Some new insights in derivative-based global sensitivity measures

Bertrand Iooss[a*], Anne-Laure Popelin[a], Géraud Blatman[b], Catalina Ciric[c], Fabrice Gamboa[d], Sylvain Lacaze[b], Matieyendou Lamboni[d]

[a]EDF R&D, Département Management des Risques Industriels, Chatou, France
[b]EDF R&D, Département Matériaux et Mécanique des Composants, Moret-sur-Loing, France
[c]EDF R&D, Laboratoire National d'Hydraulique et d'Environnement, Chatou, France
[d]Institut de Mathématiques de Toulouse, Toulouse, France

Abstract: The estimation of variance-based importance measures (i.e. the sensitivity indices 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). This problem of dimensionality can be solved by using the Derivative-based Global Sensitivity Measures (DGSM), defined as the integral of the squared derivatives of the model output. In this communication, the general inequality link between DGSM and total Sobol' indices is explained for input variables belonging in the classes of Boltzmann and log-concave probability measures. This link provides a DGSM-based maximal bound for the total Sobol' indices. In practice, this inequality allows us to develop a generic strategy to obtain global sensitivity information from DGSM and first-order Sobol indices. Indeed, these two measures are obtained at a low computational cost and do not suffer from a large number of inputs in the numerical model. Numerical tests show the performance of the bound and its usefulness in practice. An application on an aquatic prey-predator chain allows to discriminate non-influential parameters. Results are congruent to the ranking, obtained previously with the Extended FAST method.

Keywords: Computer experiment, Derivative based global sensitivity measure, Sensitivity analysis, Sobol' indices, Uncertainty.

---

1. INTRODUCTION

Computer models simulating physical phenomena and industrial systems, which are commonly used in engineering and safety studies, often take as inputs a high number of numerical and physical variables. For the development and the analyses of such computer models, the global sensitivity analysis methodology is an invaluable tool [1]. Referring to a probabilistic modeling of the model input variables, it accounts for the whole input range of variation, and tries to explain output uncertainties on the basis of input uncertainties. In particular, the powerful total Sobol' sensitivity index gives for one specified input its overall contribution, including the effects of its interactions with all the other inputs, in the variance of the model output. However, obtaining these total sensitivity indices is rather costly in terms of the number of necessary model evaluations [1].

Recently, new sensitivity indices have been proposed in order to manage the problem of the large number of inputs. The so-called Derivative-based Global Sensitivity Measures (DGSM) [2] consist, for each input, in integrating the square derivative of the model output over the domain of the inputs. This kind of indices have been shown to be easily and efficiently estimated by sampling techniques as

Monte Carlo or quasi-Monte Carlo [3]. It allows to gain a factor ranging between 10 to 100 compared to Sobol' indices estimated with the same technique. For an input following a uniform or normal probability distribution, [2] has proved an inequality link between the total Sobol' index and DGSM. In fact, the total Sobol' index is bounded by a term involving a constant and the DGSM. [4] has recently developed a generalization of this inequality for the class of Boltzmann probability measures, which includes all the classical probability measures used in practice of uncertainty and sensitivity analysis. In practice, this inequality allows us to develop a generic strategy to obtain global sensitivity information from DGSM and first-order Sobol indices. Indeed, these two measures are obtained at a low computational cost and do not suffer from a large number of inputs in the numerical model.

In the following section, we first remind the basics of Sobol' indices and DGSM. Then, the general inequality link is presented. Some numerical tests illustrate in Section 3 the bound efficiency and demonstrate the application range of the inequality. Finally, the general methodology of sensitivity analysis based on DGSM is presented and applied in Section 4 on an aquatic prey-predator chain.

