



Regularizing nested Monte Carlo Sobol' index estimators to balance the trade-off between explorations and repetitions in global sensitivity analysis of stochastic models

Henri Mermoz Kouye, Gildas Mazo

► To cite this version:

Henri Mermoz Kouye, Gildas Mazo. Regularizing nested Monte Carlo Sobol' index estimators to balance the trade-off between explorations and repetitions in global sensitivity analysis of stochastic models. 2023. hal-03814478v2

HAL Id: hal-03814478

<https://hal.science/hal-03814478v2>

Preprint submitted on 29 Sep 2023 (v2), last revised 14 Mar 2024 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Regularizing nested Monte Carlo Sobol' index estimators to balance the trade-off between explorations and repetitions in global sensitivity analysis of stochastic models

Henri Mermoz KOUYE⁽¹⁾ and Gildas MAZO⁽¹⁾

⁽¹⁾Univ. Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France

September 29, 2023

Abstract

Sobol' sensitivity index estimators for stochastic models are functions of nested Monte Carlo estimators, which are estimators built from two nested Monte Carlo loops. The outer loop explores the input space and, for each of the explorations, the inner loop repeats model runs to estimate conditional expectations. Although the optimal allocation between explorations and repetitions of one's computational budget is well-known for nested Monte Carlo estimators, it is less clear how to deal with functions of nested Monte Carlo estimators, especially when those functions have unbounded Hessian matrices, as it is the case for Sobol' index estimators. To address this problem, a regularization method is introduced to bound the mean squared error of functions of nested Monte Carlo estimators. Based on a heuristic, an allocation strategy that seeks to minimize a bias-variance trade-off is proposed. The method is applied to Sobol' index estimators for stochastic models. A practical algorithm that adapts to the level of intrinsic randomness in the models is given and illustrated on numerical experiments.

Keywords: nested Monte Carlo, Sobol' index, allocation, trade-off, regularization, global sensitivity analysis, stochastic model

1 Introduction

(Global) sensitivity analyses of stochastic models may be challenging. Indeed, stochastic models include two sources of uncertainty: the parameters' uncertainty and the intrinsic randomness of the stochastic model. The latter can be seen as a hidden additional random input, which may challenge the definition of meaningful sensitivity indices and their efficient estimation.

Many methods have been introduced to deal with stochastic models. A first approach proposed by Hart et al. (2017) considers random Sobol'-Hoeffding decompositions of stochastic model outputs and defines sensitivity indices of such models as expectations of the resulting random Sobol' indices. A second approach, widely used in applications (Courcoul et al., 2011), focuses on deterministic quantities of interest (QoIs) such as conditional expectations or conditional variances. By conditioning with

respect to the uncertain parameters, the aim is to smooth the intrinsic randomness out and hence to deal with deterministic functions of the uncertain parameters only, so that sensitivity analysis (SA) methods for deterministic models can be applied. A third approach includes recently developed methods (Fort et al., 2021; Gamboa et al., 2021; da Veiga, 2021) that see stochastic models as deterministic functions with values in probability distribution spaces. Various sensitivity indices are defined on such spaces in order to measure the contributions of the uncertain parameters. A fourth approach rests on meta-models (Castellan et al., 2020; Étoré et al., 2020; Jimenez et al., 2017; Fort et al., 2013; Zhu and Sudret, 2021; Janon et al., 2014; Panin, 2021). As a remark, let us note that that estimation of expectations of functions of conditional moments can be performed with specific methods, such as semi-parametric (da Veiga and Gamboa, 2013; da Veiga et al., 2017) or multilevel Monte Carlo (Mycek and De Lozzo, 2019) methods. Estimation in high-dimensional deterministic models is considered in de Castro and Janon (2015).

In the first three approaches described above, estimation relies on two nested Monte Carlo loops. Indeed, not only should the model be evaluated at many points in the input space (it is said that the input space is explored), but also the model should be run several times (it is said that the model is “repeated”) at each of those explorations to estimate conditional expectations. In the first approach, these repetitions are performed when approximating the expectations of the random Sobol’ indices. In the second and third approaches, model outputs are repeated when estimating the QoIs and the probability distributions, respectively. The larger the number of explorations and the number of repetitions, the better the sensitivity index estimates. However, in practice, models may be complex and computationally expensive. Therefore, getting a large number of runs may be impossible, and, therefore, finding a compromise between explorations and repetitions under the constraint of a computational cost or precision of estimates is important. For instance, Mazo (2021) proposed a choice of the number of explorations and the number of repetitions based on the minimization of some bound of the so-called mean ranking error of the estimators. This error measures the gap between the ranks of the theoretical indices and those of the estimators. However, a small mean ranking error does not necessarily imply that estimations are close to their theoretical values.

A nested Monte Carlo estimator is an estimator built from two nested Monte Carlo loops for estimating an expectation of a function of a conditional expectation. The compromise between the sizes of the outer and inner loops, coinciding, respectively, with the number of explorations and repetitions in SA, is a bias-variance trade-off. This question is sometimes referred to as the question of optimal allocation. Typically, the mean-squared error (MSE) is approximated and minimized to find the “best” asymptotic trade-off (Rainforth et al., 2018; Giles and Haji-Ali, 2019; Giorgi et al., 2020; Hong and Juneja, 2009). More precisely, if T denotes the number of model runs available to the practitioner, and $n = T^{1-\eta}$ and $m = T^\eta$, $\eta \in [0, 1)$, denote the sizes of the outer and inner loops, respectively, then it is desired to know which value of η would lead to the optimal convergence rate of the MSE.

Sobol’ index estimators are *functions* of nested Monte Carlo estimators. It is, therefore, natural to ask whether the MSE of functions of nested Monte Carlo estimators can be bounded, too. The answer will depend on the regularity of the function in question. For instance, if the function is twice continuously differentiable with a bounded Hessian matrix, then standard Taylor expansion techniques can be used. However, if the Hessian is unbounded, as it is the case of Sobol’ index estimators, then the question is more challenging and remains open, to the best of our knowledge.

We note that the question of optimal allocation of available computational resources is shared more globally by research fields interested in stochastic simulators (Chen and Zhou, 2014, 2017; Binois et al., 2018, 2019).

In this paper, we introduce a regularized version of the nested Monte Carlo Sobol' index estimator and bound its MSE. Based on a heuristic, the bound is simplified and a minimization is carried out to balance the numbers of explorations and repetitions. A practical algorithm that adapts to the model intrinsic randomness is proposed and illustrated on numerical experiments.

This paper is organized as follows. Sobol' indices for stochastic models and their estimation based on nested Monte Carlo methods are presented in Section 2. The regularization method, as well as the challenges of bounding the MSE of functions of nested Monte Carlo estimators, are introduced in Section 3. Section 4 is an application to Sobol' index estimators. Practical algorithms are presented and illustrated in Section 5. A conclusion closes the paper.

2 Sensitivity analysis of stochastic models

A stochastic model with inputs $\mathbf{X} = (X_1, \dots, X_p) \in \mathbb{R}^p$ and output $Y \in \mathbb{R}$ is modeled as a function f of \mathbf{X} and some collection of random variables, denoted by Z , independent of \mathbf{X} such that

$$Y = f(\mathbf{X}, Z). \quad (1)$$

The stochasticity of the model originates from Z since the output of the model evaluated at an input $\mathbf{X} = x$ is a random variable $f(x, Z)$. The distribution of Z is generally unknown.

In the context of SA, a way to deal with stochastic models consists in carrying out SA for deterministic models given by deterministic QoIs. This allows to switch from a stochastic model to some deterministic models for which many SA methods are studied in the literature.

We consider QoIs of the form

$$Q(\mathbf{X}) = \mathbb{E}[\varphi(\mathbf{X}, Z) \mid \mathbf{X}], \quad (2)$$

where $\varphi(\mathbf{X}, Z)$ is a function of \mathbf{X} and Z . For instance, if $\varphi = f$ then $Q(\mathbf{X})$ is the conditional expectation of the model and if $\varphi(\mathbf{X}, Z) = (f(\mathbf{X}, Z) - \mathbb{E}[f(\mathbf{X}, Z) \mid \mathbf{X}])^2$ then $Q(\mathbf{X})$ is the conditional variance, two common choices in practice.

If u is a subset of $\{1, \dots, p\}$, denote by \mathbf{X}_u the group of inputs $\{X_i, i \in u\}$ and $\mathbf{X}_{\sim u}$ the group of inputs $\{X_i, i \notin u\}$. The Sobol' and total indices of the input vector \mathbf{X}_u with respect to the function Q are defined as

$$S_u = \frac{\text{Var}(\mathbb{E}[Q(\mathbf{X}) \mid \mathbf{X}_u])}{\text{Var}(Q(\mathbf{X}))} \quad (3)$$

$$T_u = 1 - \frac{\text{Var}(\mathbb{E}[Q(\mathbf{X}) \mid \mathbf{X}_{\sim u}])}{\text{Var}(Q(\mathbf{X}))} = 1 - S_{\sim u}. \quad (4)$$

The Sobol' index S_u (and hence T_u) can be expressed in terms of a function g linking the components of some parameter vector. Let $\tilde{\mathbf{X}}$ be an independent copy of \mathbf{X} , independent of Z . Denote by $\tilde{\mathbf{X}}_{\sim u}$ the subvector of $\tilde{\mathbf{X}}$ whose components are those of $\tilde{\mathbf{X}}$ not indexed by u . (For instance, if $p = 4$ and $u = \{1, 4\}$ then $\tilde{\mathbf{X}}_{\sim u} =$

$(\tilde{X}_2, \tilde{X}_3)$.) If $\theta = (\theta_1, \theta_2, \theta_3)$ with $\theta_1 = \mathbb{E}(Q(\mathbf{X})^2)$, $\theta_2 = \mathbb{E}(Q(\mathbf{X}))$ and $\theta_3 = \mathbb{E}(\mathbb{E}[Q(\mathbf{X}) | \mathbf{X}_u]^2) = \mathbb{E}(Q(\mathbf{X})Q(\tilde{\mathbf{X}}_{\sim u}, \mathbf{X}_u)) = \mathbb{E}(Q(\mathbf{X}_{\sim u}, \mathbf{X}_u)Q(\tilde{\mathbf{X}}_{\sim u}, \mathbf{X}_u))$ then

$$S_u = g(\theta) = \frac{\theta_3 - \theta_2^2}{\theta_1 - \theta_2^2}.$$

The construction of an estimator of S_u boils down to the construction of estimators of the three quantities θ_1, θ_2 and θ_3 . To do this, nested Monte Carlo sampling is widely-used. Let $\{\mathbf{X}^{(i)}; i = 1, \dots, n\}$ and $\{\tilde{\mathbf{X}}^{(i)}; i = 1, \dots, n\}$ be independent Monte Carlo samples from the law of \mathbf{X} . For each i , denote by $\mathbf{X}_u^{(i)}$ the subvector of $\mathbf{X}^{(i)}$ whose components are those of $\mathbf{X}^{(i)}$ indexed by u . Likewise, denote by $\mathbf{X}_{\sim u}^{(i)}$ the subvector of $\mathbf{X}^{(i)}$ whose components are those of $\mathbf{X}^{(i)}$ not indexed by u , and denote by $\tilde{\mathbf{X}}_{\sim u}^{(i)}$ the subvector of $\tilde{\mathbf{X}}^{(i)}$ whose components are those of $\tilde{\mathbf{X}}^{(i)}$ not indexed by u . An estimator of S_u is given by

$$\hat{S}_u = g(\hat{\theta}) = \frac{\hat{\theta}_3 - \hat{\theta}_2^2}{\hat{\theta}_1 - \hat{\theta}_2^2} \quad (5)$$

where

$$\left. \begin{aligned} \hat{\theta}_1 &= \frac{1}{n} \sum_{i=1}^n \hat{Q}_m(\mathbf{X}^{(i)})^2 \\ \hat{\theta}_2 &= \frac{1}{n} \sum_{i=1}^n \hat{Q}_m(\mathbf{X}^{(i)}) \\ \hat{\theta}_3 &= \frac{1}{n} \sum_{i=1}^n \hat{Q}_m(\mathbf{X}^{(i)}) \tilde{Q}_m(\tilde{\mathbf{X}}_{\sim u}^{(i)}, \mathbf{X}_u^{(i)}) \end{aligned} \right\} \quad (6)$$

