Iterative estimation of Sobol' indices based on replicated designs

Laurent Gilquin[1] · Clémentine Prieur[1] · Elise Arnaud[1] · Hervé Monod[2]

Abstract

In the field of sensitivity analysis, Sobol' indices are widely used to assess the importance of the inputs of a model to its output. Among the methods that estimate these indices, the replication procedure is noteworthy for its efficient cost. A practical problem is how many model evaluations must be performed to guarantee a sufficient precision on the Sobol' estimates. The present paper tackles this issue by rendering the replication procedure iterative. The idea is to enable the addition of new model evaluations to progressively increase the accuracy of the estimates. These evaluations are done at points located in under-explored regions of the experimental designs, but preserving their characteristics. The key feature of this approach is the construction of nested space-filling designs. For the estimation of first-order indices, a nested Latin hypercube design is used. For the estimation of closed second-order indices, two constructions of a nested orthogonal array design are proposed. Regularity and uniformity properties of the nested designs are studied.

Keywords Orthogonal array · Iterative estimator · Sensitivity analysis · Sobol' indices · Space-filling designs

Mathematics Subject Classification 49Q12 · 05B15

1 Introduction

Many mathematical models encountered in applied sciences involve numerous poorly-known inputs. It is important for the practitioner to understand how the output uncertainty can be apportioned to the uncertainty in the inputs. One way to do so is to perform a global sensitivity analysis in which statistical methods allow one to calculate importance measures. Among

✉ Clémentine Prieur
clementine.prieur@univ-grenoble-alpes.fr

1 Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP (Institute of Engineering Univ. Grenoble Alpes), LJK, 38000 Grenoble, France

2 Université Paris-Saclay, INRAE, MaIAGE, 78350 Jouy-en-Josas, France

the large number of available approaches, the variance-based method introduced by Sobol' (1993) relies on the calculation of sensitivity measures called Sobol' indices. The method is based on a variance decomposition of the model output into fractions which can be attributed to sets of inputs, assuming that the uncertainty on the sets of inputs is modeled by independent probability distributions. The influences of each set are summarized by the Sobol' indices which are scalars between 0 and 1. The higher the index the more influential the set. One can distinguish first-order indices that estimate the main effect of each set of inputs from higher order indices that estimate the corresponding order of interactions between sets of inputs. Various procedures have been proposed in the literature (see Saltelli et al. (2008) for a survey) to estimate Sobol' indices. They all rely on the choice of both an estimator of the Sobol' index and a design of experiments, simply referred to as design, that contains the points on which the model is evaluated. One key limitation of these methods is a need for a significant number of model evaluations to ensure a proper exploration of the input space. This comes at the price of rapidly being prohibitive.

The present paper offers a practical solution to this limitation through the introduction of an iterative method, referred to as iterative replication procedure. This procedure is designed to control the number of points on which the model is evaluated, adding new points at each iteration and stopping as soon as the Sobol' indices estimates have reached a convergence criterion. Such a procedure is possible as the formula of the Sobol' index estimator can be written iteratively. The proposed method relies on specific experimental designs referred as replicated designs.

The notion of replicated designs was introduced by McKay et al. (1979). Later on, Mara and Joseph (2008) combine these designs with "pick-freeze" estimators (Sobol' 1993) to estimate first-order Sobol' indices. This procedure, called replication procedure, has been further studied and generalized in Tissot and Prieur (2015) to the estimation of closed second-order indices. This last generalization relies on the construction of designs called orthogonal arrays (OA) (see Hedayat et al. 1999). The replication procedure has the notable advantage of requiring the construction of only two designs, thus considerably reducing the estimation cost of more classical methods. To better control the estimation cost, the replication procedure can be rendered iterative through the addition of new points at each iteration.

Adding new points is straightforward when the initial design is composed with independent and uniformly distributed points. Even in this framework, the issue of choosing the final sample size is crucial and has been addressed by several authors in the last decade. In Sarrazin et al. (2016), the authors define quantitative criteria to assess different types of convergence of GSA results, i.e. convergence of sensitivity indices, ranking and screening. In Sheikholeslami et al. (2019) (see also references therein), the authors combine agglomerative hierarchical clustering with bootstrapping and introduce a new robustness measure that enables an objective assessment of GSA convergence, even for high-dimensional input parameter space. In Terraz et al. (2017), the authors combine iterative statistics and in transit processing to compute sensitivity indices without any intermediate storage. However, in the replication procedure as introduced in Tissot and Prieur (2015), the initial design possesses either a structure of Latin hypercube or orthogonal array whether first- or closed second-order Sobol' indices are estimated. Hence, the need for a sampling strategy that preserves these structures along the iterations.

For the estimation of first-order Sobol' indices, the proposed approach takes advantage of an algorithm for the construction of nested Latin hypercubes introduced by Qian (2009). Dealing with the case of closed second-order indices is less straightforward. It requires nested orthogonal arrays whose construction procedure is neither limited by the input space dimensions or its discretization. Up to our knowledge, such a procedure has never been

proposed as existing algorithms do not meet these requirements (Qian et al. 2009b, a; Dey 2012). Two construction approaches are discussed in the present paper, both preserving a proper exploration of the input space. Each construction starts with an initial orthogonal array and updates it sequentially by adding a fixed number of new points.