2. GLOBAL SENSITIVITY INDICES VS. DERIVATIVE-BASED SENSIVITY MEASURES

2.1. Variance-based sensitivity indices

First, let us recall some basic notions about Sobol' indices. Let define the model

$$\begin{aligned} f: \ & \mathbb{R}^p && \to \mathbb{R} \\ & \boldsymbol{X} && \mapsto Y = f(\boldsymbol{X}) \end{aligned} \tag{2.1}$$

where Y is the code output, $\boldsymbol{X} = (X_1, \ldots, X_p)$ are p independent inputs, and f is the model function. f is supposed to be square-integrable and is considered as a "black box", i.e. a function whose analytical formulation is unknown. The main idea of the variance-based sensitivity methods is to evaluate how the variance of an input or a group of input parameters contributes to the output variance of f. These contributions are described using the following sensitivity indices:

$$S_i = \frac{\mathrm{Var}\left[\mathbb{E}\left(Y|X_i\right)\right]}{\mathrm{Var}(Y)}, \quad S_{ij} = \frac{\mathrm{Var}\left[\mathbb{E}\left(Y|X_i X_j\right)\right]}{\mathrm{Var}(Y)} - S_i - S_j, \quad S_{ijk} = \ldots \tag{2.2}$$

These coefficients, namely the Sobol' indices, can be used for any complex model functions f [5]. The second order index S_{ij} expresses the model sensitivity to the interaction between the variables X_i and X_j (without the first order effects of X_i and X_j), and so on for higher orders effects. The interpretation of these indices is natural as all indices lie in $[0, 1]$ and their sum is equal to one. The larger an index value is, the greater is the importance of the variable or the group of variables related to this index.

For a model with p inputs, the number of Sobol' indices is $2^p - 1$; leading to an intractable number of indices as p increases. Thus, to express the overall output sensitivity to an input X_i, the total sensitivity index has been introduced as

$$S_{T_i} = S_i + \sum_{j \neq i} S_{ij} + \sum_{j \neq i, k \neq i, j<k} S_{ijk} + \ldots = \sum_{l \in \# i} S_l \tag{2.3}$$

where $\#i$ represents all the "non-ordered" subsets of indices containing index i. Thus, $\sum_{l \in \#i} S_l$ is the sum of all the sensitivity indices having i in their index. The estimation of these indices (Eqs. (2.2) and (2.3)) can be performed by simple Monte-Carlo simulations [5, 6]. Refined sampling designs, as Extended FAST (EFAST) and quasi-random sequences [1], allow to significantly reduce the number of required model evaluations. More complex techniques, as smoothing and metamodelling [7, 8], consist in approximating the computer model with a low-cost mathematical function, but are outside the scope of this paper.

2.2. Derivative-based sensitivity indices

Derivative-based global sensitivity method uses the second moment of model derivatives as importance measure. If $\dfrac{\partial f(\mathbf{X})}{\partial x_i}$ is square-integrable, the DGSM of X_i reads:

$$\nu_i = \mathbb{E}\left[\left(\frac{\partial f(\mathbf{X})}{\partial x_i}\right)^2\right] = \int \left(\frac{\partial f(\mathbf{x})}{\partial x_i}\right)^2 d\mu(\mathbf{x})$$

where $\mu(\mathbf{x})$ is the distribution of the input variables.

This index is motivated by the fact that a high value of the derivative of the model output with respect to any input variable leads to much variation of model output. This method is a kind of generalization of the Morris screening method, which allows to capture any small variation of the model output due to each input variable [9]. Reversely, the Morris method can be seen as a coarse finite-difference version of DGSM, with a limited number of samples.

DGSM have been largely studied in [3, 2, 10, 4]. These authors conclude that DGSM are efficiently estimated in most of the cases using quasi-Monte Carlo samples (of size 100 to 1000). The more difficult situation occurs for functions with high effective dimensions in the truncation and superposition senses (i.e. when all inputs are influent, with some interactions of high degree). In this case, the computational cost of estimating DGSM is not smaller than that of Sobol' indices.

2.3. Link between Sobol' indices and DGSM

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 [3]. [2] has established an inequality link between these two indices for the uniform and normal random variables (maximal bound for S_{T_i}). [4] has recently extended this link for any complex model and for any input variable X_i, which has a Boltzmann probability measure $\mu_i = \mu(x_i)$ on $\mathbb{R}$. A measure is said to be a Boltzmann measure if it is absolutely continuous with respect to the Lebesque measure and its density is of the form $d\mu(x) = \rho(x)dx = c\exp[-v(x)]dx$, with $v(\cdot)$ a continuous function and c a normalizing constant. Many classical continuous probability measures used in practice of uncertainty and sensitivity analysis [1] 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. Note that the uniform probability measure 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 following theorem gives the formal link between Sobol' indices and derivative-based sensitivity indices [4].

