Derivative-based global sensitivity measures: general links with Sobol' indices and numerical tests

M. Lamboni[ab], B. Iooss[c1], A.-L. Popelin[c], F. Gamboa[b]

[a] *Université Paris Descartes, 45 rue des saints Pères, F-75006, France*
[b] *IMT, F-31062, France*
[c] *EDF R&D, 6 quai Watier, F-78401, France*

Abstract

The estimation of variance-based importance measures (called Sobol' indices) of the input variables of a numerical model can require a large number of model evaluations. It turns to be unacceptable for huge model involving a large number of input variables (typically more than ten). Recently, Sobol and Kucherenko have proposed the Derivative-based Global Sensitivity Measures (DGSM), defined as the integral of the squared derivatives of the model output, showing that it can help to solve the problem of dimensionality in some cases. We provide a general inequality link between DGSM and total Sobol' indices for input variables belonging in the class of Boltzmann probability measures, extending the previous results of Sobol and Kucherenko for uniform and normal measures. The special case of log-concave measures is also described. This link provides a DGSM-based maximal bound for the total Sobol indices. Numerical tests show the performance of the bound and its usefulness in practice.

Keywords: Boltzmann measure; Derivative based global sensitivity measure; Global sensitivity analysis; Log-concave measure; Poincaré inequality; Sobol' indices

1. Introduction

With the advent of computing technology and numerical methods, computer models are now widely used to make predictions on miskwnown physical phenomena, to solve optimization problems or to perform sensitivity studies. These complex models often include hundreds or thousands uncertain variables as inputs, whose uncertainties can strongly impact the model outputs (De Rocquigny *et al.* [1], Kleijnen

[1]Corresponding author: bertrand.iooss@edf.fr, Phone: +33 1877969, Fax: +33 130878213

[2], Patelli *et al.* [3]). In fact, it is well known that, in many cases, only a small number of input variables really act in the model (Saltelli *et al.* [4]). This number is referred to the notion of the effective dimension of a function (Caflish *et al.* [5]), which is a useful way to deal with the curse of dimensionality in practical applications.

Global Sensitivity Analysis (GSA) methods (Sobol [6], Saltelli *et al.* [4]) aim to apportion model output variability into input variables and their interactions. It is also an objective way to determine the effective dimension by using the model simulations (Kucherenko *et al.* [7]). A first class of GSA methods is qualitative and is called the "screening", as it aims to deal with a large number of input variables (from tens to hundreds). An example of screening method is the Morris' method (Morris [8]). With a few model evaluations, it allows a coarse estimation of the main effects but misses out interactions among input variables. The second class of GSA methods are the popular quantitative methods, mainly based on the decomposition of the model output variance, which leads to the so-called variance-based methods and Sobol' sensitivity indices. It allows computing the main and total effects (called first order and total Sobol' indices) of each input variable, as interaction effects. However, the estimation procedures are more expensive in terms of number of required model evaluations. Then, variance-based methods can only be applied to model with a small number of input variables (no more than tens).

Recently, Sobol and Kucherenko [9, 10] have proposed a rigorous mathematical formulation of the Morris method by the use of the so-called Derivative-based Global Sensitivity Measures (DGSM). DGSM seem computationally more tractable than variance-based measures and less suffer the curse of dimensionality. They also theoretically showed an inequality linking DGSM and total Sobol' indices in the case of uniform or Gaussian input variables.

In this paper, we investigate this close relationship between total Sobol' indices and DGSM, by extending this inequality to a large class of Boltzmann probability measures. We also obtain result for the class of log-concave measures. The paper is organized as follows. Section 2 recalls some useful definitions of Sobol' indices and DGSM. Section 3 establishes an inequality between these indices for a large class of Boltzmann (resp. log-concave) probability measures. Section 4 provides some numerical simulations on two test models, illustrating how the bound can be used in practice. We conclude in Section 5.

2. Global sensitivity indices definition

2.1. Variance-based sensitivity indices

Let $Y = f(\mathbf{X})$ be a model output with d random input variables $\mathbf{X} = (X_1, \ldots, X_d)$. If the input variables are independent (assumption A1) and $\mathbb{E}\left(f^2(\mathbf{X})\right) < +\infty$ (assumption A2), we have the following unique Hoeffding decomposition (Efron and Stein [11]) of $f(\mathbf{X})$:

$$
\begin{aligned}
f(\mathbf{X}) &= f_0 + \sum_{j}^{d} f_j(X_j) + \sum_{i<j}^{d} f_{ij}(X_i, X_j) + \ldots + f_{1\ldots d}(X_1, \ldots, X_d) && (2.1)\\
&= \sum_{u \subset \{1,2,\ldots d\}} f_u(X_u), && (2.2)
\end{aligned}
$$

where $f_0 = \mathbb{E}\left[f(\mathbf{X})\right]$ corresponds to the empty subset; $f_j(X_j) = \mathbb{E}\left[f(\mathbf{X})|X_j\right] - f_0$ and $f_u(X_u) = \mathbb{E}\left[f(\mathbf{X})|X_u\right] - \sum_{v \subset u} f_v(X_v)$ for any subset $u \subset \{1, 2, \ldots, d\}$.

