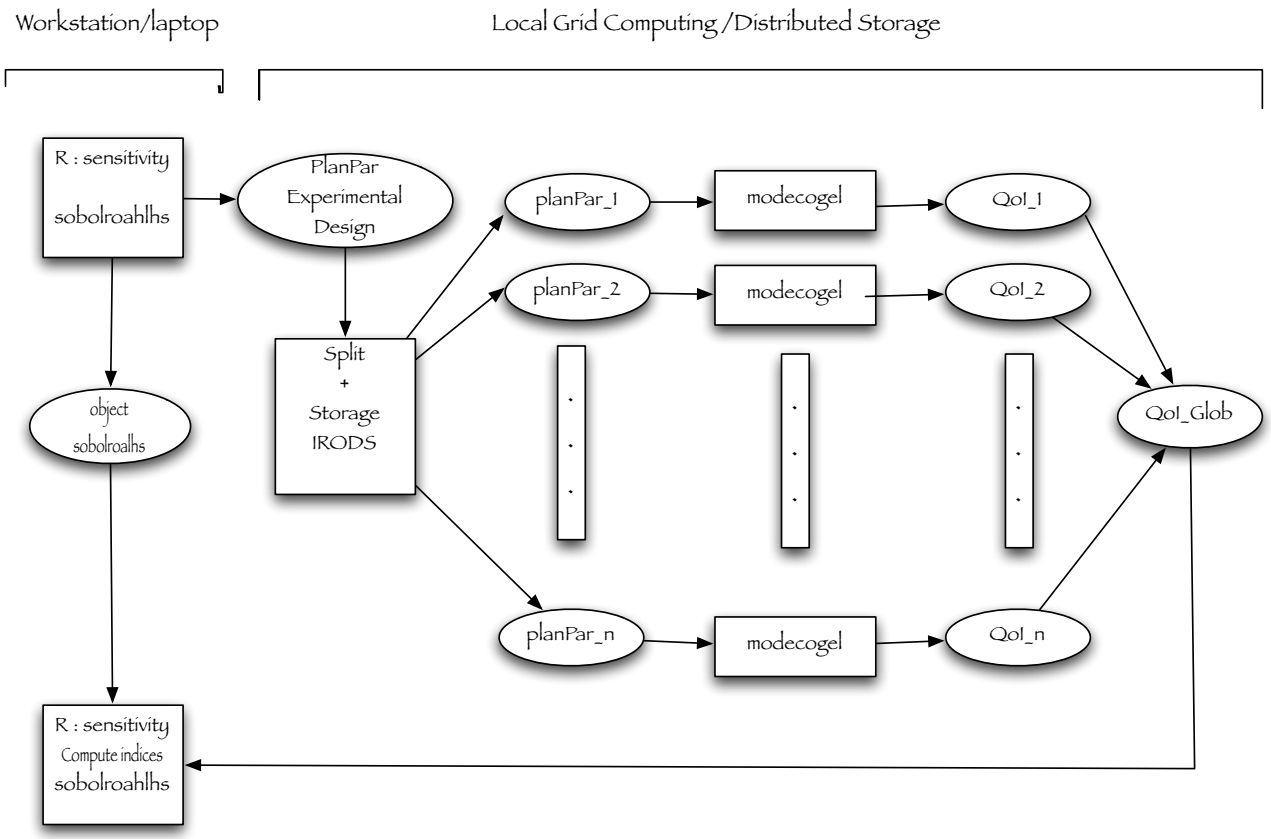
where $\mathtt{ADV}_i$ and $\mathtt{DIFF}_i$ are advection and diffusion terms, and $\mathtt{SMS}_i$ is the "source minus sink" term summing up the fluxes ($\mathtt{FLUX}(C_j \to C_i)$) between the various components of the ecosystem. We also introduce chlorophyll concentration $C_0 = \alpha(C_3 + C_4 + C_5)$.

