Estimating Sobol' indices combining Monte Carlo estimators and Latin hypercube sampling

Jean-Yves Tissot, Clémentine Prieur

Estimating Sobol' indices combining Monte Carlo estimators and Latin hypercube sampling

Jean-Yves Tissot* and Clémentine Prieur

Université de Grenoble LJK/INRIA

51, rue des Mathématiques
Campus de Saint Martin d'Hères, BP53
38041 Grenoble cedex 09 (France)

(jeanyvestissot@free.fr)
(clementine.prieur@imag.fr)

Phone: +33 (0)4 76 63 54 47
Fax: +33 (0)4 76 63 12 63

October 19, 2012

Abstract

In variance-based sensitivity analysis, the method of Sobol' (1993) allows to compute Sobol' indices using Monte Carlo integration. One of the main drawbacks of this approach is that the estimation of Sobol' indices requires the use of several samples. For example, in a d-dimensional space, the estimation of all the first-order Sobol' indices requires $d + 1$ samples. Some interesting combinatorial results have been introduced to weaken this defect, in particular by Saltelli (2002) and more recently by Owen (2012), but the quantities they estimate still require $O(d)$ samples. In this paper, we introduce a new approach to estimate for any k all the k-th order Sobol' indices by using only two samples. We establish theoretical properties of such a method for the first-order Sobol' indices and discuss the generalization to higher-order indices. As an illustration, we propose to apply this new approach to a marine ecosystem model of the Ligurian sea (northwestern Mediterranean) in order to study the relative importance of its several parameters. The calibration process of this kind of chemical simulators is well-known to be quite intricate, and a rigorous and robust — i.e. valid without strong regularity assumptions — sensitivity analysis, as the method of Sobol' provides, could be of great help.

*Corresponding author.

Keywords: global sensitivity analysis, variance-based sensitivity indices, numerical integration, orthogonal arrays

1 Introduction and notation

Sobol' indices are quantities defined by normalizing parts of variance in an ANOVA decomposition. They allow to quantify the relative importance of input factors of a function over their entire range of values. They essentially consist of integrals and as a consequence, their computation can become rapidly expensive when the number of factors increases. Many techniques have been developed to estimate these indices including Fast Amplitude Sensitivity Test (FAST) due to Cukier et al. (1978) and Saltelli et al. (1999), Random Balance Design (RBD) due to Tarantola et al. (2006) — for a recent survey see Tissot and Prieur (2012) —, polynomial chaos expansion (PCE)-based estimators developed in Sudret (2008) and Blatman and Sudret (2010) and the method of Sobol', see also Saltelli et al. (2008) for a review. Until now, spectral methods — as FAST, RBD or PCE-based methods — which exploit the spectral decomposition of the model with respect to a particular multivariate basis, are generally preferred to the method of Sobol' because the latter is too expensive. However, spectral methods provide good estimations of Sobol' indices only under assumptions on the spectral decomposition of the model itself (decay of the spectrum sufficiently fast, negligibility of high-order spectral coefficients, etc.). As a consequence, these methods are not robust to complex phenomena as high-frequency variations or discontinuities, and so the method of Sobol' appears as the only method one can trust when no strong a priori knowledge on the model of interest is available.

The general framework of ANOVA decomposition and Sobol' indices is the following. Let f be a real square integrable function defined on the unit hypercube $[0,1]^d$ and $\mathbf{X} = (X_1, \ldots, X_d)$ a random vector with independent components uniformly distributed on $[0,1]$. We consider the real random variable $Y = f(\mathbf{X})$. Note that this framework can be generalized to independent arbitrary marginal distributions $(X_i)_{i=1..d}$ by using the inverse transformation method. Then for any $\mathfrak{u} \subseteq \{1, \ldots, d\}$, denote $\mathbf{X}_\mathfrak{u}$ the random vector with components X_i, $i \in \mathfrak{u}$. The ANOVA decomposition — see Hoeffding (1948) and Efron and Stein (1981) — states that $Y = f(\mathbf{X})$ can be uniquely decomposed into summands of increasing dimensions

$$f(\mathbf{X}) = \sum_{\mathfrak{u} \subseteq \{1,\ldots,d\}} f_\mathfrak{u}(\mathbf{X}_\mathfrak{u}) \tag{1}$$

where $f_\emptyset = \mathbb{E}[Y]$ and the other components have mean zero and are mutually uncorrelated. In particular, the sum of functions

$$f_\emptyset + f_1(X_1) + f_2(X_2) + \cdots + f_d(X_d) \tag{2}$$

is the so-called additive part of f.

The Sobol' index with respect to the combination of all the variables in $\mathfrak{u} \subseteq \{1, \ldots, d\}$ — see Sobol'

(1993) — is then defined as

$$S_{\mathfrak{u}} = \frac{\sigma^2_{\mathfrak{u}}}{\sigma^2} = \frac{\mathrm{Var}\big[f_{\mathfrak{u}}(\mathbf{x}_{\mathfrak{u}})\big]}{\mathrm{Var}[Y]}$$

and the Sobol' index with respect to a subset of variables $\mathfrak{u} \subseteq \{1,\dots,d\}$ — see Homma and Saltelli (1996) — is then defined as

$$\underline{S}_{\mathfrak{u}} = \frac{\tau^2_{\mathfrak{u}}}{\sigma^2} = \frac{\mathrm{Var}\big[\sum_{\mathfrak{v}\subseteq\mathfrak{u}} f_{\mathfrak{v}}(\mathbf{x}_{\mathfrak{v}})\big]}{\mathrm{Var}[Y]} \ .$$

In practice, global sensitivity analysis focuses on the first-order — i.e. $|\mathfrak{u}| = 1$ — and the second-order — i.e. $|\mathfrak{u}| = 2$ — terms. Note that, thanks to the properties of the ANOVA decomposition, we have

$$\underline{S}_{\mathfrak{u}} = \sum_{\mathfrak{v}\subseteq\mathfrak{u}} S_{\mathfrak{v}}$$

and the Möbius inversion formula — see, e.g., Stanley (2012) — gives

$$S_{\mathfrak{u}} = \sum_{\mathfrak{v}\subseteq\mathfrak{u}} (-1)^{|\mathfrak{u}|-|\mathfrak{v}|} \underline{S}_{\mathfrak{v}} \ .$$

Concerning notation, when integrals are over a unit hypercube $[0,1]^s$, $s \leq d$, the integration set is generally omitted, and for any $\mathfrak{u} \subseteq \{1,\dots,d\}$, we denote by $\mathfrak{u}^c = \{1,\dots,d\} \setminus \mathfrak{u}$ the relative complement of $\mathfrak{u}$ with respect to $\{1,\dots,d\}$.

Section 2 provides a short review of Monte Carlo estimators of Sobol' indices and gives some notation. In Section 3 we explain how to combine Monte Carlo estimators and Latin hypercube sampling — see McKay et al. (1979) — in a basic way, and we give asymptotic and bias properties of such a technique. In Section 4 we study the method introduced in Section 3 using replicated Latin hypercubes — see McKay (1995) —, we give asymptotic and bias properties of this technique and we explain how it allows to compute all the first-order Sobol' indices using only two replicated Latin hypercubes. Potential generalization to orthogonal array-based Latin hypercubes — see Owen (1992) — is also discussed in this section. Numerical illustrations are provided in Section 5, and Section 6 has conclusions. Note that technical lemmas are given in the Appendix.

2 Review of Monte Carlo estimators

2.1 Notation

Let $\mathfrak{u}$ be a non-empty subset of $\{1,\dots d\}$, and j in $\{1,\dots n\}$, and consider

$$\mathbf{Z}^j_{\mathfrak{u}} = (X^j_1, \dots X^j_{2d-|\mathfrak{u}|}) \tag{3}$$

where the X_i^j's are independent random variables uniformly distributed on $[0,1]$. We also denote

$$\mathbf{X}_{\mathrm{u}}^j = (X_1^j, \ldots, X_{|\mathrm{u}|}^j) \tag{4}$$

$$\mathbf{X}_{\mathrm{u}^c}^{j,1} = (X_{|\mathrm{u}|+1}^j, \ldots, X_d^j) \tag{5}$$

$$\mathbf{X}_{\mathrm{u}^c}^{j,2} = (X_{d+1}^j, \ldots, X_{2d-|\mathrm{u}|}^j) \tag{6}$$

so that

$$\mathbf{Z}_{\mathrm{u}}^j = (\mathbf{X}_{\mathrm{u}}^j, \mathbf{X}_{\mathrm{u}^c}^{j,1}, \mathbf{X}_{\mathrm{u}^c}^{j,2}) \ . \tag{7}$$

Finally, for $k = 1$ and 2, consider

$$Y_{\mathrm{u}}^{j,k} = f(\mathbf{X}_{\mathrm{u}}^j, \mathbf{X}_{\mathrm{u}^c}^{j,k}) \ . \tag{8}$$

With this notation, we consider two estimators of the Sobol' indices $\underline{S}_{\mathrm{u}}$, introduced in Homma and Saltelli (1996) and Monod et al. (2006), which are functions of $(\mathbf{Z}_{\mathrm{u}}^j)_{j=1..n}$. They are defined by

$$\widetilde{\underline{S}}_{\mathrm{u},n} = \frac{\widetilde{\underline{\tau}}_{\mathrm{u},n}^2}{\widetilde{\sigma}_n^2} = \frac{\dfrac{1}{n}\displaystyle\sum_{j=1}^n Y_{\mathrm{u}}^{j,1} Y_{\mathrm{u}}^{j,2} - \left(\dfrac{1}{n}\displaystyle\sum_{j=1}^n Y_{\mathrm{u}}^{j,1}\right)\left(\dfrac{1}{n}\displaystyle\sum_{j=1}^n Y_{\mathrm{u}}^{j,2}\right)}{\dfrac{1}{n}\displaystyle\sum_{j=1}^n \left(Y_{\mathrm{u}}^{j,1}\right)^2 - \left(\dfrac{1}{n}\displaystyle\sum_{j=1}^n Y_{\mathrm{u}}^{j,1}\right)^2} \tag{9}$$

and

$$\widehat{\underline{S}}_{\mathrm{u},n} = \frac{\widehat{\underline{\tau}}_{\mathrm{u},n}^2}{\widehat{\sigma}_n^2} = \frac{\dfrac{1}{n}\displaystyle\sum_{j=1}^n Y_{\mathrm{u}}^{j,1} Y_{\mathrm{u}}^{j,2} - \left(\dfrac{1}{2n}\displaystyle\sum_{j=1}^n Y_{\mathrm{u}}^{j,1} + Y_{\mathrm{u}}^{j,2}\right)^2}{\dfrac{1}{2n}\displaystyle\sum_{j=1}^n \left(\left(Y_{\mathrm{u}}^{j,1}\right)^2 + \left(Y_{\mathrm{u}}^{j,2}\right)^2\right) - \left(\dfrac{1}{2n}\displaystyle\sum_{j=1}^n Y_{\mathrm{u}}^{j,1} + Y_{\mathrm{u}}^{j,2}\right)^2} \ , \tag{10}$$

respectively. Note that other Monte Carlo estimators exist (for a recent review, see Owen (2012)).

2.2 Statistical properties of the estimators

Asymptotic properties of both the estimators introduced in the previous section are detailed in Janon et al. (2012). $\widetilde{\underline{S}}_{\mathrm{u},n}$ and $\widehat{\underline{S}}_{\mathrm{u},n}$ are strongly consistent and asymptotically normal estimators, and $\widehat{\underline{S}}_{\mathrm{u},n}$ is, in addition, asymptotically efficient in some sense (see details in Proposition 2.5 in Janon et al. (2012)). Concerning the biases, it is easy to show that

$$\mathbb{E}\left[\widetilde{\underline{\tau}}_{\mathrm{u},n}^2\right] = \underline{\tau}_{\mathrm{u}}^2 - \frac{1}{n}\underline{\tau}_{\mathrm{u}}^2 \tag{11}$$

$$\mathbb{E}\left[\widetilde{\sigma}_n^2\right] = \sigma^2 - \frac{1}{n}\sigma^2 \tag{12}$$

and — see e.g. Owen (2012) —

$$\mathbb{E}\left[\widehat{\underline{\tau}}_{\mathrm{u},n}^2\right] = \underline{\tau}_{\mathrm{u}}^2 - \frac{1}{2n}(\sigma^2 + \underline{\tau}_{\mathrm{u}}^2) \tag{13}$$

$$\mathbb{E}\left[\widehat{\sigma}_n^2\right] = \sigma^2 - \frac{1}{2n}(\sigma^2 + \underline{\tau}_{\mathrm{u}}^2) \tag{14}$$

but as far as we know, there is no result on the global biases of $\widetilde{\underline{S}}_{\mathrm{u},n}$ and $\widehat{\underline{S}}_{\mathrm{u},n}$.

3 Monte Carlo estimators and Latin hypercube sampling

3.1 Notation and definitions

We begin with the definition of a Latin hypercube:

Definition 1. *Let d and n in $\mathbb{N}^*$, and consider Π_n the set of all the permutations of $\{1,\dots,n\}$. We say that $(\mathbf{X}^j)_{j=1..n}$ is a Latin hypercube of size n in $[0,1]^d$ — and we denote $(\mathbf{X}^j)_j \sim \mathcal{LH}(n,d)$ — if for all $j \in \{1,\dots,n\}$,*

$$\mathbf{X}^j = \left(\frac{\pi_1(j) - U_{1,\pi_1(j)}}{n}, \dots, \frac{\pi_d(j) - U_{d,\pi_d(j)}}{n} \right) \tag{15}$$

where the π_i's and the $U_{i,j}$'s are independent random variables uniformly distributed on Π_n and $[0,1]$, respectively.

Now let $\mathfrak{u}$ be a non-empty subset of $\{1,\dots d\}$, and j in $\{1,\dots n\}$, and consider

$$\dot{\mathbf{Z}}^j_{\mathfrak{u}} = (\dot{X}^j_1, \dots, \dot{X}^j_{2d-|\mathfrak{u}|}) \tag{16}$$

such that $(\dot{\mathbf{Z}}^j_{\mathfrak{u}})_j \sim \mathcal{LH}(n, 2d-|\mathfrak{u}|)$ and denote

$$\begin{aligned} \dot{\mathbf{X}}^j_{\mathfrak{u}} &= (\dot{X}^j_1, \dots, \dot{X}^j_{|\mathfrak{u}|}) \\ \dot{\mathbf{X}}^{j,1}_{\mathfrak{u}^c} &= (\dot{X}^j_{|\mathfrak{u}|+1}, \dots, \dot{X}^j_d) \\ \dot{\mathbf{X}}^{j,2}_{\mathfrak{u}^c} &= (\dot{X}^j_{d+1}, \dots, \dot{X}^j_{2d-|\mathfrak{u}|}) \end{aligned} \tag{17}$$

so that $(\dot{\mathbf{X}}^j_{\mathfrak{u}})_j \sim \mathcal{LH}(n, |\mathfrak{u}|)$ and $(\dot{\mathbf{X}}^{j,1}_{\mathfrak{u}^c})_j$, $(\dot{\mathbf{X}}^{j,2}_{\mathfrak{u}^c})_j \sim \mathcal{LH}(n, d-|\mathfrak{u}|)$. Finally, for $k=1$ and 2, we denote

$$\dot{Y}^{j,k}_{\mathfrak{u}} = f(\dot{\mathbf{X}}^j_{\mathfrak{u}}, \dot{\mathbf{X}}^{j,k}_{\mathfrak{u}^c}) \ . \tag{18}$$

As in the previous section, we consider the estimators defined in (9) and (10) but we now replace the simple random sample $(\mathbf{Z}^j_{\mathfrak{u}})_{j=1..n}$ by the stratified sample $(\dot{\mathbf{Z}}^j_{\mathfrak{u}})_{j=1..n}$. The resulting estimators are now denoted $\widetilde{\underline{S}}^{LHS}_{\mathfrak{u},n} = \widetilde{\underline{\tau}}^{2,LHS}_{\mathfrak{u},n} / \widetilde{\sigma}^{2,LHS}_n$ and $\widehat{\underline{S}}^{LHS}_{\mathfrak{u},n} = \widehat{\underline{\tau}}^{2,LHS}_{\mathfrak{u},n} / \widehat{\sigma}^{2,LHS}_n$, respectively.

3.2 Statistical properties of the estimators

The statistical properties of $\widetilde{\underline{S}}^{LHS}_{\mathfrak{u},n}$ and $\widehat{\underline{S}}^{LHS}_{\mathfrak{u},n}$ are gathered in the following result:

Proposition 1.

(i) *If f^4 is integrable then $\widetilde{\underline{S}}^{LHS}_{\mathfrak{u},n}$ et $\widehat{\underline{S}}^{LHS}_{\mathfrak{u},n}$ are strongly consistent.*

(ii) *If f^6 is integrable then $\sqrt{n}(\widetilde{\underline{S}}^{LHS}_{\mathfrak{u},n} - \underline{S}_{\mathfrak{u}})$ and $\sqrt{n}(\widehat{\underline{S}}^{LHS}_{\mathfrak{u},n} - \underline{S}_{\mathfrak{u}})$ converge in law to a zero-mean normal distribution with lower variance than the respective variance given in the central limit theorem (CLT)*

for the basic estimators $\widetilde{S}_{\mathrm{u},n}$ *and* $\widehat{S}_{\mathrm{u},n}$.