By regrouping all the terms in equation (2.1) that contain the variable X_j ($j = 1, 2, \ldots, d$) in the function called $g(\cdot)$:

$$
g(X_j, \mathbf{X}_{\sim j}) = \sum_{u \ni j} f_u(\mathbf{X}_u),
$$

we have the following decomposition:

$$
f(\mathbf{X}) = f_0 + g(X_j, \mathbf{X}_{\sim j}) + h(\mathbf{X}_{\sim j}), \qquad (2.3)
$$

where $\mathbf{X}_{\sim j}$ denotes the vector containing all variables except X_j. Further set $h(\cdot) =$

$f(\cdot) - f_0 - g(\cdot)$. Notice that this decomposition is also unique under assumptions A1 and A2. The function $g(\cdot)$, itself, suffices to compute the total sensitivity indices. Indeed, it contains all information about the variable X_j.

Definition 2.1. *Assume that A1, A2 hold, let $\mu(\mathbf{X}) = \mu(X_1, \ldots, X_d)$ be the distribution of the input variables. For any non empty subset $u \subseteq \{1, 2, \ldots, d\}$, set first*

$$D = \int f^2(\mathbf{x}) d\mu(\mathbf{x}) - f_0^2 \ ,$$

$$D_u = \int f_u^2(\mathbf{x}_u) d\mu(\mathbf{x}_u) \ ,$$

$$D_u^{tot} = \int \sum_{v \supseteq u} f_v^2(\mathbf{x}_v) d\mu(\mathbf{x}_v), \tag{2.4}$$

Futher, the first order Sobol sensitivity indices (Sobol [6]) of $\mathbf{X}_u$ is

$$S_u = \frac{D_u}{D} \ , \tag{2.5}$$

The total sensitivity Sobol index of $\mathbf{X}_u$ (Homma and Saltelli [12]) is

$$S_{T_u} = \frac{D_u^{tot}}{D} \ . \tag{2.6}$$

Note that D is nothing more than the variance of $f(\mathbf{X})$.

The following proposition gives another way to compute the total sensitivity indices.

Proposition 2.1. *Under assumptions A1 and A2, the total sensitivity indices of variable X_u with $u = \{j\}$ $(j = 1, 2, \ldots, d)$ is obtained by the following formulas:*

$$D_j^{tot} = \int g^2(x_j, \mathbf{x}_{\sim j}) d\mu(\mathbf{x}) \tag{2.7}$$

and

$$D_j^{tot} = \frac{1}{2} \int \left[f(\mathbf{x}) - f(x_j', \mathbf{x}_{\sim j}) \right]^2 d\mu(\mathbf{x}) d\mu(x_j') \ . \tag{2.8}$$

Proof 2.1. *The first formula is an obvious consequence of equation (2.3), and it is obtained by using the orthogonality constraints among terms in the decomposition of equation (2.1). Indeed, $D_j^{tot} = \int \sum_{v \supseteq j} f_v^2(\mathbf{x}_v) d\mu(\mathbf{x}_v) = \int \left[\sum_{v \supseteq j} f_v(\mathbf{x}_v) \right]^2 d\mu(\mathbf{x}) = \int g^2(x_j, \mathbf{x}_{\sim j}) d\mu(\mathbf{x})$. The later formula is proved in Sobol [13].*

2.2. Derivative-based sensitivity indices

Derivative-based global sensitivity method uses the second moment of model derivatives as importance measure. This method is motivated by the fact that a high value of the derivative of the model output with respect to some input variable means that a big variation of model output is expected for a variation of the variable. This method extends the Morris method (Morris [8]). Indeed, it allows to capture any small variation of the model output due to a single input variable.

DGSM have been first proposed in Sobol and Gresham [14]. Then, they have been largely studied in Kucherenko *et al*. [15], Sobol and Kucherenko [9] and Patelli *et al*. [16]. from now, we assume that the function f is differentiable. Two kind of DGSM are defined below:

Definition 2.2. *Assume that A1 holds and that* $\dfrac{\partial f(\mathbf{X})}{\partial x_j}$ *is square-integrable (assumption A3). Then, for* $j = 1, 2, \ldots d$*, we define the DGSM indices by:*

$$\begin{aligned} \nu_j &= \mathbb{E}\left[\left(\frac{\partial f(\mathbf{X})}{\partial x_j}\right)^2\right] \\ &= \int \left(\frac{\partial f(\mathbf{x})}{\partial x_j}\right)^2 d\mu(\mathbf{x}) \, . \end{aligned} \tag{2.9}$$