Theorem 2.1 *If $f \in L^2(\mathbb{R})$, $\dfrac{\partial f}{\partial x_i} \in L^2(\mathbb{R})$, $\mathbf{X}$ is a vector of independent random variables and the law of X_i is a Boltzmann probability measure, we have*

$$S_{T_i} \leq \Upsilon_i \tag{2.4}$$

with

$$\Upsilon_i = \frac{C(\mu_i)}{Var(Y)} \nu_i \tag{2.5}$$

and

$$C(\mu_i) = 4 \left[\sup_{x \in \mathbb{R}} \frac{\min(F_i(x), 1 - F_i(x))}{\rho_i(x)} \right]^2 , \tag{2.6}$$

where $F_i(\cdot)$ is the cumulative probability function of X_i and $\rho_i(\cdot)$ is the density function of X_i.

This result is based on a one-dimensional L^2-Poincaré inequality (see for example [11]): $\|u\|_{L^2} \leq C\|\nabla u\|_{L^2}$ for u a Sobolev' space function. It is applied to the function $g(\cdot)$, defined such that

$$f(\boldsymbol{X}) = f_0 + g(X_j, \boldsymbol{X}_{\sim j}) + h(\boldsymbol{X}_{\sim j}), \tag{2.7}$$

where $f_0 = \mathbb{E}[f(X)]$ and $\boldsymbol{X}_{\sim j}$ denotes the vector containing all variables except X_j. The function $g(\cdot)$ contains all information about the variable X_j and $h(\cdot)$ is the residual function. Notice that this decomposition is unique under the first and third assumptions of Theorem 2.1 ([5, 10]).

In the particular case where X_i follows a log-concave distribution, we have:

$$C(\mu_i) = [\exp(v(m))]^2 , \tag{2.8}$$

where m is the median of the measure μ_i. Table 1 shows the previous constants 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 $v(\cdot)$. For uniform distribution $\mathcal{U}[a\,b]$, [2] obtained via direct integral manipulations the inequality $D_i^{tot} \leq \dfrac{(b-a)^2}{\pi^2}\nu_i$. For some log-concave measures, no analytical expressions are available for the constant. In this latter case or in case of non log-concave but Boltzmann measure, we can obtain the constant by numerical simulation using Eq. (2.6).

3. NUMERICAL TESTS

To study empirically the usefulness and precision of the DGSM bound, numerical tests have been performed in [4] with various input distributions. In this section, we show the majoration of the total Sobol' indices by the DGSM on two different test functions including only uniformly distributed inputs. Our studies concentrate on the convergence of the DGSM, relatively to the total Sobol' indices.

Distribution	$v(x)$	Median m	Constant $C(\mu)$
Normal $\mathcal{N}(\mu,\sigma^2)$	$\frac{(x-\mu)^2}{2\sigma^2}+\log(\sigma)$	μ	σ^2
Exponential $\mathcal{E}(\lambda)$, $\lambda>0$	$\lambda x-\log(\lambda)$	$\log 2/\lambda$	$4/\lambda^2$
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 $\beta>0$	$\frac{x-\mu}{\beta}+\log\beta+\exp\left(-\frac{x-\mu}{\beta}\right)$	$\mu-\beta\log(\log 2)$	$\left(\frac{2\beta}{\log 2}\right)^2$
Weibull $\mathcal{W}(k,\lambda)$, shape $k\geq 1$, scale $\lambda>0$	$\log\left(\frac{\lambda}{k}\right)+(1-k)\log\left(\frac{x}{\lambda}\right)+\left(\frac{x}{\lambda}\right)^k$	$\lambda(\log 2)^{1/k}$	$\left[\frac{2\lambda(\log 2)^{(1-k)/k}}{k}\right]^2$
Uniform $\mathcal{U}[a\,b]$	Not log-concave	$\frac{a+b}{2}$	$\frac{(b-a)^2}{\pi^2}$ (see [2])

Table 1: Standard log-concave probability distributions: $v(\cdot)$ function, median and constant $C(\mu)$.

