A few insights on Global Sensitivity Analysis

Clémentine PRIEUR
Grenoble Alpes University

AMA 2019, March 11-12, 2019, Toulouse (France) inputs $\longrightarrow$ model $\longrightarrow$ output

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

Specificities:

- the model $\mathcal{M}$ is expensive to evaluate;
- the inputs space dimension is high $d >> 1$.

In a calibration framework, one may aim at **fixing non influential input variables**.

Sensitivity analysis can help in detecting "low-effective dimension".

Introduction

Many examples in different application fields:

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

MODECOGeL is a 1D 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.

▷ 87 scalar input parameters;

▷ spatio-temporal outputs.

Context:

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

Context:

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

Aim: to determine the way the output of the model reacts to variations of the inputs parameters, to fix non influential input parameters.

Context:

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

Aim: to determine the way the output of the model reacts to variations of the inputs parameters, to fix non influential input parameters.

Several analyses are possible:

▷ qualitative analyses: are there non linear effects? interactions? **screening approaches**.
▷ quantitative analyses: factors' hierarchization, statistical hypothesis testing: e.g., H_0 "the i^{th} factor has no influence on the output". **sensitivity measures**.

Introduction

Various approaches for quantitative sensitivity:

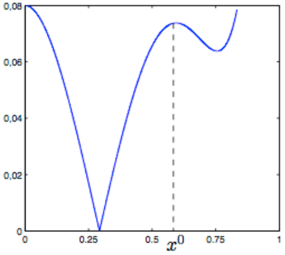
Local approaches:

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



The paradigm of Global Sensitivity Analysis (GSA):

The uncertainty on the inputs is modeled by a probability distribution, from experts' knowledge, or from observations, ...

e.g., if the inputs are independent, this probability distribution is characterized by its marginals.

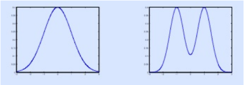


Figure: unimodal distribution (left), bimodal distribution (right)

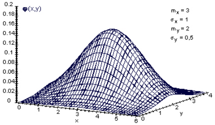


Figure: bivariate distribution.

Let $Y = \mathcal{M}(X_1, \ldots, X_d)$, with $\mathbf{X} \sim \mathcal{U}\left([0,1]^d\right)$.

We consider the discretization grid: $\Omega := \left\{0, \frac{1}{p-1}, \ldots, 1\right\}^d$.

For Δ a multiple of $1/(p-1)$, for $i = 1, \ldots, d$, define

$$\Omega_i^\Delta := \{\mathbf{x} \in \Omega \text{ s.t. } (x_1, \ldots, x_{i-1}, x_i + \Delta, x_{i+1}, \ldots, x_d) \in \Omega\}.$$

The procedure is OAT (One At a Time): we vary the input parameters one by one.

Elementary effects for input factor X_i

Let $\mathbf{x} \in \Omega_i^\Delta$,

$$d_i(\mathbf{x}) = \frac{1}{\Delta}\left\{\mathcal{M}(x_1, \ldots, x_{i-1}, x_i + \Delta, x_{i+1}, \ldots, x_d) - \mathcal{M}(\mathbf{x})\right\}.$$

For each input, there are $p^{d-1}(p - \Delta(p-1))$ elementary effects to be computed.

- we draw uniformly a sample of size m on Ω_i^{Δ} : $\mathbf{x}^1, \ldots, \mathbf{x}^m$;
- we compute $d_i(\mathbf{x}^j)$, $j = 1, \ldots, m$, $i = 1, \ldots, d$;
- we compute

$$\mu_i = \frac{1}{m} \sum_{j=1}^{m} d_i(\mathbf{x}^j)$$

$$\sigma_i^2 = \frac{1}{m} \sum_{j=1}^{m} (d_i(\mathbf{x}^j) - \mu_i)^2 .$$

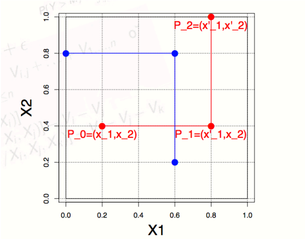
- Interpretation:

	"small" σ_i^2	"high" σ_i^2
"small" $\lvert\mu_i\rvert$	neglectable	non linearities and/or interactions
"high" $\lvert\mu_i\rvert$	influent	non linearities and/or interactions

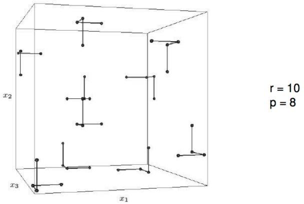
The efficiency of the method "number of elementary effects to be computed / number of runs" is equal to $(md)/(2md) = 1/2$.