Let $w(\cdot)$ *is be a bounded measurable function. A weighted version of the last indices is:*

$$\tau_j = \int \left(\frac{\partial f(\mathbf{x})}{\partial x_j}\right)^2 w(x_j) d\mu(\mathbf{x}). \tag{2.10}$$

Remark 2.1. *Sobol and Kucherenko [10] showed that, for a specific weighting function* $w(x_j) = \dfrac{1 - 3x_j + 3x_j^2}{6}$ *and for a class of linear model with respect to each input variable, we have* $\tau_j = D_j^{tot}$*.*

Remark 2.2. *By bearing in mind the decomposition in equation (2.3), we can replace in equations (2.9) and (2.10) the function* $f(\cdot)$ *by the function* $g(\cdot)$*. In general,* $g(\cdot)$ *is a* d_1 $(d_1 \leq d)$ *dimension function, and this can drastically reduce the number of model evaluations for the numerical computation of* ν *or* τ*. Thus, we have:*

$$\nu_j = \int \left(\frac{\partial g(\mathbf{x})}{\partial x_j}\right)^2 d\mu(\mathbf{x}) \, . \tag{2.11}$$

$$\tau_j = \int \left(\frac{\partial g(\mathbf{x})}{\partial x_j}\right)^2 w(x_j) d\mu(\mathbf{x}), \tag{2.12}$$

3. Variance-based sensitivity indices vs. derivative-based sensitivity indices

A formal link between total Sobol' indices and DGSM is worth interesting to control Sobol' indices and to use the DGSM in practice for factors prioritization. Indeed, DGSM estimations need much less model evaluations than total Sobol' indices estimations (Kucherenko *et al.* [15]). Sobol and Kucherenko [9] have established an inequality linking these two indices for uniform and Gaussian random variables (maximal bound for S_{T_j}). In this section, we extend the inequality for any model when the marginal distribution of input variable are Boltzmann measure on $\mathbb{R}$ (assumption A4). A measure δ on $\mathbb{R}$ is said to be a Boltzmann measure if it is absolutely continuous with respect to the Lebesgue measure and its density $d\delta(x) = \rho(x)dx = c\exp[-v(x)]dx$. Here $v(\cdot)$ is a continuous function and c a normalizing constant. Many classical continuous probability measures used in practice (see de Rocquigny *et al.* [1] and Saltelli *et al.* [4]) are Boltzmann measures.

The class of Boltzmann probability measures includes the well known class of log-concave probability measures. In this case, $v(\cdot)$ is a convex function (assumption A5). In other words, a twice differentiable probability density function $\rho(x)$ is said to be log-concave if, and only if,

$$\frac{d^2}{dx^2}[\log \rho(x)] \leq 0 \; . \tag{3.13}$$

Note that the density of the uniform probability measure on an interval is not continuous on $\mathbb{R}$. So it cannot be considered in the class of log-concave probability measure, nor in the class of Boltzmann probability measures.

The two following propositions give the formal link between Sobol' indices and derivative-based sensitivity indices.

Theorem 3.1. *Under assumptions A1, A2, A3 and A4, we have:*

$$D_j^{tot} \leq C(\mu_j)\nu_j \tag{3.14}$$

with $C(\mu_j) = 4C_1^2$ *and* $C_1 = \sup_{x\in\mathbb{R}} \dfrac{\min(F_j(x), 1 - F_j(x))}{\rho_j(x)}$ *the Cheeger constant,* $F_j(\cdot)$ *the cumulative probability function of* X_j *and* $\rho_j(\cdot)$ *the density function of* X_j*.*

We recall the four assumptions:

- A1: independence between inputs $X_1, X_2, \ldots, X_d$,
- A2: $f \in L^2(\mathbb{R})$,
- A3: $\dfrac{\partial f}{\partial x_j} \in L^2(\mathbb{R})$,
- A4: the distribution of X_j is a Boltzmann probability measure.

Proof 3.1. *The inequality (3.14) is a one-dimensional Poincaré inequality by using expressions* $D_j^{tot} = \int g^2(x_j, \mathbf{x}_{\sim j}) d\mu(\mathbf{x})$ *(equation (2.7)), and* $\nu_j = \int \left(\dfrac{\partial g(\mathbf{x})}{\partial x_j}\right)^2 d\mu(\mathbf{x})$ *(equation (2.11)). The constant is obtained in Bobkov [17], and Fougères [18] for the one-dimensional Poincaré inequality. A proof of the d-dimensional Poincaré inequality is given in Bakry et al. [19].*

Theorem 3.2. *Under assumptions A1, A2, A3 and A5, we have:*

$$D_j^{tot} \leq [\exp(v(m))]^2 \, \nu_j \ , \tag{3.15}$$

with $C_1 = \dfrac{\exp(v(m))}{2}$ *the Cheeger constant and* m *the median of the measure* μ_j *(such that* $\mu(X_j \leq m) = \mu(X_j > m)$*).*

We recall the assumption A5: the distribution of X_j is a log-concave probability measure.