QoIs

Index j	Name	Definition
1	surface maximum	$\max_t C_i(0,t)$
2	time of surface maximum	$\mathrm{argmax}_t C_i(0,t)$
3	maximum of vertical average	$\max_t \frac{1}{Z}\int_0^Z C_i(z,t)\,dz$
4	time of maximum of vertical average	$\mathrm{argmax}_t \frac{1}{Z}\int_0^Z C_i(z,t)\,dz$
5	time and vertical average	$\frac{1}{ZT}\int_0^T\int_0^Z C_i(z,t)\,dz\,dt$

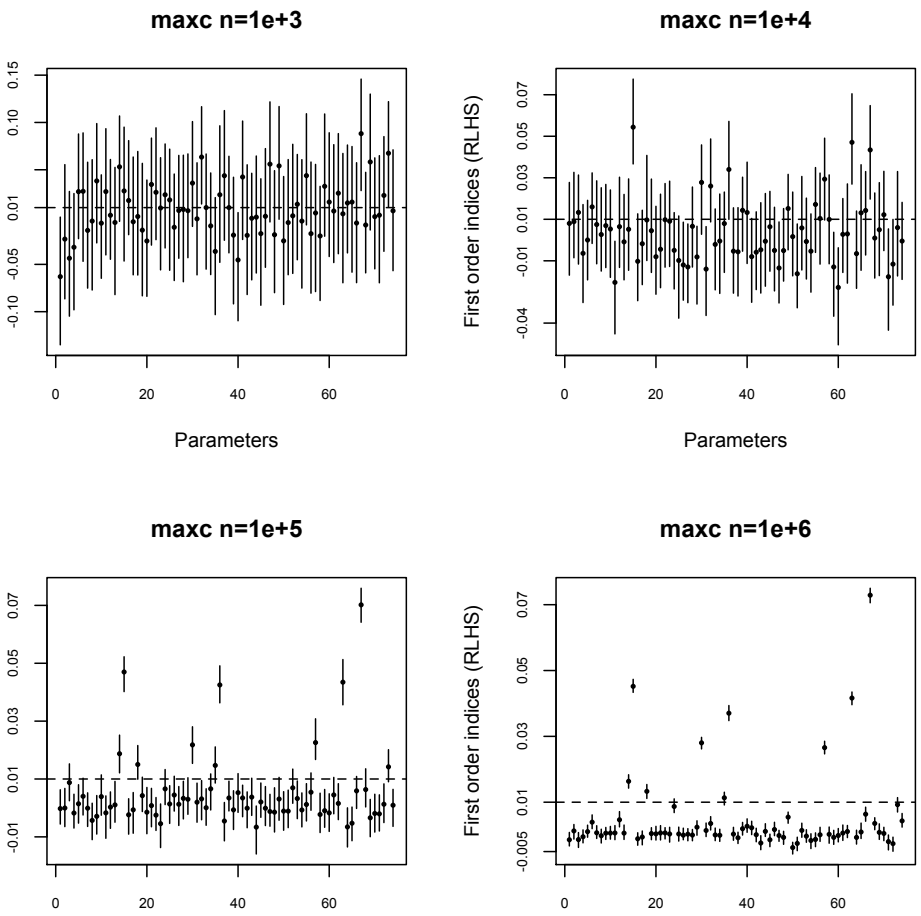
Quantities of interest Y_{ij}. The maximum depth for averaging is $Z = 40\,\mathrm{m}$, and T is the total duration of the experiment.