and

$$\begin{aligned} \hat{Q}_m(\mathbf{X}^{(i)}) &= \frac{1}{m} \sum_{k=1}^m \varphi(\mathbf{X}^{(i)}, Z^{(i,k)}) \\ \tilde{Q}_m(\tilde{\mathbf{X}}_{\sim u}^{(i)}, \mathbf{X}_u^{(i)}) &= \frac{1}{m} \sum_{k=1}^m \varphi(\tilde{\mathbf{X}}_{\sim u}^{(i)}, \mathbf{X}_u^{(i)}, \tilde{Z}^{(i,k)}); \end{aligned}$$

here the objects $\{Z^{(i,k)}, \tilde{Z}^{(i,k)}; k = 1, \dots, m; i = 1, \dots, n\}$, are independent and identically distributed random variables, independent of $\{\mathbf{X}^{(i)}, \tilde{\mathbf{X}}^{(i)}; i = 1, \dots, n\}$, representing the randomness of the user's model. If m is fixed and $n \rightarrow \infty$, then

$$\sqrt{n} \left(\hat{S}_u - S_u \left[1 - \frac{\mathbb{E} \text{Var}(\varphi(\mathbf{X}, Z) | \mathbf{X})}{\mathbb{E} \text{Var}(\varphi(\mathbf{X}, Z) | \mathbf{X}) + m \text{Var} \mathbb{E}(\varphi(\mathbf{X}, Z) | \mathbf{X})} \right] \right)$$

converges to a centered normal distribution with some variance σ_m^2 depending on m . If, moreover, $m \rightarrow \infty$ such that $\sqrt{n}/m \rightarrow 0$ then $\sqrt{n}(\hat{S}_u - S_u)$ goes to a centered normal distribution with variance $\lim_{m \rightarrow \infty} \sigma_m^2$. For more details, see Mazo (2021). A theoretically-guided choice for m and n that penalizes bad rankings of the sensitivity index estimates $\hat{S}_1, \dots, \hat{S}_p$ was given in Mazo (2021).

3 Nested Monte Carlo estimation

We say that an estimator $\hat{\theta}$ is a nested Monte-Carlo estimator if

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \phi_1 \left(\frac{1}{m} \sum_{k=1}^m \varphi_1(\xi^{(i)}, \zeta^{(i,k)}) \right) \cdots \phi_p \left(\frac{1}{m} \sum_{k=1}^m \varphi_p(\xi^{(i)}, \zeta^{(i,k)}) \right), \quad (7)$$

where $\phi_1, \phi_2, \dots, \phi_p$ are measurable functions and the sets of random vectors $\{\xi^{(i)}, \zeta^{(i,k)}, k = 1, \dots, m\}, i = 1, \dots, n$, are independent and identically distributed. The estimator (7) is called a nested Monte Carlo estimator because it is built from two nested Monte Carlo loops: an outer loop is used to simulate the random vectors $\xi^{(1)}, \dots, \xi^{(n)}$, and, for each $\xi^{(i)}$, an inner loop is used to simulate $\zeta^{(i,k)}, k = 1, \dots, m$. The random vectors $\{\zeta^{(i,k)}\}$ and $\{\xi^{(i)}\}$ are assumed to be independent. With this definition, each of the three estimators $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ in (6) is a nested Monte Carlo estimator: the outer loop simulates the “explorations” $\mathbf{X}^{(i)}, \tilde{\mathbf{X}}^{(i)}, i = 1, \dots, n$, and the inner loop simulates the “repetitions” $\varphi(\mathbf{X}^{(i)}, Z^{(i,k)}), \varphi(\tilde{\mathbf{X}}_u^{(i)}, \mathbf{X}_u^{(i)}, \tilde{Z}^{(i,k)}), k = 1, \dots, m$.

If, for each $i = 1, \dots, n$ and conditionally on $\xi^{(i)}$, the random variables $\zeta^{(i,k)}$ are independent and identically distributed, and if

$$\theta := \mathbb{E} \left(\phi_1 \left(\mathbb{E} \left[\varphi_1(\xi^{(1)}, \zeta^{(1,1)}) \mid \xi^{(1)} \right] \right) \dots \phi_p \left(\mathbb{E} \left[\varphi_p(\xi^{(1)}, \zeta^{(1,1)}) \mid \xi^{(1)} \right] \right) \right), \quad (8)$$

then we have

$$\mathbb{E} \|\hat{\theta} - \theta\|^2 = \frac{1}{n} \text{Trace}(\Sigma_m) + \|b_m\|^2, \quad (9)$$

where Σ_m is the variance-covariance matrix of

$$\phi_1 \left(\frac{1}{m} \sum_{k=1}^m \varphi_1(\xi^{(1)}, \zeta^{(1,k)}) \right) \dots \phi_p \left(\frac{1}{m} \sum_{k=1}^m \varphi_p(\xi^{(1)}, \zeta^{(1,k)}) \right).$$

The MSE in Equation (9) is a sum of two terms: a variance term which involves the ratio $1/n$ and a bias term that is function of m only.

Assumption 1. $b_m \rightarrow 0$ and $\Sigma_m \rightarrow \Sigma$ as $m \rightarrow +\infty$

Assumption 1 is fulfilled by Nested Monte Carlo estimators provided that the functions ϕ_1, \dots, ϕ_p in (7) have good properties such as boundedness, Lipschitz continuity or boundedness of derivatives (Giorgi et al., 2017, 2020). In the context of global sensitivity analysis for stochastic model, Mazo (2021) showed that Assumption 1 is satisfied by $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ (see Equation (6)) associated with Sobol’ index estimators.

If Assumption 1 is satisfied, it holds that $\lim_{n,m \rightarrow +\infty} \mathbb{E} \|\hat{\theta} - \theta\|^2 = 0$ and thereby $\hat{\theta}$ converges in quadratic mean to θ . If, moreover, the rate at which b_m goes to zero is known, then optimal rates for $n = T^{1-\eta}$ and $m = T^\eta, \eta \in [0, 1)$, can be calculated in terms of the total number of runs T . For instance, in the case $p = 1$ and ϕ_1 is a smooth function with uniformly bounded third derivative, Hong and Juneja (2009) showed that $b_m = O(1/m)$ and thus $\mathbb{E} \|\hat{\theta} - \theta\|^2 = O(1/n + 1/m^2)$, leading to $O(T^{2/3})$ reached for $n = O(T^{2/3})$ and $m = O(T^{1/3})$. If ϕ_1 is simply Lipschitz continuous, Rainforth et al. (2018) rather found that $\mathbb{E} \|\hat{\theta} - \theta\|^2 = O(1/n + 1/m)$, leading to a MSE of order $O(T^{1/2})$, ensured by the choice of $n = O(T^{1/2})$ and $m = O(T^{1/2})$.

3.1 Functions of nested Monte Carlo estimators

Let $g : \mathcal{D} \rightarrow \mathbb{R}$ be a continuous, non-constant function and let $\hat{\theta}$ be as in (6). Note that the estimator $g(\hat{\theta})$ converges in probability to $g(\theta)$ but it does not necessarily hold that its MSE $\mathbb{E} (g(\hat{\theta}) - g(\theta))^2$ converges to 0. In fact, the MSE may even be infinite, depending on the properties of g and the law of $\hat{\theta}$. Since the law of $\hat{\theta}$ is unknown, it is important to find verifiable conditions on g under which the MSE is kept under control

for all possible laws of $\hat{\theta}$. For instance, if g is Lipschitz continuous, then there exists an constant L such that

$$\mathbb{E} \left(g(\hat{\theta}) - g(\theta) \right)^2 \leq L^2 \mathbb{E} \|\hat{\theta} - \theta\|^2.$$

This bound could be minimized as in Hong and Juneja (2009), see above. If g is assumed to be a twice continuously differentiable such that its Hessian matrix denoted $\nabla^2 g$ is uniformly bounded, then, up to existence of some moments of $\hat{\theta}$, the combination of a Taylor-Lagrange expansion and convexity inequalities would yield

$$\begin{aligned} \mathbb{E} \left(g(\hat{\theta}) - g(\theta) \right)^2 &\leq 4\mathbb{E} \left(\nabla g(\mu_m)^\top (\hat{\theta} - \mu_m) \right)^2 + 2(g(\mu_m) - g(\theta))^2 \\ &\quad + L'^2 \left(\mathbb{E} \|\hat{\theta} - \mu_m\|^4 \right), \end{aligned} \quad (10)$$

with $L' > 0$ such that $\sup_{x \in \mathcal{D}} \|\nabla^2 g(x)\|_F \leq L'$ and $\|\cdot\|_F$ denotes the Frobenius norm. From this upper bound, we could derive another bound by noticing that $\mathbb{E} \left(\|\hat{\theta} - \mu_m\|^4 \right) = O(1/n^2)$. This follows from the Marcinkiewicz–Zygmund inequality (Marcinkiewicz and Zygmund, 1937), provided that suitable moments exist. It would then hold that:

$$\mathbb{E} \left(g(\hat{\theta}) - g(\theta) \right)^2 \leq 4\mathbb{E} \left(\nabla g(\mu_m)^\top (\hat{\theta} - \mu_m) \right)^2 + 2(g(\mu_m) - g(\theta))^2 + O(1/n^2). \quad (11)$$

Equation (11) highlights three terms in the right-hand side: the first term stands for a variance term, the second term represents a squared bias term and the third term is a negligible approximation error.

However, in practice, in many interesting applications, the function g does not have good enough properties such as above. For instance, uniform boundedness of derivatives of g is a very strong condition in general. In order to weaken such a condition, consider the following one:

$$\mathbb{E} \left(\sup_{\lambda \in [0,1]} \|\nabla^2 g(\lambda \hat{\theta} + (1-\lambda)\mu_m)\|_F^4 \right) = O(1) \quad \text{as } n, m \rightarrow \infty. \quad (12)$$

Under the condition (12), the decomposition (11) holds but asymptotically when both n and m go to infinity. However, whether or not this condition is true depends on the law of $\hat{\theta}$, and hence hard to check in practice. For instance, in the case of Sobol' index estimators defined in Section 2, this comes down to require that $\mathbb{E}[(\hat{\theta}_1 - \hat{\theta}_2)^{-\alpha}]$ with $\alpha > 0$ exists, whereas the probability distribution of $\hat{\theta}_1 - \hat{\theta}_2$ is unknown.

In the next section, we propose a condition, weaker than the one in Equation (12), that does not put constraints on the distribution of $\hat{\theta}$.

3.2 A regularization method to address functions with unbounded Hessian matrix

The idea is to introduce a "slight perturbation" g_h (with $h \in (0, 1)$) of the function g that approaches the true g as $h \rightarrow 0$ and such that (12) is satisfied with g_h in place of g . Introducing g_h can be thought as a way to "transport" the original estimator

$\hat{\theta}$ to regions of \mathcal{D} where control of moments of $g_h(\hat{\theta})$ is possible without additional conditions on the law of $\hat{\theta}$. The goal is to “get away” from certain regions of the parameter space where the Hessian of g may explode. For this, consider a family of functions $\{g_h : \mathcal{D} \rightarrow \mathbb{R}, h \in (0, 1)\}$ such that for all h , g_h is twice continuously differentiable and for all $x \in \mathcal{D}$, $\lim_{h \rightarrow 0} g_h(x) = g(x)$. Moreover, assume g_h satisfies the following assumption:

Assumption 2. *There exists a constant C independent of h such that, for all $h \in (0, 1)$:*

$$\limsup_{n,m \rightarrow \infty} \mathbb{E} \left(\sup_{\lambda \in [0,1]} \|\nabla^2 g_h(\lambda \hat{\theta} + (1-\lambda)\mu_m)\|_F^4 \right) \leq C.$$

The advantage of having such a family of functions is that $\mathbb{E}(g_h(\hat{\theta}) - g_h(\theta))^2$, the “perturbed MSE”, could be bounded with an approximate upper bound in the form of Equation (11) with $g = g_h$. A direct implication of this is that for h fixed, $\lim_{n,m \rightarrow \infty} \mathbb{E} \left(g_h(\hat{\theta}) - g_h(\theta) \right)^2 = 0$. Further, if the regularized estimator $g_h(\hat{\theta})$ is used for estimation of $g(\theta)$ instead of $g(\hat{\theta})$, the corresponding mean squared error is bounded in Proposition 1.