(iii) *We have*

$$\mathbb{E}\big[\widetilde{\tau}^{2,LHS}_{\mathrm{u},n}\big] = \underline{\tau}^2_{\mathrm{u}} + B_{n,1} \tag{19}$$

$$\mathbb{E}\big[\widetilde{\sigma}^{2,LHS}_{n}\big] = \sigma^2 + B_{n,2} \tag{20}$$

$$\mathbb{E}\big[\widehat{\tau}^{2,LHS}_{\mathrm{u},n}\big] = \underline{\tau}^2_{\mathrm{u}} + B_{n,3} \tag{21}$$

$$\mathbb{E}\big[\widehat{\sigma}^{2,LHS}_{n}\big] = \sigma^2 + B_{n,3} \tag{22}$$

where

$$-\frac{1}{n-1}\underline{\tau}^2_{\mathrm{u}} \leq B_{n,1} \leq 0 \tag{23}$$

$$-\frac{1}{n-1}\sigma^2 \leq B_{n,2} \leq 0 \tag{24}$$

$$-\frac{1}{2(n-1)}(\sigma^2 + \underline{\tau}^2_{\mathrm{u}}) \leq B_{n,3} \leq 0\ . \tag{25}$$

Proof.

(i) This is a consequence of the strong law of large numbers for Latin hypercube sampling given in Theorem 3 in Loh (1996).

(ii) The proof consists in translating the original proof, given for simple random sampling — see Proposition 2.2 in Janon et al. (2012) — for Latin hypercube sampling. Concerning $\widetilde{S}^{LHS}_{\mathrm{u},n}$, it is easy to show that

$$\widetilde{S}^{LHS}_{\mathrm{u},n} = \Phi(\overline{\mathbf{V}}_n) \tag{26}$$

where

$$\overline{\mathbf{V}}_n = \sum_{j=1}^{n} \mathbf{V}_j \tag{27}$$

$$\mathbf{V}_j = \Big(\big(\dot{Y}^{j,1}_{\mathrm{u}} - \mathbb{E}[Y]\big)\big(\dot{Y}^{j,2}_{\mathrm{u}} - \mathbb{E}[Y]\big)\ ,\ \dot{Y}^{j,1}_{\mathrm{u}} - \mathbb{E}[Y]\ ,\ \dot{Y}^{j,2}_{\mathrm{u}} - \mathbb{E}[Y]\ ,\ \big(\dot{Y}^{j,1}_{\mathrm{u}} - \mathbb{E}[Y]\big)^2\Big)^T \tag{28}$$

and

$$\Phi(x,y,z,t) = \frac{x - yz}{t - y^2}\ . \tag{29}$$

Then we deduce from Theorem 2 in Loh (1996) that

$$\sqrt{n}\big(\overline{\mathbf{V}}_n - \mu\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_4(0,\Gamma) \tag{30}$$

where $\mu = (\underline{\tau}^2_{\mathrm{u}}, 0, 0, \sigma^2)^T$ and Γ is the covariance matrix of $\mathbf{R}_1 = \mathbf{V}_1 - \mathbf{A}_1$ — see details in Eq. (3) in Loh (1996) — defined by

$$\forall i \in \{1,\ldots,4\},\ A_{1i} \text{ is the additive part — see (2) — of } V_{1i}\ . \tag{31}$$

Thus the Delta method — see Theorem 3.1 in Van der Vaart (1998) — gives

$$\sqrt{n}\big(\widetilde{\underline{S}}_{\mathrm{u},n}^{LHS} - \underline{S}_{\mathrm{u}}\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_1(0, g^T\Gamma g) \tag{32}$$

where $g = \nabla\Phi(\mu)$. Developing the term $g^T\Gamma g$ does not seem to provide any useful information. However, denoting σ^2_{LHS} this term, and σ^2_{IID} the analogous quantity in the CLT for simple random sampling, we can show that $\sigma^2_{LHS} \leq \sigma^2_{IID}$. Indeed we first note that, for simple random sampling, the variance given in Janon et al. (2012) reads

$$\sigma^2_{IID} = \frac{\mathrm{Var}\big[V_{11} - \underline{S}_{\mathrm{u}} V_{14}\big]}{\sigma^2} \tag{33}$$

and for Latin hypercube sampling, it is easy to show that

$$\sigma^2_{LHS} = \frac{\mathrm{Var}\big[R_{11} - \underline{S}_{\mathrm{u}} R_{14}\big]}{\sigma^2} \ . \tag{34}$$

Hence,

$$\sigma^2_{IID} = \sigma^2_{LHS} + \frac{\mathrm{Var}\big[A_{11} - \underline{S}_{\mathrm{u}} A_{14}\big]}{\sigma^2} \tag{35}$$

and the conclusion of (ii) for $\widetilde{\underline{S}}_{\mathrm{u},n}^{LHS}$ follows. Concerning $\widehat{\underline{S}}_{\mathrm{u},n}^{LHS}$, the proof follows the same lines — see Proof of (10) in Janon et al. (2012) for details.

(iii) First we have

$$\begin{aligned}
\mathbb{E}\big[\widetilde{\underline{\tau}}_{\mathrm{u},n}^{2,LHS}\big] &= \frac{n-1}{n^2}\sum_{j=1}^{n}\mathbb{E}\big[\dot{Y}_{\mathrm{u}}^{j,1}\dot{Y}_{\mathrm{u}}^{j,2}\big] - \frac{1}{n^2}\sum_{j=1}^{n}\sum_{\substack{l=1\\ l\neq j}}^{n}\mathbb{E}\big[\dot{Y}_{\mathrm{u}}^{j,1}\dot{Y}_{\mathrm{u}}^{l,2}\big] && (36)\\
&= \frac{n-1}{n}\Big(\mathbb{E}[Y]^2 + \underline{\tau}_{\mathrm{u}}^2\Big) - \frac{n-1}{n}\Big(\mathrm{Cov}(\dot{Y}_{\mathrm{u}}^{1,1}, \dot{Y}_{\mathrm{u}}^{2,2}) + \mathbb{E}[Y]^2\Big) && (37)
\end{aligned}$$

and thanks to Lemma 4 in Appendix A, it gives

$$\mathbb{E}\big[\widetilde{\underline{\tau}}_{\mathrm{u},n}^{2,LHS}\big] = \underline{\tau}_{\mathrm{u}}^2 + B_{n,1} \tag{38}$$

with

$$-\frac{1}{n-1}\underline{\tau}_{\mathrm{u}}^2 \leq B_{n,1} \leq 0 \ . \tag{39}$$

Concerning $\widetilde{\sigma}_n^{2,LHS}$, we have

$$\begin{aligned}
\mathbb{E}\big[\widetilde{\sigma}_{\underline{\mathrm{u}},n}^{2,LHS}\big] &= \frac{n-1}{n^2}\sum_{j=1}^{n}\mathbb{E}\big[(\dot{Y}_{\mathrm{u}}^{j,1})^2\big] - \frac{1}{n^2}\sum_{j=1}^{n}\sum_{\substack{l=1\\ l\neq j}}^{n}\mathbb{E}\big[\dot{Y}_{\mathrm{u}}^{j,1}\dot{Y}_{\mathrm{u}}^{l,1}\big] && (40)\\
&= \frac{n-1}{n}\mathbb{E}[Y^2] - \frac{n-1}{n}\Big(\mathrm{Cov}(\dot{Y}_{\mathrm{u}}^{1,1}, \dot{Y}_{\mathrm{u}}^{2,1}) + \mathbb{E}[Y]^2\Big) && (41)
\end{aligned}$$

and noting that

$$\mathrm{Cov}(\dot{Y}_{\mathrm{u}}^{1,1}, \dot{Y}_{\mathrm{u}}^{2,1}) = \mathrm{Cov}\big(\dot{Y}_{\{1,\ldots,d\}}^{1,1}, \dot{Y}_{\{1,\ldots,d\}}^{2,2}\big) \tag{42}$$

we conclude that

$$\mathbb{E}\big[\widetilde{\sigma}_n^{2,LHS}\big] = \sigma^2 + B_{n,2} \tag{43}$$

with

$$-\frac{1}{n-1}\sigma^2 \leq B_{n,2} \leq 0\ . \tag{44}$$

As for $\widehat{\underline{\tau}}^{2,LHS}_{\mathrm{u},n}$ and $\widehat{\sigma}^{2,LHS}_n$, we have

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\dot{Y}_{\mathrm{u}}^{j,1}+\dot{Y}_{\mathrm{u}}^{j,2}}{2}\right)^2\right] \tag{45}$$

$$= \frac{1}{4n}\mathbb{E}\big[(\dot{Y}_{\mathrm{u}}^{1,1}+\dot{Y}_{\mathrm{u}}^{1,2})^2\big] + \frac{1}{4n^2}\sum_{j=1}^{n}\sum_{\substack{l=1\\l\neq j}}^{n}\mathbb{E}\big[(\dot{Y}_{\mathrm{u}}^{j,1}+\dot{Y}_{\mathrm{u}}^{j,2})(\dot{Y}_{\mathrm{u}}^{l,1}+\dot{Y}_{\mathrm{u}}^{l,2})\big] \tag{46}$$

$$= \frac{1}{2n}\Big(\mathbb{E}[Y^2]+\underline{\tau}_{\mathrm{u}}^2+\mathbb{E}[Y]^2\Big) + \frac{n-1}{n}\mathbb{E}[Y]^2 + \frac{n-1}{2n}\Big(\mathrm{Cov}(\dot{Y}_{\mathrm{u}}^{1,1},\dot{Y}_{\mathrm{u}}^{2,1})+\mathrm{Cov}(\dot{Y}_{\mathrm{u}}^{1,1},\dot{Y}_{\mathrm{u}}^{2,2})\Big) \tag{47}$$

$$= \frac{1}{2n}\Big(\sigma^2+\underline{\tau}_{\mathrm{u}}^2\Big) + \mathbb{E}[Y]^2 + \frac{n-1}{2n}\Big(\mathrm{Cov}(\dot{Y}_{\mathrm{u}}^{1,1},\dot{Y}_{\mathrm{u}}^{2,1})+\mathrm{Cov}(\dot{Y}_{\mathrm{u}}^{1,1},\dot{Y}_{\mathrm{u}}^{2,2})\Big)\ . \tag{48}$$

Then it is easy to conclude that

$$\mathbb{E}\big[\widehat{\underline{\tau}}^{2,LHS}_{\mathrm{u},n}\big] = \underline{\tau}_{\mathrm{u}}^2 + B_{n,3} \tag{49}$$

$$\mathbb{E}\big[\widehat{\sigma}^{2,LHS}_n\big] = \sigma^2 + B_{n,3} \tag{50}$$

with

$$-\frac{1}{2(n-1)}(\sigma^2+\underline{\tau}_{\mathrm{u}}^2) \leq B_{n,3} \leq 0\ . \tag{51}$$

□

Remark 1. *Due to their intricate structure, the biases of the estimators* $\widetilde{\underline{\tau}}^{2,LHS}_{\mathrm{u},n}$, $\widetilde{\sigma}^{2,LHS}_n$, $\widehat{\underline{\tau}}^{2,LHS}_{\mathrm{u},n}$ *and* $\widehat{\sigma}^{2,LHS}_n$ *can't be easily reduced. Nevertheless we can note that these biases are asymptotically negligible, with a rate of convergence in* $O(n^{-1})$ *larger than the rate of convergence of the estimators — to their theoretical values — themselves, which is in* $O(n^{-1/2})$.

4 Monte Carlo estimators and replicated Latin hypercube sampling

4.1 Notation and definitions

We begin with the definition of replicated Latin hypercubes:

Definition 2. *Let* d *and* n *in* $\mathbb{N}^*$, *and consider* Π_n *the set of all the permutations of* $\{1,\ldots,n\}$. *We say that* $(\mathbf{X}^j)_{j=1..n}$ *and* $(\mathbf{X}'^j)_{j=1..n}$ *are two replicated Latin hypercubes of size* n *in* $[0,1]^d$ *— and we denote* $(\mathbf{X}^j,\mathbf{X}'^j)_j \sim \mathcal{RLH}(n,d)$ *— if for all* $j\in\{1,\ldots,n\}$,

$$\mathbf{X}^j = \left(\frac{\pi_1(j)-U_{1,\pi_1(j)}}{n},\ldots,\frac{\pi_d(j)-U_{d,\pi_d(j)}}{n}\right) \tag{52}$$

and

$$\mathbf{X}'^{j} = \left(\frac{\pi'_1(j) - U_{1,\pi'_1(j)}}{n}, \ldots, \frac{\pi'_d(j) - U_{d,\pi'_d(j)}}{n} \right) \tag{53}$$

where the π_i*'s, the* π'_i*'s and the* $U_{i,j}$*'s are independent random variables uniformly distributed on* Π_n*,* Π_n *and* $[0,1]$*, respectively.*

Now let $\mathfrak{u}$ be a non-empty subset of $\{1, \ldots d\}$, and j in $\{1, \ldots n\}$, and consider

$$\ddot{\mathbf{Z}}^j_{\mathfrak{u}} = (\ddot{X}^j_1, \ldots, \ddot{X}^j_{2d-|\mathfrak{u}|}) \tag{54}$$

and

$$\ddot{\mathbf{X}}^j_{\mathfrak{u}} = (\ddot{X}^j_1, \ldots, \ddot{X}^j_{|\mathfrak{u}|}) \tag{55}$$

$$\ddot{\mathbf{X}}^{j,1}_{\mathfrak{u}^c} = (\ddot{X}^j_{|\mathfrak{u}|+1}, \ldots, \ddot{X}^j_d) \tag{56}$$

$$\ddot{\mathbf{X}}^{j,2}_{\mathfrak{u}^c} = (\ddot{X}^j_{d+1}, \ldots, \ddot{X}^j_{2d-|\mathfrak{u}|}) \tag{57}$$

where $(\ddot{\mathbf{X}}^j_{\mathfrak{u}})_j \sim \mathcal{LH}(n, |\mathfrak{u}|)$ and $(\ddot{\mathbf{X}}^{j,1}_{\mathfrak{u}^c}, \ddot{\mathbf{X}}^{j,2}_{\mathfrak{u}^c})_j \sim \mathcal{RLH}(n, d - |\mathfrak{u}|)$, $(\ddot{\mathbf{X}}^j_{\mathfrak{u}})_j$ and $(\ddot{\mathbf{X}}^{j,1}_{\mathfrak{u}^c}, \ddot{\mathbf{X}}^{j,2}_{\mathfrak{u}^c})_j$ being independent. Finally, for $k = 1$ and 2, we denote

$$\ddot{Y}^{j,k}_{\mathfrak{u}} = f(\ddot{\mathbf{X}}^j_{\mathfrak{u}}, \ddot{\mathbf{X}}^{j,k}_{\mathfrak{u}^c}) \ . \tag{58}$$

As in Section 2, we consider the estimators defined in (9) and (10) but we now replace the simple random sample $(\mathbf{Z}^j_{\mathfrak{u}})_{j=1..n}$ by the stratified sample based on replicated Latin hypercubes $(\ddot{\mathbf{Z}}^j_{\mathfrak{u}})_{j=1..n}$. The resulting estimators are now denoted $\widetilde{S}^{RLHS}_{\mathfrak{u},n} = \widetilde{\tau}^{2,RLHS}_{\mathfrak{u},n}/\widetilde{\sigma}^{2,RLHS}_n$ and $\widehat{S}^{RLHS}_{\mathfrak{u},n} = \widehat{\tau}^{2,RLHS}_{\mathfrak{u},n}/\widehat{\sigma}^{2,RLHS}_n$, respectively. Note that estimators of Sobol' indices based on r replicated Latin hypercubes have already be introduced by McKay (1995) (see also the summarized presentation by Saltelli et al. (2000)), but these estimators converge to their corresponding analytical Sobol' index only as r tends to $+\infty$.

4.2 Statistical properties of the estimators

The statistical properties of $\widetilde{S}^{RLHS}_{\mathfrak{u},n}$ and $\widehat{S}^{RLHS}_{\mathfrak{u},n}$ are gathered in the following result:

Proposition 2.

(i) *If* f^4 *is integrable then* $\widetilde{S}^{RLHS}_{\mathfrak{u},n}$ *and* $\widehat{S}^{RLHS}_{\mathfrak{u},n}$ *are strongly consistent.*

(ii) *If* f^6 *is integrable then* $\sqrt{n}(\widetilde{S}^{RLHS}_{\mathfrak{u},n} - \underline{S}_{\mathfrak{u}})$ *and* $\sqrt{n}(\widehat{S}^{RLHS}_{\mathfrak{u},n} - \underline{S}_{\mathfrak{u}})$ *converge in law to a zero-mean normal distribution with the same respective variance given in CLT for the estimators* $\widetilde{S}^{LHS}_{\mathfrak{u},n}$ *and* $\widetilde{S}^{LHS}_{\mathfrak{u},n}$

.