Processing chain

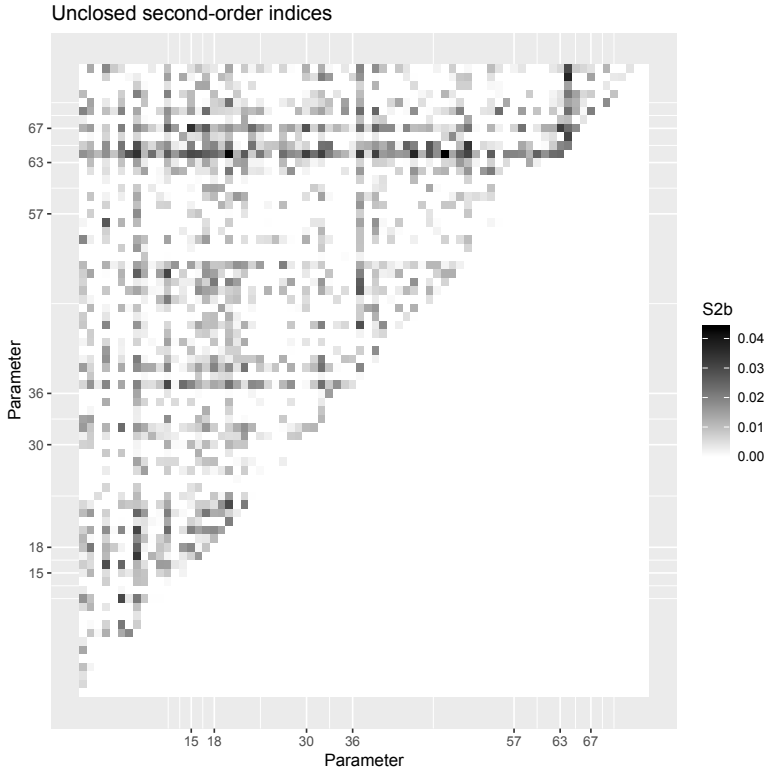


The steps for the estimation of all first-order (or all closed second-order) Sobol' indices with the `sobolroalhs` function of the R `sensitivity` package. The experimental design PlanPar is split into p sets of simulations (100 simulations each in our case). Each set of simulations is performed using MODECOGeL and the QoIs are computed for each simulation. All values for the QoI are grouped in a single file QoIGlob, which is sent to `sobolroalhs` for the actual

How the results look like?



Estimated first-order indices (y-axis) with their 95% confidence interval for the 74 model parameters (x-axis), for $n = 10^3, 10^4, 10^5$ and 10^6, in the case of the output Y_{01}. The dashed horizontal line corresponds to a threshold arbitrarily chosen to be 0.01. Confidence intervals were obtained with a bootstrap procedure and a bootstrap sample size of 100.



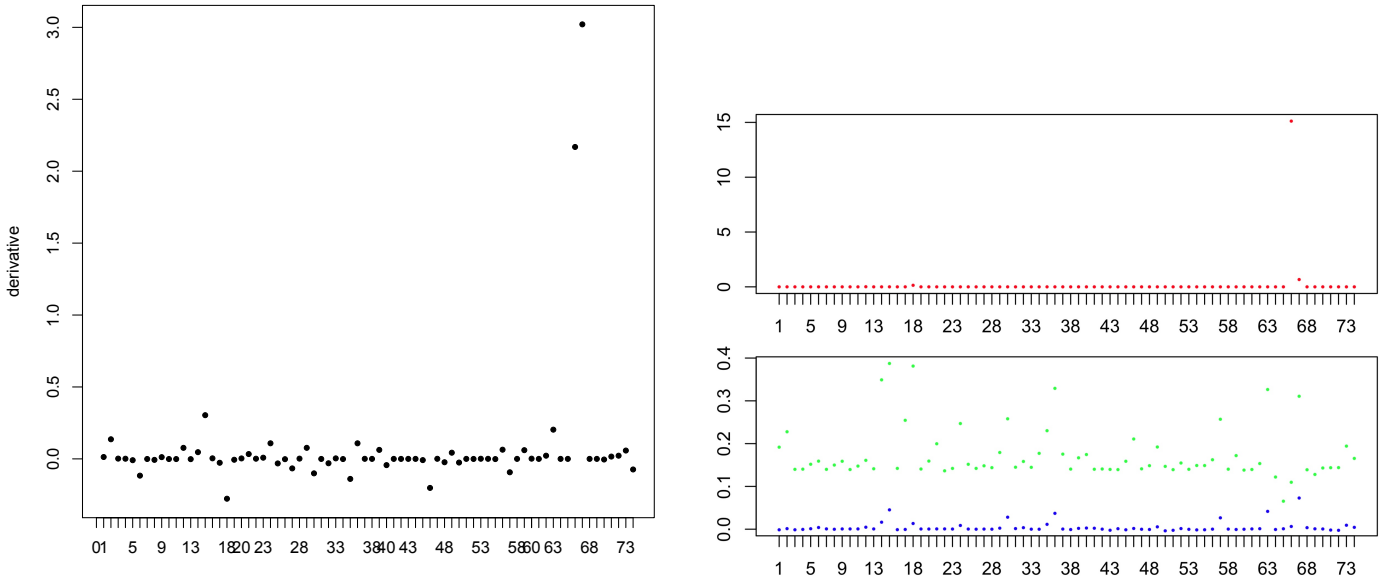
Map (74×74) of the second-order unclosed Sobol indices for QoI Y_{01}. The x and y axes correspond to the number of the parameters, and the grey scale to the value of the index. Note that the numbers indicated on the axes correspond to parameters with high first-order indices.