Proposition 1. *Under Assumptions 1 and 2, for every $h \in (0, 1)$:*

$$\mathbb{E} \left(g_h(\hat{\theta}) - g(\theta) \right)^2 \leq 4(1 + p_{n,m}(h)) (V_{n,m}(h) + B_m(h)^2), \quad (13)$$

where $\limsup_{n,m \rightarrow \infty} p_{n,m}(h) = 0$ and $V_{n,m}(h) := \mathbb{E}(\nabla g_h(\mu_m)^\top (\hat{\theta} - \mu_m))^2$ and $B_m(h) := g_h(\mu_m) - g(\theta)$.

Remark that $B_m(h) = (g_h(\mu_m) - g_h(\theta)) + (g_h(\theta) - g(\theta))$ accounts for two types of bias. The first term in the sum stands for the bias of the estimator $\hat{\theta}$; it vanishes as soon as $\mu_m \rightarrow \theta$. The second term is a regularization error that stems from the use of g_h instead of g . This term does not vanish asymptotically in general. However, it goes to zero as $h \rightarrow 0$, yielding Corollary 1.

Corollary 1. *We have $\lim_{h \rightarrow 0} \lim_{n,m \rightarrow \infty} \mathbb{E} \left(g_h(\hat{\theta}) - g(\theta) \right)^2 = 0$.*

Note that the limits in Corollary 1 cannot be interchanged. Additional conditions on the gap $\mathbb{E}(g(\hat{\theta}) - g_h(\hat{\theta}))^2$ would be needed to get the convergence of the MSE of $g(\hat{\theta})$.

4 Application to Sobol’ index estimators

Let $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ be as in (6). Recall that $\theta_1 = \mathbb{E}(Q(\mathbf{X})^2)$, $\theta_2 = \mathbb{E}(Q(\mathbf{X}))$, $\theta_3 = \mathbb{E}(\mathbb{E}[Q(\mathbf{X}) | \mathbf{X}_u]^2) = \mathbb{E}(Q(\mathbf{X})Q(\tilde{\mathbf{X}}_{\sim u}, \mathbf{X}_u))$, so that $\mu_m = (\mu_{m1}, \mu_{m2}, \mu_{m3}) = (\theta_1 + b_{m1}, \theta_2 + b_{m2}, \theta_3 + b_{m3})$ and $b_m = (b_{m1}, b_{m2}, b_{m3}) = (\mathbb{E}\text{Var}[\varphi(\mathbf{X}, Z) | \mathbf{X}] / m, 0, 0)$. Recall that $\hat{S}_u = g(\hat{\theta})$, where the function

$$g : (x_1, x_2, x_3) \mapsto (x_3 - x_2^2) / (x_1 - x_2^2) \quad (14)$$

is a twice-continuously differentiable function over its definition domain. But unfortunately, g is unbounded. The form of such a function makes the study of the MSE

of \hat{S}_u difficult unless strong conditions are imposed on the output distribution of the stochastic model. This could explain why, in spite of our efforts, it is difficult to find in the literature material on the convergence in quadratic mean of such estimators. The approach introduced in Section 3 allows to bypass the unboundedness issue. We shall apply the results of Section 3 to bound the MSE of the regularized version of the Sobol' index estimator and propose an allocation method.

Throughout this section, it is assumed that $\mathbb{E}(Q(\mathbf{X})^{16}) < +\infty$. We consider the family of function $g_h(x) := g(x + h\mathbf{u})$ with $\mathbf{u} = (1, 0, 0)$. In order to provide an upper bound for $\mathbb{E}(g_h(\hat{\theta}) - g(\theta))^2$ as in Proposition 1, it is necessary to fulfill Assumption 1 and 2. Assumption 1 is trivially satisfied.

Theorem 1. *With $\hat{\theta}$, g_h and g as above, Assumption 2 holds. There is a constant \bar{C} independent of h such that, for every $h \in (0, 1)$,*

$$\mathbb{E}(g_h(\hat{\theta}) - g(\theta))^2 \leq \bar{C} \left(\frac{1}{n} + \left(\frac{\mathbb{E}(\text{Var}[\varphi(\mathbf{X}, Z) | \mathbf{X}])}{m} \right)^2 + h^2 \right),$$

as $n, m \rightarrow \infty$.

Corollary 2. *With $\hat{\theta}$, g_h and g as above, we have $\lim_{h \rightarrow 0} \lim_{n, m \rightarrow \infty} \mathbb{E}(g_h(\hat{\theta}) - g(\theta))^2 = 0$.*

As $n, m \rightarrow +\infty$, it occurs that the smaller h is, the smaller the MSE gets. Therefore, in practice, parameter h should be taken in order to approximate the true $g(\theta)$.

To balance the trade-off between the number of repetitions and explorations, we proceed with a heuristic. For h small enough, one could assume that h^2 is negligible with respect to the quantity $\text{BVT} := 1/n + (\mathbb{E}(\text{Var}[\varphi(\mathbf{X}, Z) | \mathbf{X}]) / m)^2$. This is, therefore, the quantity we shall be interested in for optimizing the numbers of repetitions and explorations. Let $T = mn$ be the number of available samples to compute the Sobol' index estimator. From nested Monte Carlo theory (see beginning of Section 3), we already know that m and n must be of order $T^{1/3}$ and $T^{2/3}$ respectively so as to obtain optimal convergence rate.

However, an asymptotic order is not a specific value. We expect that the greater the variance of intrinsic randomness, the larger the value of optimal m should be. Thus, in this section, the goal consists in proposing a value of m that adapts to the intrinsic randomness of the stochastic model.

Under the constraint $nm = T$, the optimal convergence rate of the BVT is obtained when n_{opt} is of order $T^{2/3}$ and m_{opt} is of order $T^{1/3}$. Let $m_{\text{opt}} = \kappa T^{1/3}$ where $\kappa > 0$. Then, $n_{\text{opt}} = \kappa^{-1} T^{2/3}$. Thus:

$$\text{BVT} = \kappa T^{-2/3} + \frac{(\mathbb{E}\text{Var}[\varphi(\mathbf{X}, Z) | \mathbf{X}])^2}{\kappa^2} T^{-2/3}.$$

Coefficient κ can be chosen such that $\kappa T^{-2/3} + \frac{(\mathbb{E}\text{Var}[\varphi(\mathbf{X}, Z) | \mathbf{X}])^2}{\kappa^2} T^{-2/3}$ is the smallest over $\kappa > 0$. The minimum of such a quantity is reached at $\kappa_{\text{opt}} = (2 (\mathbb{E}\text{Var}[\varphi(\mathbf{X}, Z) | \mathbf{X}])^2)^{1/3}$. Therefore:

$$m_{\text{opt}}(T) = (2 (\mathbb{E}\text{Var}[\varphi(\mathbf{X}, Z) | \mathbf{X}])^2)^{1/3} T^{1/3}. \quad (15)$$

Therefore, the number of repetitions suggested above ensures first that the BVT converges at optimal rate and second adapts to the model's intrinsic randomness because

that $m_{opt}(T)$ depends on $\mathbb{E}\text{Var}[\varphi(\mathbf{X}, Z) \mid \mathbf{X}]$, which quantifies the part of the total variance $\text{Var}(\varphi(\mathbf{X}, Z))$ that is not attributed to the inputs \mathbf{X} ; and so, that measures the influence of the intrinsic noise of the stochastic model $\varphi(\mathbf{X}, Z)$. Finally, it also appears that $m_{opt}(T)$ depends on both T and the function φ . The dependence with respect to T guarantees that $m_{opt}(T)$ grows as T gets large. Besides, the dependence with respect to φ means that even if $m_{opt}(T)$ remains proportional to $T^{1/3}$, it also varies with respect to the chosen QoI of the stochastic model f .

5 An algorithm of estimation of Sobol' indices for stochastic models

5.1 Algorithms

This section is devoted to the practical implementation of the bias-variance trade-off strategy when performing SA for some QoI of a stochastic model. Recall that f is a stochastic model as in (1) and we are interested in carrying out SA of a QoI under the form (2), that is, $Q(\mathbf{X}) = \mathbb{E}[\varphi(\mathbf{X}, Z) \mid \mathbf{X}]$ in order to measure the impact of l groups of inputs $u_i \subset \{1, \dots, p\}$, $i = 1, \dots, l$. In other words, we are interested in estimating S_{u_1}, \dots, S_{u_l} . We shall use at most $T \times (l + 1)$ evaluations of $\varphi(\mathbf{X}, Z)$. Under the constraint $nm = T$, the number of repetitions m_{opt} found in (15) depends on $\rho := \mathbb{E}\text{Var}[\varphi(\mathbf{X}, Z) \mid \mathbf{X}]$. However, in practice, ρ is often unknown. So, before sensitivity index estimation, ρ needs to be estimated. Notice that ρ can be rewritten $\rho = \mathbb{E}(\varphi(\mathbf{X}, Z) - \varphi(\mathbf{X}, \tilde{Z}))^2$, which suggests the following estimation procedure.

Consider r_0 i.i.d. samples of \mathbf{X} , denoted by $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(r_0)}$, and generate two outputs at each sample $\mathbf{X}^{(i)}$: $(\varphi(\mathbf{X}^{(1)}, Z^{(1,1)}), \varphi(\mathbf{X}^{(1)}, Z^{(1,2)}))$, \dots , $(\varphi(\mathbf{X}^{(r_0)}, Z^{(r_0,1)}), \varphi(\mathbf{X}^{(r_0)}, Z^{(r_0,2)}))$. Thus:

$$\hat{\rho} = \frac{1}{r_0} \sum_{i=1}^{r_0} \left(\varphi(\mathbf{X}^{(i)}, Z^{(i,1)}) - \varphi(\mathbf{X}^{(i)}, Z^{(i,2)}) \right)^2$$

is a consistent and unbiased estimator of ρ . It appears that the estimation of ρ requires $2r_0$ evaluations of the model $\varphi(\mathbf{X}, Z)$. However, the maximal number of evaluations is $T \times (l + 1)$. So, for index estimation procedure, at most $T \times (l + 1) - 2r_0$ model evaluations are allowed.

Therefore, our strategy consists in leveraging the model outputs used to estimate ρ and then plugging and completing those outputs in order to compute sensitivity index estimates. This strategy relies on two algorithms: Algorithm 1 and Algorithm 2. Algorithm 1 enables to generate complementary outputs in addition to outputs already available after estimation of ρ . This allows to satisfy the constraint of the maximal number of model evaluations $T \times (l + 1)$. This part helps to optimize the whole estimation procedure by using the model outputs already generated. Regarding Algorithm 2, it effectively estimates indices in three steps based on pick-freeze procedure. First, it estimates ρ and thereby compute m_{opt} and $n_{opt} = T/m_{opt}$. Then, in the second step, by relying on Algorithm 1, complementary outputs required for estimations are generated. In the final step, sensitivity index estimates are computed with respect to inputs or groups of inputs specified by the user.