Morris (91) also propose a tricky design of experiment which yields an efficiency equal to $d/(d+1)$.

Morris' design, projection in two-dimension (X_1, X_2), with $p = 6$, $\Delta = p/[2(p-1)] = 3/5$, $N = r \times (d+1)$ with $r = 2$ and $d = 2$.



Morris' design in three-dimension with $p = 8$, $\Delta = p/[2(p-1)] = 4/7$, $N = r \times (d+1)$ with $r = 10$ and $d = 3$.



A toy example

Reaction-diffusion-advection equation with Dirichlet boundary conditions :

$$\begin{cases} \frac{\partial u}{\partial t} = -r.u - a\frac{\partial u}{\partial x} + \lambda \frac{\partial^2 u}{\partial x^2} + f \quad x \in [0, L],\ t \in [0, T] \\ u(x = 0, t) = \Psi_1(t) \quad t \in [0, T] \\ u(x = L, t) = \Psi_2(t) \quad t \in [0, T] \\ u(x, t = 0) = g(x) \quad x \in (0, L)\,. \end{cases}$$

Quantity of Interest: energy norm of the solution at time $t = T$.

Sensitivity of the QoI to parameters (a, r, λ)? The uncertainty on input parameters is modeled by independent random variables: $a, r \sim \mathcal{U}([0.4, 0.6])$, $\lambda \sim \mathcal{U}([0.04, 0.06])$.

Adams-Moulton scheme with 2 steps, sample of size 2^{13}.

A toy example

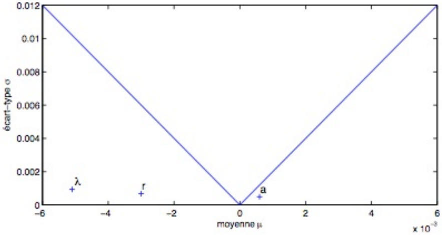


Figure: Morris screening with $p = 50$, $\Delta = 25/49$.

$$S_a = 0.0188,\ S_\lambda = 0.7299,\ S_r = 0.2488,\ S_a + S_\lambda + S_r = 0.988.$$

Sensitivity measures based on linear regression:

Let $\mathbf{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d)$. Recall that $Y = \mathcal{M}(X_1, \ldots, X_d)$.

- Linear correlation

$$\rho_i = \rho(X_i, Y) = \frac{\text{Cov}(X_i, Y)}{\sqrt{\text{Var}(X_i)}\sqrt{\text{Var}(Y)}}.$$

If $Y = \sum_{i=1}^{d} \beta_i X_i$, and if inputs are independent, $\sum_{i=1}^{d} \rho^2(X_i, Y) = 1$.

- Partial correlation

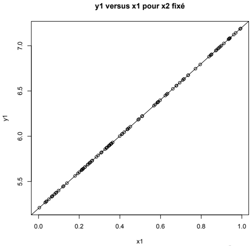
If inputs are correlated , it might be more suitable to compute

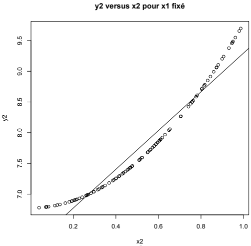
$$PCC_i = PCC(X_i, Y) = \rho\left(Y - \widehat{Y}(\mathbf{X}_{-i}), X_i - \widehat{X_i}(\mathbf{X}_{-i})\right).$$

with $\widehat{Y}(\mathbf{X}_{-i})$ the regression of Y on $\mathbf{X}_{-i}$ and $\widehat{X_i}(\mathbf{X}_{-i})$ the one of X_i on $\mathbf{X}_{-i}$.

Assessment of linear model?

Toy example : $Y = 2X_1 + 3{X_2}^2 + 5$, $X_i \sim \mathcal{U}([0,1])$, $i = 1, 2$, $X_1 X_2$.





We can approximate this model by a linear model :
$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_0 + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

Learning sample : $y_k = \mathcal{M}(x_{1,k}, \ldots, x_{d,k})$, $k = 1, \ldots, m$

$$\Rightarrow \quad \hat{y} = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_0 = 2.06x1 + 3.15x2 + 4.34.$$

Which measure to assess the fit of this model?