Top eight ranking of the local derivative $\partial Y / \partial X_j$, and first-order and total Sobol' indices S_j and $\mathrm{S}_j^{\mathrm{tot}}$.

j	2	14	15	18	30	35	36	46	57	63	66	67
$\partial Y / \partial X_j$	8th		3rd	4th		7th		6th		5th	2nd	1st
$\mathrm{S}_{\{j\}}$		7th	2nd	8th	5th		4th		6th	3rd		1st
$\mathrm{S}_{\{j\}}^{\mathrm{tot}}$		3rd	1st	2nd	7th		4th		8th	5th		6th

We can normalize local derivatives

$$\mathrm{S}_j^{\mathrm{loc}} = \frac{V[X_j]}{V[Y]} \left(\frac{\partial Y}{\partial X_j} \right)^2 .$$



Derivative $\dfrac{\partial Y_{01}}{\partial X_j}$ (left), non dimensional derivative $\mathrm{S}_j^{\mathrm{loc}} = \dfrac{V[X_j]}{\mathrm{Var}(Y_{01})}\left(\dfrac{\partial Y_{01}}{\partial X_j}\right)^2$ (right, upper panel), and first order and total Sobol' indices (right, lower panel) as functions of the number of the parameter (x-axis). The derivatives are computed for $(x_1, \ldots, x_d) = (\mathrm{E}(X_1), \ldots, \mathrm{E}(X_d))$.

Summary of statistics

Estimation of Sobol' indices	sobolSalt n=10^5		roalhs n=10^3	roalhs n=10^4	roalhs n=10^5	roalhs n=10^6	roalhs q=227
	S1	ST	S1	S1	S1	S1	S2
Estimated Error							
maximum value	0.0076	0.0064	0.077	0.025	0.0077	0.0025	0.024
mean value	0.0065	0.0047	0.062	0.020	0.0064	0.0020	0.018
standard deviation	$2.0\,10^{-7}$	$2.3\,10^{-7}$	$2.7\,10^{-5}$	$3.4\,10^{-6}$	$5.7\,10^{-8}$	$3.2\,10^{-8}$	$2.7\,10^{-7}$
Number of evaluations	$7.6\,10^6$		$2\,10^3$	$2\,10^4$	$2\,10^5$	$2\,10^6$	$q^2 \simeq 10^5$

Statistics (maximum and mean values, standard deviation) related to the estimated error over all 74 parameters, and number of model runs required for the estimation of the Sobol' indices.

Some references I

B. Broto, F. Bachoc, and M. Depecker. Variance Reduction for Estimation of Shapley Effects and Adaptation to Unknown Input Distribution. *SIAM/ASA Journal on Uncertainty Quantification*, 8 (2):693–716, 2020.

S. Chatterjee. A new coefficient of correlation. *Journal of the American Statistical Association*, 0 (0):1–21, 2020.

R. I. Cukier, C. M. Fortuin, K. E. Shuler, A. G. Petschek, and J. H. Schaibly. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients: Theory. *Journal of Chemical Physics*, 59:3873–3878, 1973.