Algorithm 1 Completing model evaluations

Inputs: $n, m, \varphi, l, (\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(T)})$

Data: $(\varphi(\mathbf{X}^{(1)}, Z^{(1,1)}), \varphi(\mathbf{X}^{(1)}, Z^{(1,2)})), \dots, (\varphi(\mathbf{X}^{(r_0)}, Z^{(r_0,1)}), \varphi(\mathbf{X}^{(r_0)}, Z^{(r_0,2)}))$

```
if  $n \geq r_0$  then
  if  $m \geq 2$  then
    for  $i = 1, \dots, r_0$  do
      for  $k = 3, \dots, m$  do
        Compute  $\varphi(\mathbf{X}^{(i)}, Z^{(i,k)})$ 
      end for
    end for
    for  $i = r_0 + 1, \dots, n$  do
      for  $k = 1, \dots, m$  do
        Compute  $\varphi(\mathbf{X}^{(i)}, Z^{(i,k)})$ 
      end for
    end for
  end if
  if  $m = 1$  then
    for  $i = r_0 + 1, \dots, n - r_0 - \lceil r_0/(l+1) \rceil$  do
      Compute  $\varphi(\mathbf{X}^{(i)}, Z^{(i,1)})$ 
    end for
  end if
end if
if  $n < r_0$  then
  if  $m > 2 + 2\lceil 1/(l+1) * (-1 + r_0/n) \rceil$  then
    for  $i = 1, \dots, n$  do
      for  $k = 3, \dots, m - 2\lceil 1/(l+1) * (-1 + r_0/n) \rceil$  do
        Compute  $\varphi(\mathbf{X}^{(i)}, Z^{(i,k)})$ 
      end for
    end for
  end if
  if  $m \leq 2 + 2\lceil 1/(l+1) * (-1 + r_0/n) \rceil$  then
    Exit: Budget already consumed
  end if
end if
```

Algorithm 2 Estimation of Sobol' indices

Inputs: $h, r_0, T, \varphi, w = \{u_1, \dots, u_l\}, (\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(T)}), (\tilde{\mathbf{X}}^{(1)}, \dots, \tilde{\mathbf{X}}^{(T)})$

- 1: **for** $i = 1, \dots, r_0$ **do**
- 2: **for** $k = 1, 2$ **do**
- 3: Compute $\varphi(\mathbf{X}^{(i)}, Z^{(i,k)})$
- 4: **end for**
- 5: **end for**
- 6: Compute $\hat{\rho} \leftarrow \frac{1}{r_0} \sum_{s=1}^{r_0} (\varphi(\mathbf{X}^{(i)}, Z^{(i,1)}) - \varphi(\mathbf{X}^{(i)}, Z^{(i,2)}))^2$
- 7: Compute \hat{m}_{opt} according to Equation (15)
- 8: $\hat{n}_{opt} \leftarrow \lfloor T / \hat{m}_{opt} \rfloor$
- 9: Run Algorithm 1 with $m = \hat{m}_{opt}, n = \hat{n}_{opt}$ to complete samples $\varphi(\mathbf{X}^{(1)}, Z^{(1,1)}), \dots, \varphi(\mathbf{X}^{(r_0)}, Z^{(r_0,2)})$.
- 10: **for** $j = 1, \dots, l$ **do**
- 11: **for** $i = 1, \dots, \hat{n}_{opt}$ **do**
- 12: **for** $k = 1, \dots, \hat{m}_{opt}$ **do**
- 13: Compute $\varphi((\tilde{\mathbf{X}}_{u_j}^{(i)}, \mathbf{X}_{u_j}^{(i)}), \tilde{Z}^{(i,k)})$
- 14: **end for**
- 15: **end for**
- 16: **end for**
- 17: Compute sensitivity index estimates or their regularized estimates

Algorithm 2 requires: h, r_0, T, φ, w and input samples. The transformation φ of the stochastic model is supplied as well as w the set of inputs or groups of inputs whose indices are estimated. In practice, r_0, T and h must be chosen. We recommend to take r_0 small with respect to T so as not to waste too much of the budget in the first stage of Algorithm 2. Indeed, the estimator $\hat{\rho}$ is unbiased and consistent and often provides good estimates even for small r_0 . Regarding T , it follows T should be taken as large as possible depending on the computational cost of a run of both φ and the original model f . Furthermore, to ensure that the MSE has a precision $\varepsilon \in (0, 1)$ with $h^2 \ll \varepsilon$, T must be roughly chosen larger than $\varepsilon^{-3/2}$ since the MSE is $O(T^{-2/3})$. This provides approximations for practical choice of T .

6 Illustrations

This section presents the performance of the estimators of first-order and total indices computed by Algorithm 2 in the case of two toy stochastic models for which analytical values of indices are known: a linear model $f(X_1, X_2, Z) = 1 + X_1 + 2X_2 + \sigma Z$ with $\sigma > 0$ and a stochastic version of the Ishigami function $f'(X_1, X_2, X_3, Z) = \sin X_1 + a \sin^2 X_2 + b X_3^4 (\sin X_1) Z^2$ with $a, b > 0$ (Ishigami and Homma (1990)). For each value of $T = nm$, the estimators of Algorithm 2 are compared with two other arbitrary choices, namely, $(n, m) = (T/5, 5)$ and $(n, m) = (T^{1/2}, T^{1/2})$.

These two choices above represent two different situations. The choice $(n, m) = (T/5, 5)$ presents a case where the number of repetitions is constant and independent of T . This illustrates the situation where the bias does not get reduced so that it disturbs estimations no matter how large T is. Regarding $(n, m) = (T^{1/2}, T^{1/2})$, it shows that the case where the variance is not sufficiently reduced.

For illustrations, the product $T = mn$ is chosen in the set $T \in \{10^3, 10^4, \dots, 10^7\}$.

For each choice of the couple (n, m) , $N = 100$ replications of estimations are carried out. The tuning parameter r_0 is set to 10. Thus, $2r_0 = 20$ model evaluations are used to get the estimates $\hat{\rho}$, \hat{n}_{opt} , \hat{m}_{opt} in the first part of Algorithm 2, and then $T - 2r_0 \in \{10^3 - 20, 10^4 - 20, \dots, 10^7 - 20\}$ model evaluations are used to get the sensitivity index estimators with $n = \hat{n}_{opt}$ and $m = \hat{m}_{opt}$. For both toy stochastic models, the QoI considered is the conditional expectation so that $\varphi = f$ or $\varphi = f'$ depending on the model. For each value of T , the boxplots of sensitivity index estimates are plotted for each of the three choices.

Linear model

Let $f(X_1, X_2, Z) = 1 + X_1 + 2X_2 + \sigma Z$ where $\sigma > 0$ and X_1, X_2 and Z are i.i.d. under the standard normal distribution. Such model includes two uncertain parameters X_1 and X_2 with respective first-order Sobol' indices $S_1 = 1/5$ and $S_2 = 4/5$. Two values of σ are considered: $\sigma = 1$ and $\sigma = 5$.

Figure 1 shows that the estimations obtained with Algorithm 2 are more efficient as T gets large because both bias and variance are efficiently reduced. Boxplots highlight that the strategies $m = 5$ and $m = T^{1/2}$ suffer respectively from bias and variance. Notice that in the case of the linear model under study, $\mathbb{E}(\text{Var}(f(X_1, X_2, Z) | (X_1, X_2))) = 2\sigma^2$; so the bias depends on σ . This explains why in the case $\sigma = 5$ (see Figure 2), even for large value of T , estimations resulting of the choice $m = 5$ do not converge to their true targets.

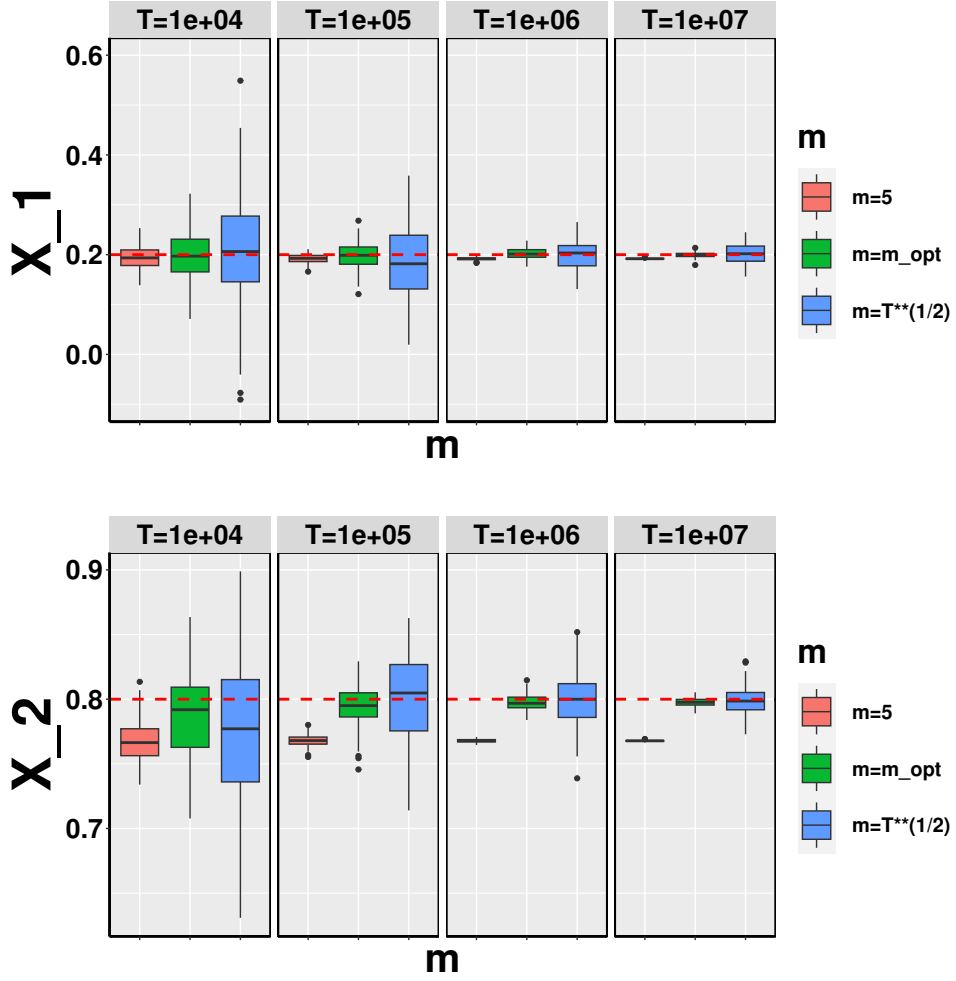


Figure 1: Boxplots of sensitivity index estimates of X_1 and X_2 with respect to the linear model and for different values of T . The standard deviation is $\sigma = 1$ and regularization parameter $h = 10^{-2}$. Three strategies of choice of m are compared: $m = 5$ (in red), $m = m_{opt}$ given by the trade-off strategy of Algorithm 2 (in green) and $m = T^{1/2}$ (in blue). The red dashed lines showed the true sensitivity index values, in particular, in this setting $S_1 = 0.2$ and $S_2 = 0.8$

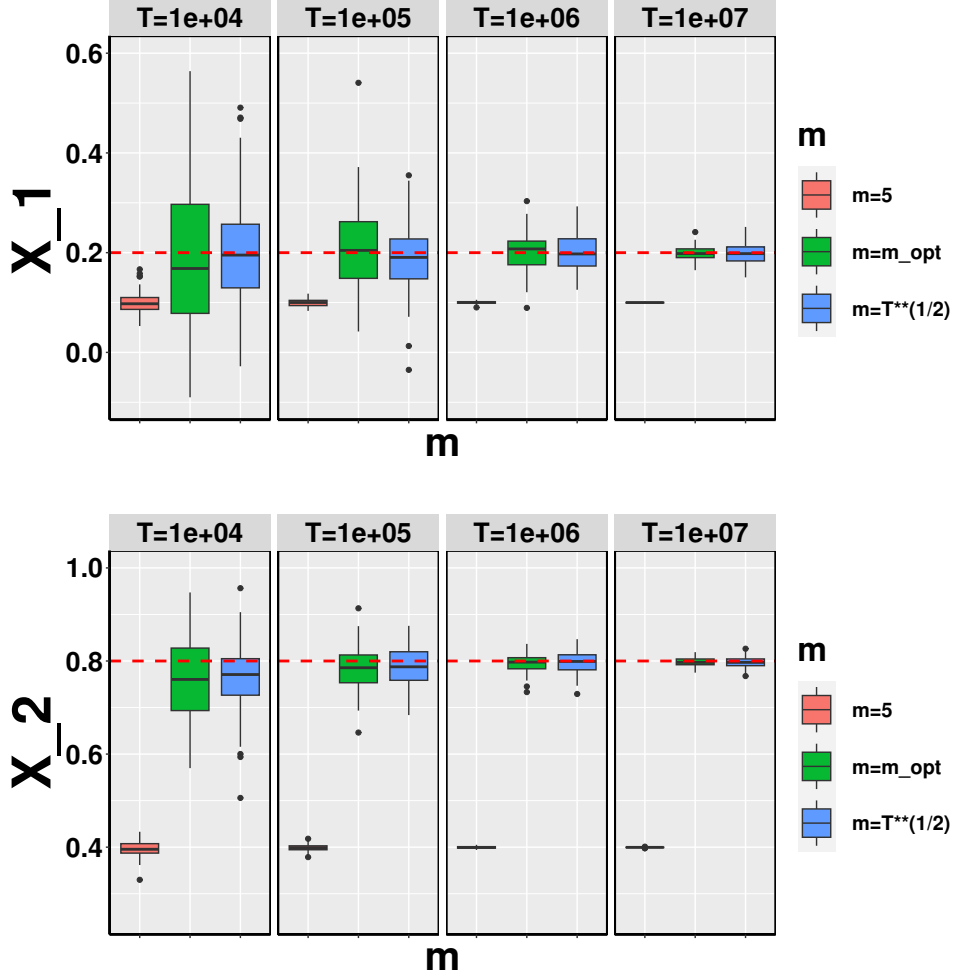


Figure 2: Boxplots of sensitivity index estimates of X_1 and X_2 with respect to the linear model and for different values of T . The standard deviation and regularization parameter are $\sigma = 5$ and $h = 10^{-2}$. Three strategies of choice of m are compared: $m = 5$ (in red), $m = m_{opt}$ given by the trade-off strategy of Algorithm 2 (in green) and $m = T^{1/2}$ (in blue). The red dashed lines showed the true sensitivity index values, in particular, in this setting $S_1 = 0.2$ and $S_2 = 0.8$