Coefficient R^2

$$R^2 = \frac{SCE}{SCT} = \frac{\sum_{k=1}^{m} \left(\widehat{y}_k - \overline{y}\right)^2}{\sum_{k=1}^{m} \left(y_k - \overline{y}\right)^2},$$

$\widehat{y}_k = \sum_{i=1}^{d} \hat{\beta}_i x_{i,k}$, $\overline{y} = \frac{1}{m} \sum_{k=1}^{m} y_k$.

Coefficient R^2

$$R^2 = \frac{SCE}{SCT} = \frac{\sum_{k=1}^{m} (\widehat{y}_k - \overline{y})^2}{\sum_{k=1}^{m} (y_k - \overline{y})^2},$$

$\widehat{y}_k = \sum_{i=1}^{d} \hat{\beta}_i x_{i,k}$, $\overline{y} = \frac{1}{m} \sum_{k=1}^{m} y_k$.

Prediction error, e.g. cross-validation

$$\frac{1}{m} \frac{\sum_{k=1}^{m} \left(\widehat{y}_k^{-(k)} - y_k \right)^2}{\frac{1}{m} \sum_{k=1}^{m} (y_k - \overline{y})^2},$$

$\widehat{y}_k^{-(k)} = \sum_{i=1}^{d} \widehat{\beta}_i^{-(k)} x_{i,k}$, $\widehat{\beta}_i^{-(k)}$ inferred from

$$(y_j, \mathbf{x}_j),\ j = 1, \ldots, k-1, k+1, \ldots, m.$$

If the relationship input/output is no more linear but simply monotonic, we work with ranks.

y_k, $x_{i,k}$, $k = 1, \ldots, m$, $i = 1, \ldots, d$

$r_{i,k}$ rank of $x_{i,k}$ in $(x_{i,1}, \ldots, x_{i,m})$, r_k rank of y_k in $(y_1, \ldots, y_m)$

- $\rho_i^S = \frac{\sum_{k=1}^{m}(r_{i,k}-\bar{r}_i)(r_k-\bar{r})}{\sqrt{\sum_{k=1}^{m}(r_{i,k}-\bar{r}_i)^2}\sqrt{\sum_{k=1}^{m}(r_k-\bar{r})^2}}$
- idem for pcc_i

We now focus on variance based sensitivity measures. We assume $Y \in \mathbb{R}$ and X_i i.i.d. $\sim \mathcal{U}([0,1])$ (these assumptions are discussed further).

In this part of the talk, inputs are assumed to be **independent**.

Towards Sobol' sensitivity indices:

Does the output Y vary more or less when fixing one of its inputs?

$\mathrm{Var}(Y|X_i = x_i)$, how to choose x_i?

We now focus on variance based sensitivity measures. We assume $Y \in \mathbb{R}$ and X_i i.i.d. $\sim \mathcal{U}([0,1])$ (these assumptions are discussed further).

In this part of the talk, inputs are assumed to be **independent**.

Towards Sobol' sensitivity indices:

Does the output Y vary more or less when fixing one of its inputs?

$\text{Var}(Y|X_i = x_i)$, how to choose x_i? $\Rightarrow \mathbb{E}[\text{Var}(Y|X_i)]$

We now focus on variance based sensitivity measures. We assume $Y \in \mathbb{R}$ and X_i i.i.d. $\sim \mathcal{U}([0,1])$ (these assumptions are discussed further).

In this part of the talk, inputs are assumed to be **independent**.

Towards Sobol' sensitivity indices:

Does the output Y vary more or less when fixing one of its inputs?

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

The smaller $\mathbb{E}[\mathrm{Var}(Y|X_i)]$, *the more influential the* i^{th} *input*, X_i.

We now focus on variance based sensitivity measures. We assume $Y \in \mathbb{R}$ and X_i i.i.d. $\sim \mathcal{U}([0,1])$ (these assumptions are discussed further).

In this part of the talk, inputs are assumed to be **independent**.

Towards Sobol' sensitivity indices:

Does the output Y vary more or less when fixing one of its inputs?

$\text{Var}(Y|X_i = x_i)$, how to choose x_i? $\Rightarrow \mathbb{E}[\text{Var}(Y|X_i)]$

The smaller $\mathbb{E}[\text{Var}(Y|X_i)]$, *the more influential the* i^{th} *input,* X_i.

From the total variance theorem, $\text{Var}(Y) = \text{Var}[\mathbb{E}(Y|X_i)] + \mathbb{E}[\text{Var}(Y|X_i)]$. Define the first-order Sobol' index associated to X_i as $S_i = \text{Var}[\mathbb{E}(Y|X_i)]/\text{Var}[Y]$.