Proof 3.2. *See proof 3.1.*

Table 1 shows the Cheeger constant for some probability distributions that are log-concave and are used in practice of uncertainty and sensitivity analysis. We also give their medians and the functions $v(\cdot)$. We obtain the same results for the normal distribution $\mathcal{N}(\mu, \sigma^2)$ than Sobol and Kucherenko [9] but we prove them in another way (in this case, $v(m) = \log(\sigma)$). For uniform distribution $\mathcal{U}[a\, b]$, Sobol and Kucherenko [9] obtained via direct integral manipulations the inequality $D_j^{tot} \leq$

Distribution	$v(x)$	m	C_1
Normal $\mathcal{N}(\mu, \sigma^2)$	$\frac{(x-\mu)^2}{2\sigma^2} + \log(\sigma)$	μ	$\frac{\sigma}{2}$
Exponential $\mathcal{E}(\lambda)$, $\lambda > 0$	$\lambda x - \log(\lambda)$	$\frac{\log 2}{\lambda}$	$\frac{1}{\lambda}$
Beta $\mathcal{B}(\alpha, \beta)$, $\alpha, \beta \geq 1$	$\log\left[x^{1-\alpha}(1-x)^{1-\beta}\right]$	No expression	—
Gamma $\Gamma(\alpha, \beta)$, scale $\alpha \geq 1$, shape $\beta > 0$	$\log\left(x^{1-\alpha}\Gamma(\alpha)\right) + \frac{x}{\beta} + \alpha \log \beta$	No expression	—
Gumbel $\mathcal{G}(\mu, \beta)$, scale $\beta > 0$	$\frac{x-\mu}{\beta} + \log \beta + \exp\left(-\frac{x-\mu}{\beta}\right)$	$\mu - \beta \log(\log 2)$	$\frac{\beta}{\log 2}$
Weibull $\mathcal{W}(k, \lambda)$, shape $k \geq 1$, scale $\lambda > 0$	$\log\left(\frac{\lambda}{k}\right) + (1-k)\log\left(\frac{x}{\lambda}\right) + \left(\frac{x}{\lambda}\right)^k$	$\lambda(\log 2)^{1/k}$	$\frac{\lambda(\log 2)^{(1-k)/k}}{k}$

Table 1: $v(\cdot)$ Standard log-concave probability distributions: $v(\cdot)$ function, median m and Cheeger constant C_1 (see Theorem 3.2).

$\frac{(b-a)^2}{\pi^2}\nu_j$. This relation is a the classical Poincaré or Writtinger inequality (Ane *et al.* [20]).

For general log-concave measures, no analytical expressions are available for the Cheeger constant. In this latter case or in case of non log-concave but Boltzmann measure, we can obtain the Cheeger constant by numerical simulation using the expression $\sup_{x\in\mathbb{R}} \frac{\min(F_j(x), 1-F_j(x))}{\rho_j(x)}$.

4. Numerical tests

4.1. Derivative sensitivity indices estimates

For the DGSM, a classical estimator of DGSM indices is the empirical one and is given below:

$$\widehat{\nu}_j = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial f(\mathbf{X}^{(i)})}{\partial x_j}\right)^2. \tag{4.16}$$

Experimental convergence properties of this estimator are given in Sobol and Kucherenko [9].

From definition (2.3), we know that $\frac{\partial f(\mathbf{X}^{(i)})}{\partial x_j} = \frac{\partial g(\mathbf{X}^{(i)})}{\partial x_j}$. Estimator of D_j^{tot} (see equation (2.7)) and estimator (4.16) are based on the same function $g(\cdot)$ and it seems that estimations of these two indices will require approximately the same number of model evaluations in order to converge towards their respective values.

Computations of DGSM and Sobol' indices can be done by using Monte Carlo algorithm or any variates from it, such as Latin Hypercube Sampling, quasi-Monte Carlo and Monte Carlo Markov Chain sampling. Kucherenko *et al.* [15] have shown that quasi-Monte Carlo outperforms Monte Carlo when model has a low effective dimension. Computation of DGSM needs model gradient estimation. For complex models, model gradient computation can easily be obtained by finite difference method. Patelli and Pradlwarter [3] propose a Monte Carlo estimation of gradient in high dimension. They used an unbiased estimator for gradients and have shown that the number of Monte Carlo evaluations $n \leq d$ is sufficient for gradient computations. In the worst case, their procedure requires the same number of model evaluations than the finite difference method. The method is very efficient when the model has a low effective dimension. When there are many dominate gradient values, an orthogonal linear transformation allows to be in a new space with a few dominate variables.