A stochastic Ishigami function

Let $f'(X_1, X_2, X_3, Z) = \sin X_1 + a \sin^2 X_2 + b X_3^4 (\sin X_1) Z^2$ such that with $a, b > 0$, X_1, X_2, X_3 and Z are independent with X_1, X_2, X_3 distributed under $\mathcal{U}([-\pi, \pi])$ and $Z \sim \mathcal{N}(0, 1)$. The model f' is a modified version of benchmark function known as the Ishigami function in SA. For this model, first-order Sobol' indices of inputs X_1, X_2 , and X_3 for the QoI $\mathbb{E}(f'(X_1, X_2, X_3, Z) \mid X_1, X_2, X_3)$ are respectively given by $S_1 = \frac{1}{2} \frac{(1 + \frac{b\pi^4}{5})^2}{\frac{a^2}{8} + \frac{b\pi^4}{5} + \frac{b^2\pi^8}{18} + \frac{1}{2}}$, $S_2 = \frac{\frac{a^2}{8}}{\frac{a^2}{8} + \frac{b\pi^4}{5} + \frac{b^2\pi^8}{18} + \frac{1}{2}}$ and $S_3 = 0$. Parameters a and b are chosen with respect to Sobol' and Levitan (1999): $a = 7$, $b = 0.05$ and Marrel et al.

(2009): $a = 7, b = 0.1$.

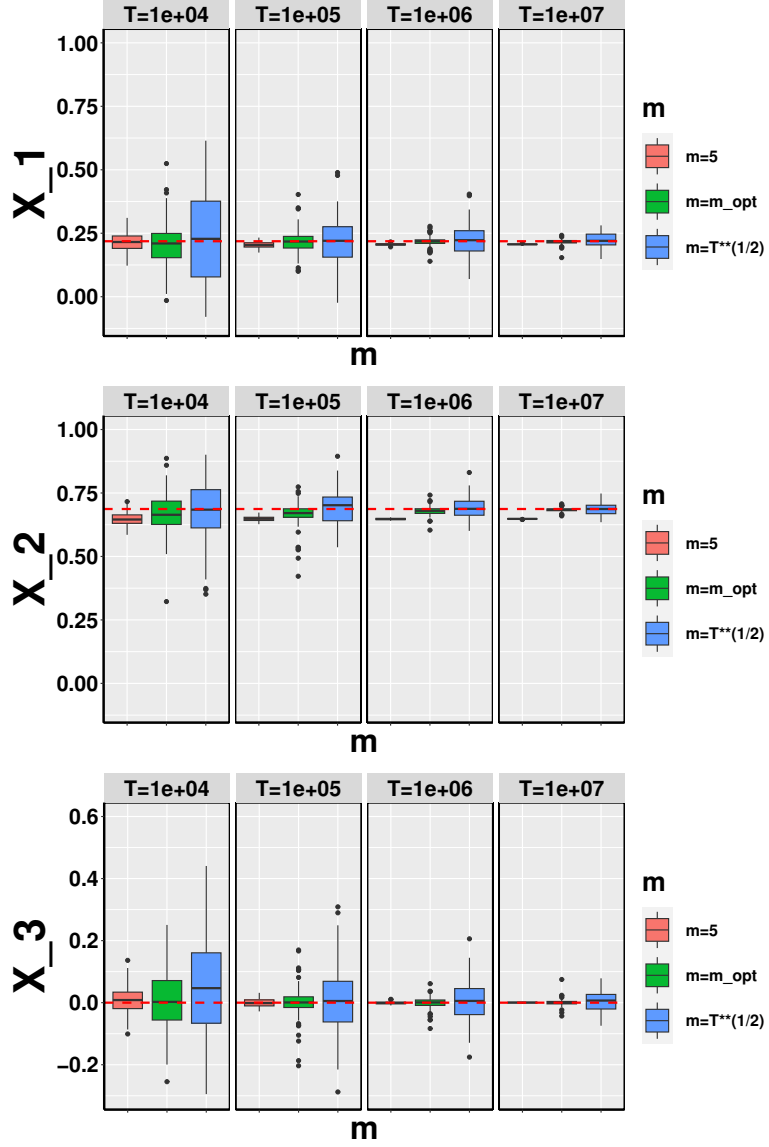


Figure 3: Boxplots of sensitivity index estimates of X_1 , X_2 and X_3 with respect to the stochastic version of Ishigami function (with $b = 0.05$) for different values of T . Three strategies of choice of m are compared: $m = 5$ (in red), $m = m_{opt}$ given by the trade-off strategy of Algorithm 2 (in green) and $m = T^{1/2}$ (in blue).

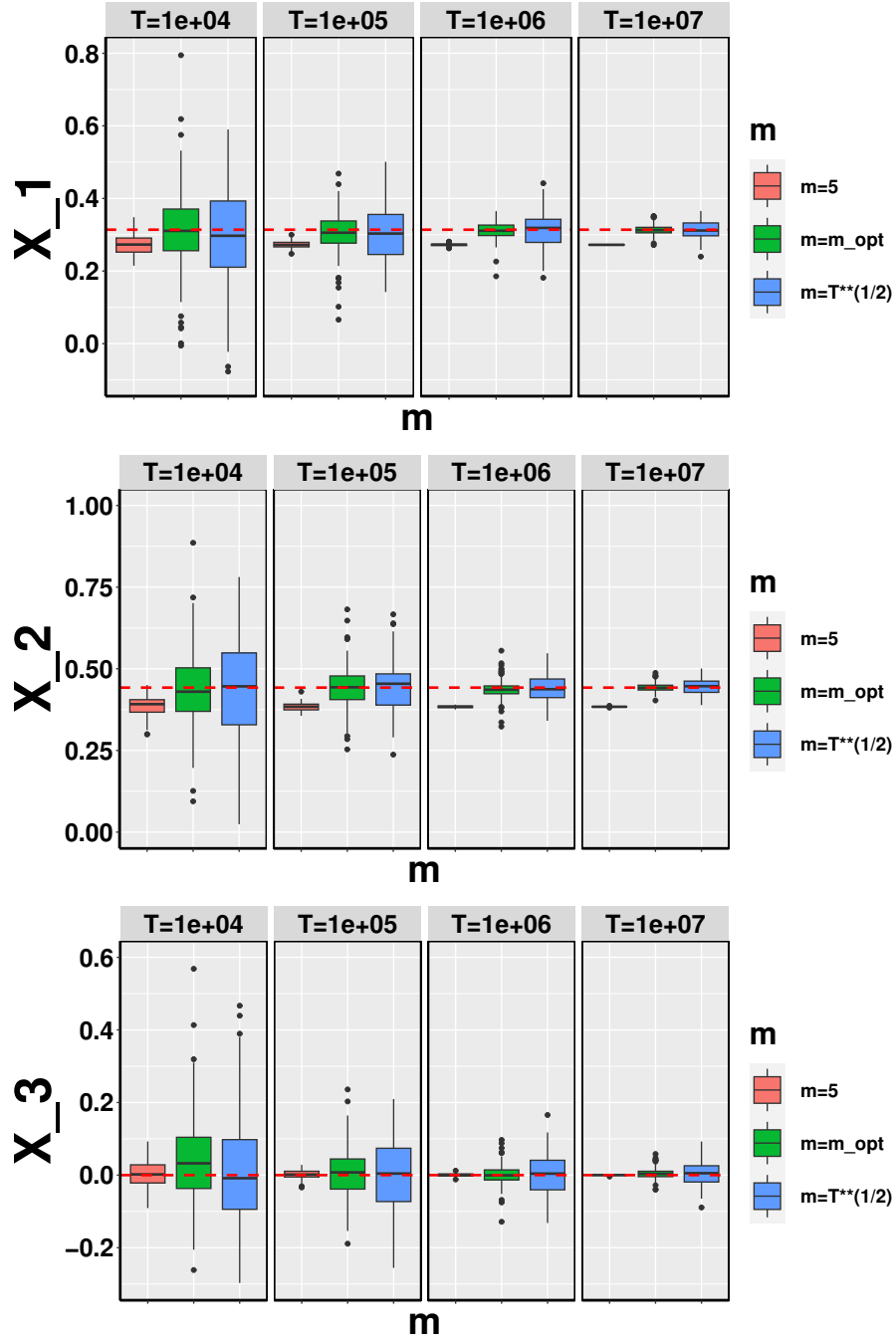


Figure 4: Boxplots of sensitivity index estimates of X_1 , X_2 and X_3 with respect to the stochastic version of Ishigami function (with $b = 0.1$) for different values of T . Three strategies of choice of m are compared: $m = 5$ (in red), $m = m_{opt}$ given by the trade-off strategy of Algorithm 2 (in green) and $m = T^{1/2}$ (in blue).

Figures 3 and 4 also reveal that estimations obtained by using Algorithm 2 are

more efficient for large T . Besides, remark that the term $bX_3^4 \sin X_1$ that multiplies the intrinsic noise term Z^2 includes b so that $|bX_3^4 \sin X_1| \leq b\pi^4$. Then, b allows to control the variance of the intrinsic noise term of the model. In both cases, the strategy implemented in Algorithm 2 has better results.

Overall, it appears that the strategy of Algorithm 2 provides better estimations and estimators it generates converge faster. In the particular case of $m = 5$, it is noticeable that errors do not decrease when T gets larger but rather they are quite constant. This is explained by the fact that the bias is constant since m is constant. This illustrates the importance of varying the number of repetition when the total computational budget grows. Regarding the case $m = T^{1/2}$, it turns out that estimators do not converge with optimal rate compared to the case $m = m_{opt}$ due the variance part of the mean-squared error. Indeed, with $m = T^{1/2}$, the variance converges to 0 at rate $T^{1/2}$ while the squared-bias converges at rate T . Then, the global convergence rate of the MSEs is $T^{1/2}$ that is slower than the rate $T^{2/3}$ of estimators built by Algorithm 2. These two cases clearly illustrate the bias-variance trade-off problem in Sobol' index estimation for stochastic models and they allow to show that the strategy proposed in this paper performs well.

7 Conclusion

To balance the tradeoff between repetitions and explorations in the estimation of Sobol' indices for stochastic models, a regularized estimator was devised and its mean-squared error bounded, approximated and minimized. A connection to functions of nested Monte Carlo estimators was made, for which generic results were obtained, especially useful when the Hessian matrix is unbounded. To estimate Sobol' indices in practice, an algorithm that adapts to the intrinsic randomness of the stochastic model was designed and illustrated on simulations, where the importance of a good balance between explorations and repetitions was observed. In the theoretical results, the regularization parameter, though arbitrary, has to be fixed. In the future, it would be of primary interest to let it vanish as the number of explorations and repetitions go to infinity, allowing one to get a genuine convergence rate for the mean-squared error of the regularized estimator. A further step would be to get a minimax low bound. Also, it would be interesting to compare the results with multilevel Monte Carlo methods (Mycek and De Lozzo (2019); Giles and Haji-Ali (2019)). Furthermore, it could be interesting to couple the iterative estimation approach of Gilquin et al. (2021) to the algorithms implemented in this study in order to build an adaptive version which could perform estimation with respect to a given precision. Another estimation approach worth considering is that based on Chatterjee's empirical correlation coefficient (Chatterjee, 2021; Gamboa et al., 2022).