(iii) *We have*

$$\mathbb{E}\big[\widetilde{\underline{\tau}}^{2,RLHS}_{\mathfrak{u},n}\big] = \underline{\tau}^2_{\mathfrak{u}} - \frac{1}{n}\underline{\tau}^2_{\mathfrak{u}} + B_{n,1} + B_{|\mathfrak{u}|,n} \tag{59}$$

$$\mathbb{E}\big[\widetilde{\sigma}^{2,RLHS}_{n}\big] = \sigma^2 + B_{n,3} \tag{60}$$

$$\mathbb{E}\big[\widehat{\underline{\tau}}^{2,RLHS}_{\mathfrak{u},n}\big] = \underline{\tau}^2_{\mathfrak{u}} - \frac{1}{2n}\underline{\tau}^2_{\mathfrak{u}} + B_{n,1} + B_{n,2} + B_{|\mathfrak{u}|,n} \tag{61}$$

$$\mathbb{E}\big[\widehat{\sigma}^{2,RLHS}_{n}\big] = \sigma^2 - \frac{1}{2n}\underline{\tau}^2_{\mathfrak{u}} + B_{n,1} + B_{n,2} \tag{62}$$

where

$$|B_{n,1}| \leq \left(\frac{d+1}{n} + 2\right)\left(\frac{d+1}{n}\right)\mathbb{E}[Y^2] \tag{63}$$

$$|B_{n,2}| \leq \frac{\sigma^2}{2n} \tag{64}$$

$$-\frac{1}{n-1}\sigma^2 \leq B_{n,3} \leq 0 \tag{65}$$

$$|B_{|\mathfrak{u}|,n}| \leq \left(\frac{d-|\mathfrak{u}|+1}{n} + 2\right)\left(\frac{d-|\mathfrak{u}|+1}{n-1}\right)\mathbb{E}[Y^2]\ . \tag{66}$$

Proof.

(i) The proof is divided into two parts. In the first one, we only consider continuous functions, and in the second one, we extend the result to the whole class of functions such that f^4 is integrable.

First part: Consistency is obvious as in Proposition 1, except for the term

$$\frac{1}{n}\sum_{j=1}^{n} \ddot{Y}^{j,1}_{\mathfrak{u}}\ddot{Y}^{j,2}_{\mathfrak{u}}\ . \tag{67}$$

So denote $\overline{\mathbf{X}}^{j,2}_{\mathfrak{u}^c}$ the Latin hypercube defined by

$$\overline{\mathbf{X}}^{j,2}_{\mathfrak{u}^c} = \frac{\lfloor n\ddot{\mathbf{X}}^{j,2}_{\mathfrak{u}^c}\rfloor + \mathbf{U}_j}{n} \tag{68}$$

where the $\mathbf{U}_j$'s are independent random vectors uniformly distributed in $[0,1[^{d-|\mathfrak{u}|}$ independent from all the permutations and shifts in the definition of $(\ddot{\mathbf{Z}}^j_{\mathfrak{u}})_j$, and $\lfloor\cdot\rfloor$ is the floor function. We can write

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n} \ddot{Y}^{j,1}_{\mathfrak{u}}\ddot{Y}^{j,2}_{\mathfrak{u}} &= \frac{1}{n}\sum_{j=1}^{n} f(\ddot{\mathbf{X}}^j_{\mathfrak{u}}, \ddot{\mathbf{X}}^{j,1}_{\mathfrak{u}^c})f(\ddot{\mathbf{X}}^j_{\mathfrak{u}}, \ddot{\mathbf{X}}^{j,2}_{\mathfrak{u}^c}) \\
&= \frac{1}{n}\sum_{j=1}^{n} f(\ddot{\mathbf{X}}^j_{\mathfrak{u}}, \ddot{\mathbf{X}}^{j,1}_{\mathfrak{u}^c})f(\ddot{\mathbf{X}}^j_{\mathfrak{u}}, \overline{\mathbf{X}}^{j,2}_{\mathfrak{u}^c}) \\
&+ \frac{1}{n}\sum_{j=1}^{n} f(\ddot{\mathbf{X}}^j_{\mathfrak{u}}, \ddot{\mathbf{X}}^{j,1}_{\mathfrak{u}^c})\Big(f(\ddot{\mathbf{X}}^j_{\mathfrak{u}}, \ddot{\mathbf{X}}^{j,2}_{\mathfrak{u}^c}) - f(\ddot{\mathbf{X}}^j_{\mathfrak{u}}, \overline{\mathbf{X}}^{j,2}_{\mathfrak{u}^c})\Big)\ .
\end{aligned} \tag{69}$$

The first term on the right-hand side is an estimator as described in Section 3 since we note that $(\ddot{\mathbf{X}}^j_{\mathfrak{u}}, \ddot{\mathbf{X}}^{j,1}_{\mathfrak{u}^c}, \overline{\mathbf{X}}^{j,2}_{\mathfrak{u}^c})_{j=1..n} \sim \mathcal{LH}(n, 2d-|\mathfrak{u}|)$; so it converges to $\mathbb{E}[Y]^2 + \underline{\tau}^2_{\mathfrak{u}}$ almost surely. The second term on the right-hand side converges to 0 since as f is bounded — by continuity on a compact — it is bounded by

$$\frac{\sup|f|}{n}\sum_{j=1}^{n}\Big|f(\ddot{\mathbf{X}}^j_{\mathfrak{u}}, \ddot{\mathbf{X}}^{j,2}_{\mathfrak{u}^c}) - f(\ddot{\mathbf{X}}^j_{\mathfrak{u}}, \overline{\mathbf{X}}^{j,2}_{\mathfrak{u}^c})\Big| \tag{70}$$

and by uniform continuity of f — due to Heine-Cantor theorem — this quantity tends to 0 as n tends to $+\infty$. Thus the sum in the right-hand side, i.e. $\frac{1}{n}\sum_{j=1}^{n} \ddot{Y}_{\mathfrak{u}}^{j,1}\ddot{Y}_{\mathfrak{u}}^{j,2}$, converges to $\mathbb{E}[Y]^2+\underline{\tau}_{\mathfrak{u}}^2$ almost surely.

Second part: Since the space of continuous functions on $[0,1]^d$ — denoted $\mathcal{C}\big([0,1]^d\big)$ — is dense in $L^4\big([0,1]^d\big)$, let $(f_m)_{m\in\mathbb{N}^*}$ be a sequence in $\mathcal{C}\big([0,1]^d\big)$ such that $\mathbb{E}\big[|f_m(\mathbf{X})-f(\mathbf{X})|^4\big]$ converges to 0 as m tends to $+\infty$, where $\mathbf{X}$ is uniformly distributed on $[0,1]^d$.

Now let $\epsilon>0$ and $M=M(\epsilon)\in\mathbb{N}^*$ such that

$$\mathbb{E}\Big[\big(f_M(\mathbf{X})-f(\mathbf{X})\big)^2\Big] < \frac{\epsilon^2}{65\ \mathbb{E}[f^2(\mathbf{X})]}. \tag{71}$$

We can write

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n} \ddot{Y}_{\mathfrak{u}}^{j,1}\ddot{Y}_{\mathfrak{u}}^{j,2} &= \frac{1}{n}\sum_{j=1}^{n} f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\overline{\mathbf{X}}_{\mathfrak{u}^c}^{j,2}) \\
&+ \frac{1}{n}\sum_{j=1}^{n} f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})\Big(f_M(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\overline{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})\Big) \\
&+ \frac{1}{n}\sum_{j=1}^{n} f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})\Big(f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})\Big) \\
&+ \frac{1}{n}\sum_{j=1}^{n} f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})\Big(f_M(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\overline{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\overline{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})\Big)
\end{aligned} \tag{72}$$

As noted in the proof of (i) in Proposition 1, the first term on the right-hand side of (72) converges to $\underline{\tau}_{\mathfrak{u}}^2+\mathbb{E}[Y]^2$ almost surely as n tends to $+\infty$ i.e.

$$\mathbb{P}\Big(\forall\varepsilon>0,\ \exists N_1\in\mathbb{N}^*,\ \forall n>N_1,\ \Big|\frac{1}{n}\sum_{j=1}^{n} f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\overline{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})-\underline{\tau}_{\mathfrak{u}}^2-\mathbb{E}[Y]^2\Big|<\frac{\varepsilon}{4}\Big)=1\ . \tag{73}$$

Since f_M is uniformly continuous on $[0,1]^d$, we have that

$$A_n=\sup_{1\le j\le n}\big|f_M(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\overline{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})\big| \tag{74}$$

converges almost surely to 0 as n tends to $+\infty$. Moreover, since f is integrable, we have that $\frac{1}{n}\sum_{j=1}^{n}|f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})|$ converges to $\mathbb{E}\big[|Y|\big]$ as n tends to $+\infty$. Hence

$$\mathbb{P}\Big(\forall\varepsilon>0,\ \exists N_1\in\mathbb{N}^*,\ \forall n>N_1,\ \Big|\frac{1}{n}\sum_{j=1}^{n} f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})\Big(f_M(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\overline{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})\Big)\Big|<\frac{\varepsilon}{4}\Big) \tag{75}$$

$$\begin{aligned}
&\ge \mathbb{P}\Big(\forall\varepsilon>0,\ \exists N_2\in\mathbb{N}^*,\ \forall n>N_2,\ A_n\frac{1}{n}\sum_{j=1}^{n}\big|f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})\big|<\frac{\varepsilon}{4}\Big) \\
&= 1\ .
\end{aligned} \tag{76}$$

For the third and the fourth terms on the right-hand side of (72), we apply twice the same proof. First the Cauchy-Schwartz inequality gives

$$\mathbb{P}\Big(\forall\varepsilon>0,\ \exists N_3\in\mathbb{N}^*,\ \forall n>N_3,\ \Big|\frac{1}{n}\sum_{j=1}^{n} f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})\Big(f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})\Big)\Big|<\frac{\varepsilon}{4}\Big) \tag{77}$$

$$\geq \mathbb{P}\Big(\forall\varepsilon>0,\ \exists N_3\in\mathbb{N}^*,\ \forall n>N_3,\ \Big(\frac{1}{n}\sum_{j=1}^{n} f^2(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})\Big)^{1/2}\Big(\frac{1}{n}\sum_{j=1}^{n}\big(f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})\big)^2\Big)^{1/2}<\frac{\varepsilon}{4}\Big). \tag{78}$$

Then note that $\frac{1}{n}\sum_{j=1}^{n} f^2(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})$ and $\frac{1}{n}\sum_{j=1}^{n}\big(f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})\big)^2$ converge almost surely to $\mathbb{E}[Y^2]$ and $\mathbb{E}[(f_M(\mathbf{X})-f(\mathbf{X}))^2]$ — where $\mathbf{X}$ is uniformly distributed on $[0,1]^d$ — respectively. And deduce that there exists $N_4\in\mathbb{N}^*$ such that for all $n>N_4$, we have $\frac{1}{n}\sum_{j=1}^{n} f^2(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})<2\ \mathbb{E}[Y^2]$ and $\frac{1}{n}\sum_{j=1}^{n}\big(f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})\big)^2<2\ \mathbb{E}[(f_M(\mathbf{X})-f(\mathbf{X}))^2]$ almost surely. As a consequence, deduce from Eq. (71) that

$$\mathbb{P}\Big(\forall\varepsilon>0,\ \exists N_3\in\mathbb{N}^*,\ \forall n>N_3,\ \Big|\frac{1}{n}\sum_{j=1}^{n} f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})\Big(f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})\Big)\Big|<\frac{\varepsilon}{4}\Big) \tag{79}$$

$$\begin{aligned}&\geq \mathbb{P}\Big(\forall\varepsilon>0,\ \exists N_3>N_4,\ \forall n>N_3,\ \varepsilon\sqrt{\frac{4}{65}}<\frac{\varepsilon}{4}\Big)\\ &= 1\end{aligned} \tag{80}$$

Finally, Eqs. (73–80) gives

$$\mathbb{P}\Big(\forall\varepsilon>0,\ \exists N\in\mathbb{N}^*,\ \forall n>N,\ \Big|\frac{1}{n}\sum_{j=1}^{n}\ddot{Y}_{\mathfrak{u}}^{j,1}\ddot{Y}_{\mathfrak{u}}^{j,2}\Big|<\varepsilon\Big)=1 \tag{81}$$

and we have the conclusion.

(ii) As in (i), the only term to treat is

$$\frac{1}{n}\sum_{j=1}^{n}\ddot{Y}_{\mathfrak{u}}^{j,1}\ddot{Y}_{\mathfrak{u}}^{j,2}\ , \tag{82}$$

so asymptotic normality is shown in the same way by using the decomposition in (69). We always obtain the sum of a term already considered in Section 3 which converges in law to a normal distribution and a term which converges to 0 in probability, and the conclusion follows from Slutsky's lemma. We only detail the proof for $\underline{\widetilde{S}}_{\mathfrak{u},n}^{RLHS}$, it is exactly the same for $\underline{\widehat{S}}_{\mathfrak{u},n}^{RLHS}$. So note that following the proof of (ii) in Proposition 1 and the notation above, it is sufficient to show that

$$\sqrt{n}\Big(\frac{1}{n}\sum_{j=1}^{n}\big(f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})-\mathbb{E}[Y]\big)\big(f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\overline{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})\big)\Big)\xrightarrow[n\to\infty]{\mathbb{P}}0 \tag{83}$$

to prove the asymptotic normality of $\underline{\widetilde{S}}_{\mathfrak{u},n}^{RLHS}$.

So consider $\varepsilon,\eta>0$ and prove that there exists $N\in\mathbb{N}^*$ such that for all $n>N$, the quantity

$$P=\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{j=1}^{n}\big(f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})-\mathbb{E}[Y]\big)\big(f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\overline{\mathbf{X}}_{\mathfrak{u}^c}^{j,2})\big)\Big|>\varepsilon\Big) \tag{84}$$

is less than η. First as f^6 is integrable, there exists a constant $K>0$ such that $\mathbb{P}(|f(\ddot{\mathbf{X}}_{\mathfrak{u}}^{j},\ddot{\mathbf{X}}_{\mathfrak{u}^c}^{j,1})|>K)<$

$\eta/4$. Hence

$$\begin{aligned} P &\leq \mathbb{P}\Big(\Big(\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\big(f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,1})-\mathbb{E}[Y]\big)\big(f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})\big)\Big|>\varepsilon\Big)\bigcap\big(|f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,1})|\leq K\big)\Big) \\ &+ \mathbb{P}\Big(\Big(\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\big(f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,1})-\mathbb{E}[Y]\big)\big(f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})\big)\Big|>\varepsilon\Big)\bigcap\big(|f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,1})|>K\big)\Big) \\ &< \mathbb{P}\Big(\frac{K+|\mathbb{E}[Y]|}{\sqrt{n}}\sum_{j=1}^{n}\big|f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})\big|>\varepsilon\Big)+\frac{\eta}{4}. \end{aligned} \tag{85}$$

Now note that the space of continuous functions on $[0,1]^d$, denoted by $\mathcal{C}([0,1]^d)$, is dense in $L^6([0,1]^d)$ and let $(f_m)_{m\in\mathbb{N}^*}$ be a sequence in $\mathcal{C}([0,1]^d)$ such that $\mathbb{E}[|f_m(\mathbf{X})-f(\mathbf{X})|^6]$ converges to 0 as m tends to $+\infty$ where $\mathbf{X}$ is uniformly distributed on $[0,1]^d$. It is easy to note that there exists $M=M(n)$ such that $\mathbb{P}(|f_M(\mathbf{X})-f(\mathbf{X})|>1/n)<\eta/4$. Thus we get from Eq. (85) that

$$\begin{aligned} P &< \sum_{i=1}^{4}\mathbb{P}\Big(\Big(\frac{K+|\mathbb{E}[Y]|}{\sqrt{n}}\sum_{j=1}^{n}\big(|f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})|+|f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})| \\ &+ |f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})|\big)\Big)\bigcap A_i\Big)+\frac{\eta}{4} \end{aligned} \tag{86}$$

where

$$A_1 = \big(|f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})|>\frac{1}{n}\big)\cap\big(|f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})|>\frac{1}{n}\big) \tag{87}$$

$$A_2 = \big(|f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})|>\frac{1}{n}\big)\cap\big(|f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})|<\frac{1}{n}\big) \tag{88}$$

and A_3 and A_4 are the complementary events of A_1 and A_2, respectively. So we deduce

$$\begin{aligned} P &< \mathbb{P}\Big(\Big(\frac{K+|\mathbb{E}[Y]|}{\sqrt{n}}\sum_{j=1}^{n}\big(|f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})|+|f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})| \\ &+ |f_M(\overline{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})|\big)\Big)\bigcap A_3\Big)+\mathbb{P}(A_1)+\mathbb{P}(A_2)+\mathbb{P}(A_4)+\frac{\eta}{4} \\ &< \mathbb{P}\Big(\frac{K+|\mathbb{E}[Y]|}{\sqrt{n}}\Big(2+\sum_{j=1}^{n}|f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})|\Big)>\varepsilon\Big)+\eta. \end{aligned} \tag{89}$$

Now by an other density argument, note that there exists a sequence of Lipschitz continuous functions with constant 1, denoted $(f_{M,q})_{q\in\mathbb{N}^*}$, such that $\sup_{[0,1]^d}|f_{M,q}(\mathbf{x})-f_M(\mathbf{x})|$ converges to 0 as q tends to $+\infty$. Then there exists $Q=Q(n)\in\mathbb{N}^*$ such that $\sup_{[0,1]^d}|f_{M,Q}(\mathbf{x})-f_M(\mathbf{x})|<1/n$ and deduce that

$$\begin{aligned} P &< \mathbb{P}\Big(\frac{K+|\mathbb{E}[Y]|}{\sqrt{n}}\Big(2+\sum_{j=1}^{n}\big(|f_{M,Q}(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f_{M,Q}(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})|+|f_{M,Q}(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\ddot{\mathbf{X}}_{\mathrm{u}^c}^{j,2})| \\ &+ |f_{M,Q}(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})-f_M(\ddot{\mathbf{X}}_{\mathrm{u}}^{j},\overline{\mathbf{X}}_{\mathrm{u}^c}^{j,2})|\big)\Big)>\varepsilon\Big)+\eta \\ &< \mathbb{P}\Big(\frac{5(K+|\mathbb{E}[Y]|)}{\sqrt{n}}>\varepsilon\Big)+\eta. \end{aligned} \tag{90}$$

and the conclusion follows.