3.1. Morris function

We consider the reduced Morris function with four inputs:

$$f(\boldsymbol{x})=\sum_{i=1}^{4}b_ix_i+\sum_{i\leq j}^{4}b_{ij}x_ix_j+\sum_{i\leq j\leq k}^{4}b_{ijk}x_ix_jx_k \tag{3.9}$$

$$\text{with } b_i=\begin{bmatrix}0.05\\0.59\\10\\0.21\end{bmatrix},\quad b_{ij}=\begin{bmatrix}0&80&60&40\\0&30&0.73&0.18\\0&0&0.64&0.93\\0&0&0&0.06\end{bmatrix},\quad b_{ij4}=\begin{bmatrix}0&10&0.98&0.19\\0&0&0.49&50\\0&0&0&1\\0&0&0&0\end{bmatrix}.$$

The indices b_{ijk} $\forall\, k\neq 4$ are null. The four input variables x_i $(i=1,\dots,4)$ follow uniform distribution on $[0,1]$. This function allows to control importance of each variable, through the choice of the matrices, not only by their own weight but also through their 2 or 3-order interaction with the others.

DGSM are computed with Monte-Carlo sampling of size n. With n ranging from 10 to 200, Figure 1 (left) shows that DGSM bounds Υ_j are greater than the total Sobol' indices S_{T_i} (for $i=1,2,3,4$) as expected, except for $n<20$ which is a too small sample size. The differences between the Υ_i and S_{T_i} seem to be proportional to S_{T_i}. For small S_{T_i}, Υ_i is close to this value. It confirms that DGSM bounds are first useful for screening exercises. Using Monte Carlo samples, Figure 1 (right) shows that DGSM estimates converge much faster than total Sobol' index ones ([6]'s algorithm). Moreover, DGSM estimate curve is much less chaotic than the Sobol' one.

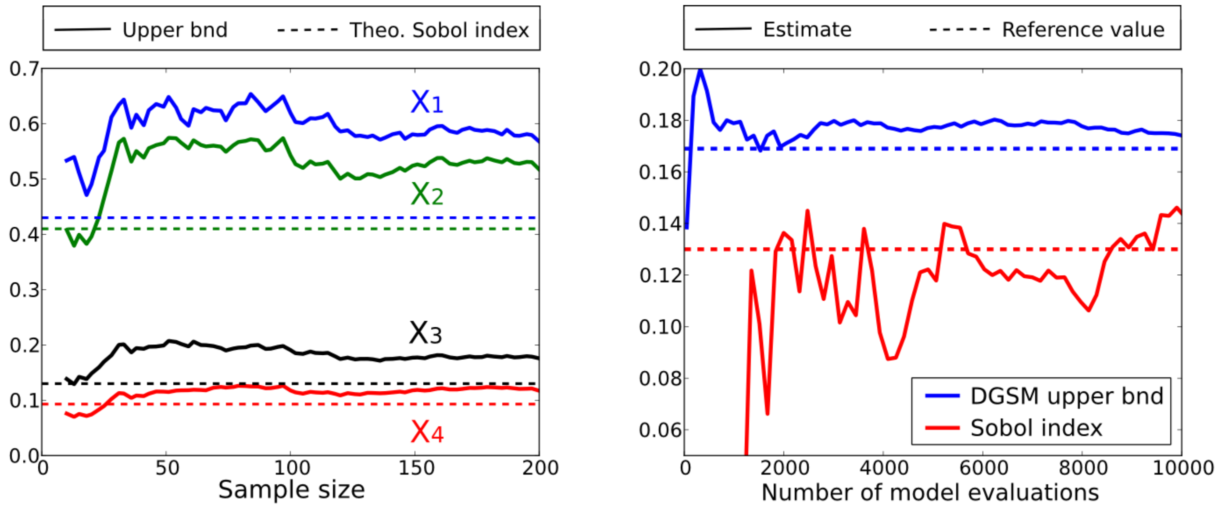

Figure 1: Convergence of the DGSM bound estimates in function of the sample size for the reduced Morris' function. Left: Comparison to theoretical values of total Sobol' indices S_{T_i}. Right: Comparison to Monte Carlo S_{T_i} estimates for the input variable X_4.