In the following Sections, we compare the estimates of the Sobol indices (S_j and

S_{T_j}) and the upper bound of S_{T_j} (see inequality (3.14)). let denote Υ_j, the total sensitivity upper bound:

$$\Upsilon_j = K \frac{\nu_j}{D} \,, \tag{4.17}$$

where D is the variance of the model output $f(\mathbf{X})$ and $K = 4C_1^2$. The goal of our numerical tests is just to compare the differences in terms of ranking and not to study the speed of convergence of the estimates.

4.2. Test on the Morris function

As a first test, we consider the Morris function (Morris [8]) that includes 20 independent and uniform input variables. The Morris function is defined by the following equation:

$$y = \beta_0 + \sum_{i=1}^{20} \beta_i w_i + \sum_{i<j}^{20} \beta_{i,j} w_i w_j + \sum_{i<j<l}^{20} \beta_{i,j,l} w_i w_j w_l + \sum_{i<j<l<s}^{20} \beta_{i,j,l,s} w_i w_j w_l w_s \,, \tag{4.18}$$

where $w_i = 2\left(x_i - \frac{1}{2}\right)$ except for $i = 3, 5, 7$ where $w_i = 2\left(1.1 \frac{x_i}{x_i + 1} - \frac{1}{2}\right)$. The coefficient values are:

$\beta_i = 20$ for $i = 1, 2, \ldots, 10$,

$\beta_{i,j} = -15$ for $i, j = 1, 2, \ldots, 6$, $i < j$

$\beta_{i,j,l} = -10$ for $i, j, l = 1, 2, \ldots, 5$, $i < j < l$

and $\beta_{1,2,3,4} = 5$.

The remaining first and second order coefficients were generated independently from the normal distribution $\mathcal{N}(0,\, 1)$ and the remaining third and fourth coefficient were set to 0.

We replace the uniform distributions associated to several input variables by different log-concave measures of the Table 1 in order to show how the bounds can be used in practical sensitivity analysis. Table 2 shows the probability distributions associated to each input of the Morris function.

We have performed some simulations that allow computing the DGSM indices

Input	Probability distribution	Input	Probability distribution
$X1$	$\mathcal{U}[0,1]$	$X11$	$\mathcal{U}[0,1]$
$X2$	$\mathcal{N}(0.5,0.1)$	$X12$	$\mathcal{N}(0.5,0.1)$
$X3$	$\mathcal{E}(4)$	$X13$	$\mathcal{E}(4)$
$X4$	$\mathcal{G}(0.2,0.2)$	$X14$	$\mathcal{G}(0.2,0.2)$
$X5$	$\mathcal{W}(2,0.5)$	$X15$	$\mathcal{W}(2,0.5)$
$X6$	$\mathcal{U}[0,1]$	$X16$	$\mathcal{U}[0,1]$
$X7$	$\mathcal{U}[0,1]$	$X17$	$\mathcal{U}[0,1]$
$X8$	$\mathcal{U}[0,1]$	$X18$	$\mathcal{U}[0,1]$
$X9$	$\mathcal{U}[0,1]$	$X19$	$\mathcal{U}[0,1]$
$X10$	$\mathcal{U}[0,1]$	$X20$	$\mathcal{U}[0,1]$

Table 2: Probability distributions of the input variables of the Morris function

and the Sobol' indices for the 20 independent factors. The goal here is not to compare their algorithmic performances in terms of simulation number, but just to look at the inputs' ranking. Sobol' indices S_j and S_{Tj} are obtained with the method of Saltelli [21], using two initial Monte Carlo samples of size 5×10^4. With 20 input variables, it leads to 1.1×10^6 model evaluations. The total Sobol' indices are used in this paper as a reference. It shows that only the first 10 inputs have some influence. Model derivatives are evaluated for each input on a Monte Carlo sample of size 1×10^4 by the finite-difference method (perturbation of 0.01%). Then, DGSM ν_j require 2.1×10^5 model evaluations. Υ_j is then computed using equation (4.17) where the variance of the Morris function is estimated to $D = 991.521$. The results are available in Table 3.