The paper is organized as follows. Backgrounds on Sobol' indices are given in Sect. 2. The estimator formula and its iterative version are detailed. The replication procedure is described in Sect. 3, both for the estimation of first- and closed second-order indices. Section 4 is dedicated to the iterative replication procedure. A focus on the construction of space filling designs, both nested and replicated, is made: nested Latin hypercube for first-order indices, nested orthogonal arrays for second-order indices. The last section presents the numerical analysis. First, regularity and uniformity properties of the nested designs are studied. Then, the iterative replication procedure is illustrated on a toy example and an engineering application.

2 Iterative estimation of Sobol' indices

2.1 Definition of Sobol' indices

Consider the following model defined from a black box perspective:

$$f : \begin{cases} \mathbb{R}^d & \to & \mathbb{R} \\ \boldsymbol{x} = (x_1, \ldots, x_d) & \mapsto & y = f(\boldsymbol{x}) \end{cases} \tag{1}$$

where y is the output of the model f, $\boldsymbol{x}$ the input vector and d the dimension of the input space. Denote by $\subsetneq$ the proper (strict) inclusion symbol and by $\subseteq$ the inclusion symbol.

Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space. The uncertainty on the inputs is modeled by a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ whose components are independent. Let $u \subseteq \{1, \ldots, d\}$, $\boldsymbol{X}_u$ denotes a vector with components X_j, $j \in u$. Let $P_X = P_{X_1} \otimes \cdots \otimes P_{X_d}$ be the distribution of $\boldsymbol{X}$. Assuming that $f \in \mathbb{L}^2(P_X)$, the model f can then be uniquely decomposed into summands of increasing dimensions (functional ANOVA decomposition Sobol' 1993; Hoeffding 1948),

$$f(\boldsymbol{X}) = f_0 + \sum_j f_j(X_j) + \sum_{k<l} f_{k,l}(X_k, X_l) + \cdots + f_{1,\ldots,d}(X_1, \ldots, X_d), \tag{2}$$

where $\mathrm{E}[f_u(\boldsymbol{X}_u) f_w(\boldsymbol{X}_w)] = 0$, $\forall (u, w) \subseteq \{1, \ldots, d\}^2$, $u \neq w$. Denote $Y = f(\boldsymbol{X})$, this implies that $f_0 = \mathrm{E}[Y]$ and that the components are mutually orthogonal with respect to P_X. Let $u \subseteq \{1, \ldots, d\}$, each component is defined by:

$$f_u(\boldsymbol{X}_u) = \mathrm{E}[Y|\boldsymbol{X}_u] - \sum_{v \subsetneq u} f_v(\boldsymbol{X}_v).$$

The functional decomposition can be used to measure the global sensitivity of the output Y to $\boldsymbol{X}_u$. Let $\sigma^2 = \mathrm{Var}[Y]$. By squaring and integrating Eq. (2), due to orthogonality constraints, one gets:

$$\sigma^2 = \sum_j \sigma_j^2 + \sum_{k<l} \sigma_{k,l}^2 + \cdots + \sigma_{1,\ldots,d}^2, \tag{3}$$

where:

$$\sigma_u^2 = \mathrm{Var}[f_u(\boldsymbol{X}_u)] = \mathrm{Var}[\mathrm{E}[Y|\boldsymbol{X}_u]] - \sum_{v \subsetneq u} \sigma_v^2.$$

Then, it is natural to define, for each $u \subseteq \{1, \ldots, d\}$, the Sobol' index S_u as:

$$S_u = \frac{\sigma_u^2}{\sigma^2}.$$

Let $|u|$ denote the cardinal of u. The Sobol' index S_u measures the contribution to σ^2 of the interaction of order $|u|$ between the X_j, $j \in u$. Dividing Eq. (3) by σ^2, the following equality is obtained:

$$1 = \sum_{u \subseteq \{1, \ldots, d\}, u \neq \varnothing} S_u. \tag{4}$$

A straightforward implication of Eq. (4) is the following: if the sum of the first-order indices is close to 1 then the model is free of interaction effects. One can also define for each $u \subseteq \{1, \ldots, d\}$ the closed Sobol' index $\underline{S}_u$ by:

$$\underline{S}_u = \frac{\text{Var}[\text{E}[Y|X_u]]}{\sigma^2}.$$

The closed Sobol' index $\underline{S}_u$ measures the contribution of the X_j, $j \in u$, by themselves or in interaction with each other, to the variance σ^2 of the model output. As an example, if $u = \{k, l\}, k \neq l$, then,

$$\underline{S}_{k,l} = S_{k,l} + S_k + S_l. \tag{5}$$

Most of the time, no explicit formulation of Sobol' indices is available. These last must therefore be estimated.

2.2 Estimation of Sobol' indices

This sections briefly reviews the estimation procedure for closed Sobol' indices $\underline{S}_u$, $u \subseteq \{1, \ldots, d\}$ (see, e.g., Sobol' 1993). Consider X and X' two independent vectors distributed as the input vector. Let $u \subseteq \mathscr{D}$ and denote by $-u$ its complementary set in $\mathscr{D}$. The hybrid point $W = (X_u : X'_{-u})$ is defined by $W_j = X_j$ if $j \in u$ and $W_j = X'_j$ otherwise. The associated model outputs are defined by $Y = f(X)$, $Y_u = f(X_u : X'_{-u})$.