3.2. Sobol function

We consider the Sobol function :

$$f(\boldsymbol{x}) = \prod_{i=1}^{8} \frac{|4x_i - 2| + a_i}{1 + a_i} \tag{3.10}$$

where $a = [0, 1, 4.5, 9, 99, 99, 99, 99]$ and the inputs x_i $(i = 1, \dots, 8)$ follow uniform distribution on $[0, 1]$. This function allows to control the importance of each input variable through the choice of the vector $\boldsymbol{a}$. The smaller is a_i, the greater is the variation rank of g_i, increasing the influence of X_i on Y.

Results on Figure 2 (left) show on this case that DGSM related to X_1 and X_2 are very rough upper bounds of their corresponding Sobol' indices since they are greater than 1. Thus they provide no valuable information about the sensitivity indices, except perhaps for the ranking. Figure 2 (right) shows more interesting results for less important variables such as $X_5 - X_8$: DSGM bounds are smaller than 10^{-4} while correctly upper bounding the Sobol indices. Hence, the DGSM make it possible to identify that these variables are negligible. As a conclusion of these tests, we argue that the DGSM bound is well-suited for a screening purpose, and has good convergence properties.

4. APPLICATION OF A DGSM-BASED GENERAL METHODOLOGY ON A PREY-PREDATOR CHAIN

In this section, derivative-based global sensitivity analysis is applied on an aquatic prey-predator chain model. Results on variables screening (by Morris' method) and variance-based sensitivity analysis (using EFAST) have been previously obtained in [12], and will serve as a comparison basis.

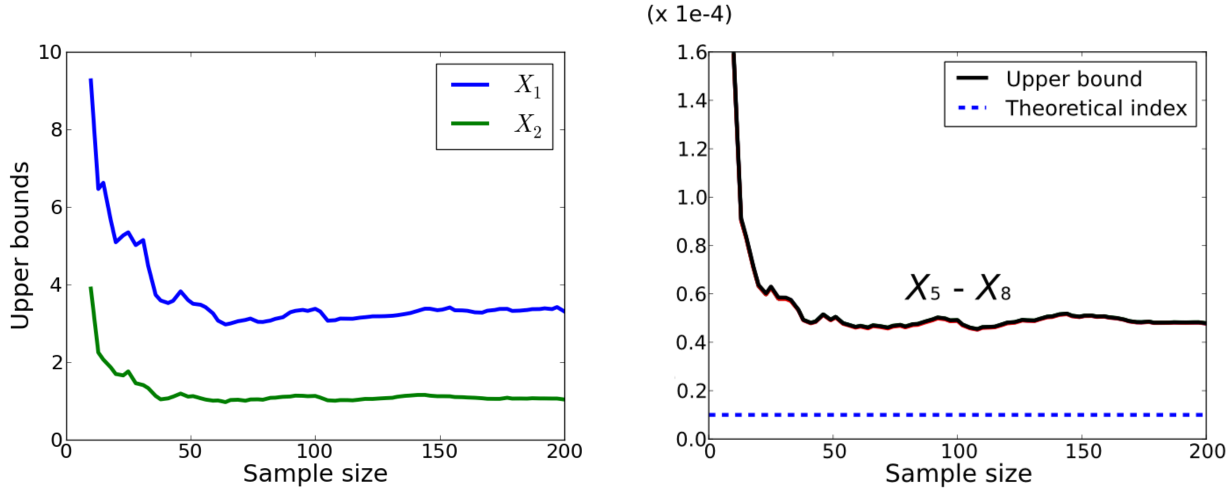

Figure 2: DGSM bound estimates in function of the Monte Carlo sample size.