In Table 3, we can first observe that the total sensitivity upper bounds Υ_j are always greater than the total sensitivity indices as expected. For each input, we distinguish several situations that can occur:

1. First order and total Sobol' indices are negligible (inputs $X11$ to $X20$). In this case, we observe that the bound Υ_j is always negligible. For all the inputs, this test shows the high efficiency of the bound: a negligible bound warrants that the input has no influence.
2. First order and total Sobol' indices significantly differ from zero and have

Input	S_j	sd	S_{Tj}	sd	ν_j	K	Υ_j
X1	0.046	0.008	0.172	0.004	2043.820	0.101	0.209
X2	0.010	0.009	0.029	0.001	2856.580	0.01	0.029
X3	0.070	0.008	0.166	0.003	31653.270	0.250	7.981
X4	0.006	0.010	0.134	0.002	2025.950	0.333	0.680
X5	0.037	0.009	0.054	0.002	4203.060	0.360	1.526
X6	0.040	0.009	0.114	0.003	1337.100	0.101	0.137
X7	0.070	0.008	0.068	0.002	6605.960	0.101	0.675
X8	0.157	0.008	0.155	0.003	1826.390	0.101	0.187
X9	0.191	0.008	0.191	0.004	2249.770	0.101	0.230
X10	0.149	0.008	0.147	0.004	1730.400	0.101	0.177
X11	0.003	0.009	0.002	0.000	22.630	0.101	0.002
X12	0.003	0.009	0.000	0.000	23.940	0.01	0.000
X13	0.003	0.009	0.001	0.000	17.670	0.250	0.004
X14	0.004	0.009	0.003	0.000	42.850	0.333	0.014
X15	0.003	0.009	0.001	0.000	19.870	0.360	0.007
X16	0.003	0.009	0.002	0.000	18.860	0.101	0.002
X17	0.003	0.009	0.002	0.000	21.400	0.101	0.002
X18	0.003	0.009	0.002	0.000	19.950	0.101	0.002
X19	0.006	0.009	0.005	0.001	54.380	0.101	0.006
X20	0.004	0.009	0.003	0.001	42.250	0.101	0.004

Table 3: Sensitivity indices (Sobol' and DGSM) for the Morris function. For the Sobol' indices S_j and S_{Tj}, 20 replicates has been used to get the standard deviation (sd).

approximately the same value (inputs $X7$ to $X10$). It means that the input has some influence but no interacts with other inputs. In this case, the bound is clearly relevant (except for $X7$). The interpretation of the bound gives a useful information about the total influence of the input.

3. First order Sobol' index is negligible while total Sobol' index significantly differs from zero (inputs $X1$ to $X6$). In this case, the bound Υ_j largely overstimates the total Sobol' index S_{Tj} for $X3$, $X4$ and $X5$. However, for X_4, we have $\Upsilon_4 < 1$ and this coarse information is still usefull. For the three other inputs, the bound is relevant.

For two inputs ($X3$ and $X5$), results can be judged as strongly unsatisfactory as the bound is useless (larger than 1 which is the maximal value for a sensitivity index). We suspect that these results come from:

- the model non linearity with respect to these inputs (see equation (4.18)),
- the input distributions (exponential and Weibull).

The second explanation seems to be the more convincing as these types of distribution can provide larger values during Monte Carlo simulations. In this case, departures from the central part of the input domain leads to uncontrolled derivative values of the Morris function. Indeed, it can be seen that ν_j is particularly large for $X3$ and $X5$, because of high derivative values in the estimation samples.

As a conclusion of this first test, we argue that the bound Υ_j is well-suited for a screening purpose. Moreover, coupling Υ_j interpretation with first order Sobol' indices S_j (estimated at low cost using a smoothing technique or a metamodel, see [4, 22]) can bring useful information about the presence or absence of interaction. For inputs following uniform or normal and exponential distributions, bound seems to be extremely efficient. In these particular cases, the bound is the best one and cannot be improved.

4.3. A case study: a flood model

To illustrate how the Cheeger constant can be used for factors prioritization, when we use the DGSM, we consider a simple application model that simulates the height of a river compared to the height of a dyke. When the height of a river is over the height of the dyke, flooding occurs. This academic model is used as a pedagogical example in Iooss [22]. The model is based on a crude simplification of the 1D hydro-dynamical equations of SaintVenant under the assumptions of uniform and constant flowrate and large rectangular sections. It consists of an equation that involves the characteristics of the river stretch:

$$S = Z_v + H - H_d - C_b \quad \text{with} \quad H = \left(\frac{Q}{BK_s\sqrt{\frac{Z_m - Z_v}{L}}} \right)^{0.6}, \tag{4.19}$$

with S the maximal annual overflow (in meters) and H the maximal annual height of the river (in meters).

The model has 8 input variables, each one follows a specific probability distribution (see Table 4). Among the input variables of the model, H_d is a design parameter. The randomness of the other variables is due to their spatio-temporal variability, our ignorance of their true value or some inaccuracies of their estimation. We suppose that the input variables are independent.