As underlined in Janon et al. (2014, Lemma 1.2), the Sobol' index $\underline{S}_u$ can be expressed as a correlation coefficient between Y and Y_u,

$$\underline{S}_u = \frac{\text{Cov}(Y, Y_u)}{\text{Var}[Y]}. \tag{6}$$

Most of the estimation procedures proposed in the literature are based on Eq. (6), replacing both the numerator and the denominator by one of their numerous estimators (see Prieur and Tarantola 2017 for a survey). The present paper focuses on the estimator introduced in Homma and Saltelli (1996), as it was proven in Janon et al. (2014) to have optimal asymptotic variance properties. More precisely, consider two designs of size n,

$$\mathscr{P} = \{X_i\}_{i=1}^{n}, \qquad \mathscr{P}' = \{X'_i\}_{i=1}^{n}.$$

$\mathscr{P}$ (resp .$\mathscr{P}'$) is a matrix where each row is a point $X_i = (X_{i,1}, \ldots, X_{i,d})$ (resp. X'_i) of the input space and each column contains n realizations $X_{i,j}$ of each input X_j, $j = 1, \ldots, d$. A third design $\mathscr{P}^u = \{X_{i,u} : X'_{i,-u}\}_{i=1}^{n}$ is constructed from $\mathscr{P}$ and $\mathscr{P}'$ by columns substitution. By evaluating the model with $\mathscr{P}$ and $\mathscr{P}^u$, n realizations of Y and Y_u are obtained, denoted

by $\{Y_i\}_{i=1}^n$ and $\{Y_{i,u}\}_{i=1}^n$. Then, the estimation of $\underline{S}_u$ reads:

$$
\widehat{\underline{S}}_u = \frac{\frac{1}{n}\sum_{i=1}^n Y_i Y_{i,u} - \left(\frac{1}{n}\sum_{i=1}^n Y_i\right)\left(\frac{1}{n}\sum_{i=1}^n Y_{i,u}\right)}{\frac{1}{n}\sum_{i=1}^n (Y_i)^2 - \left(\frac{1}{n}\sum_{i=1}^n Y_i\right)^2}. \tag{7}
$$

The main drawback of the aforementioned procedure is the high number of model evaluations required. Estimating all first-order (resp. all closed second-order) Sobol' indices costs $n(d+1)$ (resp. $n(\binom{d}{2}+1)$) model evaluations. The larger n, the more accurate the estimation of the indices.

In Saltelli (2002, Theorem 2), subtle combinatorial arguments are used, which allow the estimation of all first-order and closed second-order indices (among others) at a linear cost of $n(2d+2)$ model evaluations.

More recently, the replication procedure–procedure based on replicated designs—has been proposed to get rid of the dependence in the input space dimension d. All first-order and all closed second-order Sobol' indices are estimated with $4n$ model evaluations (see Mara and Joseph 2008; Tissot and Prieur 2015).

2.3 Iterative scheme

For the sake of both storage ability and computational cost, it is interesting to design an iterative scheme for the estimation of Sobol' indices. The underlying procedure consists in sequentially adding new points to perform new model evaluations and stopping once the estimates have reached a convergence criterion. The iterative scheme relies on an iterative version of Eq. (7) and operates as follows.

Consider a pair of nested designs $\left(\mathscr{P}_\ell, \mathscr{P}'_\ell\right)_{l\geq 0}$. Denote by B_ℓ, $\ell \geq 0$, a block, that is a set of points on which the model is evaluated. The two nested designs are built by blocks concatenation, according to the following scheme:

$$
\begin{cases} \mathscr{P}_{-1} = \varnothing \\ \mathscr{P}_\ell = \mathscr{P}_{\ell-1} \cup B_\ell \end{cases}, \quad \begin{cases} \mathscr{P}'_{-1} = \varnothing \\ \mathscr{P}'_\ell = \mathscr{P}'_{\ell-1} \cup B'_\ell \end{cases}, \quad \ell \geq 0. \tag{8}
$$

At iteration ℓ, the new blocks B_ℓ and B'_ℓ contain each m_ℓ new points and $n_\ell = \sum_{k=0}^\ell m_k$ denotes the common size of $\mathscr{P}_\ell$ and $\mathscr{P}'_\ell$. Keeping the notation introduced in Sect. 2.2,

$$
\mathscr{P}_\ell = \{X_i\}_{i=1}^{n_\ell}, \ \ \mathscr{P}'_\ell = \{X'_i\}_{i=1}^{n_\ell}, \ \ \mathscr{P}^u_\ell = \{X_{i,u} : X'_{i,-u}\}_{i=1}^{n_\ell} \text{ for } u \subseteq \{1, \ldots, d\}.
$$

The model evaluations obtained with $\mathscr{P}_\ell$ and $\mathscr{P}^u_\ell$ read:

$$
\bigcup_{k=0}^\ell \{Y^i\}_{i=n_{k-1}+1}^{n_k}, \qquad \bigcup_{k=0}^\ell \{Y^i_u\}_{i=n_{k-1}+1}^{n_k},
$$

with $n_{-1} = 0$. Then, $\underline{S}_u$ is estimated with:

$$
\widehat{\underline{S}}_u^{(\ell)} = \frac{\phi_\ell - \psi_\ell \xi_\ell}{V_\ell}, \tag{9}
$$