4.1. Presentation of the DGSM-based general methodology

This methodology is based on the estimation for each input of the first order Sobol' index S_i, the DGSM ν_i, the maximal bound Υ_i (of S_{T_i}), and the variance σ_i^2 of $\partial f/\partial X_i$ (as in the Morris' method [9] where σ_i is used). The estimations of S_i $(i = 1, \ldots, p)$ are done via a n-size sample of $(X_1, \ldots, X_p, Y)$, using local polynomial technique [7]. The estimations of Υ_i and σ_i^2 $(i = 1, \ldots, p)$ are done via a n-size sample of $\left(X_1, \ldots, X_p, \dfrac{\partial Y}{\partial X_1}, \ldots, \dfrac{\partial Y}{\partial X_p}\right)$, using the empirical estimator of Eq (2.4). For each X_i $(i = 1, \ldots, p)$, the procedure is the following:

1. If S_i is judged significantly different from zero (for example $S_i > 0.05$), we know that X_i is influent and has some individual effect on Y. Then:
   (a) if σ_i^2 is significantly smaller than ν_i, we conclude that $S_{T_i} \sim S_i$ (because X_i acts only linearly on Y);
   (b) else, Υ_i is compared to S_i, to have an overall idea of the total influence of X_i with respect to the other inputs.
2. If $S_i \sim 0$, we look at the bound Υ_i to see if there are some possible interactions between X_i and other inputs:
   (a) if $\Upsilon_i \sim 0$, then $S_{T_i} \sim 0$ and X_i can be considered as a non-influent input;
   (b) else, Υ_i is considered as a maximal bound of S_{T_i}, in order to judge the potentiality of X_i to be influent.

In the cases 1(b) and 2(b), we propose to perform a global sensitivity analysis on the model derivatives in order to detect which inputs interact with X_i. From the sample $\left(X_1, \ldots, X_p, \dfrac{\partial Y}{\partial X_1}, \ldots, \dfrac{\partial Y}{\partial X_p}\right)$, we use the same local polynomial technique than for S_i to estimate the first order Sobol' indices for $j = 1, \ldots, p$:

$$S_i^{'j} = \frac{\mathrm{Var}\left[\mathbb{E}\left(\frac{\partial Y}{\partial X_j}|X_i\right)\right]}{\mathrm{Var}\left(\frac{\partial Y}{\partial X_j}\right)} . \tag{4.11}$$

For $i \neq j$, $S_i^{'j}$ gives the interaction effect between X_i and X_j, while for $i = j$, $S_i^{'i}$ gives the non linear effect of X_i. The analysis of these first order Sobol indices of model derivatives is done as usual. For each X_i, if their sum is far from one ($\sum_{j=1}^{p} S_i^{'j} << 1$), it tells us that interaction effects with degree larger than two are influent in the model.

4.2. Presentation of the numerical model

An aquatic ecosystem model, MELODY[1], was built in order to simulate the functioning of aquatic mesocosms as well as the impact of toxic substances on the dynamics of their populations. This model had a total of 13 compartments (variables) and 219 uncertain parameters. For the purpose of a comparative study between two sensitivity analysis methods (Morris and EFAST) we used only two compartments, the Periphyton and the Grazers, that form a prey-predator chain.

The Periphyton-Grazers sub-model is representative of processes involved in dynamics of primary producers and primary consumers, i.e. photosynthesis, excretion, respiration, egestion, mortality, sloughing and predation. Photosynthesis was assumed to be limited by temperature, light and nutrient levels. Respiration and mortality were exponentially increased when water temperature exceeded the maximum acceptable temperature of organisms. The predation was influenced by temperature and by the biomass of both predators and preys. Egestion and excretion were modeled by a constant coeffcient. In our application, initial values for periphyton and grazers biomasses were deduced from mesocosm experimental data. The forcing variable of the model was temperature (measured once a week during mesocosm experiments).

For the DGSM method we will study one model output (the periphyton biomass) at only one reference time, day 60 of simulations, which corresponds to the period of maximum periphyton biomass and a growth phase for grazers, according to experimental data.

This Periphyton-Grazers sub-model contains a total number of $p = 20$ input parameters. As indicated in the introduction, most of them are highly uncertain (due to natural variability and insufficient data). In order to conduct sensitivity analysis, it was previously necessary to define a range of extreme potential values for each parameter. Based on an extensive literature review, we obtained numerous values for the model parameters. In absence of any expert opinion or relevant information related to the probability density function (PDFs) form, the less informative one was selected, i.e., each parameter was described by a uniform PDF. Although open to criticism, the choice of a uniform distribution was motivated by our intention to perform a screening sensitivity analysis, i.e. to identify non-influential parameters.