(iii) The proof is given in Appendix B. □

4.3 Estimating all the first-order Sobol' indices using only two replicated Latin hypercube

First note that for any independent random permutations π_1 and π_2 uniformly distributed in Π_n, we have that π_1 and $\pi_1 \circ \pi_2$ are independent.

Then, let $(\mathbf{D}^{j,1}, \mathbf{D}^{j,2})_j \sim \mathcal{RLH}(n,d)$ be a design of experiments of $2n$ points in dimension d defined as in Definition 2. Note that, by keeping the notation used in Definition 2, for any $i \in \{1,\ldots,d\}$, we have:

(i) $(D_i^{j,1})_j \sim \mathcal{LH}(n,1)$

(ii) for all $j \in \{1,\ldots,n\}$, $D_i^{j,1} = D_i^{\pi_i'^{-1}\circ\pi_i(j),2}$

(iii) $\left(\mathbf{D}_{\{i\}^c}^{i,1}, \mathbf{D}_{\{i\}^c}^{\pi_i'^{-1}\circ\pi_i(j),2}\right)_j \sim \mathcal{RLH}(n,d-1)$

(iv) $(D_i^{i,1})_j$ and $\left(\mathbf{D}_{\{i\}^c}^{i,1}, \mathbf{D}_{\{i\}^c}^{\pi_i'^{-1}\circ\pi_i(j),2}\right)_j$ are independent.

As a consequence, we can estimate all the S_i's with the two replicated Latin hypercubes $(\mathbf{D}^{j,1}, \mathbf{D}^{j,2})_j$ by considering successively as in Eqs. (55–58), for any $i \in \{1,\ldots,d\}$ and $j \in \{1,\ldots,n\}$,

$$\ddot{X}_{\{i\}}^{j} = D_i^{i,1} = D_i^{\pi_i'^{-1}\circ\pi_i(j),2} \tag{91}$$

$$\ddot{\mathbf{X}}_{\{i\}^c}^{j,1} = \mathbf{D}_{\{i\}^c}^{i,1} \tag{92}$$

$$\ddot{\mathbf{X}}_{\{i\}^c}^{j,2} = \mathbf{D}_{\{i\}^c}^{\pi_i'^{-1}\circ\pi_i(j),2} \tag{93}$$

and for $k = 1$ and 2,

$$\ddot{Y}_{\{i\}}^{j,k} = f\left(\ddot{\mathbf{X}}_{\{i\}^c}^{j,k}\right). \tag{94}$$

4.4 Construction with Latin hypercubes based on general orthogonal arrays

We first begin with the definition of an orthogonal array (OA):

Definition 3. *An orthogonal array in dimension d, with q levels, strength $t \leq d$ and index λ is a matrix with $n = \lambda q^t$ rows and d columns such that in every n-by-t submatrix each of the q^t possible rows — i.e. the distinct t-tuples $(l_1,\ldots,l_t)$ where the l_i's take their values in the set of the q levels — occurs exactly the same number λ of times.*

We now recall the definition of OA-based Latin hypercubes — see Owen (1992) — and introduce the general notion of replicated OA-based Latin hypercubes.

Definition 4. *Let $(A_i^j)_{i=1..d,j=1..n}$ be an orthogonal array in dimension d, with n points and q levels in $\{1,\ldots,q\}$, and consider Π_q the set of all the permutations of $\{1,\ldots,q\}$. We say that $(\mathbf{X}^j)_{j=1..n}$ is a Latin hypercube based on the orthogonal array $(\mathbf{A}^j)_{j=1..n}$ — and we denote $(\mathbf{X}^j)_j \sim \mathcal{LH}\big((\mathbf{A}^j)_j\big)$ — if for all $j \in \{1,\ldots,n\}$,*

$$\mathbf{X}^j = \left(\frac{\pi_1(A_1^j) - U_{1,\pi_1(A_1^j)}}{q}, \ldots, \frac{\pi_d(A_d^j) - U_{d,\pi_d(A_d^j)}}{q}\right) \tag{95}$$

where the π_i's and the $U_{i,j}$'s are independent random variables uniformly distributed on Π_q and $[0,1]$, respectively.

Definition 5. *Let $(A_i^j)_{i=1..d,j=1..n}$ an orthogonal array in dimension d, with n points and q levels in $\{1,\ldots,q\}$, and consider Π_q the set of all the permutations of $\{1,\ldots,q\}$. We say that $(\mathbf{X}^j)_{j=1..n}$ and $(\mathbf{X}'^j)_{j=1..n}$ are two replicated Latin hypercubes based on the orthogonal array $(\mathbf{A}^j)_{j=1..n}$ — and we denote $(\mathbf{X}^j, \mathbf{X}'^j)_j \sim \mathcal{RLH}\big((\mathbf{A}^j)_j\big)$ — if for all $j \in \{1,\ldots,n\}$,*

$$\mathbf{X}^j = \left(\frac{\pi_1(A_1^j) - U_{1,\pi_1(A_1^j)}}{q}, \ldots, \frac{\pi_d(A_d^j) - U_{d,\pi_d(A_d^j)}}{q}\right) \tag{96}$$

and

$$\mathbf{X}'^j = \left(\frac{\pi'_1(A_1^j) - U_{1,\pi'_1(A_1^j)}}{q}, \ldots, \frac{\pi'_d(A_d^j) - U_{d,\pi'_d(A_d^j)}}{q}\right) \tag{97}$$

where the π_i's, the π'_i's and the $U_{i,j}$'s are independent random variables uniformly distributed on Π_q, Π_q and $[0,1]$, respectively.

Note that in the particular case of the orthogonal array $(\mathbf{A}^j)_{j=1}$ with strength 1 and index unity defined by

$$\forall i \in \{1,\ldots d\},\ \forall j \in \{1,\ldots n\},\quad A_i^j = j, \tag{98}$$

these definitions are exactly Definitions 1 and 2.

Now the designs of experiments introduced in Definition 5 allow to estimate all the $\underline{S}_{\mathfrak{u}}$'s with $|\mathfrak{u}| = t$ where t is the strength of the underlying orthogonal array of the replicated Latin hypercube. More precisely, let $(\mathbf{A}^j)_j$ be an orthogonal array with q levels in $\{1,\ldots,q\}$, strength t and index unity, and consider $(\mathbf{D}^{j,1}, \mathbf{D}^{j,2})_j \sim \mathcal{RLH}\big((\mathbf{A}^j)_j\big)$. For any $\mathfrak{u} \subseteq \{1,\ldots,d\}$, with $\mathfrak{u} = \{i_1,\ldots,i_t\}$, and any $k = 1$ or 2, define $(\mathbf{D}^{j(\mathfrak{u}),k})_j$ as the set of points $(\mathbf{D}^{j,k})_j$ ranked in increasing lexicographic order with respect to the i_1-, …, i_t-th coordinates. Then we can define estimators of the $\underline{S}_{\mathfrak{u}}$'s by considering those in Eqs. (9) and (10) and by replacing (8) by

$$Y_{\mathfrak{u}}^{j,k} = f\big(\mathbf{D}^{j(\mathfrak{u}),k}\big)\ . \tag{99}$$

Remark 2. *Theoretical properties of the estimators for this generalisation remain open issues and will consist of a further work. The first step for strong consistency will be to state a strong law of large*

numbers for OA-based Latin hypercubes with strength $t > 1$ since, as far as we know, such a result does not exist. Asymptotic normality has already been proved for OA-based Latin hypercube with strength $t = 2$ under smoothness conditions — see Loh (2008) *— but it is not sufficient to conclude in the case of replicated OA-based Latin hypercubes since formulas as in* (172) *and* (173) *are necessary. As for the biases of the estimators, it will be necessary to study covariances in OA-based Latin hypercubes with strength $t > 1$ in order to state formulas as in* (172) *and* (173) *as well.*

5 Numerical illustrations

5.1 Application to an analytical test-case

5.1.1 Main experiment

In this section, we apply the new method proposed in Section 4 to the Ishigami function, see Ishigami and Homma (1990):

$$f(X_1, X_2, X_3) = \sin(X_1) + 7\sin^2(X_2) + 0.1X_3^4\sin(X_1) \tag{100}$$

where the X_i's are independent random variable uniformly distributed on $[-\pi, \pi]$. Analytical values of Sobol' indices of this model are

$$\underline{S}_1 = 0.3139,\ \ \underline{S}_2 = 0.4424,\ \ \underline{S}_3 = 0,\ \ \underline{S}_{12} = 0.7563,\ \ \underline{S}_{23} = 0.4424,\ \ \underline{S}_{13} = 0.5575 \text{ and } \underline{S}_{123} = 1. \tag{101}$$

We are interested in comparing the new method, with the classic one based on crude Monte Carlo method and which need $d + 1$ samples to estimate all the first-order Sobol' indices, and $2d + 2$ samples to estimate all the second-order Sobol' indices, see Saltelli (2002)). Here, both methods are compared at the same sample size n in order to investigate the estimators themselves, but keep in mind that the new method is definitely more efficient since only two samples are needed to estimate all the first-order Sobol' indices or all the second-order Sobol' indices. In the experiment, we focus on the empirical coverage — i.e. the empirical proportion of confidence interval containing the analytical value of the Sobol' index — of both estimators at different sample size between 10^2 and 10^5, and for $r = 100000$ replicates. We first investigate estimators $\widehat{\underline{S}}_{\{i\},n}$ and $\widehat{\underline{S}}^{RLHS}_{\{i\},n}$, $i \in \{1, \ldots, d\}$ and in both cases, we provide asymptotic confidence intervals from the estimation of the asymptotic variance given in Janon et al. (2012) (see end of the proof of Prop. 2.2). Indeed, as we know that this asymptotic variance is:

$$\sigma^2_{IID,\mathrm{u}} = \frac{\mathrm{Var}\big[(Y^1_{\mathrm{u}} - \mathbb{E}[Y^1_{\mathrm{u}}])(Y^2_{\mathrm{u}} - \mathbb{E}[Y^1_{\mathrm{u}}]) - \underline{S}_{\mathrm{u}}/2\big((Y^1_{\mathrm{u}} - \mathbb{E}[Y^1_{\mathrm{u}}])^2\big)\big(Y^2_{\mathrm{u}} - \mathbb{E}[Y^1_{\mathrm{u}}]\big)\big]}{\mathrm{Var}[Y]^2} \geq \sigma^2_{RLHS,\mathrm{u}}, \tag{102}$$

we can provide an estimator of the asymptotic confidence interval for the classic method

$$I_{IID,\mathrm{u},\alpha} = \Big[\underline{S}_{\mathrm{u}} - \frac{\sigma^2_{IID,\mathrm{u}}\ u_{\alpha/2}}{\sqrt{n}}, \underline{S}_{\mathrm{u}} + \frac{\sigma^2_{IID,\mathrm{u}}\ u_{\alpha/2}}{\sqrt{n}}\Big]$$

and an other one for the new method

$$I_{RLHS,\mathrm{u},\alpha} = \left[\underline{S}_{\mathrm{u}} - \frac{\sigma^2_{RLHS,\mathrm{u}}\ u_{\alpha/2}}{\sqrt{n}}, \underline{S}_{\mathrm{u}} + \frac{\sigma^2_{RLHS,\mathrm{u}}\ u_{\alpha/2}}{\sqrt{n}}\right]$$

where $u_{\alpha/2}$ is the normal quantile at the significance level α. By using the estimator of the asymptotic variance given in (102) in both cases, the confidence interval lengths of the classic and the new estimators are the same. More specifically, the estimated length of the new estimator is greater or equal than its optimal value. Thus the asymptotic value of the empirical coverage of the new method is greater or equal than the expected one. However at the moment, we do not know how to estimate correctly $\sigma^2_{RLHS,\mathrm{u}}$ because of its singular expression (see Proof of (ii) in Proposition 1 in Section 3.2). We just say few words about it in the next subsection and more fundamentally, it should consist of a further work.

We also investigate estimators $\widehat{\underline{S}}_{\{i,j\},n}$ and $\widehat{\underline{S}}^{OA2-RLHS}_{\{i,j\},n}$, $i \neq j \in \{1,\ldots,d\}$, where the notation $OA2 - RLHS$ refers to the generalization to replicated latin hypercube based on orthogonal array of strength 2 presented in Section 4.4. In this case, we conjecture that the Central Limit theorem established in (ii) in Proposition 2 is also true under some smoothness assumption — note that, here, Ishigami function is $\mathcal{C}^\infty$. Results are gathered in Figures 1 to 4. For the second-order Sobol' indices,

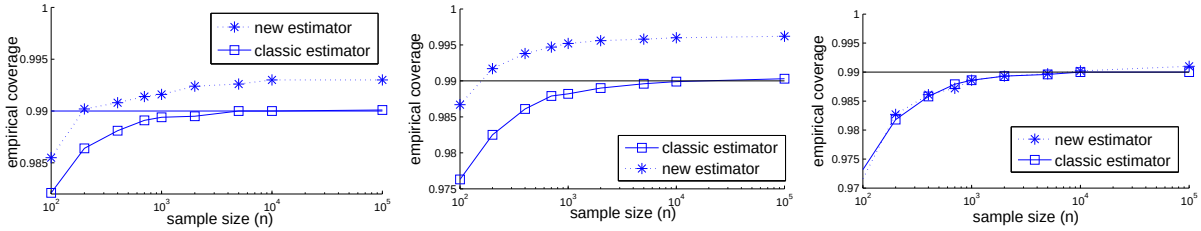

Figure 1: Empirical coverage of confidence intervals for $\underline{S}_1$ (left), $\underline{S}_2$ (center) and $\underline{S}_3$ (right).

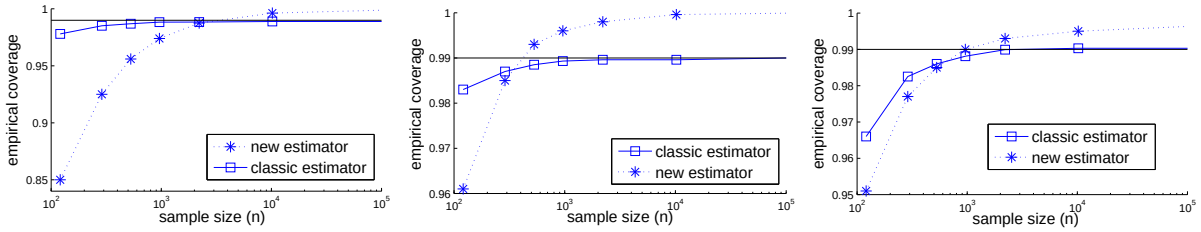

Figure 2: Empirical coverage of confidence intervals for $\underline{S}_{12}$ (left), $\underline{S}_{13}$ (center) and $\underline{S}_{23}$ (right).

we can observe that the bivariate stratification has a bad effect on the new estimator at very low sample size, but we can notice its good properties as the number of simulations increases.

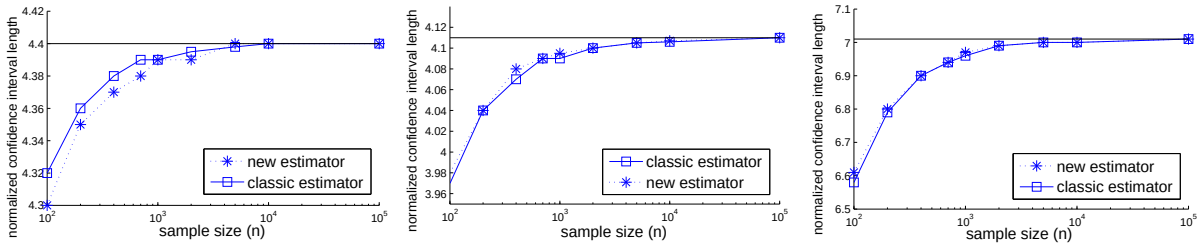

Figure 3: Normalized ($\times\sqrt{n}$) length of the empirical interval for $\underline{S}_1$ (left), $\underline{S}_2$ (center) and $\underline{S}_3$ (right).

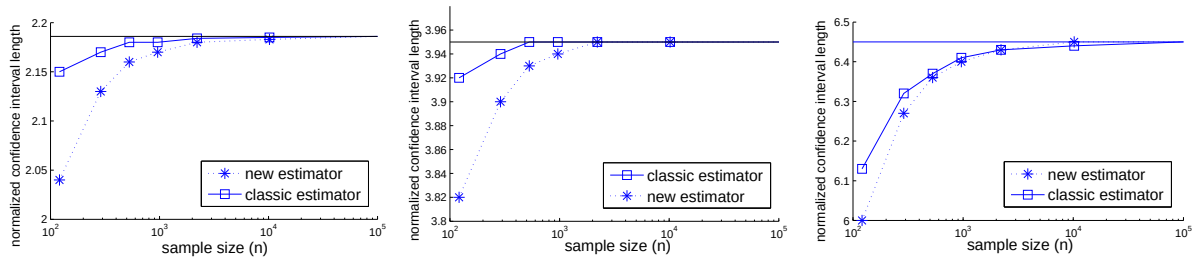

Figure 4: Normalized ($\times\sqrt{n}$) length of the empirical interval for $\underline{S}_{12}$ (left), $\underline{S}_{13}$ (center) and $\underline{S}_{23}$ (right).