The larger $0 \leq S_i \leq 1$, *the more influential the* i^{th} *input,* X_i.

Remark: if $Y = \sum_{i=1}^{d} \beta_i X_i$, one gets $S_i = \beta_i^2 \mathrm{Var}[X_i] / \mathrm{Var}[Y] = \rho_i^2$, with ρ_i the linear correlation coefficient.

Remark: if $Y = \sum_{i=1}^{d} \beta_i X_i$, one gets $S_i = \beta_i^2 \mathrm{Var}[X_i]/\mathrm{Var}[Y] = \rho_i^2$, with ρ_i the linear correlation coefficient.

More generally, it is possible to define second-order, third-order. . . Sobol' indices.

Hoeffding decomposition ([Hoeffding, 1948, Sobol', 1993])

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

$\mathcal{M}$ admits a unique decomposition of the form

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

- *$\mathcal{M}_0$ constant,*
- $\forall\, 1 \le s \le d$, $\forall\, 1 \le i_1 < \ldots < i_s \le d$, $\forall\, 1 \le p \le s$
  $\int_0^1 \mathcal{M}_{i_1,\ldots,i_s}(x_{i_1}, \ldots, x_{i_s}) dx_{i_p} = 0$

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

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

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

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

With the orthogonality constraints, we get:

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

$$\mathbb{E}(Y - \mathcal{M}_0)^2 = \mathbb{E}(Y - \mathbb{E}(Y))^2 = \mathrm{Var}[Y]$$
$$= \sum_{i=1}^d \mathrm{Var}[\mathcal{M}_i(X_i)] + \ldots + \mathrm{Var}[\mathcal{M}_{1,\ldots,d}(X_1, \ldots, X_d)].$$

First-order Sobol' indices: $\forall\, i = 1, \ldots, d$

$$S_i = \frac{\mathrm{Var}\left(\mathcal{M}_i(X_i)\right)}{\mathrm{Var}(Y)} = \frac{\mathrm{Var}\left[\mathbb{E}\left(Y|X_i\right)\right]}{\mathrm{Var}(Y)}$$

Second-order Sobol' indices: $\forall i \neq j = 1, \ldots, d$

$$\begin{aligned} S_{i,j} = & \frac{\mathrm{Var}\left[\mathcal{M}_{i,j}(X_i, X_j)\right]}{\mathrm{Var}[Y]} \\ = & \frac{\mathrm{Var}\left[\mathbb{E}\left(Y|X_i, X_j\right)\right] - \mathrm{Var}\left[\mathbb{E}\left(Y|X_i\right)\right] - \mathrm{Var}\left[\mathbb{E}\left(Y|X_j\right)\right]}{\mathrm{Var}[Y]} \end{aligned}$$

Higher-order Sobol' indices ... $\forall\, \mathbf{u} \subset \{1, \ldots, d\}$

$$S_{\mathbf{u}} = \frac{\mathrm{Var}\left[\mathcal{M}_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}})\right]}{\mathrm{Var}[Y]} = \frac{\sum_{\mathbf{v} \subset \mathbf{u}} (-1)^{|u|-|v|} \mathrm{Var}\left[\mathbb{E}\left(Y|\mathbf{X}_{\mathbf{v}}\right)\right]}{\mathrm{Var}[Y]}$$

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

Total-effect Sobol' indices:

$$\forall\, i = 1, \ldots, d \quad S_i^{\text{tot}} = \sum_{\mathbf{u} \subset \{1,\ldots,d\}\,,\, \mathbf{u} \neq \emptyset\,,\, i \in \mathbf{u}} S_{\mathbf{u}}$$

Example: $d = 3$, $S_1^{\text{tot}} = S_1 + S_{1,2} + S_{1,3} + S_{1,2,3}$.