4.3. Application of the methodology - Results

In this case, each input parameter X_i follows a uniform distribution on $[a_i, b_i]$, then $C(\mu_i) = \dfrac{(b_i - a_i)^2}{\pi^2}$. A space-filling design (Latin Hypercube Sample with low-centered L^2 discrepancy [8]) of 100 points has been created. On each point of this design, the output Y of the numerical model is computed. All first order Sobol' indices S_i are estimated using local polynomial technique [7]. Derivatives $\frac{\partial Y}{\partial X_i}$ with respect to each input X_i is also computed using first-order finite-difference scheme and a perturbation

[1]modelling MESocosm structure and functioning for representing LOtic DYnamic ecosystems

of 1% of the input. Therefore, $n = 100(p+1) = 2100$ model evaluations have been finally performed to estimate the DGSM (ν_i, σ_i^2 and the total sensitivity upper bound Υ_i).

Estimation results of the sensitivity indices are listed in Tables 2. Using Υ_i values, a clear ranking appears for the influent inputs even if these high Υ_i values (larger than one for the first five inputs) are useless bounds. On the other hand, Υ_i values are fully usefull in a screening context. If we consider that input parameters with a $\Upsilon_i < 0.01$ are non-influential, we detect eight such inputs: kp_{peri}, km_{peri}, r_{peri}, $kExcr_{peri}$, $Tmax_{gr}$, km_{gr}, $kExcr_{gr}$, r_{gr}. These parameters are mainly involved in the respiration, excretion an mortality processes of both outputs. This implies that these processes had little impact on the biomass dynamics of the outputs. This result is reassuring as we suspected that the driving processes of the system were the growth rate of periphyton and the grazing rate of grazers.

Variables	Description	S_i	ν_i	σ_i^2	Υ_i
$Pmax_{peri}$	Maximum photosynthesis rate	0.141	4.175e-02	3.994e-02	24.363
$Cmax_{gr}$	Maximum consumption rate	0.070	4.048e-02	3.908e-02	6.208
Im_{peri}	Ligh saturation level	0.057	2.015e-06	1.932e-06	4.486
$Topt_{peri}$	Optimum temperature	0.018	7.651e-05	7.320e-05	1.662
kn_{peri}	Half-saturation for nitrogen	0.016	7.926e-02	7.610e-02	1.032
$Topt_{gr}$	Optimum temperature	0.015	1.669e-05	1.632e-05	0.518
$Q10_{gr}$	Rate of change per 10C	0.0131	5.168e-04	5.016e-04	0.459
h_{peri_gr}	Handling of periphyton by grazers	0.024	2.660e-03	2.609e-03	0.378
$Q10_{peri}$	Rate of change per 10C	0.013	3.852e-04	3.687e-04	0.216
$Tmax_{peri}$	Maximum survival temperature	0.017	7.927e-07	7.626e-07	0.056
w_{peri_gr}	Grazers preference for periphyton	0.017	1.972e-03	1.935e-03	0.067
a_{peri_gr}	Assimilation rate of periphyton by grazers	0.003	1.438e-04	1.424e-04	0.011
$Tmax_{gr}$	Maximum survival temperature	0.013	9.902e-08	9.523e-08	0.007
km_{peri}	Intrinsic mortality rate	0.002	4.951e-01	4.701e-01	0.001
$kExcr_{peri}$	Excretion rate	0.021	4.953e-01	4.702e-01	0.001
km_{gr}	Intrinsic mortality rate	0.001	4.270e-02	4.167e-02	0.001
$kExcr_{gr}$	Excretion rate	0.012	4.270e-02	4.168e-02	9e-04
r_{peri}	Basal respiration rate	0.006	7.532e-02	7.197e-02	2e-04
r_{gr}	Basal respiration rate	0.005	1.254e-03	1.217e-03	2e-05
kp_{peri}	Half-saturation for phosphorus	0.0257	5.660e-12	5.657e-12	2e-12

Table 2: Sensitivity indices for the output of the reduced aquatic ecosystem model, ranking by decreasing Υ_i. Subscript $_{peri}$ denotes a periphyton variable and subscript $_{gr}$ denotes a grazer variable.