Acknowledgment

We thank a anonymous reviewer for their helpful comments. We thank Cl  mentine Prieur for reading the first version of this paper and for the pertinent remarks and references she suggested for its improvement.

Appendix

A Proof of Proposition 1

Using convexity inequality, for all $h \in (0, 1)$ and $m \geq 1$, it holds:

$$\mathbb{E} \left(g_h(\hat{\theta}) - g(\theta) \right)^2 \leq 2\mathbb{E} \left(g_h(\hat{\theta}) - g_h(\mu_m) \right)^2 + 2 \left(g_h(\mu_m) - g(\theta) \right)^2.$$

Applying a Taylor-Lagrange expansion to g_h at points $\hat{\theta}$ and μ_m yields:

$$\begin{aligned} & \mathbb{E} \left(g_h(\hat{\theta}) - g_h(\mu_m) \right)^2 \\ &= \mathbb{E} \left(\nabla g_h(\mu_m)^\top (\hat{\theta} - \mu_m) + \frac{1}{2} (\hat{\theta} - \mu_m)^\top \nabla^2 g_h(\lambda\hat{\theta} + (1-\lambda)\mu_m) (\hat{\theta} - \mu_m) \right)^2 \\ &\leq 2\mathbb{E} \left(\nabla g_h(\mu_m)^\top (\hat{\theta} - \mu_m) \right)^2 + \frac{1}{2} \mathbb{E} \left((\hat{\theta} - \mu_m)^\top \nabla^2 g_h(\lambda\hat{\theta} + (1-\lambda)\mu_m) (\hat{\theta} - \mu_m) \right)^2, \end{aligned}$$

for some $\lambda \in (0, 1)$. Thus:

$$\frac{\mathbb{E} \left(g_h(\hat{\theta}) - g_h(\mu_m) \right)^2}{\mathbb{E} \left(\nabla g_h(\mu_m)^\top (\hat{\theta} - \mu_m) \right)^2} \leq 2 + \frac{1}{2} \frac{\mathbb{E} \left((\hat{\theta} - \mu_m)^\top \nabla^2 g_h(\lambda\hat{\theta} + (1-\lambda)\mu_m) (\hat{\theta} - \mu_m) \right)^2}{\mathbb{E} \left(\nabla g_h(\mu_m)^\top (\hat{\theta} - \mu_m) \right)^2}$$

Let

$$p_{n,m}(h) := \sup_{\lambda \in [0,1]} \frac{\mathbb{E} \left((\hat{\theta} - \mu_m)^\top \nabla^2 g_h(\lambda\hat{\theta} + (1-\lambda)\mu_m) (\hat{\theta} - \mu_m) \right)^2}{\mathbb{E} \left(\nabla g_h(\mu_m)^\top (\hat{\theta} - \mu_m) \right)^2}, \quad (16)$$

then

$$\mathbb{E} \left(g_h(\hat{\theta}) - g_h(\mu_m) \right)^2 \leq (2 + \frac{1}{2} p_{n,m}(h)) V_{n,m}(h) \leq 2(1 + p_{n,m}(h)) V_{n,m}(h).$$

Now, let us show that $p_{n,m}(h) \rightarrow 0$ as $n, m \rightarrow \infty$. Using Cauchy-Schwarz inequality yields that:

$$p_{n,m}(h) \leq \sqrt{\sup_{\lambda \in [0,1]} \mathbb{E} \left(\left\| \nabla^2 g_h(\lambda\hat{\theta} + (1-\lambda)\mu_m) \right\|_F^4 \right)} \times \frac{\sqrt{\mathbb{E} \left(\|\hat{\theta} - \mu_m\|^8 \right)}}{\mathbb{E} \left(\nabla g_h(\mu_m)^\top (\hat{\theta} - \mu_m) \right)^2}.$$

As $n, m \rightarrow \infty$, the first term of the product above is bounded by a constant independent of h , uniformly in λ , by Assumption 2. The second term is given by:

$$\frac{\sqrt{\mathbb{E}(\|\hat{\theta} - \mu_m\|^8)}}{\mathbb{E}(\nabla g(\mu_m + h\mathbf{u})^\top (\hat{\theta} - \mu_m))^2} \quad (17)$$

$$\begin{aligned} &= \frac{\sqrt{\mathbb{E}(\|\hat{\theta} - \mu_m\|^8)}}{\mathbb{E}(\nabla g(\mu_m)^\top (\hat{\theta} - \mu_m))^2} \times \frac{\mathbb{E}(\nabla g(\mu_m)^\top (\hat{\theta} - \mu_m))^2}{\mathbb{E}(\nabla g(\mu_m + h\mathbf{u})^\top (\hat{\theta} - \mu_m))^2} \\ &\leq \frac{\sqrt{\mathbb{E}(\|\hat{\theta} - \mu_m\|^8)}}{\mathbb{E}(\nabla g(\mu_m)^\top (\hat{\theta} - \mu_m))^2} \times \sup_{h \in (0,1)} \frac{\mathbb{E}(\nabla g(\mu_m)^\top (\hat{\theta} - \mu_m))^2}{\mathbb{E}(\nabla g(\mu_m + h\mathbf{u})^\top (\hat{\theta} - \mu_m))^2}. \end{aligned} \quad (18)$$

Lemma A.1.

$$\frac{\sqrt{\mathbb{E}(\|\hat{\theta} - \mu_m\|^8)}}{\mathbb{E}(\nabla g(\mu_m)^\top (\hat{\theta} - \mu_m))^2} = o(1), \quad (19)$$

as $n, m \rightarrow +\infty$

The first factor in the right-hand side in (18) does not depend on h and is of order $o(1)$ as stated in Lemma A.1. Moreover, since ∇g is continuous in parameter h , we have that the second factor in the right-hand side of (18) is $O(1)$ as $n, m \rightarrow \infty$. Therefore, $p_{n,m}(h) = o(1)$, and hence

$$\mathbb{E}(g_h(\hat{\theta}) - g(\theta))^2 \leq 4(1 + p_{n,m}(h))(V_{n,m}(h) + B_m(h)^2),$$

where $B_m(h) = g_h(\mu) - g(\theta)$ and $\lim_{n,m \rightarrow +\infty} p_{n,m}(h) = 0$.

Proof of Lemma A.1

First, let us bound the numerator of the ratio in Equation (19); we have

$$\mathbb{E}(\|\hat{\theta} - \mu_m\|^8) \leq 27 \sum_{j=1}^3 \mathbb{E}(\|\hat{\theta}_j - \mu_{mj}\|^8),$$

where $\hat{\theta}_j$ and μ_{mj} denote the j th component of $\hat{\theta}$ and μ_m , respectively. By Marcinkiewicz and Zygmund (1937) and Jensen inequalities, we have for every j that

$$\mathbb{E}(\|\hat{\theta}_j - \mu_{mj}\|^8) \leq \frac{B_8}{n^4} \mathbb{E}(|\hat{\theta}_{mj}^{(1)} - \mu_{mj}|^8),$$

where here B_8 is a universal constant.

The case $j = 2$ is the simplest. Notice that μ_{m2} does not depend on m , then the expansion of $\mathbb{E}(|\hat{Q}_m(\mathbf{X}^{(1)}) - \mu_{m2}|^8)$ through Newton formula yields terms of the form $\mu_{m2}^k \mathbb{E}(\hat{Q}_m^{8-k})$, $k = 0, \dots, 8$. Using Lemma C.1 in Section C provides that those terms are polynomial in m^{-1} .

Let us deal with the case $j = 1$. Expanding the power 8 through Newton's formula and bounding its terms yields

$$\mathbb{E} \left(\left| \widehat{Q}_m(\mathbf{X}^{(1)})^2 - \mu_{m1} \right|^8 \right) \leq (\mu_{m1}^8 \vee 1) \binom{8}{4} \left(\mathbb{E} \left(\widehat{Q}_m(\mathbf{X}^{(1)})^{16} \right) + 1 \right). \quad (20)$$

Denoting $\varphi(\mathbf{X}^{(1)}, Z^{(1,k)}) = Y^{(1,k)}$, we have

$$\begin{aligned} \mathbb{E} \left(\left| \widehat{Q}_m(\mathbf{X}^{(1)}) \right|^{16} \right) &= \mathbb{E} \left(\left| \frac{1}{m} \sum_{k=1}^m Y^{(1,k)} \right|^{16} \right) \\ &= \frac{1}{m^{16}} \sum_{k_1, \dots, k_{16}=1}^m \mathbb{E} \left(Y^{(1,k_1)} \dots Y^{(1,k_{16})} \right). \end{aligned}$$

The expectation in the right-hand side is symmetric in k_1, \dots, k_{16} , and hence, from Lemma C.1, the sum is a polynomial in m of degree 16. Therefore, the right-hand side in (20) is bounded uniformly in m .

Let us deal with the case $j = 3$. Proceeding as in (20), we have

$$\begin{aligned} &\mathbb{E} \left(\left| \widehat{Q}_m(\mathbf{X}^{(1)}) \widetilde{Q}_m(\widetilde{\mathbf{X}}_{\sim u}^{(1)}, \mathbf{X}_u^{(1)}) - \mu_{m3} \right|^8 \right) \\ &\leq (\mu_{m3}^8 \vee 1) \binom{8}{4} \left(\mathbb{E} \left(\left| \widehat{Q}_m(\mathbf{X}^{(1)}) \widetilde{Q}_m(\widetilde{\mathbf{X}}_{\sim u}^{(1)}, \mathbf{X}_u^{(1)}) \right|^8 \right) + 1 \right) \\ &\leq (\mu_{m3}^8 \vee 1) \binom{8}{4} \left(\mathbb{E} \left(\frac{1}{2} \widehat{Q}_m(\mathbf{X}^{(1)})^{16} + \frac{1}{2} \widetilde{Q}_m(\widetilde{\mathbf{X}}_{\sim u}^{(1)}, \mathbf{X}_u^{(1)})^{16} \right) + 1 \right), \end{aligned}$$

and this is also bounded uniformly in m . (Again by Lemma C.1.)

We now deal with the root of the denominator of the ratio in Equation (19). We have

$$\begin{aligned} &\mathbb{E} \left(\nabla g(\mu_m)^T (\widehat{\theta} - \mu_m) \right)^2 \\ &= \sum_{j_1, j_2=1}^3 \nabla g(\mu_m)_{j_1} \nabla g(\mu_m)_{j_2} \mathbb{E}(\widehat{\theta} - \mu_m)_{j_1} (\widehat{\theta} - \mu_m)_{j_2} \\ &= \frac{1}{n} \sum_{j_1, j_2=1}^3 \nabla g(\mu_m)_{j_1} \nabla g(\mu_m)_{j_2} \left(\mathbb{E} \widehat{\theta}_{mj_1}^{(1)} \widehat{\theta}_{mj_2}^{(1)} - \mu_{mj_1} \mu_{mj_2} \right). \quad (21) \end{aligned}$$

The infimum of the sum in (21) is reached for some m and greater than zero. Therefore, the numerator in Equation (19) is less than $1/n^4$ times a constant not depending on m or n and the denominator is equal to $1/n^2$ times a quantity greater than zero. Therefore, the supremum over m of the ratio in Equation (19) is of order $O(n^{-2})$. The proof is complete.