Let $\mathbf{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d)$. We have:

$$S_i^{\text{tot}} = \frac{\mathbb{E}\left[\text{Var}\left(Y|\mathbf{X}_{-i}\right)\right]}{\text{Var}[Y]} = 1 - \frac{\text{Var}\left[\mathbb{E}\left(Y|\mathbf{X}_{-i}\right)\right]}{\text{Var}[Y]}$$

More generally, $\forall\, \mathbf{u} \subset \{1, \ldots, d\}$, for $\mathbf{X}_{-\mathbf{u}} = (X_i\,,\; i \notin \mathbf{u})$,

$$\begin{aligned} S_{\mathbf{u}}^{\text{tot}} &= \frac{\mathbb{E}\left[\text{Var}\left(Y|\mathbf{X}_{-\mathbf{u}}\right)\right]}{\text{Var}[Y]} = 1 - \frac{\text{Var}\left[\mathbb{E}\left(Y|\mathbf{X}_{-\mathbf{u}}\right)\right]}{\text{Var}[Y]} \\ &= \sum_{\mathbf{v} \subset \{1,\ldots,d\}, \mathbf{v} \cap \mathbf{u} \neq \emptyset} S_{\mathbf{v}}\,. \end{aligned} \tag{1}$$

If the inputs are dependent, there exist some alternatives to allocate parts of variance: hierarchical Hoeffding decomposition, Shapley effects,...

Note that if we define, $\forall\, \mathbf{u} \subset \{1, \dots, d\}$

$$S_{\mathbf{u}}^{\text{dep}} = \frac{\text{Cov}\left(\mathcal{M}_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}), \mathcal{M}(\mathbf{X})\right)}{\text{Var}\left(\mathcal{M}(\mathbf{X})\right)},$$

then Eq. (1) still holds true (see, e.g., [Hart and Gremaud, 2018]) with $S_{\mathbf{u}}$, $S_{\mathbf{v}}$ replaced by $S_{\mathbf{u}}^{\text{dep}}$, $S_{\mathbf{v}}^{\text{dep}}$.

If $S_{\mathbf{u}}^{\text{tot}}$ is small, it is reasonable to propose a metamodel of $\mathcal{M}$ taking as inputs only the input parameters indexed by $i \notin \mathbf{u}$.

Hear, e.g., the SIAM Conference on UQ MiniSymposium *Advances in Global Sensitivity Analysis*

`https://www.pathlms.com/siam/courses/7376/sections/10632`

Assume the input parameters are independent.

Let $\mathbf{X}^1$ and $\mathbf{X}^2$ be two independent copies of $\mathbf{X}$.

For $i = 1, \ldots, d$, we define:

$$\mathbf{Z}^i = (X_1^2, \ldots, X_{i-1}^2, X_i^1, X_{i+1}^2, \ldots, X_d^2)$$

Let $Y = \mathcal{M}(\mathbf{X}^1)$ and, for $i = 1, \ldots, d$, $Y^i = \mathcal{M}(\mathbf{Z}^i)$.

If the random vector $\mathbf{X}$ has independent components, then we deduce:

$$S_i = \frac{\mathrm{Cov}\left(Y, Y^i\right)}{\mathrm{Var}[Y]}.$$

For any $i \in \{1, \ldots, d\}$, let $X_i^{1,j}$ and $X_i^{2,j}$, $j = 1, \ldots, n$ be two independent samples of size n of the parameter X_i.

We define:

$$\mathbf{X}^{1,j} = (X_1^{1,j}, \ldots, X_{i-1}^{1,j}, X_i^{1,j}, X_{i+1}^{1,j}, \ldots, X_d^{1,j}) \quad j = 1, \ldots, n$$

$$\mathbf{Z}^{i,j} = (X_1^{2,j}, \ldots, X_{i-1}^{2,j}, X_i^{1,j}, X_{i+1}^{2,j}, \ldots, X_d^{2,j}) \quad j = 1, \ldots, n\,, \; i = 1, \ldots, d$$

We evaluate the model $(1 + d)n$ times:

$$Y^j = \mathcal{M}(\mathbf{X}^{1,j}) \quad j = 1, \ldots, n$$

$$Y^{i,j} = \mathcal{M}(\mathbf{Z}^{i,j}) \quad j = 1, \ldots, n\,, \; i = 1, \ldots, d.$$

Monte Carlo estimator: [Monod et al., 2006, Janon et al., 2014]

$$\hat{\mathrm{S}}_{i,n} = \frac{\frac{1}{n}\sum_{j=1}^{n} Y^j Y^{i,j} - \left(\frac{1}{n}\sum_{j=1}^{n}\frac{Y^j + Y^{i,j}}{2}\right)^2}{\frac{1}{n}\sum_{j=1}^{n}\frac{(Y^j)^2 + (Y^{i,j})^2}{2} - \left(\frac{1}{n}\sum_{j=1}^{n}\frac{Y^j + Y^{i,j}}{2}\right)^2}$$

Total and higher order interaction indices can also be estimated.