Algorithm 1 Iterative estimation of Sobol' indices

1: $\ell \leftarrow 0, \widehat{\underline{S}}_u^{(0)} \leftarrow 0, \texttt{test} \leftarrow \texttt{true}$
2: $\mathscr{P}_0 \leftarrow B_0, \mathscr{P}'_0 \leftarrow B'_0$
3: **while** test **do**
4:    **for** $u \subset \mathscr{D}$ **do**
5:        Compute $\{Y^i\}_{i=n_{l-1}+1}^{n_l}$ and $\{Y_u^i\}_{i=n_{l-1}+1}^{n_l}$ from B_l and B'_l
6:        Evaluate $\widehat{\underline{S}}_u^{(\ell)}$ with (9)
7:    **end for**
8:    $\texttt{test} \leftarrow$ stopping criterion
9:    $\mathscr{P}_{\ell+1} \leftarrow \mathscr{P}_\ell \cup B_{\ell+1}$
       $\mathscr{P}'_{\ell+1} \leftarrow \mathscr{P}'_\ell \cup B'_{\ell+1}$
10:   $\ell \leftarrow \ell + 1$
11: **end while**
12: Return the Sobol' estimates

---

where $\phi_\ell, \psi_\ell, \xi_\ell$ and V_ℓ are defined by the following recursive formulae:

$$
\begin{cases}
n_\ell = n_{\ell-1} + m_\ell, \\
n_\ell \phi_\ell = n_{\ell-1} \phi_{\ell-1} + \displaystyle\sum_{i=n_{\ell-1}+1}^{n_\ell} Y^i Y_u^i, \\
n_\ell \psi_\ell = n_{\ell-1} \psi_{\ell-1} + \displaystyle\sum_{i=n_{\ell-1}+1}^{n_\ell} Y^i, \\
n_\ell \xi_\ell = n_{\ell-1} \xi_{\ell-1} + \displaystyle\sum_{i=n_{\ell-1}+1}^{n_\ell} Y_u^i, \\
n_\ell V_\ell = n_{\ell-1}(V_{\ell-1} + \psi_{\ell-1}^2) + \displaystyle\sum_{i=n_{\ell-1}+1}^{n_\ell} (Y^i)^2 - n_\ell \psi_\ell^2,
\end{cases}
$$

and $n_{-1} = 0, \phi_{-1} = 0, \psi_{-1} = 0, \xi_{-1} = 0, V_{-1} = 0$. It is important to note that the expressions of $\phi_\ell, \psi_\ell, \xi_\ell$ and V_ℓ only involve the model evaluations obtained with the latest couple of replicated blocks. Hence, the opportunity to store only the last subset of model evaluations instead of the whole set.

Algorithm 1 summarizes the main steps of the iterative scheme. The set $\mathscr{D}$ at step 4 equals either $\{1, \ldots, d\}$ or $\{(k, l) \in \{1, \ldots, d\}^2; \ k < l\}$ whether first-order or closed second-order indices are estimated. In any case, the cost of this iterative procedure equals $2 \times \sum_{k=0}^{K} m_k$ if stopped at step K.

Stopping criterion   The form of the stopping criterion, variable test in Algorithm 1, is based on a reasonable heuristic convergence criterion. At each iteration $\ell \geq 1$, the following quantity is evaluated:

$$
\mathbf{e}^{(\ell)} = \left| \widehat{\underline{S}}_u^{(\ell)} - \widehat{\underline{S}}_u^{(\ell-1)} \right|, \tag{10}
$$

where $|.|$ denotes the absolute value function. Equation (10) reads as the positive difference between two consecutive estimation of a Sobol' index. Let ℓ_0 and $\ell_{\max}$ be two positive integers. The former defines a number of consecutive iterations, the latter defines the maximum number of iterations allowed in Algorithm 1. Let $\varepsilon > 0$, the algorithm stops at then end of step $\ell \geq \ell_0$ if:

$$
\ell = \ell_{\max} - 1 \text{ or } \forall u \in \mathscr{D} : \mathbf{e}^{(\ell-(\ell_0-1))} < \varepsilon, \ldots, \mathbf{e}^{(\ell)} < \varepsilon. \tag{11}
$$

In other words, Eq. (11) checks if all the Sobol' estimates have stopped significantly varying over the last ℓ_0 iterations or if a maximum number of iterations has been reached. The parameters ε, ℓ_0 and $\ell_{\max}$ have to be tuned adequately.

At this point, the iterative scheme is generic enough that it could be combined with most of the classical estimation procedures. The main contribution of this paper is to adapt the present iterative scheme for the replication procedure, resulting in an iterative version of the replication procedure. In this particular framework, the iterative scheme needs to be tuned so that at each iteration, the designs $\mathscr{P}_\ell$ and $\mathscr{P}'_\ell$ preserve the structure of the designs used in the replication procedure. This is the topic of Sects. 3 and 4. The former briefly reviews the original replication procedure while the latter introduces its iterative version.

3 Replication procedure

Without loss of generality, the inputs $X_1, \ldots, X_d$ are assumed to be independent random variables uniformly distributed on $[0, 1]$. Note that the replication procedure can easily be generalized to any marginal distributions through many procedures, for instance the Smirnov transform (see Devroye 1986 for others approaches). The replication procedure owns its name to the nature of the designs it relies on, the so-called replicated designs:

Definition 1 Let $\mathscr{P} = \{X_i\}_{i=1}^n$ and $\mathscr{P}' = \{X'_i\}_{i=1}^n$ be two non-identical designs in $[0, 1]^d$. $\mathscr{P}$ and $\mathscr{P}'$ are two replicated designs of order p, if for any $u \subset \{1, \ldots, n\}$ such that $|u| = p$, there exists a permutation π_u of $\{1, \ldots, n\}$ such that $\forall i \in \{1, \ldots, n\}$, $x_{i,u} = x'_{\pi_u(i),u}$.