5.1.2 Remark on the confidence interval length of the new estimator

Concerning the estimation of the right confidence interval length of the new estimators, note that if the asymptotic empirical coverage — estimated using Formula (102) — is $1-\alpha'$ instead of the expected value $1-\alpha$, then it means that the true asymptotic confidence interval should be $u_{\alpha/2}/u_{\alpha'/2}$ time as long, where $u_.$ denote the normal quantiles. More specifically in our first application, we obtain in this way the true asymptotic normalized ($\times\sqrt{n}$) confidence interval length of $\underline{S}_1$, $\underline{S}_2$, $\underline{S}_{12}$, $\underline{S}_{13}$ and $\underline{S}_{23}$; they are gathered in Table 1. Moreover considering these right normalized confidence interval lengths, we can observe on Figures 5 and 6 that the empirical coverage of the new estimator converges to the expected level 0.99 as n increases, and so we confirm the reliability of the empirical confidence intervals constructed with the true asymptotic length. Unfortunately, evaluating the true asymptotic confidence interval length is infeasible in practice since it requires a lot of replications to estimate the empirical coverage. So the issue related to the construction of optimal confidence intervals remains open.

	$\underline{S}_1$	$\underline{S}_2$	$\underline{S}_{12}$	$\underline{S}_{13}$	$\underline{S}_{23}$
estimated lengths using (102)	4.40	4.15	2.19	3.95	6.45
right lengths	3.96	3.28	1.53	2.37	5.16

Table 1: Comparison between confidence interval lengths estimated using (102) and the right lengths for $\underline{S}_1$, $\underline{S}_2$, $\underline{S}_{12}$, $\underline{S}_{13}$ and $\underline{S}_{23}$

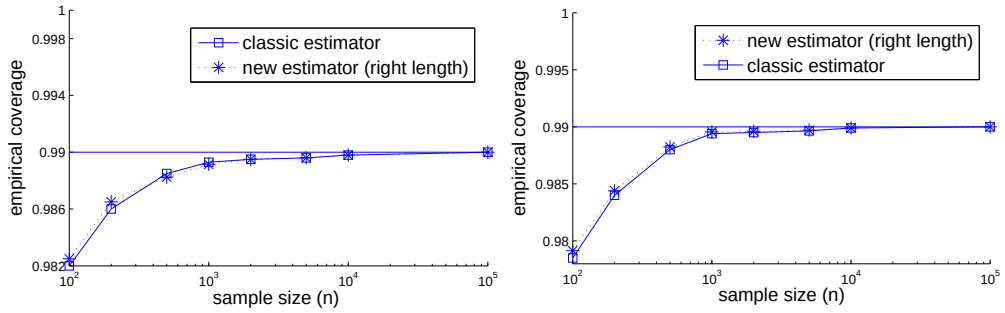

Figure 5: Empirical coverage of confidence intervals for $\underline{S}_1$ (left) and $\underline{S}_2$ (right).

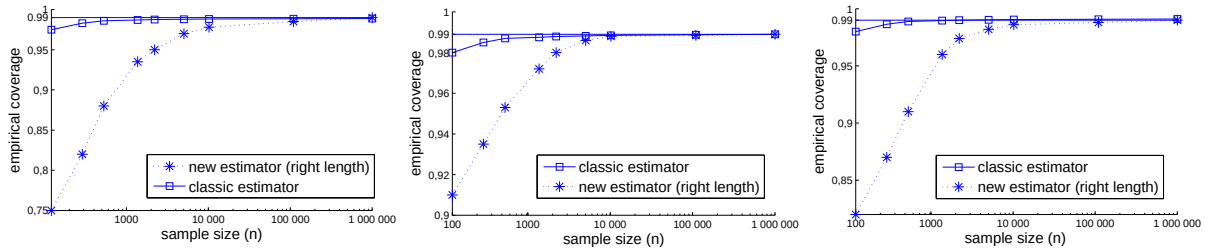

Figure 6: Empirical coverage of confidence intervals for $\underline{S}_{12}$ (left), $\underline{S}_{13}$ (center) and $\underline{S}_{23}$ (right).

5.2 Application to a marine ecosystem simulator

We now illustrate the new method to a one-dimensional coupled hydrodynamical– biological model developed and applied to the Ligurian Sea (northwestern Mediterranean). This ecosystem simulator, MODèle d'ÉCOsystème du GHER et du LOBEPM[1] (MODECOGeL), combines a 1D (vertical) version of the 3D GHER model which takes into account momentum and heat surface fluxes computed from a real meteorological data set, and a biogeochemical model defined by a nitrogen cycle of 12 biological state variables (see Figure 7) controlled by 87 input parameters, see Lacroix and Nival (1998). Here we focus on the chlorophyll-a concentration which is defined as a function of time and depth

$$\mathtt{chla}(t,z) = 1.59 * \big(\mathtt{pp}(t,z) + \mathtt{np}(t,z) + \mathtt{mp}(t,z)\big) \tag{103}$$

where pp, np and mp are the phyto-, nano- and microphytoplankton biomasses, respectively. The behavior of these three state variables are modeled by the following reaction-diffusion and reaction-advection-diffusion equation

$$\frac{\partial\,\mathtt{pp}}{\partial t} = \frac{\partial}{\partial z}\left(\lambda\frac{\partial\,\mathtt{pp}}{\partial z}\right) + \big((1-exud_{\mathtt{pp}})\mu_{\mathtt{pp}} - mort_{\mathtt{pp}}\big)\mathtt{pp} - ing_{\mathtt{pp},\mathtt{nz}}\mathtt{nz} \tag{104}$$

$$\frac{\partial\,\mathtt{np}}{\partial t} = \frac{\partial}{\partial z}\left(\lambda\frac{\partial\,\mathtt{np}}{\partial z}\right) + \big((1-exud_{\mathtt{np}})\mu_{\mathtt{np}} - mort_{\mathtt{np}}\big)\mathtt{np} - ing_{\mathtt{np},\mathtt{miz}}\mathtt{miz} \tag{105}$$

$$\frac{\partial\,\mathtt{mp}}{\partial t} = \frac{\partial}{\partial z}\left(\lambda\frac{\partial\,\mathtt{mp}}{\partial z}\right) + \big((1-exud_{\mathtt{mp}})\mu_{\mathtt{mp}} - mort_{\mathtt{mp}}\big)\mathtt{mp} - ing_{\mathtt{mp},\mathtt{mez}}\mathtt{mez} - sin_{\mathtt{mp}}\frac{\partial\mathtt{mp}}{\partial z} \tag{106}$$

[1]GHER: GeoHydrodynamics and Environment Research, Université de Liège, Belgium. LOBEPM: Laboratoire d'Océanologie Biologique et d'Écologie du Plancton Marin, Université Pierre et Marie Curie, France

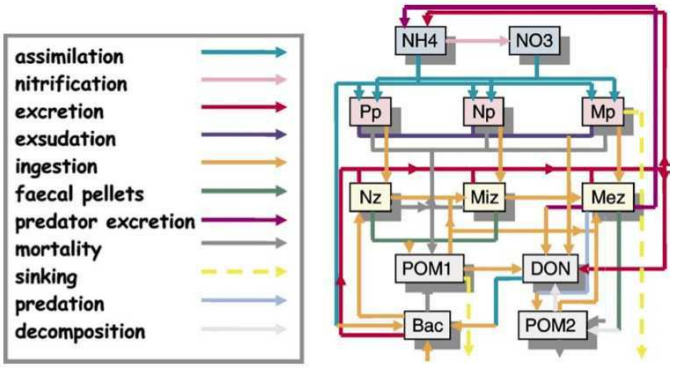

Figure 7: Biogeochemical model (NH4: Ammonium; NH3: nitrate; Pp, Np, Mp: pico-, nano-, microphytoplankton; Nz, Miz, Mez: nano-, micro-, mesozooplankton; PON1, PON2: type 1 and 2 particulate organic nitrogen; Bac: bacteria; DON: dissolved organic nitrogen).

where `nz`, `miz` and `mez` are the nano-, micro- and mesozooplankton biomasses, respectively, and the other notations are

λ	vertical turbulent diffusivity ($\mathrm{m}^2.\mathrm{s}^{-1}$)
$exud_A$	exudation of A (percentage)
μ_A	growth rate of A (day^{-1})
$mort_A$	mortality rate of A (day^{-1})
$ing_{A,B}$	ingestion rate of A by predator B (mgChl)
$sin_{\mathtt{mp}}$	sinking velocity of microphytoplankton ($\mathrm{m.day}^{-1}$)

In our experiment, we focus on two different outputs: the annual maximum of chlorophyll-a concentration in surface water Y_{surf} and the annual maximum of the mean of cholorphyll-a concentration between 20 and 50 meters in depth Y_{depth}. These are practical indicators of biological activity. We are interested in the influence of eight parameters among the 87 input factors. On the one hand, we consider 6 a priori sensitive parameters $\mu_{max\mathtt{pp}}$, $\mu_{max\mathtt{np}}$, $\mu_{max\mathtt{mp}}$, $I_{opt\mathtt{pp}}$, $I_{opt\mathtt{np}}$ and $I_{opt\mathtt{mp}}$ where μ_{maxA} and I_{optA} denote the maximum growth rate of A and the optimum insolation for A, respectively. These input factors are directly related to the growth rate of A, μ_A (see details in Appendix C). On the other hand, we consider the maximum growth rate of bacteria $\mu_{max\mathtt{bac}}$ and the sinking velocity of particulate organic nitrogen (type 1) $sin_{\mathtt{pon1}}$ which have a priori a negligible effect on chlorophyll-a concentration since they do not act directly on `pp`, `np` and `mp` but on the state variables `bac` and `pon1`. We take these eight parameters to be independent gamma distributed random variables with parameters given in Table 2. We estimate all first- and second-order Sobol' indices of both outputs Y_{surf} and Y_{depth} by using the estimators defined in Sections 4.3 and 4.4 with sample sizes $n = 65536$ and $n = 66049$, respectively.

The first-order Sobol' indices are estimated by using nested replicated latin hypercubes following Qian's construction Qian (2009). They allow to visualize empirical convergence of the estimated indices as shown in Figure 8. The estimated indices at the biggest sample size ($n = 65536$) are reported in

Tables 3 and 4; we can notice that both outputs do not define an additive model since in both cases, the sum of the first-order Sobol' indices are less than sixty percents. We also notice that $\mu_{max\texttt{pp}}$ is important in both outputs, while three other a priori important parameters — $\mu_{max\texttt{np}}$, $I_{opt\texttt{np}}$ and $I_{opt\texttt{mp}}$— have actually no effect. At last, it is surprising to observe that the parameter $\mu_{max\texttt{bac}}$, which does not act directly on both outputs, has non-zero values.

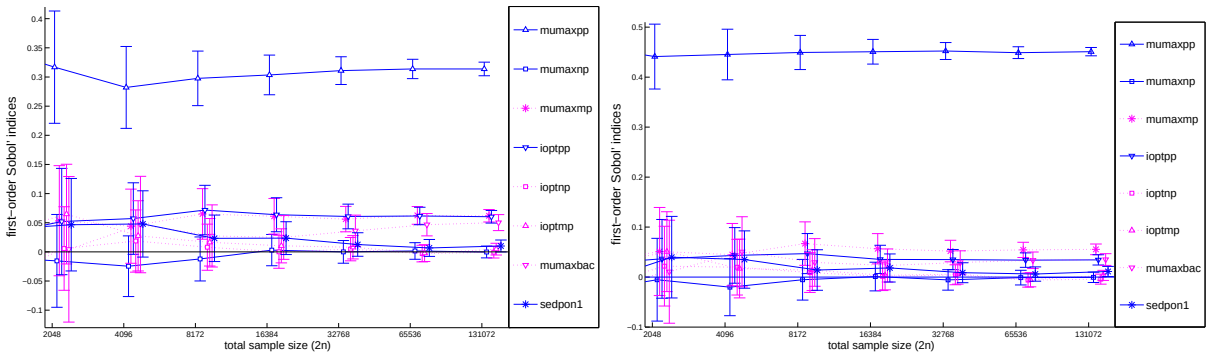

Figure 8: Plots of first-order Sobol' indices with error bars — 99% confidence interval — for both outputs Y_{surf} (left) and Y_{depth} (right).

The second-order Sobol' indices are estimated by using a replicated latin hypercube based on an orthogonal array with 257 levels, index 1 and strength 2 — i.e. $n = 66049$ — following Bush's construction, see Bose (1938). The results are reported in Tables 5 and 6; they confirm that $\mu_{max\texttt{pp}}$ has the main role in both outputs since the non-negligible second-order Sobol' indices are all related to the latter. As a conclusion, we can notice that both outputs are extremely complex and contain, without any doubt, interactions of order more than or equal to 3. Such an analysis with the Monte Carlo estimator of Sobol' indices would be less efficient without the new approach we proposed in this paper. More precisely, both order 1 and order 2 analysis using the classic Monte Carlo estimator — i.e. estimating all the Sobol' indices of order 1 or 2 — only could use a sample size of 30000 instead of 132000 since this classic approach needs 9 independent samples while the new one only needs 2 for the order 1 analysis and 18 independent samples while the new one only needs 4 for the order 2 analysis, see Saltelli (2002).

6 Conclusion

We have introduced a new method to estimate all the k-th order Sobol' indices by using only 2 samples, for any k. This outperforms existing methods including the combinatorial results established by Saltelli (2002). We derive theoretical results in the particular case of first-order Sobol' indices from the work by Janon et al. (2012) on asymptotical properties of Sobol' indices and from the work by Loh (1996) on asymptotical properties of LHS. Further works will consist in deriving these theoretical results to

	label	k	θ	mean	standard deviation
$\mu_{max\texttt{pp}}$ (day^{-1})	1	9	0.33	3	1
$\mu_{max\texttt{np}}$ (day^{-1})	2	9	0.28	2.5	0.83
$\mu_{max\texttt{mp}}$ (day^{-1})	3	9	0.22	2	0.67
$I_{opt\texttt{pp}}$ (W.m^{-2})	4	9	1.11	10	3.33
$I_{opt\texttt{np}}$ (W.m^{-2})	5	9	1.67	15	5
$I_{opt\texttt{mp}}$ (W.m^{-2})	6	9	2.22	20	6.67
$\mu_{max\texttt{bac}}$ (day^{-1})	7	9	0.22	2	0.67
$sin_{\texttt{pon1}}$ (m.day^{-1})	8	9	0.17	1.5	0.5

Table 2: Distributions of variables using gamma density $f(x;k,\theta) = x^{k-1}\exp(-x/\theta)/(\Gamma(\theta)\theta^k)$, where $\Gamma(\cdot)$ is the gamma function.

	$\underline{S}_{\{1\}}$	$\underline{S}_{\{2\}}$	$\underline{S}_{\{3\}}$	$\underline{S}_{\{4\}}$	$\underline{S}_{\{5\}}$	$\underline{S}_{\{6\}}$	$\underline{S}_{\{7\}}$	$\underline{S}_{\{8\}}$
estimated index	0.314	0	0.061	0.060	0	0.003	0.051	0.010
estimated error	0.010	0.011	0.012	0.011	0.010	0.010	0.013	0.012

Table 3: Estimation of first-order Sobol' indices for the output Y_{surf}. The estimated error is the radius of the 99% confidence interval.

	$\underline{S}_{\{1\}}$	$\underline{S}_{\{2\}}$	$\underline{S}_{\{3\}}$	$\underline{S}_{\{4\}}$	$\underline{S}_{\{5\}}$	$\underline{S}_{\{6\}}$	$\underline{S}_{\{7\}}$	$\underline{S}_{\{8\}}$
estimated index	0.451	0	0.055	0.034	0	0	0.035	0.011
estimated error	0.009	0.010	0.010	0.010	0.010	0.010	0.012	0.010

Table 4: Estimation of first-order Sobol' indices for the output Y_{depth}. The estimated error is the radius of the 99% confidence interval.

	$\underline{S}_{\{1,2\}}$	$\underline{S}_{\{1,3\}}$	$\underline{S}_{\{1,4\}}$	$\underline{S}_{\{1,5\}}$	$\underline{S}_{\{1,6\}}$	$\underline{S}_{\{1,7\}}$	$\underline{S}_{\{1,8\}}$	$\underline{S}_{\{2,3\}}$	$\underline{S}_{\{2,4\}}$	$\underline{S}_{\{2,5\}}$
estimated index	0.374	0.479	0.424	0.339	0.324	0.400	0.318	0.069	0.066	0.016
estimated error	0.012	0.011	0.013	0.011	0.011	0.010	0.011	0.011	0.011	0.011

	$\underline{S}_{\{2,6\}}$	$\underline{S}_{\{2,7\}}$	$\underline{S}_{\{2,8\}}$	$\underline{S}_{\{3,4\}}$	$\underline{S}_{\{3,5\}}$	$\underline{S}_{\{3,6\}}$	$\underline{S}_{\{3,7\}}$	$\underline{S}_{\{3,8\}}$	$\underline{S}_{\{4,5\}}$	$\underline{S}_{\{4,6\}}$
estimated index	0.015	0.069	0.015	0.125	0.074	0.075	0.128	0.072	0.077	0.070
estimated error	0.010	0.015	0.010	0.011	0.011	0.011	0.013	0.011	0.011	0.011

	$\underline{S}_{\{4,7\}}$	$\underline{S}_{\{4,8\}}$	$\underline{S}_{\{5,6\}}$	$\underline{S}_{\{5,7\}}$	$\underline{S}_{\{5,8\}}$	$\underline{S}_{\{6,7\}}$	$\underline{S}_{\{6,8\}}$	$\underline{S}_{\{7,8\}}$
estimated index	0.121	0.066	0.017	0.055	0.014	0.056	0.009	0.050
estimated error	0.013	0.011	0.010	0.015	0.010	0.014	0.010	0.015

Table 5: Estimation of second-order Sobol' indices for the output Y_{surf}. The estimated error is the radius of the 99% confidence interval.