B Proof of Theorem 1

The following lemma will be needed:

Lemma B.1. For all $\alpha > 0$ and all $h \in (0, 1)$,

$$\lim_{n, m \rightarrow +\infty} \mathbb{E} \left[\frac{1}{\left(h + \hat{\theta}_1 - \hat{\theta}_2^2 \right)^\alpha} \right] = \frac{1}{\left(h + \text{Var}(Q(\mathbf{X})) \right)^\alpha} \leq \text{Var}(Q(\mathbf{X}))^{-\alpha}. \quad (22)$$

Note that the function g is infinitely differentiable over its domain \mathcal{D} and then its gradient is given by:

$$\nabla g(\theta_1, \theta_2, \theta_3) = \left(-\frac{\theta_3 - \theta_2^2}{(\theta_1 - \theta_2^2)^2}, \frac{2\theta_2(\theta_3 - \theta_1)}{(\theta_1 - \theta_2^2)^2}, \frac{1}{\theta_1 - \theta_2^2} \right)^\top.$$

Furthermore, the hessian matrix of g yields:

$$\nabla^2 g(\theta_1, \theta_2, \theta_3) = \begin{pmatrix} \frac{2(\theta_3 - \theta_2^2)}{(\theta_1 - \theta_2^2)^3} & \frac{2\theta_2(\theta_1 - 2\theta_3 + \theta_2^2)}{(\theta_1 - \theta_2^2)^3} & \frac{-1}{(\theta_1 - \theta_2^2)^2} \\ \frac{2\theta_2(\theta_1 - 2\theta_3 + \theta_2^2)}{(\theta_1 - \theta_2^2)^3} & \frac{2(\theta_3 - \theta_1)(\theta_1 + 3\theta_2^2)}{(\theta_1 - \theta_2^2)^3} & \frac{2\theta_2}{(\theta_1 - \theta_2^2)^2} \\ \frac{-1}{(\theta_1 - \theta_2^2)^2} & \frac{2\theta_2}{(\theta_1 - \theta_2^2)^2} & 0 \end{pmatrix}.$$

For any $(\theta_1, \theta_2, \theta_3) \in \mathcal{D}$, the matrix $\nabla^2 g(\theta_1, \theta_2, \theta_3)$ is under the form $\nabla^2 g(\theta_1, \theta_2, \theta_3) = B(\theta_1, \theta_2, \theta_3)/(\theta_1 - \theta_2^2)^3$ where $B(\theta_1, \theta_2, \theta_3)$ is the matrix:

$$B(\theta_1, \theta_2, \theta_3) = \begin{pmatrix} 2(\theta_3 - \theta_2^2) & 2\theta_2(\theta_1 - 2\theta_3 + \theta_2^2) & -(\theta_1 - \theta_2^2) \\ 2\theta_2(\theta_1 - 2\theta_3 + \theta_2^2) & 2(\theta_3 - \theta_1)(\theta_1 + 3\theta_2^2) & 2\theta_2(\theta_1 - \theta_2^2) \\ -(\theta_1 - \theta_2^2) & 2\theta_2(\theta_1 - \theta_2^2) & 0 \end{pmatrix}.$$

Notice that $B(\theta_1, \theta_2, \theta_3)$ includes only multivariate polynomials of variables θ_1, θ_2 and θ_3 .

Let us check Assumption 2. Let $\lambda \in [0, 1]$ and $h \in (0, 1)$. We have

$$\nabla^2 g \left(\lambda \hat{\theta} + (1 - \lambda)\mu_m + h\mathbf{u} \right) = \frac{B \left(\lambda \hat{\theta} + (1 - \lambda)\mu_m + h\mathbf{u} \right)}{\left(h + \lambda \hat{\theta}_1 + (1 - \lambda)\mu_{m1} - \left(\lambda \hat{\theta}_2 + (1 - \lambda)\mu_{m2} \right)^2 \right)^3}.$$

Thus:

$$\begin{aligned}
& \mathbb{E} \left(\sup_{\lambda} \|\nabla^2 g \left(\lambda \hat{\theta} + (1 - \lambda) \mu_m + h \mathbf{u} \right)\|_F^4 \right) \\
&= \mathbb{E} \left(\sup_{\lambda} \frac{\|B \left(\lambda \hat{\theta} + (1 - \lambda) \mu_m + h \mathbf{u} \right)\|_F^4}{\left(h + \lambda \hat{\theta}_1 + (1 - \lambda) \mu_{m1} - \left(\lambda \hat{\theta}_2 + (1 - \lambda) \mu_{m2} \right)^2 \right)^{12}} \right) \\
&\leq \sqrt{\mathbb{E} \left(\sup_{\lambda} \frac{1}{\left(h + \lambda \hat{\theta}_1 + (1 - \lambda) \mu_{m1} - \left(\lambda \hat{\theta}_2 + (1 - \lambda) \mu_{m2} \right)^2 \right)^{24}} \right)} \\
&\quad \times \sqrt{\mathbb{E} \left(\sup_{\lambda} \|B \left(\lambda \hat{\theta} + (1 - \lambda) \mu_m + h \mathbf{u} \right)\|_F^8 \right)} \\
&\leq \sqrt{\mathbb{E} \left(\sup_{\lambda} \frac{1}{\left(h + \lambda \left(\hat{\theta}_1 - \hat{\theta}_2^2 \right) + (1 - \lambda) (\mu_{m1} - \mu_{m1}^2) \right)^{24}} \right)} \quad (\text{by convexity inequality}) \\
&\quad \times \sqrt{\mathbb{E} \left(\sup_{\lambda} \|B \left(\lambda \hat{\theta} + (1 - \lambda) \mu_m + h \mathbf{u} \right)\|_F^8 \right)} \\
&\leq \sqrt{\mathbb{E} \left(\sup_{\lambda} \frac{\lambda}{\left(h + \left(\hat{\theta}_1 - \hat{\theta}_2^2 \right) \right)^{24}} \right) + \sup_{\lambda} \frac{1 - \lambda}{\left(h + \mu_{m1} - \mu_{m1}^2 \right)^{24}}} \quad (\text{by convexity inequality}) \\
&\quad \times \sqrt{\mathbb{E} \left(\sup_{\lambda} \|B \left(\lambda \hat{\theta} + (1 - \lambda) \mu_m + h \mathbf{u} \right)\|_F^8 \right)} \\
&\leq \sqrt{\mathbb{E} \left(\frac{1}{\left(h + \left(\hat{\theta}_1 - \hat{\theta}_2^2 \right) \right)^{24}} \right) + \frac{1}{\left(h + \mu_{m1} - \mu_{m1}^2 \right)^{24}}} \\
&\quad \times \sqrt{\mathbb{E} \left(\sup_{\lambda} \|B \left(\lambda \hat{\theta} + (1 - \lambda) \mu_m + h \mathbf{u} \right)\|_F^8 \right)} \\
&\leq \sqrt{\mathbb{E} \left(\frac{1}{\left(h + \left(\hat{\theta}_1 - \hat{\theta}_2^2 \right) \right)^{24}} \right) + \sup_{h \in (0,1)} \frac{1}{\left(h + \mu_{m1} - \mu_{m1}^2 \right)^{24}}} \\
&\quad \times \sqrt{\sup_{h \in (0,1)} \mathbb{E} \left(\sup_{\lambda} \|B \left(\lambda \hat{\theta} + (1 - \lambda) \mu_m + h \mathbf{u} \right)\|_F^8 \right)}.
\end{aligned}$$

One should remark that $\sup_{h \in (0,1)} \frac{1}{\left(h + \mu_{m1} - \mu_{m1}^2 \right)^{24}} \leq \frac{1}{(\text{Var}(\mathbb{E}[\varphi(\mathbf{X}, Z) | \mathbf{X}]))^{24}} < +\infty$.

Moreover, the matrix B is composed with polynomials of three variables. Since $\mathbb{E} \left(Q(\mathbf{X})^{16} \right) < +\infty$ then by using Lemma C.1 and by continuity of polynomial functions, it yields that

$\sup_{h \in (0,1)} \mathbb{E} \left(\sup_{\lambda \in [0,1]} \|B \left(\lambda \hat{\theta} + (1-\lambda)\mu_m + h\mathbf{u} \right)\|_F^8 \right)$ is bounded. Finally, by relying on Lemma B.1, $\mathbb{E}(h + \hat{\theta}_1 - \hat{\theta}_2^2)^{-24}$ is a bounded by $\frac{1}{(\text{Var}(\mathbb{E}[\varphi(\mathbf{X}, Z) | \mathbf{X}]))^{24}}$ as $n, m \rightarrow +\infty$. Therefore, Assumption 2 is satisfied.

Proof of Lemma B.1

Let $h \in (0, 1)$ be fixed. The function $\beta_h : x \mapsto 1/(h + x)^\alpha$ is continuously differentiable such that its first derivative is uniformly bounded on \mathbb{R}_+ by $1/h^{\alpha+1}$ then it is Lipschitz. Therefore:

$$\begin{aligned} \mathbb{E} \left(\beta_h(\hat{\theta}_1 - \hat{\theta}_2^2) - \beta_h(\theta_1 - \theta_2^2) \right)^2 &\leq \frac{1}{h^{2\alpha+2}} \mathbb{E} \left(\hat{\theta}_1 - \hat{\theta}_2^2 - \theta_1 + \theta_2^2 \right)^2 \\ &\leq \frac{2}{h^{2\alpha+2}} \left(\text{Var}(\hat{\theta}_1) + \text{Var}(\hat{\theta}_2^2) + \left(\mathbb{E}(\hat{\theta}_2^2) - \theta_2^2 \right)^2 \right), \end{aligned}$$

using convexity inequality. Based on Marcinkiewicz-Zygmund inequality (see Theorem D.1) and Lemma C.1, it follows that

$$\lim_{n, m \rightarrow \infty} \left(\text{Var}(\hat{\theta}_1) + \text{Var}(\hat{\theta}_2^2) + \left(\mathbb{E}(\hat{\theta}_2^2) - \theta_2^2 \right)^2 \right) = 0.$$

Straightforwardly:

$$\lim_{n, m \rightarrow \infty} \mathbb{E} \left(\frac{1}{(h + \hat{\theta}_1 - \hat{\theta}_2^2)^\alpha} \right) = \frac{1}{(h + \theta_1 - \theta_2^2)^\alpha} \leq \frac{1}{\text{Var}(Q(\mathbf{X}))^\alpha}.$$

C A lemma

Lemma C.1. Let $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}$ be n i.i.d. copies of \mathbf{X} and $Z^{(1,1)}, \dots, Z^{(n,m)}$ be $n \times m$ i.i.d. copies of Z such that $(\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)})$ and $(Z^{(1,1)}, \dots, Z^{(n,m)})$ are independent. Then, for all $q \in \mathbb{N}$: $m^{-q} \mathbb{E} \left(\sum_{k=1}^m \varphi(\mathbf{X}^{(1)}, Z^{(1,k)}) \right)^q$ is polynomial in m^{-1} of degree $q - 1$ with constant $\mathbb{E}(\mathbb{E}[\varphi(\mathbf{X}, Z) | \mathbf{X}])^q$.

It holds that:

$$\begin{aligned} \mathbb{E} \left(\frac{1}{m} \sum_{k=1}^m \varphi(\mathbf{X}^{(1)}, Z^{(1,k)}) \right)^q &= \frac{1}{m^q} \mathbb{E} \left(\sum_{k_1=1}^m \dots \sum_{k_q=1}^m \varphi(\mathbf{X}^{(1)}, Z^{(1,k_1)}) \dots \varphi(\mathbf{X}^{(1)}, Z^{(1,k_q)}) \right) \\ &= \frac{1}{m^q} \sum_{k_1=1}^m \dots \sum_{k_q=1}^m \mathbb{E} \left(\varphi(\mathbf{X}^{(1)}, Z^{(1,k_1)}) \dots \varphi(\mathbf{X}^{(1)}, Z^{(1,k_q)}) \right). \end{aligned}$$

Denote by $\lambda : \{1, \dots, m\}^q \rightarrow \mathbb{N}$ the map which with each $\mathbf{k} := (k_1, \dots, k_q)$ associates the number of distinct indices among k_1, \dots, k_q . If $1 \leq l \leq q$ then denote by $\rho_l : \lambda^{-1}(l) \rightarrow \{1, \dots, q\}^l$ the map which with each $\mathbf{k} \in \lambda^{-1}(l)$ associates (r_1, \dots, r_l) , where $r_i = |\{j : k_j = k_{j_i}\}|$ for every $i = 1, \dots, l$ and k_{j_1}, \dots, k_{j_l} are the distinct

indices found among k_1, \dots, k_q . Obviously, $r_1 + \dots + r_l = q$. We have

$$\begin{aligned}
& \sum_{k_1=1}^m \dots \sum_{k_q=1}^m \mathbb{E} \left(\varphi(\mathbf{X}^{(1)}, Z^{(1,k_1)}) \dots \varphi(\mathbf{X}^{(1)}, Z^{(1,k_q)}) \right) \\
&= \sum_{k_1=1}^m \dots \sum_{k_q=1}^m f(\mathbf{k}) \\
&= \sum_{l=1}^q \left(\sum_{(r_1, \dots, r_l) \in \{1, \dots, q\}^l : r_1 + \dots + r_l = q} \left(\sum_{\mathbf{k} \in \lambda^{-1}(l) : \rho_l(\mathbf{k}) = (r_1, \dots, r_l)} f(\mathbf{k}) \right) \right). \quad (23)
\end{aligned}$$