Input	Description	Unit	Probability distribution
Q	Maximal annual flowrate	m^3/s	Truncated Gumbel $\mathcal{G}(1013, 558)$ on $[500, 3000]$
K_s	Strickler coefficient	-	Truncated normal $\mathcal{N}(30, 8)$ on $[15, +\infty[$
Z_v	River downstream level	m	Triangular $\mathcal{T}(49, 50, 51)$
Z_m	River upstream level	m	Triangular $\mathcal{T}(54, 55, 56)$
H_d	Dyke height	m	Uniform $\mathcal{U}[7, 9]$
C_b	Bank level	m	Triangular $\mathcal{T}(55, 55.5, 56)$
L	Length of the river stretch	m	Triangular $\mathcal{T}(4990, 5000, 5010)$
B	River width	m	Triangular $\mathcal{T}(295, 300, 305)$

Table 4: Input variables of the flood model and their probability distributions

We also consider another model output: the associated cost (in million euros) of the dyke presence:

$$C_p = \mathbb{1}_{S>0} + \left[0.2 + 0.8\left(1 - \exp^{-\frac{1000}{S^4}}\right)\right] \mathbb{1}_{S\leq 0} + \frac{1}{20}\left(H_d \mathbb{1}_{H_d>8} + 8\mathbb{1}_{H_d\leq 8}\right), \quad (4.20)$$

with $\mathbb{1}_A(x)$ the indicator function which is equal to 1 for $x \in A$ and 0 otherwise. In this equation, the first term represents the cost due to a flooding ($S > 0$) which is 1 million euros, the second term corresponds to the cost of the dyke maintenance ($S \leq 0$) and the third term is the investment cost related to the construction of the dyke. The latter cost is constant for a height of dyke less than 8 m and is growing proportionally with respect to the dyke height otherwise.

Global sensitivity analysis was performed with 2×10^6 model evaluations in order to compute Sobol' indices (first order indices S_j and total indices S_{T_j}). We use Sobol sequences, as input samples, to simulate the input values. For estimating the DGSM

(ν_j, weighted DGSM τ_j and the total sensitivity upper bound Υ_j), a Sobol sequence is also used with 1×10^4 model evaluations.

Results of global sensitivity analysis and derivative-based global sensitivity analysis for respectively the overflow S and the cost C_p outputs are listed in Tables 5 and 6. Global sensitivity indices show small interaction among input variables for the overflow and the cost outputs. Four input variables (Q, H_d, K_s, Z_v) drive the overflow and the cost outputs. This variable classification will serve as reference for comparison issue.

Input	S_j	S_{Tj}	ν_j	τ_j	Υ_j
Q	0.345	0.353	1.296e-06	1.072	2.807
K_s	0.134	0.142	3.286e-03	1.033	0.198
Z_v	0.190	0.189	1.123e+00	1377.41	0.561
Z_m	0.003	0.003	2.279e-02	33.742	0.011
H_d	0.284	0.283	8.389e-01	23.77	0.340
C_b	0.035	0.034	8.389e-01	1268.90	0.105
L	0.000	0.000	2.147e-08	0.268	0.000
B	0.000	0.000	2.386e-05	1.070	0.000

Table 5: Sensitivity indices for the overflow output of the flood model.

Input	S_j	S_{Tj}	ν_j	τ_j	Υ_j
Q	0.361	0.478	1.3906e-06	2.013	3.011e+00
K_s	0.160	0.249	8.5307e-03	1.926	5.129e-01
Z_v	0.172	0.219	1.3891e+00	1715.89	6.932e-01
Z_m	0.007	0.004	4.6038e-02	68.17	2.29e-02
H_d	0.121	0.172	1.5366e+00	44.04	6.227e-01
C_b	0.033	0.036	9.4628e-01	1428.69	1.180e-01
L	0.003	-0.003	4.0276e-08	0.503	2.009e-06
B	0.003	-0.003	4.4788e-05	2.007	5.587e-04

Table 6: Sensitivity indices for the cost ouput of the flood model.

Based on derivative sensitivity indices (ν_j) or weighted derivative sensitivity indices (τ_j) we have obtained another subset of the most influential variables such as Z_v, C_b, H_d, Z_m. These results mean that, for example, the maximum annual flowrate (Q) does not have any impact on the overflow and the cost output. If we compare these results to the global sensitivity indices, the results are obviously wrong. This is easily explained by the fact that the input variables have different unities and that the indices ν_j and τ_j do not take into account the probability distribution of X_j. However, the difference between total sensitivity S_{T_j} and weighted derivative sensitivity τ_j seems to suggest the non-linearity between model output and its input variables.

By looking at the total sensitivity upper bound Υ_j, the most influential variables are the following: Q, Z_v, H_d, K_s for the overflow output and for the cost output. It gives the same subset of the most influential variables with some slight differences for the prioritization of the most influential variables. In conclusion, we confirm that in practice, if Sobol' indices cannot be estimated because of a cpu time expensive model, Υ_j can provide correct information on input variance-based sensitivities.