Monte Carlo estimator: [Monod et al., 2006, Janon et al., 2014]

$$\hat{S}_{i,n} = \frac{\frac{1}{n}\sum_{j=1}^{n} Y^j Y^{i,j} - \left(\frac{1}{n}\sum_{j=1}^{n} \frac{Y^j + Y^{i,j}}{2}\right)^2}{\frac{1}{n}\sum_{j=1}^{n} \frac{(Y^j)^2 + (Y^{i,j})^2}{2} - \left(\frac{1}{n}\sum_{j=1}^{n} \frac{Y^j + Y^{i,j}}{2}\right)^2}$$

Total and higher order interaction indices can also be estimated.

Main issue: the cost is prohibitive.

- $(1+d)n$ model evaluations for all first-order Sobol' indices;
- $(\binom{d}{2}+1)n$ for all second-order Sobol' indices.

▷ with combinatorial tricks, a cost of $(2d+2)n$ model eval. for double estimates of all first-order, second-order and total Sobol' indices [Saltelli, 2002];

▷ with replicated orthogonal arrays, a cost of $2q^2$ model eval. for a single estimate of all second-order, and $q \times q!$ estimates of all first-order, with $q \geq d-1$ a prime number [Gilquin et al., 2018].

▷ with combinatorial tricks, a cost of $(2d+2)n$ model eval. for double estimates of all first-order, second-order and total Sobol' indices [Saltelli, 2002];

▷ with replicated orthogonal arrays, a cost of $2q^2$ model eval. for a single estimate of all second-order, and $q \times q!$ estimates of all first-order, with $q \geq d-1$ a prime number [Gilquin et al., 2018].

That cost may still be prohibitive, thus the necessity to learn a metamodel, such as:

▷ polynomial chaos expansion (Bruno Sudret's talk),

▷ Gaussian Process emulators (talks by Miguel Munoz Zuniga and Daniel Williamson).

Procedure:

▷ learn the metamodel from a sample of moderate size $(\mathbf{x}^j, y^j)_{j=1,\ldots,n}$

▷ compute Sobol' indices by running the metamodel.

What about vectorial outputs [Lamboni et al., 2011]:
We assume $Y \in \mathbb{R}^p$.

One defines, for $i = 1, \ldots, d$

$$GS_i = \sum_{k=1}^{p} \frac{\mathrm{Var}[Y_k]}{\sum_{j=1}^{p} \mathrm{Var}[Y_j]} S_i(Y_k)\,.$$

What about vectorial outputs [Lamboni et al., 2011]:

We assume $Y \in \mathbb{R}^p$. Let Σ denote the variance-covariance matrix of Y. The PCA decomposition of Y is based on the expansion

$$\Sigma = \sum_{k=1}^{p} \mu_k \mathbf{v}_k \mathbf{v}_k^T$$

with $\mu_1 \geq \ldots \geq \mu_p$ the eigenvalues of Σ and $\mathbf{v}_1, \ldots, \mathbf{v}_p$ a set of normalized and mutually orthogonal eigenvectors associated to these eigenvalues. One has

$$Y = \mathbb{E}Y + \sum_{k=1}^{p} \left((Y - \mathbb{E}Y)^T \mathbf{v}_k \right) \mathbf{v}_k = \mathbb{E}Y + \sum_{k=1}^{p} h_k \, \mathbf{v}_k \, .$$

One gets, for $i = 1, \ldots, d$

$$GS_i = \sum_{k=1}^{p} \frac{\mu_k}{\text{trace } (\Sigma)} S_i(h_k) = \frac{\text{trace}(C_i)}{\text{trace}(\Sigma)}$$

with C_i the variance-covariance matrix of $\mathbb{E}(Y|X_i)$.

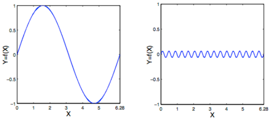
"Globalized" local approaches: e.g., (1) $\mathbb{E}_X \left[\left. \frac{\partial \mathcal{M}}{\partial x_i} \right|_{\mathbf{x}} \right]$, or (2) $\nu_i = \mathbb{E}_X \left[\left(\left. \frac{\partial \mathcal{M}}{\partial x_i} \right|_{\mathbf{x}} \right)^2 \right]$.

"Globalized" local approaches: e.g., (1) $\mathbb{E}_X\left[\frac{\partial \mathcal{M}}{\partial x_i}\Big|_{\mathbf{x}}\right]$, or (2) $\nu_i = \mathbb{E}_X\left[\left(\frac{\partial \mathcal{M}}{\partial x_i}\Big|_{\mathbf{x}}\right)^2\right]$.