	$\underline{S}_{\{1,2\}}$	$\underline{S}_{\{1,3\}}$	$\underline{S}_{\{1,4\}}$	$\underline{S}_{\{1,5\}}$	$\underline{S}_{\{1,6\}}$	$\underline{S}_{\{1,7\}}$	$\underline{S}_{\{1,8\}}$	$\underline{S}_{\{2,3\}}$	$\underline{S}_{\{2,4\}}$	$\underline{S}_{\{2,5\}}$
estimated index	0.506	0.593	0.510	0.455	0.450	0.515	0.447	0.056	0.034	0.005
estimated error	0.010	0.009	0.010	0.009	0.009	0.008	0.009	0.011	0.011	0.011

	$\underline{S}_{\{2,6\}}$	$\underline{S}_{\{2,7\}}$	$\underline{S}_{\{2,8\}}$	$\underline{S}_{\{3,4\}}$	$\underline{S}_{\{3,5\}}$	$\underline{S}_{\{3,6\}}$	$\underline{S}_{\{3,7\}}$	$\underline{S}_{\{3,8\}}$	$\underline{S}_{\{4,5\}}$	$\underline{S}_{\{4,6\}}$
estimated index	0.008	0.055	0.009	0.087	0.057	0.064	0.109	0.063	0.041	0.043
estimated error	0.010	0.014	0.010	0.011	0.011	0.011	0.013	0.011	0.010	0.010

	$\underline{S}_{\{4,7\}}$	$\underline{S}_{\{4,8\}}$	$\underline{S}_{\{5,6\}}$	$\underline{S}_{\{5,7\}}$	$\underline{S}_{\{5,8\}}$	$\underline{S}_{\{6,7\}}$	$\underline{S}_{\{6,8\}}$	$\underline{S}_{\{7,8\}}$
estimated index	0.082	0.041	0.009	0.040	0.007	0.046	0.006	0.041
estimated error	0.013	0.010	0.010	0.014	0.010	0.014	0.010	0.014

Table 6: Estimation of second-order Sobol' indices for the output Y_{depth}. The estimated error is the radius of the 99% confidence interval.

higher-order Sobol' indices and in improving the method by studying how we can estimate correctly the asymptotic variance of the new estimator.

Acknowledgments

the authors thank Pierre Brasseur, Jean-Michel Brankart and Eric Blayo for valuable discussions on the simulator MODECOGeL and more generally on marine ecosystem models. Thanks also to Art Owen for his helpful comments. This work has been partially supported by French National Research Agency (ANR) through COSINUS program (project COSTA-BRAVA n° ANR-09-COSI-015).

7 Lemmas for Proposition 1

Let $\mathbf{X}^1$ and $\mathbf{X}^2$ two distinct points of a Latin hypercube of size n in $[0,1]^d$. For any function f defined on $[0,1]^d$, consider $Y^1 = f(\mathbf{X}^1)$ et $Y^2 = f(\mathbf{X}^2)$. In Theorem 1 in Stein (1987), Stein gives the following result

Theorem 1. *If f is a square integrable function then as n tends to $+\infty$, we have*

$$\mathrm{Cov}(Y^1, Y^2) = -\frac{1}{n}\sum_{i=1}^{d}\sigma_i^2 + o(n^{-1}) \ . \tag{107}$$

In this section, we prove an analogous result with more general settings and without the asymptotic assumption on n (see Lemma 4).

7.1 Notation and definitions

For s and n in $\mathbb{N}^*$, define the partition of $[0,1)^s$ in elementary hypercubes of side $1/n$,

$$\mathcal{Q}_s(n) = \left\{ Q \subseteq [0,1)^s \mid Q = \prod_{i=1}^{s} [\alpha_i, \beta_i),\ \alpha_i \in \left\{0, \frac{1}{n}, \ldots, \frac{n-1}{n}\right\},\ \beta_i = \alpha_i + \frac{1}{n} \right\}. \tag{108}$$

For any square integrable function g defined on $[0,1)^s$, $s \leq d$, define the sequence with general term

$$u_n(g) = n^s \sum_{Q \in \mathcal{Q}_s(n)} \left(\int_Q g(\mathbf{x}) d\mathbf{x} \right)^2,\ n \in \mathbb{N}. \tag{109}$$

7.2 Preliminary results

The first lemma is the analogous result for Lebesgue integrability of a result given in Equation (A.4) in Stein (1987) for Riemann integrability. The second one gives an important inequality which allows to work without asymptotic assumption on n. The last one consists in simplifying integrals under Latin hypercube sampling using the ANOVA decomposition.

Lemma 1. *If g is a square integrable function, the sequence $\left(u_n(g)\right)$ converges to $\int g^2(\mathbf{x})d\mathbf{x}$ as n tends to $+\infty$.*

Proof. Noting that

$$u_n(g) = \int g_n(\mathbf{x}) d\mathbf{x} \tag{110}$$

where

$$\forall \mathbf{x} \in [0,1)^s,\ g_n(\mathbf{x}) = \sum_{Q \in \mathcal{Q}_s(n)} \left(n^s \int_Q g(\mathbf{y}) d\mathbf{y} \right)^2 \mathbf{1}_Q(\mathbf{x}) \tag{111}$$

Lemma 1 is a straightforward consequence of the dominated convergence theorem. So let us prove that there exists an integrable function h such that for all $n \in \mathbb{N}^*$, $|g_n| \leq h$ almost surely, and g_n converges pointwise to g^2, and the conclusion will follow.

First since g is a square integrable function, we have $|g(\mathbf{x})| \leq M$ a.s., and by their definition, the g_n's are as well. Hence their exists an integrable function $(h : \mathbf{x} \mapsto M)$ such that $|g_n| \leq h$ almost surely. Concerning the pointwise convergence, let us prove that for any $\mathbf{x} \in [0,1)^s$,

$$\forall \varepsilon > 0,\ \exists\ N > 0,\ \forall n \geq N,\ \left| \left(n^s \int_{Q_{\mathbf{x}}} g(\mathbf{y}) d\mathbf{y} \right)^2 - g^2(\mathbf{x}) \right| < \varepsilon \tag{112}$$

where $Q_{\mathbf{x}}$ is the set Q in $\mathcal{Q}_s(n)$ containing $\mathbf{x}$. This is obvious if g^2 is a simple function and we easily generalize the result to any g^2 since any measurable function is a pointwise limit of simple functions. □

Lemma 2. *The sequence $\left(u_n(g)\right)$ is dominated by $\int g^2(\mathbf{x})d\mathbf{x}$.*

Proof. Let $n \in \mathbb{N}^*$, the result is proved by showing that the sequence of general term $v_k(g) = u_{2^k n}(g)$ is increasing. In this case, by Lemma 1, we have $\lim v_k(g) = \int g^2(\mathbf{x})d\mathbf{x}$, and since v_k is increasing, all the terms of this sequence are dominated by $\int g^2(\mathbf{x})d\mathbf{x}$, hence $v_0(g) = u_n(g) \leq \int g^2(\mathbf{x})d\mathbf{x}$. To prove that the sequence $\big(v_k(g)\big)$ is increasing, note that

$$v_{k+1}(g) = (2^{k+1}n)^s \sum_{Q \in \mathcal{Q}_s(2^{k+1}n)} \left(\int_Q g(\mathbf{x})d\mathbf{x} \right)^2 \tag{113}$$

$$= (2^k n)^s \sum_{Q \in \mathcal{Q}_s(2^k n)} \left(2^s \sum_{P \in \mathcal{P}(Q, 2^{k+1}n)} \left(\int_P g(\mathbf{x})d\mathbf{x} \right)^2 \right) \tag{114}$$

where $\mathcal{P}(Q, 2^{k+1}n) = \mathcal{Q}(2^{k+1}n) \cap Q$. Then by Jensen inequality, we have

$$2^s \sum_{P \in \mathcal{P}(Q, 2^{k+1}n)} \left(\int_P g(\mathbf{x})d\mathbf{x} \right)^2 \geq \left(\int_Q g(\mathbf{x})d\mathbf{x} \right)^2 \tag{115}$$

and we conclude that $v_{k+1}(g) \geq v_k(g)$. □

For $0 \leq x_1,\ x_2 \leq 1$ define

$$r_n(x_1, x_2) = 1 \text{ if } \lfloor nx_1 \rfloor = \lfloor nx_2 \rfloor \tag{116}$$

$$= 0 \text{ otherwise }, \tag{117}$$

where $\lfloor \cdot \rfloor$ is the floor function. We now end with the following result

Lemma 3. *Let $\mathfrak{v}$ be a subset of $\{1, \ldots, d\}$, we have*

$$\int f(\mathbf{x}_1) f(\mathbf{x}_2) \prod_{i \in \mathfrak{v}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_1 d\mathbf{x}_2 = \int \sum_{\mathfrak{w} \subseteq \mathfrak{v}} f_{\mathfrak{w}}(\mathbf{x}_{1\mathfrak{w}}) f_{\mathfrak{w}}(\mathbf{x}_{2\mathfrak{w}}) \prod_{i \in \mathfrak{v}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_1 d\mathbf{x}_2 \ . \tag{118}$$

Proof. By the ANOVA decomposition — see (1) — we have

$$\int f(\mathbf{x}_1) f(\mathbf{x}_2) \prod_{i \in \mathfrak{v}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_1 d\mathbf{x}_2 = \int \sum_{\mathfrak{w}_1 \subseteq \{1,..,d\}} \sum_{\mathfrak{w}_2 \subseteq \{1,..,d\}} f_{\mathfrak{w}_1}(\mathbf{x}_{1\mathfrak{w}_1}) f_{\mathfrak{w}_2}(\mathbf{x}_{2\mathfrak{w}_2}) \prod_{i \in \mathfrak{v}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_1 d\mathbf{x}_2 \ . \tag{119}$$

Then note that a certain number of terms in the member on the right-hand side vanishe. If $(\mathfrak{w}_1 \cap \mathfrak{v}^c) \cup (\mathfrak{w}_2 \cap \mathfrak{v}^c) \neq \emptyset$ then suppose without loss of generality that there exists $k \in \mathfrak{w}_1 \setminus \mathfrak{v}$; we have

$$\int f_{\mathfrak{w}_1}(\mathbf{x}_{1\mathfrak{w}_1}) f_{\mathfrak{w}_2}(\mathbf{x}_{2\mathfrak{w}_2}) \prod_{i \in \mathfrak{v}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_1 d\mathbf{x}_2 = \int \underbrace{\left(\int f_{\mathfrak{w}_1}(\mathbf{x}_{1\mathfrak{w}_1}) dx_{1k} \right)}_{I_1} f_{\mathfrak{w}_2}(\mathbf{x}_{2\mathfrak{w}_2}) \prod_{i \in \mathfrak{v}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_{1\{k\}^c} d\mathbf{x}_2 \tag{120}$$

and note that, by a basic property of the ANOVA decomposition, $I_1 = 0$. If $(\mathfrak{w}_1 \cap \mathfrak{v}^c) \cup (\mathfrak{w}_2 \cap \mathfrak{v}^c) = \emptyset$ and $\mathfrak{w}_1 \neq \mathfrak{w}_2$, then suppose without loss of generality that there exists $k \in \mathfrak{w}_1 \setminus \mathfrak{w}_2$. In this case, we have

$$\int f_{\mathfrak{w}_1}(\mathbf{x}_{1\mathfrak{w}_1}) f_{\mathfrak{w}_2}(\mathbf{x}_{2\mathfrak{w}_2}) \prod_{i \in \mathfrak{v}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_1 d\mathbf{x}_2 \tag{121}$$

$$= \int \underbrace{\Big(\int f_{\mathfrak{w}_1}(\mathbf{x}_{1\mathfrak{w}_1}) r_n(x_{1k}, x_{2k}) dx_{1k} dx_{2k} \Big)}_{I_2} f_{\mathfrak{w}_2}(\mathbf{x}_{2\mathfrak{w}_2}) \Big(\prod_{i \in \mathfrak{v} \setminus \{k\}} r_n(x_{1i}, x_{2i}) \Big) d\mathbf{x}_{1\{k\}^c} d\mathbf{x}_{2\{k\}^c} \tag{122}$$

and note that by the definition of r_n, we have

$$\int f_{\mathfrak{w}_1}(\mathbf{x}_{1\mathfrak{w}_1}) r_n(x_{1k}, x_{2k}) dx_{1k} dx_{2k} = \int f_{\mathfrak{w}_1}(\mathbf{x}_{1\mathfrak{w}_1}) dx_{1k} \tag{123}$$

and thus $I_2 = 0$. The conclusion of the lemma follows. □

7.3 Main result

Let $\mathfrak{u}$ be a non-empty subset of $\{1, \dots d\}$ and consider $(\dot{\mathbf{Z}}^j_{\mathfrak{u}})_j \sim \mathcal{LH}(n, 2d - |\mathfrak{u}|)$. For any function f defined on $[0,1]^d$, consider

$$\dot{Y}^{1,1}_{\mathfrak{u}} = f(\dot{\mathbf{X}}^1_{\mathfrak{u}}, \dot{\mathbf{X}}^{1,1}_{\mathfrak{u}^c}) \text{ and } \dot{Y}^{2,2}_{\mathfrak{u}} = f(\dot{\mathbf{X}}^2_{\mathfrak{u}}, \dot{\mathbf{X}}^{2,2}_{\mathfrak{u}^c}) \ . \tag{124}$$

We have the following result

Lemma 4. *If f is a square integrable function then we have*

$$- \sum_{\substack{\emptyset \neq \mathfrak{w} \subseteq \mathfrak{u} \\ |\mathfrak{w}| \ odd}} \frac{\sigma^2_{\mathfrak{w}}}{(n-1)^{|\mathfrak{w}|}} \leq \mathrm{Cov}(\dot{Y}^{1,1}_{\mathfrak{u}}, \dot{Y}^{2,2}_{\mathfrak{u}}) \leq \sum_{\substack{\emptyset \neq \mathfrak{w} \subseteq \mathfrak{u} \\ |\mathfrak{w}| \ even}} \frac{\sigma^2_{\mathfrak{w}}}{(n-1)^{|\mathfrak{w}|}} \ . \tag{125}$$

Proof. Recall that for $0 \leq x_1, \ x_2 \leq 1$,

$$\begin{aligned} r_n(x_1, x_2) &= 1 \text{ if } \lfloor n x_1 \rfloor = \lfloor n x_2 \rfloor & (126) \\ &= 0 \text{ otherwise} , & (127) \end{aligned}$$

where $\lfloor \cdot \rfloor$ is the floor function. For $\mathbf{x}_1 = (x_{11}, \dots, x_{1d})$ in $[0,1)^d$, define $\mathbf{x}_{1\mathfrak{v}} = (x_{1i_1}, \dots, x_{1i_{|\mathfrak{v}|}})$ where $\mathfrak{v} = \{i_1, \dots, i_{|\mathfrak{v}|}\}$. Due to the joint density of $(\dot{\mathbf{X}}^1_{\mathfrak{u}}, \dot{\mathbf{X}}^2_{\mathfrak{u}})$ under Latin hypercube sampling — see McKay et al. (1979) or Stein (1987) — and by Lemma 3, we have

$$\mathrm{Cov}(\dot{Y}^{1,1}_{\mathfrak{u}}, \dot{Y}^{2,2}_{\mathfrak{u}}) + \Big(\int f(\mathbf{x}) d\mathbf{x} \Big)^2 = \int f(\mathbf{x}_1) f(\mathbf{x}_2) \Big(\frac{n}{n-1} \Big)^{|\mathfrak{u}|} \prod_{i \in \mathfrak{u}} \big(1 - r_n(x_{1i}, x_{2i}) \big) d\mathbf{x}_1 d\mathbf{x}_2 \tag{128}$$

$$\begin{aligned} &= \Big(\frac{n}{n-1} \Big)^{|\mathfrak{u}|} \sum_{\mathfrak{v} \subseteq \mathfrak{u}} (-1)^{|\mathfrak{v}|} \int f(\mathbf{x}_1) f(\mathbf{x}_2) \prod_{i \in \mathfrak{v}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_1 d\mathbf{x}_2 \\ &= \Big(\frac{n}{n-1} \Big)^{|\mathfrak{u}|} \sum_{\mathfrak{v} \subseteq \mathfrak{u}} (-1)^{|\mathfrak{v}|} \int \sum_{\mathfrak{w} \subseteq \mathfrak{v}} f_{\mathfrak{w}}(\mathbf{x}_{1\mathfrak{w}}) f_{\mathfrak{w}}(\mathbf{x}_{2\mathfrak{w}}) \prod_{i \in \mathfrak{v}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_1 d\mathbf{x}_2 \\ &= \Big(\frac{n}{n-1} \Big)^{|\mathfrak{u}|} \sum_{\mathfrak{v} \subseteq \mathfrak{u}} (-1)^{|\mathfrak{v}|} \sum_{\mathfrak{w} \subseteq \mathfrak{v}} \Big(\frac{1}{n} \Big)^{|\mathfrak{v}| - |\mathfrak{w}|} \int f_{\mathfrak{w}}(\mathbf{x}_{1\mathfrak{w}}) f_{\mathfrak{w}}(\mathbf{x}_{2\mathfrak{w}}) \prod_{i \in \mathfrak{w}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_{1\mathfrak{w}} d\mathbf{x}_{2\mathfrak{w}} \end{aligned} \tag{129}$$