R. I. Cukier, J. H. Schaibly, and K. E. Shuler. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients: Analysis of the approximations. *Journal of Chemical Physics*, 63:1140–1149, 1975.

R. I. Cukier, H. B. Levine, and K. E. Shuler. Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics*, 26:1–42, 1978.

S. Da Veiga, F. Gamboa, B. Iooss, C. Prieur, Society for Industrial, and Applied Mathematics. *Basics and Trends in Sensitivity Analysis: Theory and Practice in R*. Computational science and engineering. Society for Industrial and Applied Mathematics, 2021.

L. Devroye, D. Schäfer, L. Györfi, and H. Walk. The estimation problem of minimum mean squared error. *Statistics & Decisions*, 21(1):15–28, 2003.

L. Devroye, L. Györfi, G. Lugosi, and H. Walk. A nearest neighbor estimate of the residual variance. *Electronic Journal of Statistics*, 12(1):1752–1778, 2018.

Some references II

R. Faivre, D. Makowski, S. Mahévas, and B. Iooss. Analyse de sensibilité et exploration de modèles: application aux sciences de la nature et de l'environnement. *Analyse de sensibilité et exploration de modèles*, pages 1–352, 2013.

F. Gamboa, P. Gremaud, T. Klein, and A. Lagnoux. Global sensitivity analysis: a new generation of mighty estimators based on rank statistics. *to appear in Bernoulli*.

F. Gamboa, A. Janon, T. Klein, and A. Lagnoux. Sensitivity analysis for multidimensional and functional outputs. *Electronic Journal of Statistics*, 8(1):575–603, 2014.

F. Gamboa, A. Janon, T. Klein, A. Lagnoux, and C. Prieur. Statistical inference for sobol pick-freeze monte carlo method. *Statistics*, 50(4):881–902, 2016.

L. Gilquin, E. Arnaud, C. Prieur, and A. Janon. Making the best use of permutations to compute sensitivity indices with replicated orthogonal arrays. *Reliability Engineering & System Safety*, 187:28–39, 2019.

W. F. Hoeffding. A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics*, 19:293–325, 1948.

A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. Asymptotic normality and efficiency of two sobol index estimators. *ESAIM: Probability and Statistics*, 18:342–364, 2014.

W. Mauntz. Global sensitivity analysis of general nonlinear systems. *Master's Thesis, Imperial College. Supervisors: C. Pantelides and S. Kucherenko*, 2002.

Some references III

C. Prieur and S. Tarantola. Variance-based sensitivity analysis: Theory and estimation algorithms. *Handbook of uncertainty quantification*, pages 1217–1239, 2017.

C. Prieur, L. Viry, E. Blayo, and J-M Brankart. A global sensitivity analysis approach for marine biogeochemical modeling. *Ocean Modelling*, 139:101402, 2019.

A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145:280–297, 2002.

A. Saltelli, K. Chan, and E. M. Scott. *Sensitivity Analysis*. John Wiley & Sons, 2000.

I. M. Sobol'. Sensitivity analysis for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1:407–414, 1993.

S. Tarantola, D. Gatelli, and T. A. Mara. Random balance designs for the estimation of first-order global sensitivity indices. *Reliability Engineering and System Safety*, 91:717–727, 2006.

J. Y. Tissot and C. Prieur. Bias correction for the estimation of sensitivity indices based on random balance designs. *Reliability Engineering and System Safety*, 107:205–213, 2012a.

J. Y. Tissot and C. Prieur. Variance-based sensitivity analysis using harmonic analysis. `http://hal.archives-ouvertes.fr/docs/00/68/07/25/PDF`, 2012b.

J. Y. Tissot and C. Prieur. A randomized orthogonal array-based procedure for the estimation of first- and second-order sobol' indices. *Journal of Statistical Computation and Simulation*, 85: 1358–1381, 2015.

Thanks for your attention!