Example   Consider the two designs:

$$\mathscr{P} = \begin{pmatrix} 0.08 & 0.46 & 0.21 \\ 0.15 & 0.77 & 0.43 \\ 0.89 & 0.30 & 0.05 \\ 0.70 & 0.23 & 0.95 \end{pmatrix}, \qquad \mathscr{P}' = \begin{pmatrix} 0.89 & 0.30 & 0.95 \\ 0.15 & 0.23 & 0.21 \\ 0.70 & 0.46 & 0.43 \\ 0.08 & 0.77 & 0.05 \end{pmatrix}.$$

$\mathscr{P}$ and $\mathscr{P}'$ are two replicated designs of order 1. Indeed, for any $j \in \{1, \ldots, d\}$, the j-th columns of $\mathscr{P}$ and $\mathscr{P}'$ share the same unordered set of values. In particular, the permutation $\pi_1 = (4, 2, 1, 3)$ orders the first column of $\mathscr{P}'$ into the first column of $\mathscr{P}$.

The key point of the replication procedure is to use the permutation π_u of Definition 1 to mimic the vector $\{Y_{i,u}\}_{i=1}^n$ of Eq. (7). This is done as follows. Denote by $\{Y_i\}_{i=1}^n$ and $\{Y'_i\}_{i=1}^n$ the two sets of model evaluations obtained with $\mathscr{P}$ and $\mathscr{P}'$, respectively. From Definition 1, it results that,

$$Y'_{\pi_u(i)} = f(X'_{\pi_u(i),u} : X'_{\pi_u(i),-u}), = f(X_{i,u} : X'_{\pi_u(i),-u}).$$

Hence, while Y_i is evaluated on X_i, $Y'_{\pi_u(i)}$ is evaluated on $X'_{\pi_u(i)}$, where the components indexed by $i \in u$ are frozen and the $d - |u|$ other components are resampled. Therefore, the Sobol' index $\underline{S}_u$ can be estimated by applying Eq. (7) with $Y'_{\pi_u(i)}$ in place of $Y_{i,u}$ without requiring further model evaluations. Applying this combinatorial trick with all u, it is possible to estimate all closed Sobol' indices $\{\underline{S}_u\}_{u \subseteq \{1,\ldots,d\}, |u|=p}$ by evaluating the model on the set of $2n$ points contained in $\mathscr{P}' \cup \mathscr{P}$.

In Tissot and Prieur (2015), the authors suggest the use of two replicated Latin hypercube designs to estimate all first-order Sobol' indices, and the use of two replicated orthogonal arrays of strength two to estimate all closed second-order Sobol' indices. The choice and description of such designs in discussed below.

3.1 Estimation of first-order indices

Different strategies can be applied to build two replicated designs of order 1. In Mara and Joseph (2008), $\mathscr{P}$ and $\mathscr{P}'$ are composed with i.i.d points. In Tissot and Prieur (2015), the authors propose to use Latin hypercube designs insuring most of the time a better exploration of the input space:

Definition 2 (*Latin hypercube design*) Denote by Π_n the set of all the permutations of $\{1, \ldots, n\}$ and let $\pi_1, \ldots, \pi_d$ be d independent random variables uniformly distributed on Π_n. $\mathscr{P} = \{X_i\}_{i=1}^n$ is a Latin hypercube design if:

$$X_i = \left(\frac{\pi_1(i) - U_{i,1}}{n}, \ldots, \frac{\pi_d(i) - U_{i,d}}{n} \right), \tag{12}$$

where the $U_{i,j}$ are independent random variables uniformly distributed on $[0, 1]$ and independent of the π_j.

To estimate all first-order Sobol' indices, the authors in Tissot and Prieur (2015) first create a Latin hypercube design $\mathscr{P}$, in the sense of Definition 2 above. They then create a replicated design, $\mathscr{P}'$, by permuting independently the values of each column of $\mathscr{P}$. A nice property is that the replicated design $\mathscr{P}'$ is also a Latin hypercube design. The designs $\mathscr{P}$ and $\mathscr{P}'$ are replicated designs of order 1.

3.2 Estimation of closed second-order indices

The generalization of the replication procedure to the estimation of closed second-order indices was introduced in Tissot and Prieur (2015). It requires the construction of two replicated designs of order two. The construction in Tissot and Prieur (2015) relies on orthogonal arrays of strength two. The definition of an orthogonal array given in Hedayat et al. (1999, Definition 1.1) is recalled hereafter:

Definition 3 (*Orthogonal array*) A $n \times d$ array $A = \{A_i\}_{i=1}^n$, $A_i = (A_{i,1}, \ldots, A_{i,d})$, with values from a set S of cardinality q is said to be an orthogonal array with q levels, strength t ($0 \leq t \leq d$) and index λ if every $n \times t$ sub-array of A contains each t-tuple based on S exactly λ times as a row. The orthogonal array A satisfies $n = \lambda q^t$. It is denoted by $OA_\lambda(q, d, t)$.

The construction of two replicated designs $\mathscr{P}$ and $\mathscr{P}'$ of order two proposed in Tissot and Prieur (2015) starts with the construction of an orthogonal array A with q levels, strength $t = 2$ and index $\lambda = 1$. The q levels are then substituted by $1, \ldots, q$. Then, A' is obtained by permuting independently the values of each column of A. Denote by $\lozenge$ the operator achieving this rearrangement:

$$A' = \lozenge(A, \{\pi_1, \ldots, \pi_d\}) \Leftrightarrow A_i' = \left(\pi_1(A_{i,1}), \ldots, \pi_d(A_{i,d}) \right), \; i = 1, \ldots, n. \tag{13}$$

Construction procedures for orthogonal arrays can be found, e.g., in Hedayat et al. (1999).