Then note that for any function of $\mathfrak{w}$ denoted by $A(\mathfrak{w})$, we have

$$\begin{aligned}
&\sum_{\mathfrak{v}\subseteq\mathfrak{u}}(-1)^{|\mathfrak{v}|}\sum_{\mathfrak{w}\subseteq\mathfrak{v}}\Big(\frac{1}{n}\Big)^{|\mathfrak{v}|-|\mathfrak{w}|}A(\mathfrak{w}) \\
&\qquad = \sum_{\mathfrak{v}\subseteq\mathfrak{u}}\Big(-\frac{1}{n}\Big)^{|\mathfrak{v}|}\sum_{\mathfrak{w}\subseteq\mathfrak{v}}\Big(\frac{1}{n}\Big)^{-|\mathfrak{w}|}A(\mathfrak{w}) \\
&\qquad = \sum_{\mathfrak{w}\subseteq\mathfrak{u}}\left(\sum_{k=0}^{|\mathfrak{u}|-|\mathfrak{w}|}\binom{|\mathfrak{u}|-|\mathfrak{w}|}{k}\Big(-\frac{1}{n}\Big)^{k+|\mathfrak{w}|}\right)\Big(\frac{1}{n}\Big)^{-|\mathfrak{w}|}A(\mathfrak{w}) \\
&\qquad = \sum_{\mathfrak{w}\subseteq\mathfrak{u}}\Big(\frac{n-1}{n}\Big)^{|\mathfrak{u}|-|\mathfrak{w}|}(-1)^{|\mathfrak{w}|}A(\mathfrak{w}) \qquad (130)
\end{aligned}$$

Hence, we deduce that

$$\mathrm{Cov}(\dot{Y}_{\mathfrak{u}}^{1,1},\dot{Y}_{\mathfrak{u}}^{2,2}) = \sum_{\substack{\mathfrak{w}\subseteq\mathfrak{u}\\ \mathfrak{w}\neq\emptyset}}\Big(\frac{n}{n-1}\Big)^{|\mathfrak{w}|}(-1)^{|\mathfrak{w}|}\int f_{\mathfrak{w}}(\mathbf{x}_{1\mathfrak{w}})f_{\mathfrak{w}}(\mathbf{x}_{2\mathfrak{w}})\prod_{i\in\mathfrak{w}}r_n(x_{1i},x_{2i})d\mathbf{x}_{1\mathfrak{w}}d\mathbf{x}_{2\mathfrak{w}}\ . \qquad (131)$$

Finally by the definition of r_n, we have

$$0\leq\int f_{\mathfrak{w}}(\mathbf{x}_{1\mathfrak{w}})f_{\mathfrak{w}}(\mathbf{x}_{2\mathfrak{w}})\prod_{i\in\mathfrak{w}}r_n(x_{1i},x_{2i})d\mathbf{x}_{1\mathfrak{w}}d\mathbf{x}_{2\mathfrak{w}}\leq\sum_{Q\in\mathcal{Q}_{|\mathfrak{w}|}(n)}\left(\int_Q f_{\mathfrak{w}}(\mathbf{x}_{1\mathfrak{w}})d\mathbf{x}_{1\mathfrak{w}}\right)^2 \qquad (132)$$

and by Lemma 2, this gives

$$0\leq\int f_{\mathfrak{w}}(\mathbf{x}_{1\mathfrak{w}})f_{\mathfrak{w}}(\mathbf{x}_{2\mathfrak{w}})\prod_{i\in\mathfrak{w}}r_n(x_{1i},x_{2i})d\mathbf{x}_{1\mathfrak{w}}d\mathbf{x}_{2\mathfrak{w}}\leq\frac{\sigma_{\mathfrak{w}}^2}{n^{|\mathfrak{w}|}}\ . \qquad (133)$$

The latter inequalities and (131) lead to Lemma 4. □

8 Proof of (iii) in Proposition 2

We first give three lemmas. The proof of (iii) in Proposition 2 is given in Section B.2.

8.1 Preliminary results

Lemma 5. *Let $d\in\mathbb{N}^*$, if $n\geq\frac{d^2}{2}$ then*

$$\left(1+\frac{1}{n}\right)^d-1\leq\frac{d+1}{n}\ . \qquad (134)$$

Proof. If $d=1$, the result is obvious. Otherwise, for any $x>0$, consider the function g_d defined by

$$g_d(x)=\left(1+\frac{1}{x}\right)^d-1-\frac{d+1}{x}\ . \qquad (135)$$

We show that

(1) if there exists $x_0>0$ such that $g_d(x_0)\leq 0$ then for all $x\geq x_0$, $g_d(x)\leq 0$

(2) $g_d(d^2/2) \leq 0$

and the conclusion follows. Concerning (1) note that

$$g_d(x) = 1 + \frac{d}{x} + O(x^{-2}) - 1 - \frac{d}{x} - \frac{1}{x} = -\frac{1}{x} + O(x^{-2}) \tag{136}$$

and then that g_d is negative as x tends to $+\infty$. Moreover for any $d > 1$, g_d is first decreasing and then increasing. Indeed, we have

$$g_d'(x) = -\frac{d}{x^2}\left(1 + \frac{1}{x}\right)^{d-1} + \frac{d+1}{x^2} \tag{137}$$

and we deduce that $g_d'(x_0) = 0$ with

$$x_0 = \frac{1}{\left(\frac{d+1}{d}\right)^{1/(d-1)} - 1} > 0 \tag{138}$$

and is negative on the left side and positive on the right side. The conclusion of (1) follows. Concerning (2), it is easy to check that it is true for $d = 1$ and 2, and for $d \geq 3$ we have

$$\begin{aligned}
g_d\left(\frac{d^2}{2}\right) &= \sum_{k=0}^{d} \binom{d}{k}\left(\frac{2}{d^2}\right)^k - 1 - \frac{2}{d} - \frac{2}{d^2} && (139)\\
&= -\frac{2}{d^3} + \sum_{k=3}^{d} \binom{d}{k}\left(\frac{2}{d^2}\right)^k && (140)\\
&\leq -\frac{2}{d^3} + \sum_{k=3}^{d} \frac{1}{k!}\left(\frac{2}{d}\right)^k && (141)\\
&\leq -\frac{2}{d^3} + \frac{1}{d^3} + \frac{1}{3d^3} + \frac{2}{3}\sum_{k=4}^{d} \frac{1}{d^k} && (142)\\
&\leq -\frac{2}{d^3} + \sum_{k=3}^{d} \frac{1}{d^k} + \frac{1}{3d^3}\left(1 - \sum_{k=1}^{d-3} \frac{1}{d^k}\right) && (143)\\
&\leq -\frac{2}{d^3} + \sum_{k=3}^{d} \frac{1}{d^k} && (144)\\
&\leq -\frac{2}{d^3} + \frac{2}{d^3} && (145)
\end{aligned}$$

and the conclusion follows. □

With the same notation as at the beginning of Section A.3, we have the following result

Lemma 6. *If f is a square integrable function, we have*

$$\mathbb{E}\big[f(\dot{\mathbf{X}}^1_{\mathfrak{u}}, \dot{\mathbf{X}}^{1,1}_{\mathfrak{u}^c}) f(\dot{\mathbf{X}}^1_{\mathfrak{u}}, \dot{\mathbf{X}}^{2,1}_{\mathfrak{u}^c})\big] = \mathbb{E}[Y]^2 + \underline{\tau}^2_{\mathfrak{u}} + B_{\mathfrak{u},n} \tag{146}$$

where

$$-\mathbb{E}[Y^2] \sum_{\substack{\emptyset \neq \mathfrak{v} \subseteq \mathfrak{u}^c \\ |v| \ odd}} \frac{1}{(n-1)^{|v|}} \leq B_{\mathfrak{u},n} \leq \mathbb{E}[Y^2] \sum_{\substack{\emptyset \neq \mathfrak{v} \subseteq \mathfrak{u}^c \\ |v| \ even}} \frac{1}{(n-1)^{|v|}} \ . \tag{147}$$

Proof. First, due to the joint density of $(\dot{\mathbf{X}}^{1,1}_{\mathfrak{u}^c}, \dot{\mathbf{X}}^{2,1}_{\mathfrak{u}^c})$ under Latin hypercube sampling — see McKay et al. (1979) or Stein (1987) — we have

$$\mathbb{E}\big[f(\dot{\mathbf{X}}^1_{\mathfrak{u}}, \dot{\mathbf{X}}^{1,1}_{\mathfrak{u}^c}) f(\dot{\mathbf{X}}^1_{\mathfrak{u}}, \dot{\mathbf{X}}^{2,1}_{\mathfrak{u}^c})\big] = \int f(\mathbf{x}, \mathbf{x}_1) f(\mathbf{x}, \mathbf{x}_2) \Big(\frac{n}{n-1}\Big)^{d-|\mathfrak{u}|} \prod_{i \in \mathfrak{u}^c} \big(1 - r_n(x_{1i}, x_{2i})\big) d\mathbf{x} d\mathbf{x}_1 d\mathbf{x}_2 \tag{148}$$

$$= \Big(\frac{n}{n-1}\Big)^{d-|\mathfrak{u}|} \int \underbrace{\Big(\sum_{\mathfrak{v} \subseteq \mathfrak{u}^c} (-1)^{|\mathfrak{v}|} \int f(\mathbf{x}, \mathbf{x}_1) f(\mathbf{x}, \mathbf{x}_2) \prod_{i \in \mathfrak{v}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_1 d\mathbf{x}_2 \Big)}_{I(\mathbf{x})} d\mathbf{x} \tag{149}$$

We now denote $f_{\mathbf{x}} : \mathbf{y} \mapsto f(\mathbf{x}, \mathbf{y})$ and then by (129) and (130) we have

$$I(\mathbf{x}) = \sum_{\mathfrak{v} \subseteq \mathfrak{u}^c} (-1)^{|\mathfrak{v}|} \int f_{\mathbf{x}}(\mathbf{x}_1) f_{\mathbf{x}}(\mathbf{x}_2) \prod_{i \in \mathfrak{v}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_1 d\mathbf{x}_2 \tag{150}$$

$$= \sum_{\mathfrak{v} \subseteq \mathfrak{u}^c} (-1)^{|\mathfrak{v}|} \sum_{\mathfrak{w} \subseteq \mathfrak{v}} \Big(\frac{1}{n}\Big)^{|\mathfrak{v}|-|\mathfrak{w}|} \int f_{\mathbf{x},\mathfrak{w}}(\mathbf{x}_{1\mathfrak{w}}) f_{\mathbf{x},\mathfrak{w}}(\mathbf{x}_{2\mathfrak{w}}) \prod_{i \in \mathfrak{w}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_{1\mathfrak{w}} d\mathbf{x}_{2\mathfrak{w}} \tag{151}$$

$$= \sum_{\mathfrak{w} \subseteq \mathfrak{u}^c} (-1)^{|\mathfrak{w}|} \Big(\frac{n-1}{n}\Big)^{d-|\mathfrak{u}|-|\mathfrak{w}|} \int f_{\mathbf{x},\mathfrak{w}}(\mathbf{x}_{1\mathfrak{w}}) f_{\mathbf{x},\mathfrak{w}}(\mathbf{x}_{2\mathfrak{w}}) \prod_{i \in \mathfrak{w}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_{1\mathfrak{w}} d\mathbf{x}_{2\mathfrak{w}} \ . \tag{152}$$

Hence by (133) we have for all $\mathfrak{w} \neq \emptyset$,

$$0 \leq \int f_{\mathbf{x},\mathfrak{w}}(\mathbf{x}_{1\mathfrak{w}}) f_{\mathbf{x},\mathfrak{w}}(\mathbf{x}_{2\mathfrak{w}}) \prod_{i \in \mathfrak{w}} r_n(x_{1i}, x_{2i}) d\mathbf{x}_{1\mathfrak{w}} d\mathbf{x}_{2\mathfrak{w}} \leq \frac{\int f^2_{\mathbf{x},\mathfrak{w}}(\mathbf{x}_{1\mathfrak{w}}) d\mathbf{x}_{1\mathfrak{w}}}{n^{|\mathfrak{w}|}} \leq \frac{\int f^2_{\mathbf{x}}(\mathbf{x}_1) d\mathbf{x}_1}{n^{|\mathfrak{w}|}} \tag{153}$$

and note that

$$\int f_{\mathbf{x},\emptyset}(\mathbf{x}_{1\emptyset}) f_{\mathbf{x},\emptyset}(\mathbf{x}_{2\emptyset}) \prod_{i \in \emptyset} r_n(x_{1i}, x_{2i}) d\mathbf{x}_{1\emptyset} d\mathbf{x}_{2\emptyset} = f^2_{\mathbf{x},\emptyset} \ . \tag{154}$$

Finally, note that

$$\int f^2_{\mathbf{x},\emptyset} d\mathbf{x} = \underline{\tau}^2_{\mathfrak{u}} + \mathbb{E}[Y]^2 \tag{155}$$

and

$$\int \int f^2_{\mathbf{x}}(\mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x} = \mathbb{E}\big[Y^2\big] \tag{156}$$

and conclude that

$$\mathbb{E}\big[f(\dot{\mathbf{X}}^1_{\mathfrak{u}}, \dot{\mathbf{X}}^{1,1}_{\mathfrak{u}^c}) f(\dot{\mathbf{X}}^1_{\mathfrak{u}}, \dot{\mathbf{X}}^{2,1}_{\mathfrak{u}^c})\big] = \Big(\frac{n}{n-1}\Big)^{d-|\mathfrak{u}|} \int I(\mathbf{x}) d\mathbf{x} = \underline{\tau}^2_{\mathfrak{u}} + \mathbb{E}[Y]^2 + B_{\mathfrak{u},n} \tag{157}$$

with

$$-\mathbb{E}[Y^2] \sum_{\substack{\emptyset \neq \mathfrak{v} \subseteq \mathfrak{u}^c \\ |v| \text{ odd}}} \frac{1}{(n-1)^{|v|}} \leq B_{\mathfrak{u},n} \leq \mathbb{E}[Y^2] \sum_{\substack{\emptyset \neq \mathfrak{v} \subseteq \mathfrak{u}^c \\ |v| \text{ even}}} \frac{1}{(n-1)^{|v|}} \ . \tag{158}$$

□

Lemma 7. *The inequalities in Equation* (147) *imply that*

$$\Big| \sum_{\mathfrak{w} \subseteq u^c} \Big(\frac{1}{n}\Big)^{|\mathfrak{w}|} B_{\mathfrak{u} \cup \mathfrak{w},n} \Big| \leq \Big(\frac{d-|\mathfrak{u}|+1}{n} + 1\Big)\Big(\frac{d-|\mathfrak{u}|+1}{n-1}\Big) \mathbb{E}\big[Y^2\big] \ . \tag{159}$$