Pros: it takes into account the inputs' distribution, the cost is independent of the dimension in case an adjoint is available .

Cons:

(1) & (2) are not enough discriminant



(2) is known as Derivative-based Global Sensitivity Measures , see Sobol' & Gresham (1995), Sobol' & Kucherenko (2009). This index is more appropriate for screening than for hierarchization (see Lamboni *et al.*, 2013).

Link with active subspaces [Constantine and Diaz, 2017]:

Assume $\mathbf{x} \sim \mathcal{U}\left([-1,1]^d\right)$. Define $\rho(\mathbf{x}) = 2^{-d}$ for $\mathbf{x} \in [-1,1]^d$.

Active subspaces are based on the eigendecomposition of

$$H = \int \nabla \mathcal{M}(\mathbf{x}) \nabla \mathcal{M}(\mathbf{x})^T \rho(\mathbf{x}) d\mathbf{x} = W \Lambda W^T$$

with $W = [\mathbf{w}_1, \dots, \mathbf{w}_d]$ the orthogonal matrix of eigenvectors, and $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_d)$ the diagonal matrix of eigenvalues in decreasing order.

For any $i = 1, \dots, d$, one has $H_{i,i} = \mathbb{E}_X \left[\left(\left. \frac{\partial \mathcal{M}}{\partial x_i} \right|_{\mathbf{x}} \right)^2 \right] = \nu_i$.

One has

$$\lambda_i = \mathbf{w}_i^T H \mathbf{w}_i = \int \left(\nabla \mathcal{M}(\mathbf{x})^T \mathbf{w}_i \right)^2 \rho(\mathbf{x}) d\mathbf{x}\,.$$

It implies [$\lambda_i = 0$ iff $\mathcal{M}$ is constant along the direction $\mathbf{w}_i$].

Dimension reduction:

Let $1 \leq r \leq d-1$. We define

$$\Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix}, \; W = [W_1 \; W_2],$$

where Λ_1 contains the first r eigenvalues and W_1 the corresponding eigenvectors.

If $\lambda_{r+1}, \ldots, \lambda_d$ are sufficiently small, then it seems reasonable to use a surrogate $\mathcal{M}(\mathbf{x}) \approx g(W_1^T \mathbf{x})$, with $g : \mathbb{R}^r \to \mathbb{R}$.

One has

$$S_i^{\text{tot}} \leq \frac{1}{4\pi^2 \text{Var}(\mathcal{M})} \, \nu_i \, .$$

Let $\alpha_i(r) = \sum_{j=1}^r \lambda_j w_{i,j}^2$. Note that $\alpha_i(r) \leq \alpha_i(d) = \nu_i$.
Moreover,

$$S_i^{\text{tot}} \leq \frac{1}{4\pi^2 \text{Var}(\mathcal{M})} \left(\alpha_i(r) + \lambda_{r+1}\right) \, .$$

- It is possible to extend to non uniform distributions, as far as Poincaré constants are known.

$C(\mu_i)$ is a Poincaré constant for μ_i if for any g such that $\int g d\mu_i = 0$,

$$\int g^2 d\mu_i \leq C(\mu_i) \int g'^2 d\mu_i \,.$$

- For the estimation procedure, automatic differentiation algorithms may be used. The number of required function evaluations still remains to be proportional to the number of inputs. However, this dependence can be greatly reduced using an approach based on algorithmic differentiation in the adjoint or reverse mode.

- Thus, the ν_i, $i = 1, \ldots, d$, may be computed for screening purposes [Lamboni et al., 2013].

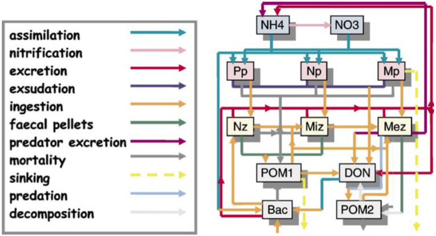
- These results may be extended to vectorial outputs.

R

With the R software, most of the methodologies presented above are implemented in the package `sensitivity`.