In some sense, A and A' are replicated designs of order two, except that they are not valued in $[0, 1]^d$ but in $\{1, \ldots, q\}^d$. Both designs $\mathscr{P}$ and $\mathscr{P}'$ are obtained from A and A' by a randomization procedure, described in Definition 4 below.

Definition 4 (*Randomized replicated orthogonal arrays*) Let $A = \{A_i\}_{i=1}^{q^t}$ be an $OA_1(q, d, t)$. Denote by Π_q the set of all the permutations of $\{1, \ldots, q\}$ and let $\pi_1, \ldots, \pi_d$ be d independent random variables uniformly distributed on Π_q. $\mathscr{P} = \{X_i\}_{i=1}^{q^t}$ and $\mathscr{P}' = \{X'_i\}_{i=1}^{q^t}$ are

two replicated orthogonal arrays if:

$$X_i = \left(\frac{A_{i,1} - U_{A_{i,1},1}}{q}, \ldots, \frac{A_{i,d} - U_{A_{i,d},d}}{q} \right),$$

$$X'_i = \left(\frac{A'_{i,1} - U_{A'_{i,1},1}}{q}, \ldots, \frac{A'_{i,d} - U_{A'_{i,d},d}}{q} \right), \tag{14}$$

where the $U_{i,j}$ are independent random variables uniformly distributed on [0, 1] and independent of the π_j.

Let $\mathscr{P}$ and $\mathscr{P}'$ be two randomized replicated orthogonal arrays built on $A = \{A_i\}_{i=1}^{q^2}$, an $OA_1(q, d, 2)$. Based on $\mathscr{P}$ and $\mathscr{P}'$, the replication procedure can be applied to estimate all closed second-order Sobol' indices with $2n$ model evaluations, with $n = q^2$.

Then, both sets of all first-order and all closed second-order indices can be estimated at a cost of $2n + 2q^2$ model evaluations.

4 Iterative replication procedure using replicated pair of nested designs

This section describes how the iterative scheme of Sect. 2.3 (algorithm 1) can be applied within the replication framework ; at the cost that, at each step ℓ, the pair of designs $\left(\mathscr{P}_\ell, \mathscr{P}'_\ell \right)$ has to be replicated. For the estimation of first-oder indices, a replicated pair of nested Latin hypercube designs is used. Its construction relies on an algorithm proposed by Qian (2009) and is detailed in Sect. 4.1. As for the estimation of closed second-order indices, a replicated pair of nested orthogonal arrays of strength two is used. Its construction is presented in Sect. 4.2.

4.1 Latin hypercube designs, replicated and nested

Recent developments in Latin Hypercube Sampling (LHS) enable sample size extension using Hierarchical Latin Hypercube Sampling (HLHS) (Sarrazin et al. 2016; Sallaberry et al. 2008; Vorechovsky et al. 2013). Given a LHS of size n, the strata of each sample component are further divided into $t + 1$ strata (one containing the original sample) and the new components are randomly paired as in a typical LHS implementation. Parameter t is referred to as the refinement factor and the sample size after r sample size extensions is $n \times (t + 1)^r$.

In the present paper, we rather focus on the algorithm introduced in Qian et al. (2009, Section 5), which offers a practical solution to iteratively augment the number of points of a Latin hypercube design while preserving its Latin hypercube structure. The resulting design, called nested Latin hypercube design, is partitioned into blocks that define multiple layers (see also Rennen et al. 2010 for the construction of nested maximin LHS).

As an illustration, a two-dimensional nested Latin hypercube design with three layers is presented in Fig. 1. Each layer possesses a Latin hypercube structure in a grid progressively refined. An important point to note is that the number of layers (therefore blocks) partitioning the design and the layer size have to be specified beforehand. This intrinsically defines the maximum number of iterations $\ell_{\max}$ in Algorithm 1 as well as the discretization of the input space at each iteration.

For the estimation of first-order indices, the two nested designs $\mathscr{P}_\ell$ and $\mathscr{P}'_\ell$ of Eq. (8) are created from block concatenation and replicated as follows. First, the block B_ℓ is constructed
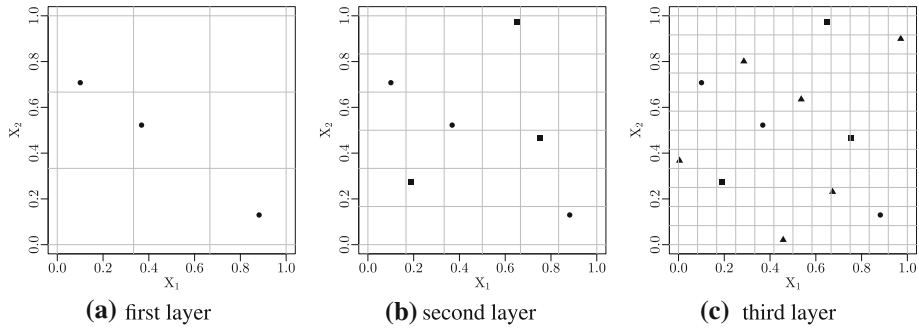
Fig. 1 Nested Latin hypercube design with three layers (**a–c**). The symbols (circle, square, triangle) identify the new points (i.e the blocks) added at each iteration

using the algorithm in Qian (2009). Then, the associated block B'_ℓ is replicated following the process described in Sect. 3.1, that is by permuting independently the values in each column of B_ℓ. This single operation guarantees that at each iteration $\ell \geq 0$, the two designs $\mathscr{P}_\ell$ and $\mathscr{P}'_\ell$ possess a structure of Latin hypercube and are two replicated designs of order 1. In the present paper, the size considered for $\mathscr{P}_\ell$ and $\mathscr{P}'_\ell$ is a power of two. Thus, the smallest size one can select for $\mathscr{P}_\ell$ and $\mathscr{P}'_\ell$ equals $2^{\ell+1}$.