Proof. By (147), we have

$$\sum_{\mathfrak{w}\subseteq\mathfrak{u}^c}\left(\frac{1}{n}\right)^{|\mathfrak{w}|}B_{\mathfrak{u}\cup\mathfrak{w},n} \leq \mathbb{E}\big[Y^2\big]\sum_{\mathfrak{w}\subseteq\mathfrak{u}^c}\left(\frac{1}{n}\right)^{|\mathfrak{w}|}\sum_{\emptyset\neq\mathfrak{v}\subseteq(\mathfrak{u}\cup\mathfrak{w})^c}\frac{1}{(n-1)^{|\mathfrak{w}|}} \tag{160}$$

$$\leq \mathbb{E}\big[Y^2\big]\sum_{\mathfrak{w}\subseteq\mathfrak{u}^c}\left(\frac{1}{n}\right)^{|\mathfrak{w}|}\left(\left(1+\frac{1}{n-1}\right)^{d-|\mathfrak{u}|-|\mathfrak{w}|}-1\right) \tag{161}$$

$$\leq \mathbb{E}\big[Y^2\big]\left[\left(1+\frac{1}{n-1}\right)^{d-|\mathfrak{u}|}\sum_{\mathfrak{w}\subseteq\mathfrak{u}^c}\left(\frac{1}{n}\right)^{|\mathfrak{w}|}-\sum_{\mathfrak{w}\subseteq\mathfrak{u}^c}\left(\frac{1}{n}\right)^{|\mathfrak{w}|}\right] \tag{162}$$

$$\leq \mathbb{E}\big[Y^2\big]\left[\left(1+\frac{1}{n-1}\right)^{d-|\mathfrak{u}|}\left(1+\frac{1}{n}\right)^{d-|\mathfrak{u}|}-\left(1+\frac{1}{n}\right)^{d-|\mathfrak{u}|}\right] \tag{163}$$

and the conclusion follows by applying twice Lemma 5. □

8.2 Proof of (iii) in Proposition 2

Proof. First note that by the definition of $(\ddot{\mathbf{Z}}^j_{\mathfrak{u}})_{j=1..n}$ we have

$$\ddot{Y}^{j,1}_{\mathfrak{u}} = f(\ddot{\mathbf{X}}^j_{\mathfrak{u}}, \ddot{\mathbf{X}}^{j,1}_{\mathfrak{u}^c}) \text{ and } \ddot{Y}^{j,2}_{\mathfrak{u}} = f(\ddot{\mathbf{X}}^j_{\mathfrak{u}}, \ddot{\mathbf{X}}^{j,2}_{\mathfrak{u}^c}) \tag{164}$$

with

$$\ddot{\mathbf{X}}^{j,1}_{\mathfrak{u}^c} = \left(\frac{\pi_1(j)-U_{1,\pi_1(j)}}{n},\dots,\frac{\pi_d(j)-U_{d,\pi_d(j)}}{n}\right) \tag{165}$$

and

$$\ddot{\mathbf{X}}^{j,2}_{\mathfrak{u}^c} = \left(\frac{\pi'_1(j)-U_{1,\pi'_1(j)}}{n},\dots,\frac{\pi'_d(j)-U_{d,\pi'_d(j)}}{n}\right) \tag{166}$$

where the π_i's, the π'_i's and the $U_{i,j}$'s are independent random variables uniformly distributed on Π_n — see Definition 1 —, Π_n and $[0,1]$, respectively. Moreover note that if for an index $i \in \mathfrak{u}^c$, we have $\pi_i(j) = \pi'_i(j)$ then $\ddot{X}^{j,1}_i = \ddot{X}^{j,2}_i$; and if $\pi_i(j) \neq \pi'_i(j)$ then $U_{i,\pi_i(j)}$ and $U_{i,\pi'_i(j)}$ are independent and therefore $\ddot{X}^{j,1}_i$ and $\ddot{X}^{j,2}_i$ are two distinct points of a Latin hypercube of size n in $[0,1]$. For $j \in \{1,\dots,n\}$, denote by $\mathfrak{e}(j)$ the set of integers $i \in \mathfrak{u}^c$ such that $\pi_i(j) = \pi'_i(j)$. Thus we have

$$\mathbb{E}\big[\ddot{Y}^{j,1}_{\mathfrak{u}}\ddot{Y}^{j,2}_{\mathfrak{u}}\big] = \frac{1}{n^{2d-|\mathfrak{u}|}}\sum_{\mathfrak{w}\subseteq\mathfrak{u}^c}\sum_{\substack{\boldsymbol{\pi}(j)\in\\ \{1,\dots,n\}^d}}\sum_{\substack{\boldsymbol{\pi}'_{\mathfrak{u}^c}(j)\in\\ \{1,\dots,n\}^{d-|\mathfrak{u}|}}}\mathbf{1}_{\{\mathfrak{e}(j)=\mathfrak{w}\}}\cdots \tag{167}$$

$$\cdots \int f\left(\frac{\pi_1(j)-u_{11}}{n},\dots,\frac{\pi_d(j)-u_{1d}}{n}\right)f\left(\frac{\pi'_1(j)-u_{21}}{n},\dots,\frac{\pi'_d(j)-u_{2d}}{n}\right)d\mathbf{u}_1 d\mathbf{u}_{2(\mathfrak{u}\cup\mathfrak{w})^c} \tag{168}$$

where for all $i \in \mathfrak{u}\cup\mathfrak{e}(j)$, $\pi_i(j) = \pi'_i(j)$ and $u_{1i}(j) = u_{2i}(j)$. And noting that

$$\frac{1}{(n-1)^{d-|\mathfrak{u}|-|\mathfrak{w}|}n^d}\sum_{\substack{\boldsymbol{\pi}(j)\in\\ \{1,\dots,n\}^d}}\sum_{\substack{\boldsymbol{\pi}'_{\mathfrak{u}^c}(j)\in\\ \{1,\dots,n\}^{d-|\mathfrak{u}|}}}\mathbf{1}_{\{\mathfrak{e}(j)=\mathfrak{w}\}}\cdots \tag{169}$$

$$\cdots\int f\left(\frac{\pi_1(j)-u_{11}}{n},\dots,\frac{\pi_d(j)-u_{1d}}{n}\right)f\left(\frac{\pi'_1(j)-u_{21}}{n},\dots,\frac{\pi'_d(j)-u_{2d}}{n}\right)d\mathbf{u}_1 d\mathbf{u}_{2(\mathfrak{u}\cup\mathfrak{w})^c} \tag{170}$$

is equal to $\mathbb{E}\big[f(\dot{\mathbf{X}}^{1}_{\mathfrak{u}\cup\mathfrak{w}},\dot{\mathbf{X}}^{1,1}_{(\mathfrak{u}\cup\mathfrak{w})^c})f(\dot{\mathbf{X}}^{1}_{\mathfrak{u}\cup\mathfrak{w}},\dot{\mathbf{X}}^{2,1}_{(\mathfrak{u}\cup\mathfrak{w})^c})\big]$ where $(\dot{\mathbf{X}}^{j}_{\mathfrak{u}\cup\mathfrak{w}},\dot{\mathbf{X}}^{j,1}_{(\mathfrak{u}\cup\mathfrak{w})^c})_j \sim \mathcal{LH}(n,d)$, Lemma 6 gives

$$\mathbb{E}\big[\ddot{Y}^{j,1}_{\mathfrak{u}}\ddot{Y}^{j,2}_{\mathfrak{u}}\big] = \sum_{\mathfrak{w}\subseteq\mathfrak{u}^c}\left(\frac{1}{n}\right)^{|\mathfrak{w}|}\Big(\mathbb{E}[Y]^2 + \underline{\tau}^2_{\mathfrak{u}\cup\mathfrak{w}} + B_{\mathfrak{u}\cup\mathfrak{w},n}\Big)\ . \tag{171}$$

By Lemmas 5 and 7, and noting that $\mathbb{E}[Y]^2 + \underline{\tau}^2_{\mathfrak{u}\cup\mathfrak{w}} \le \mathbb{E}[Y^2]$, we obtain

$$\mathbb{E}\big[\ddot{Y}^{j,1}_{\mathfrak{u}}\ddot{Y}^{j,2}_{\mathfrak{u}}\big] = \mathbb{E}[Y]^2 + \underline{\tau}^2_{\mathfrak{u}} + B_{|\mathfrak{u}|,n} \tag{172}$$

where

$$\big|B_{|\mathfrak{u}|,n}\big| \le \left(\frac{d-|\mathfrak{u}|+1}{n}+2\right)\left(\frac{d-|\mathfrak{u}|+1}{n-1}\right)\mathbb{E}[Y^2]\ . \tag{173}$$

Following the same proof, it is easy to show that for $j \neq l$, we have

$$\mathbb{E}\big[\ddot{Y}^{j,1}_{\mathfrak{u}}\ddot{Y}^{l,2}_{\mathfrak{u}}\big] = \mathbb{E}[Y]^2 + B_{n,1} \tag{174}$$

where

$$\big|B_{n,1}\big| \le \left(\frac{d+1}{n}+2\right)\left(\frac{d+1}{n-1}\right)\mathbb{E}[Y^2]\ . \tag{175}$$

Thus noting that

$$\mathbb{E}\big[\underline{\tau}^{2,RLHS}_{\mathfrak{u},n}\big] = \frac{n-1}{n}\mathbb{E}[\ddot{Y}^{1,1}_{\mathfrak{u}}\ddot{Y}^{1,2}_{\mathfrak{u}}] - \frac{n-1}{n}\mathbb{E}[\ddot{Y}^{1,1}_{\mathfrak{u}}\ddot{Y}^{2,2}_{\mathfrak{u}}] \tag{176}$$

we conclude that

$$\mathbb{E}\big[\widetilde{\underline{\tau}}^{2,RLHS}_{\mathfrak{u},n}\big] = \underline{\tau}^2_{\mathfrak{u}} - \frac{1}{n}\underline{\tau}^2_{\mathfrak{u}} + \frac{n-1}{n}\big(B_{n,1}+B_{|\mathfrak{u}|,n}\big) \tag{177}$$

where the biases are $O(n^{-1})$ as specified above. Concerning $\widetilde{\sigma}^{2,RLHS}_n$, note that $\widetilde{\sigma}^{2,RLHS}_n = \widetilde{\sigma}^{2,LHS}_n$ and the conclusion follows from (iii) in Proposition 1. Concerning $\widehat{\underline{\tau}}^{2,RLHS}_{\mathfrak{u},n}$ and $\widehat{\sigma}^{2,RLHS}_n$, we have

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\ddot{Y}^{j,1}_{\mathfrak{u}}+\ddot{Y}^{j,2}_{\mathfrak{u}}}{2}\right)^2\right] = \frac{1}{4n}\mathbb{E}\big[(\ddot{Y}^{1,1}_{\mathfrak{u}}+\ddot{Y}^{1,2}_{\mathfrak{u}})^2\big] + \frac{1}{4n^2}\sum_{j=1}^{n}\sum_{\substack{l=1\\l\neq j}}^{n}\mathbb{E}\big[(\ddot{Y}^{j,1}_{\mathfrak{u}}+\ddot{Y}^{j,2}_{\mathfrak{u}})(\ddot{Y}^{l,1}_{\mathfrak{u}}+\ddot{Y}^{l,2}_{\mathfrak{u}})\big] \tag{178}$$

$$= \frac{1}{2n}\Big(\mathbb{E}\big[(\ddot{Y}^{1,1}_{\mathfrak{u}})^2\big] + \mathbb{E}[\ddot{Y}^{1,1}_{\mathfrak{u}}\ddot{Y}^{1,2}_{\mathfrak{u}}]\Big) + \frac{n-1}{2n}\Big(\mathbb{E}[\ddot{Y}^{1,1}_{\mathfrak{u}}\ddot{Y}^{2,1}_{\mathfrak{u}}] + \mathbb{E}[\ddot{Y}^{1,1}_{\mathfrak{u}}\ddot{Y}^{2,2}_{\mathfrak{u}}]\Big). \tag{179}$$

Then using notation in (16–18), note that

$$\mathbb{E}\big[\ddot{Y}^{1,1}_{\mathfrak{u}}\ddot{Y}^{2,1}_{\mathfrak{u}}\big] = \mathbb{E}\big[\dot{Y}^{1,1}_{\mathfrak{u}}\dot{Y}^{2,1}_{\mathfrak{u}}\big] = \mathrm{Cov}\big(\dot{Y}^{1,1}_{\mathfrak{u}},\dot{Y}^{2,1}_{\mathfrak{u}}\big) + \mathbb{E}[Y]^2 = \mathrm{Cov}\big(\dot{Y}^{1,1}_{\{1,\ldots,d\}},\dot{Y}^{2,2}_{\{1,\ldots,d\}}\big) + \mathbb{E}[Y]^2 \tag{180}$$

and by (173), (175) and Lemma 4, we deduce

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\ddot{Y}^{j,1}_{\mathfrak{u}}+\ddot{Y}^{j,2}_{\mathfrak{u}}}{2}\right)^2\right] = \frac{1}{2n}\underline{\tau}^2_{\mathfrak{u}} + \mathbb{E}[Y]^2 + B_{n,1} + B_{n,2}\ . \tag{181}$$

where

$$|B_{n,2}| \le \frac{\sigma^2}{2n} \tag{182}$$

and $B_{n,1}$ is specified in (175). Then it is easy to conclude that

$$\mathbb{E}\big[\widehat{\underline{\tau}}^{2,RLHS}_{\mathfrak{u},n}\big] = \underline{\tau}^2_{\mathfrak{u}} - \frac{1}{2n}\underline{\tau}^2_{\mathfrak{u}} + B_{n,1} + B_{n,2} + \frac{n-1}{n}B_{|\mathfrak{u}|,n} \tag{183}$$

$$\mathbb{E}\big[\widehat{\sigma}^{2,RLHS}_{n}\big] = \sigma^2 - \frac{1}{2n}\underline{\tau}^2_{\mathfrak{u}} + B_{n,1} + B_{n,2} \tag{184}$$

where the biases are $O(n^{-1})$ as specified above. □

parameter	definition
μ_A	growth rate of A
$\mu_{max\mathrm{A}}$	maximum growth rate of A
$\mathrm{lim}_{\mathrm{NO_3A}}$	limitation by NO_3 for A
$\mathrm{lim}_{\mathrm{NH_4A}}$	limitation by NH_4 for A
$\mathrm{K}_{\mathrm{NO_3A}}$	half-saturation coefficient of NO_3 for A
$\mathrm{K}_{\mathrm{NH_4A}}$	half-saturation coefficient of NH_4 for A
Ψ	inhibition coefficient by NH_4
NO_3	NO_3 concentration
NH_4	NH_4 concentration
$\mathrm{lim}_{\mathrm{IA}}$	limitation by light for A
β_{IA}	shape factor for photoinhibition curve
$I_{opt\mathrm{A}}$	optimum insolation for A
PAR	photosynthetic active radiation
$\mathrm{lim}_{\mathrm{TA}}$	limitation by temperature for A
β_{TA}	shape factor for thermoinhibition curve
$T_{opt\mathrm{A}}$	optimum temperature for A
$T_{let\mathrm{A}}$	lower lethal temperature for A
T	temperature

Table 7: Parameters of the phytoplankton growth model

9 Phytoplankton growth model

The phytoplankton growth is given by the five following equations, where A stands for `pp`, `np` or `mp`.

$$\mu_A = \mu_{max\mathrm{A}} \mathrm{lim}_{\mathrm{IA}} \mathrm{lim}_{\mathrm{TA}} (\mathrm{lim}_{\mathrm{NO_3A}} + \mathrm{lim}_{\mathrm{NH_4A}}) \tag{185}$$

$$\mathrm{lim}_{\mathrm{NO_3A}} = \left(\frac{\mathrm{NO_3}}{\mathrm{NO_3} + \mathrm{K}_{\mathrm{NO_3A}}} \right) \exp(-\Psi \mathrm{NH_4}) \tag{186}$$

$$\mathrm{lim}_{\mathrm{NH_4A}} = \left(\frac{\mathrm{NH_4}}{\mathrm{NH_4} + \mathrm{K}_{\mathrm{NH_4A}}} \right) \tag{187}$$

$$\mathrm{lim}_{\mathrm{IA}} = \frac{2(1+\beta_{\mathrm{IA}}) \frac{\mathrm{PAR}}{I_{opt\mathrm{A}}}}{\left(\frac{\mathrm{PAR}}{I_{opt\mathrm{A}}}\right)^2 + 2\beta_{\mathrm{IA}} \frac{\mathrm{PAR}}{I_{opt\mathrm{A}}} + 1} \tag{188}$$

$$\mathrm{lim}_{\mathrm{TA}} = \max \left(\frac{2(1+\beta_{\mathrm{TA}}) \frac{T - T_{let\mathrm{A}}}{T_{opt\mathrm{A}} - T_{let\mathrm{A}}}}{\left(\frac{T - T_{let\mathrm{A}}}{T_{opt\mathrm{A}} - T_{let\mathrm{A}}}\right)^2 + 2\beta_{\mathrm{IA}} \frac{T - T_{let\mathrm{A}}}{T_{opt\mathrm{A}} - T_{let\mathrm{A}}} + 1}, 0 \right) \tag{189}$$

where the parameters are defined in the following table

References

Blatman, G. and Sudret, B. (2010). Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. *Reliability Engineering and System Safety*, 95:1216–1229.

Bose, R. (1938). On the application of the theory of galois fields to the problem of construction of hyper-graeco-latin squares. *Sankhya*, 3:323–338.

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

Efron, B. and Stein, C. (1981). The jackknife estimate of variance. *The Annals of Statistics*, 9(3):586–596.

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

Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17.

Ishigami, T. and Homma, T. (1990). An importance quantification technique in uncertainty analysis for computers models. *First International Symposium on Uncertainty Modeling and Analysis Proceedings*, pages 398–403.

Janon, A., Klein, T., Lagnoux, A., Nodet, M., and Prieur, C. (2012+). Asymptotic normality and efficiency of two sobol index estimators. *Preprint available at http://hal.inria.fr/docs/00/66/50/48/PDF/ArtAsymptSobol.pdf*.

Lacroix, G. and Nival, P. (1998). Influence of meteorogical variability on primary production dynamics in the ligurian sea (nw mediterranean sea) with 1d hydrodynamic/biological model. *Journal of Marine Systems*, 37:229–258.

Loh, W. L. (1996). On latin hypercube sampling. *The Annals of Statistics*, 24(5):2058–2080.

Loh, W. L. (2008). A multivariate central limit theorem for randomized orthogonal array sampling designs in computer experiments. *The Annals of Statistics*, 36:1983–2023.

McKay, M. D. (1995). Evaluating prediction uncertainty. *Technical Report NUREG/CR-6311, US Nuclear Regulatory Commission and Los Alamos National Laboratory*, pages 1–79.

McKay, M. D., Conover, W. J., and Beckman, R. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.

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

Owen, A. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica*, 2:439–452.

Owen, A. B. (2012+). Variance components and generalized sobol' indices. *Preprint available at http://statistics.stanford.edu/∼ckirby/techreports/GEN/2012/2012-07.pdf*.

Qian, P. Z. G. (2009). Nested latin hypercube sampling. *Biometrika*, 96(4):957–970.

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145:208–297.

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

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Saisana, D. G. M., and Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*. John Wiley & Sons.

Saltelli, A., Tarantola, S., and Chan, K. P. S. (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41:39–56.

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

Stanley, R. P. (2012). *Enumerative Combinatorics, Volume 1 (2nd edition)*. Cambridge University Press.

Stein, M. (1987). Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):143–151.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, 93:964–979.

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

Tissot, J. Y. and Prieur, C. (2012+). Variance-based sensitivity analysis using harmonic analysis. *Preprint avalaible at http://hal.archives-ouvertes.fr/docs/00/68/07/25/PDF/FAST_RBD_revisited.pdf*.

Van der Vaart, A. W. (1998). *Asymptotics Statistics*. Cambridge University Press.