See the link
`https://cran.r-project.org/web/packages/sensitivity/`

MODECOGeL



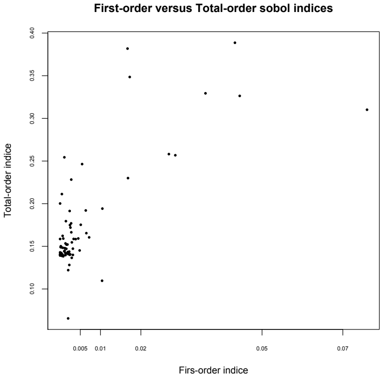
• hydrodynamic model: 1-D vertical simplification of primitive equations for the ocean, 5 state variables;
• ecosystem model: marine biogeochemistry, 12 biological state variables.

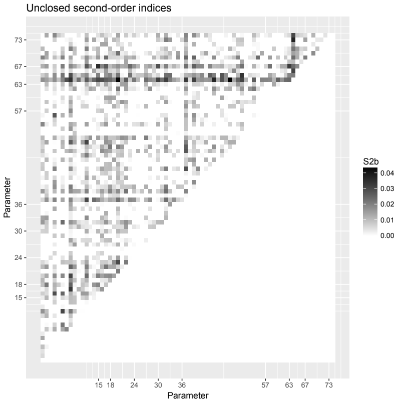
Inputs/Outputs:

▷ 87 scalar input parameters;

▷ spatio-temporal outputs.

Here the QoI is the annual maximum of surface chlorophyll concentration. We are interested in the sensitivity of the QoI to many parameters, among which the parameterization of excretion for bacteria, of grazing and ingestion for mesozooplankton.





The full set of second-order indices were estimated with 103 058 model runs, deployed on a grid environment (see [Prieur et al., 2018] for more details).

Conclusion:

- GSA is a nice framework for independent inputs.
- However, the estimation of Sobol' indices requires a large number of model evaluations.
- DGSM (or active subspaces), even if less informative, provide an alternative, at a lower cost.
- If one is interested in surrogate models, SA can be thought as a dimension reduction procedure for approximation.

To go further:

- goal-oriented sensitivity measures, not necessarily based on variance;
- handle some applications with high dimensional input space, making use, e.g., of a grid deployment [Prieur et al., 2018];
- *in situ* estim., e.g., [Gilquin et al., 2016, Ribes et al., 2019];
- visualization in high dimension; ...

Thanks for your attention!

Some references I

Constantine, P. G. and Diaz, P. (2017).
Global sensitivity metrics from active subspaces.
Reliability Engineering & System Safety, 162:1–13.

Gilquin, L., Arnaud, E., Prieur, C., and Janon, A. (2018).
Making the best use of permutations to compute sensitivity indices with replicated orthogonal arrays.
To appear in *Reliability Engineering & System Safety*.
https://hal.inria.fr/hal-01558915.

Gilquin, L., Arnaud, E., Prieur, C., and Monod, H. (2016).
Recursive estimation procedure of Sobol' indices based on replicated designs.
https://hal.inria.fr/hal-01291769.

Hart, J. and Gremaud, P. (2018).
An approximation theoretic perspective of sobol' indices with dependent variables.
International Journal for Uncertainty Quantification, 8 (6).

Some references II

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

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

Lamboni, M., Iooss, B., Popelin, A.-L., and Gamboa, F. (2013).
Derivative-based global sensitivity measures: general links with Sobol' indices and numerical tests.
Mathematics and Computers in Simulation, 87:45–54.

Lamboni, M., Monod, H., and Makowski, D. (2011).
Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models.
Reliability Engineering and System Safety, 96(4):450–459.

Some references III

Monod, H., Naud, C., and Makowski, D. (2006).
Uncertainty and sensitivity analysis for crop models.
In Wallach, D., Makowski, D., and Jones, J. W., editors, *Working with Dynamic Crop Models: Evaluation, Analysis, Parameterization, and Applications*, chapter 4, pages 55–99. Elsevier.

Morris, M. (1991).
Factorial sampling plans for preliminary computational experiments.
Technometrics, 33:161–174.

Prieur, C., Viry, L., Blayo, E., and Brankart, J.-M. (2018).
A global sensitivity analysis approach for marine biogeochemical modeling.
working paper or preprint.

Ribes, A., Terraz, T., Iooss, B., Fournier, Y., and Raffin, B. (2019).
Large scale in transit computation of quantiles for ensemble runs.
working paper or preprint.

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

Some references IV

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

Sobol', I. M. and Gresham, A. (1995).
On an alternative global sensitivity estimators.
Proceedings of SAMO, Belgirate, pages 40–42.

Sobol', I. M. and Kucherenko, S. (2009).
Derivative based global sensitivity measures and the link with global sensitivity indices.
Mathematics and Computers in Simulation, 79:3009–3017.