4.2 Randomized orthogonal arrays of strength two, replicated and nested

As mentioned in Sect. 3.2, the replication procedure requires two replicated orthogonal arrays of strength two for the estimation of closed second-order indices. Therefore, a nested version of these designs has to be devised to use the iterative scheme of Eq. (8). Various construction strategies have already been proposed in the literature, in Qian et al. (2009a, b) and Dey (2012) notably. The strategies proposed in these papers have at least one of the following restrictions:

- The size of the initial design is rather large, hence at each step, only a too large number of new points can be added.
- The construction deals only with specific values of the input space dimension d.
- The discretization is not the same in each dimension, more precisely only one dimension is finely discretized at the expense of the others.

An alternative approach is presented in this section, free from all these limitations. The nested orthogonal array built is an $OA_\lambda(q, d, 2)$ where $\lambda > 1$. It can be partitioned into λ $OA_1(q, d, 2)$, from which one can sample λ blocks by applying the randomization process detailed in Definition 4.

As in the construction of nested Latin Hypercube, the idea of the present approach is to fix the final input space discretization beforehand. In other words, this defines the final d-hypercube that will be progressively filled with λ row-wised distinct $OA_1(q, d, 2)$. By row-wised distinct it is to be understood that if two $OA_1(q, d, 2)$ among the λ are chosen then they do not share a common row. More precisely, two rows are said distinct if they differ in at least one component. A visual interpretation of a row is proposed in Fig. 2, where each row of an orthogonal array is viewed as a sub-hypercube. This figure illustrates the example of an $OA_3(3, 3, 2)$, partitioned into three distinct $OA_1(3, 3, 2)$, each, respectively, plotted in each subfigure.
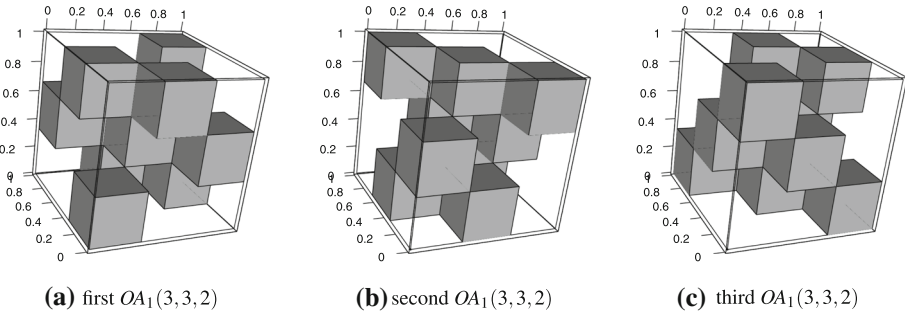
(a) first $OA_1(3,3,2)$ **(b)** second $OA_1(3,3,2)$ **(c)** third $OA_1(3,3,2)$

Fig. 2 $OA_3(3,3,2)$ that can be partitioned into three distinct $OA_1(3,3,2)$. **a** First $OA_1(3,3,2)$. **b** Second $OA_1(3,3,2)$. **c** Third $OA_1(3,3,2)$

Algorithm 2 *Accept–reject* method for the construction of A_ℓ

1: Set `bool` $\leftarrow$ false
2: **while** !`bool` **do**
3:     Sample $\pi_1, \ldots, \pi_d$ in Π_q
4:     Construct $A_\ell = \Diamond(A_0, \{\pi_1, \ldots, \pi_d\})$ with (13)
5:     **for** $k = 0, \ldots, \ell - 1$ **do**
6:         `bool`$_k \leftarrow rows(A_\ell) \cap rows(A_k) == \varnothing$
7:     **end for**
8:     `bool` $\leftarrow \forall k : $ `bool`$_k$
9: **end while**

The present approach to build the sequence of nested randomized orthogonal array of strength two $\mathscr{P}_\ell$ and $\mathscr{P}'_\ell$ follows the following steps:

1. First, construct an initial $OA_1(q, d, 2)$ noted A_0.
2. Then at each step $\ell \geq 1$, a new $OA_1(q, d, 2)$, noted A_ℓ, is constructed from A_0. Two methods may be used to ensure that A_ℓ is distinct from A_k, $0 \leq k < \ell$. They will be referred as the *accept–reject* and the *algebraic* methods and are detailed below.
3. Substituting A_ℓ for A in Definition 4, i.e. by randomization of the OA, one can obtain the two blocks B_ℓ and B'_ℓ. These two blocks define the new points on which the model will be evaluated. The fact that the OA's are distinct ensures that these points are located in unexplored regions of the input space.