5. Conclusion

Global sensitivity analysis, that allows exploring numerically complex model and setting factors prioritization, requires a large number of model evaluations. Derivative-based global sensitivity method needs a much smaller number of model evaluations (gain factor of 10 to 100). The reduction of the number of model evaluations becomes more significant when the model output is controlled by a small number of input variables and when the model does not include much interaction among input variables. This is often the case in practice.

In this paper, we have produced an inequality linking the total Sobol' index and a derivative-based sensitivity measure for a large class of probability distributions (Boltzmann measures). The new sensitivity index Υ_j, which is defined as a constant times the crude derivative-based sensitivity, is a maximal bound of the total Sobol' index. It improves factor prioritization by using derivative-based sensitivities instead of variance-based sensitivities.

Two numerical tests have confirmed that the bound Υ_j is well-suited for a screening purpose. When total Sobol' indices cannot be estimated because of a cpu time expensive model, Υ_j can provide correct information on input sensitivities. Previous studies have shown that estimating DGSM with a small derivatives' sample (with size from tens to hundreds) allows to detect non influent inputs. In subsequent works, we propose to use jointly DGSM and first order Sobol' indices. With these information, an efficient methodology of global sensitivity analysis can be applied and brings useful information about the presence or absence of interaction (see Iooss et al. [23]).

6. Acknowlegments

Part of this work has been backed by French National Research Agency (ANR) through COSINUS program (project COSTA BRAVA noANR-09-COSI-015). We thank Jean-Claude Fort for helpful discussions.

References

[1] E. de Rocquigny, N. Devictor, S. Tarantola (Eds.), Uncertainty in industrial practice, Wiley, 2008.

[2] J. Kleijnen, Design and analysis of simulation experiments, Springer, 2008.

[3] E. Patelli, H. Pradlwarter, Monte Carlo gradient estimation in high dimensions, International Journal for Numerical Methods in Engineering 81 (2) (2010) 172–188.

[4] A. Saltelli, K. Chan, E. Scott (Eds.), Sensitivity analysis, Wiley Series in Probability and Statistics, Wiley, 2000.

[5] R. Caflisch, W. Morokoff, A. Owen, Valuation of mortgage backed securities using brownian bridges to reduce effective dimension, Journal of Computational Finance 1 (1997) 27–46.

[6] I. Sobol, Sensitivity analysis for non-linear mathematical models, Mathematical Modelling and Computational Experiments 1 (1993) 407–414.

[7] S. Kucherenko, B. Feilb, N. Shaha, W. Mauntz, The identification of model effective dimensions using global sensitivity analysis, Reliability Engineering and System Safety 96 (4) (2011) 440 – 449.

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

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

[10] I. Sobol, S. Kucherenko, A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices, Computer Physics Communications 181 (7) (2010) 1212 – 1217.

[11] B. Efron, C. Stein, The jacknife estimate of variance, The Annals of Statistics 9 (1981) 586–596.

[12] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of non linear models, Reliability Engineering and System Safety 52 (1996) 1–17.

[13] I. Sobol, Global sensitivity indices for non linear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation 55 (2001) 271–280.

[14] I. Sobol, A. Gresham, On an alternative global sensitivity estimators, in: Proceedings of SAMO 1995, Belgirate, 1995, pp. 40–42.

[15] S. Kucherenko, M. Rodriguez-Fernandez, C. Pantelides, N. Shah, Monte Carlo evaluation of derivative-based global sensitivity measures, Reliability Engineering and System Safety 94 (2009) 1135–1148.

[16] E. Patelli, H. J. Pradlwarter, G. I. Schuëller, Global sensitivity of structural variability by random sampling, Computer Physics Communications 181 (12) (2010) 2072–2081.

[17] S. Bobkov, Isoperimetric and analytic inequalities for log-concave probability measures, The Annals of Probability 27 (1999) 1903–1921.

[18] P. Fougères, Spectral gap for log-concave probability measures on the real line, in: Séminaire de Probabilités XXXVIII, Vol. 1857 of Lecture Notes in Math., Springer, Berlin, 2005, pp. 95–123.

[19] D. Bakry, F. Barthe, P. Cattiaux, A. Guillin, A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case, Electron. Commun. Probab. 13 (2008) 60–66.

[20] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, G. Scheffer, Sur les inégalités de Sobolev logarithmiques, Vol. 10 of Panoramas et Synthèses, Société Mathématique de France, 2000.

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

[22] B. Iooss, Revue sur l'analyse de sensibilité globale de modèles numériques, Journal de la Société Française de Statistique 152 (2011) 1–23.

[23] B. Iooss, A.-L. Popelin, G. Blatman, C. Ciric, F. Gamboa, S. Lacaze, M. Lamboni, Some new insights in derivative-based global sensitivity measures, in: ESREL 2012 Conference, submitted, Helsinki, Finland, 2012.