With the methodology of section 4.1, we are in the case 1(b) (non negligible S_i and $\sigma_i^2 \sim \nu_i$) for inputs $Pmax_{peri}$, $Cmax_{gr}$ and Im_{peri}. For the 17 other inputs, we are in the case 2(b). In each case, Υ_i is useful to offer a classification of the inputs. It is clear that all potentially influent inputs (those with $\Upsilon_i > 0.01$) have strong interactions with other inputs.

Considering the first order Sobol' indices $S_i^{'j}$ of the model derivatives $\frac{\partial Y}{\partial X_j}$, we see (not shown here) that all $S_i^{'j}$ ($i = 1, \ldots, p$ and $j = 1, \ldots, p$) are negligible (< 0.05), except for $X_i = Pmax_{peri}$. All the inputs interact with this input, especially $Tmax_{peri}$ (19%), $Topt_{peri}$ (17%), kn_{peri} (16%), km_{peri}

(15%) and r_{peri} (15%). We suppose that some influent higher order interactions (degree three or more) are present in this model for the other inputs.

The previous sensitivity analysis study that was carried out on the Periphyton-Grazers chain using Morris [9] and EFAST [12] allowed the authors to consistently identify eight non-influential parameters for both outputs. These parameters were exactly the same than those obtained by the DGSM method. While taking into account the interactions between the inputs, the Morris and the EFAST methods did not compute sensitivity indices associated to the interactions between the inputs, so there is no comparison possible for the interaction results of the DGSM method. Additional tests could be carried out using the improved version of the Morris method that computes the second-order interactions [13]. The advantage of the DGSM method was the fact that it only required 2,100 evaluations of the model while Morris used 8,400 (because it required 40 replications for a sufficient precision) and EFAST 51,880 model evaluations.

5. CONCLUSION

The DGSM-based maximal bound for the total Sobol' indices has been explained for quite general input distributions Numerical tests have shown that we can discriminate non-influential parameters using DGSM. Using jointly DGSM and first order Sobol' indices, a new methodology of global sensitivity analysis has been proposed and succeedly tested. Indeed, these two measures are obtained at a low computational cost and do not suffer from a large number of inputs in the numerical model. We have also introduced a new idea in order to identify second-order interactions, using the first order Sobol' indices of the model output derivatives. Using this methodology, an application on an aquatic prey-predator chain allows to discriminate non-influential parameters and to detect some second-order interactions. Promising applications of these tools are the computer codes involving a large number of inputs (hundreds) and proposing the ajoint code to compute output derivatives. Indeed, in this case, the derivative calculations are independent of the number of input parameters.

ACKNOWLEDGEMENTS

Part of this work has been backed by French National Research Agency (ANR) through COSINUS program (project COSTA BRAVA noANR-09-COSI-015).

REFERENCES

[1] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis: the Primer*. Wiley, New York, 2008.

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

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

[4] M. Lamboni, B. Iooss, A-L. Popelin, and F. Gamboa. Derivative-based global sensitivity measures: general links with sobol' indices and numerical tests. 2012, submitted. Available at URL: http://fr.arxiv.org/abs/1202.0943.

[5] I.M. Sobol. Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiments*, 1:407–414, 1993.

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

[7] S. Da Veiga, F. Wahl, and F. Gamboa. Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics*, 51(4):452–463, 2009.

[8] K-T. Fang, R. Li, and A. Sudjianto. *Design and modeling for computer experiments*. Chapman & Hall/CRC, 2006.

[9] M.D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33:161–174, 1991.

[10] I.M. Sobol and 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):1212 – 1217, 2010.

[11] P. Fougères. Spectral gap for log-concave probability measures on the real line. In *Séminaire de Probabilités XXXVIII*, volume 1857 of *Lecture Notes in Math.*, pages 95–123. Springer, 2005.

[12] C. Ciric, P Ciffroy, and S. Charles. Use of sensitivity analysis to discriminate non-influential and influential parameters within an aquatic ecosystem model. submitted.

[13] F. Campolongo and R. Braddock. The use of graph theory in the sensitivity analysis of the model output: a second order screening method. *Reliability Engineering and System Safety*, 64:1–12, 1999.