Now, since

$$\begin{aligned}
& \mathbb{E} \left(\varphi(\mathbf{X}^{(1)}, Z^{(1,k_1)}) \dots \varphi(\mathbf{X}^{(1)}, Z^{(1,k_q)}) \right) \\
&= \mathbb{E} \left(\varphi(\mathbf{X}^{(1)}, Z^{(1,k_{j_1})})^{r_1} \dots \varphi(\mathbf{X}^{(1)}, Z^{(1,k_{j_l})})^{r_l} \right) \\
&= \mathbb{E} \left(\prod_{s=1}^l \mathbb{E} \left[\varphi(\mathbf{X}^{(1)}, Z^{(1,k_{j_s})})^{r_s} \mid \mathbf{X}^{(1)} \right] \right)
\end{aligned}$$

is symmetric in r_1, \dots, r_l , it holds that

$$\begin{aligned}
& \sum_{(r_1, \dots, r_l) \in \{1, \dots, q\}^l : r_1 + \dots + r_l = q} \left(\sum_{\mathbf{k} \in \lambda^{-1}(l) : \rho_l(\mathbf{k}) = (r_1, \dots, r_l)} f(\mathbf{k}) \right) \\
&= c(l, (r_1, \dots, r_l), m) \mathbb{E} \left(\prod_{s=1}^l \mathbb{E} \left[\varphi(\mathbf{X}^{(1)}, Z^{(1,k_{j_s})})^{r_s} \mid \mathbf{X}^{(1)} \right] \right)
\end{aligned}$$

where

$$c(l, (r_1, \dots, r_l), m) = \binom{q}{r_1} \binom{q-r_1}{r_2} \dots \binom{q-r_1-\dots-r_{l-1}}{r_l} m(m-1) \dots (m-l+1). \quad (24)$$

Notice that the expression in the right-hand side of (24) is invariant by permutation of r_1, \dots, r_l . Therefore, the sum (23) is a polynomial in m of degree q with constant zero and hence $\mathbb{E} \left(\frac{1}{m} \sum_{k=1}^m \varphi(\mathbf{X}^{(1)}, Z^{(1,k)}) \right)^q$ is a polynomial in $\frac{1}{m}$ of degree $q-1$ with constant $\lim_{m \rightarrow +\infty} \mathbb{E} \left(\frac{1}{m} \sum_{k=1}^m \varphi(\mathbf{X}^{(1)}, Z^{(1,k)}) \right)^q = \mathbb{E} (\mathbb{E} [\varphi(\mathbf{X}, Z) \mid \mathbf{X}])^q$.

D The Marcinkiewicz-Zygmund inequality

Theorem D.1 (Marcinkiewicz and Zygmund (1937)). *Let U_1, \dots, U_n be i.i.d. random variables such that $\mathbb{E}(U_1) = 0$ and $\mathbb{E}|U_1|^q < +\infty$, where $1 \leq q < +\infty$. There exist A_q and B_q depending only on q such that:*

$$A_q \mathbb{E} \left(\left(\sum_{i=1}^n |U_i|^2 \right)^{\frac{q}{2}} \right) \leq \mathbb{E} \left(\left| \sum_{i=1}^n U_i \right|^q \right) \leq B_q \mathbb{E} \left(\left(\sum_{i=1}^n |U_i|^2 \right)^{\frac{q}{2}} \right)$$

Furthermore, there exists C_q independent from n such that:

$$\mathbb{E} \left(\left| \frac{1}{n} \sum_{i=1}^n U_i \right|^q \right) \leq \frac{C_q}{n^{\frac{q}{2}}}. \quad (25)$$

References

- Mickaël Binois, Robert B. Gramacy, and Mike Ludkovski. Practical Heteroscedastic Gaussian Process Modeling for Large Simulation Experiments. *Journal of Computational and Graphical Statistics*, 27(4):808–821, 2018. doi: 10.1080/10618600.2018.1458625.
- Mickaël Binois, Jiangeng Huang, Robert B. Gramacy, and Mike Ludkovski. Replication or Exploration? Sequential Design for Stochastic Simulation Experiments. *Technometrics*, 61(1):7–23, 2019. doi: 10.1080/00401706.2018.1469433.
- G. Castellan, A. Cousien, and V.C. Tran. Non-parametric adaptive estimation of order 1 Sobol’ indices in stochastic models, with an application to Epidemiology. *Electronic Journal of Statistics*, 14(1):50 – 81, 2020. doi: 10.1214/19-EJS1627.
- Sourav Chatterjee. A New Coefficient of Correlation. *Journal of the American Statistical Association*, 116(536):2009–2022, October 2021. ISSN 0162-1459. doi: 10.1080/01621459.2020.1758115. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/01621459.2020.1758115>.
- Xi Chen and Qiang Zhou. Sequential experimental designs for stochastic kriging. In *Proceedings of the Winter Simulation Conference 2014*, pages 3821–3832, 2014. doi: 10.1109/WSC.2014.7020209.
- Xi Chen and Qiang Zhou. Sequential design strategies for mean response surface metamodeling via stochastic kriging with adaptive exploration and exploitation. *European Journal of Operational Research*, 262(2):575–585, 2017. ISSN 0377-2217. doi: <https://doi.org/10.1016/j.ejor.2017.03.042>.
- A. Courcoul, H. Monod, M. Nielen, D. Klinkenberg, L. Hogerwerf, F. Beaudeau, and E. Vergu. Modelling the effect of heterogeneity of shedding on the within herd coxiella burnetii spread and identification of key parameters by sensitivity analysis. *Journal of Theoretical Biology*, 284(1):130–141, 2011. ISSN 0022-5193. doi: <https://doi.org/10.1016/j.jtbi.2011.06.017>.
- S. da Veiga. Kernel-based ANOVA decomposition and Shapley effects – Application to global sensitivity analysis, 2021. arXiv:2101.05487.
- S. da Veiga and F Gamboa. Efficient estimation of sensitivity indices. *Journal of Non-parametric Statistics*, 25(3):573–595, 2013. doi: 10.1080/10485252.2013.784762.
- S. da Veiga, J-M Loubes, and M. Solís. Efficient estimation of conditional covariance matrices for dimension reduction. *Communications in Statistics - Theory and Methods*, 46(9):4403–4424, 2017. doi: 10.1080/03610926.2015.1083109.
- Yohann de Castro and Alexandre Janon. Randomized pick-freeze for sparse Sobol’ indices estimation in high dimension. *ESAIM: PS*, 19:725–745, 2015. doi: 10.1051/ps/2015013.
- P. Étoré, C. Prieur, D. K. Pham, and L. Li. Global sensitivity analysis for models described by Stochastic Differential Equations. *Methodology and Computing in Applied Probability*, 22(2):803–831, June 2020. ISSN 1387-5841, 1573-7713. doi: 10.1007/s11009-019-09732-6.

- J-C. Fort, T. Klein, A. Lagnoux, and B. Laurent. Estimation of the Sobol' indices in a linear functional multidimensional model. *Journal of Statistical Planning and Inference*, 143(9):1590–1605, 2013. ISSN 0378-3758. doi: <https://doi.org/10.1016/j.jspi.2013.04.007>.
- J-C. Fort, T. Klein, and A. Lagnoux. Global sensitivity analysis and Wasserstein spaces. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):880–921, 2021. doi: 10.1137/20M1354957.
- F. Gamboa, T. Klein, A. Lagnoux, and L. Moreno. Sensitivity analysis in general metric spaces. *Reliability Engineering & System Safety*, 212:107611, 2021.
- Fabrice Gamboa, Pierre Gremaud, Thierry Klein, and Agnès Lagnoux. Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics. *Bernoulli*, 28(4):2345–2374, November 2022. ISSN 1350-7265. Publisher: Bernoulli Society for Mathematical Statistics and Probability.
- M. B. Giles and A.-L. Haji-Ali. Multilevel nested simulation for efficient risk estimation. *SIAM/ASA Journal on Uncertainty Quantification*, 7(2):497–525, 2019.
- L. Gilquin, C. Prieur, E. Arnaud, and H. Monod. Iterative estimation of Sobol' indices based on replicated designs. *Computational and Applied Mathematics*, 40(1):18, January 2021. ISSN 1807-0302. doi: 10.1007/s40314-020-01402-5.
- D. Giorgi, V. Lemaire, and G. Pagès. Limit theorems for weighted and regular Multilevel estimators. *Monte Carlo Methods and Applications*, 23(1):43–70, 2017. doi: 10.1515/mcma-2017-0102.
- D. Giorgi, V. Lemaire, and G. Pagès. Weak Error for Nested Multilevel Monte Carlo. *Methodology and Computing in Applied Probability*, 22(3):1325–1348, September 2020. ISSN 1573-7713. doi: 10.1007/s11009-019-09751-3.
- J. L. Hart, A. Alexanderian, and P. A. Gremaud. Efficient computation of Sobol' Indices for stochastic models. *SIAM Journal on Scientific Computing*, 39(4):A1514–A1530, 2017. doi: 10.1137/16M106193X.
- L. J. Hong and S. Juneja. Estimating the mean of a non-linear function of conditional expectation. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pages 1223–1236, 2009. doi: 10.1109/WSC.2009.5429428.
- T. Ishigami and T. Homma. An importance quantification technique in uncertainty analysis for computer models. [1990] *Proceedings. First International Symposium on Uncertainty Modeling and Analysis*, pages 398–403, 1990.
- Alexandre Janon, Thierry Klein, Agnes Lagnoux-Renaudie, Maëlle Nodet, and Clémentine Prieur. Asymptotic normality and efficiency of two Sobol' index estimators. *ESAIM: Probability and Statistics*, 18:342–364, October 2014. doi: 10.1051/ps/2013040.
- M. Navarro Jimenez, O. P. Le Maître, and O. M. Knio. Nonintrusive polynomial chaos expansions for sensitivity analysis in stochastic differential equations. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):378–402, 2017. doi: 10.1137/16M1061989.

- J. Marcinkiewicz and A. Zygmund. Sur les fonctions indépendantes. *Fundamenta Mathematicae*, 29(1):60–90, 1937.
- A. Marrel, B. Iooss, B. Laurent, and O. Roustant. Calculations of Sobol’ indices for the Gaussian process metamodel. *Reliability Engineering & System Safety*, 94(3): 742–751, 2009. ISSN 0951-8320. doi: <https://doi.org/10.1016/j.ress.2008.07.008>.
- G. Mazo. A trade-off between explorations and repetitions for estimators of two global sensitivity indices in stochastic models induced by probability measures. *SIAM/ASA Journal on Uncertainty Quantification*, 9(4):1673–1713, 2021. doi: 10.1137/19M1272706.
- P. Mycek and M. De Lozzo. Multilevel Monte Carlo covariance estimation for the computation of Sobol’ Indices. *SIAM/ASA Journal on Uncertainty Quantification*, 7(4):1323–1348, 2019. doi: 10.1137/18M1216389.
- I. Panin. Risk of estimators for Sobol’ sensitivity indices based on metamodels. *Electronic Journal of Statistics*, 15(1):235 – 281, 2021. doi: 10.1214/20-EJS1793.
- T. Rainforth, R. Cornish, H. Yang, A. Warrington, and F. Wood. On Nesting Monte Carlo Estimators. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 4267–4276. PMLR, 10–15 Jul 2018.
- I.M. Sobol’ and Yu.L. Levitan. On the use of variance reducing multipliers in Monte Carlo computations of a global sensitivity index. *Computer Physics Communications*, 117(1):52–61, 1999. ISSN 0010-4655. doi: [https://doi.org/10.1016/S0010-4655\(98\)00156-8](https://doi.org/10.1016/S0010-4655(98)00156-8).
- X. Zhu and B. Sudret. Global sensitivity analysis for stochastic simulators based on generalized lambda surrogate models. *Reliability Engineering & System Safety*, 214: 107815, 2021. ISSN 0951-8320. doi: <https://doi.org/10.1016/j.ress.2021.107815>.