As a result, $\mathscr{P}_\ell$ and $\mathscr{P}'_\ell$ both have a structure of $OA_\ell(q, d, 2)$ and are replicated designs of order 2. Additionally, their size equal $(\ell + 1) \times q^2$. The *accept–reject* and the *algebraic* methods differ on the way A_ℓ is constructed. This last step is detailed hereafter for each method.

Method 1:   *accept–reject* The pseudo-code described in Algorithm 2 details the construction of A_ℓ. It uses the operator $\Diamond$ defined in Eq. (13). The idea is to randomly construct a new orthogonal array from A_0 using $\Diamond$ and to test if its rows are distinct from those of the previous orthogonal arrays constructed; namely $A_{\ell-1}, A_{\ell-2}, \ldots, A_0$. Note that this test become computationally expensive for small input space dimensions as the probability of acceptation decreases faster.

Method 2: *Algebraic method* From now on, the set S of the q levels of an orthogonal array (see Definition 3) is identified with the Galois field of order q, denoted by $GF(q)$, where q is a prime number or a power of a prime number ($q = p^{\alpha}$, p prime and $\alpha \in \mathbb{N}$).

Define the following set:

$$C = \left\{ g = (0, 0, g_3, \ldots, g_d) \mid \forall i \geq 3, \ g_i \in GF(q) \right\} \subsetneq GF(q)^d.$$

The pseudo-code described in Algorithm 3 below details the construction of A_{ℓ}. $\oplus$ denotes the addition in $GF(q)^d$.

Algorithm 3 *Algebraic* method for the construction of A_{ℓ}

1: Choose $g_{\ell} \in C$
2: Construct $A_{\ell} = g_{\ell} A_0 = \left\{ g_{\ell} \oplus A_{0i} \right\}_{i=1}^{q^2}$, $A_{0i} = (A_{0i,1}, \ldots, A_{0i,d})$
3: $C \leftarrow C \setminus \{g_{\ell}\}$

The idea of the *algebraic* method is to construct a partition of the discretized input space and select A_{ℓ} from this partition. A_{ℓ} is viewed as a coset of A_0 and is drawn from the set C. Proposition 1 below guarantees that A_{ℓ} constructed in Variant 3 is an $OA_1(q, d, 2)$.

Proposition 1 *Consider A_0 an $OA_1(q, d, 2)$ based on $GF(q)^d$. The results are as follows:*

i) $\forall g \in GF(q)^d$, $g A_0$ is an $OA_1(q, d, 2)$
ii) $\forall g, g' \in C$, such that $g \neq g'$, $g A_0 \cap g' A_0 = \varnothing$. In other words, the sets $\{g A_0\}$ form a partition of $GF(q)^d$.

Proof (i) Let $g = (g_1, \ldots, g_d) \in GF(q)^d$. Consider A_{0k}, A_{0l} two columns of A_0. Denote by E the group $(GF(q), +)$. Since $g_k E \times g_l E$ is isomorph to $E \times E$, the 2-tuples $(A_{0i,k} + g_k, A_{0i,l} + g_l)$ obtained after addition are all two by two distinct.
(ii) The proof can be found in Stinson and Massey (1995) where an orthogonal array is regarded as a "systematic linear code". $\blacksquare$

The main advantage of the *algebraic* method is that the maximum number of blocks one can construct is known beforehand. Indeed, as a consequence of Proposition 1-(ii), the maximum number of blocks one can construct equals the cardinality of C, that is q^{d-2}. If this upper bound is reached, the blocks $A_0, A_1, \ldots, A_{q^{d-2}-1}$ form a partition of the discretized input space.

5 Numerical tests and illustrations

The first part of this section focuses on the space-filling properties of the nested design used in our iterative replication procedure. Then, two applications of the proposed algorithm are detailed, on a test function and on an engineering example.

5.1 Space-filling properties

Three criteria are selected to study the properties of nested designs: maximin (Jonshon et al. 1990), emst (euclidean minimal spanning tree Franco et al. 2009) and L^2 star discrepancy

Fig. 3 (Color online) Interpretation graph of the emst criterion. The i.i.d. uniform sampling is used as the reference distribution
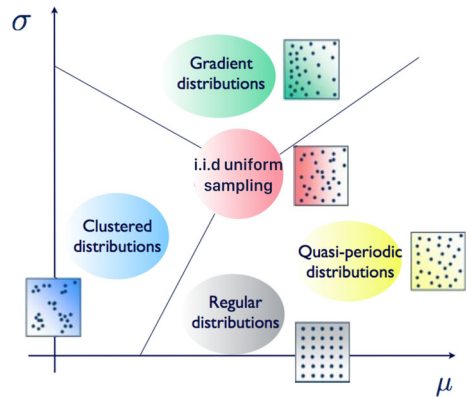
(Morokoff and Caflisch 1994). The maximin criterion returns the minimum distance among all pairs of points of a design. It can be interpreted as follows: the higher the value, the more regular the scattering of design points. The emst criterion can be interpreted using a (μ, σ) graph (see Fig. 3), called interpretation graph. A minimal spanning tree is constructed from the design, then mean (μ) and standard deviation (σ) of the tree edges lengths are evaluated. A value of the emst criterion is represented as a point in the (μ, σ) graph. The i.i.d. uniform sampling is used as a reference. A design having a higher value for μ and a smaller value for σ than those of a uniform design is said more regular. Maximin and emst criteria provide together a good estimation of a design regularity. The L^2 star discrepancy criterion measures the uniformity property of a design. The smaller the value, the more uniform the design.

5.1.1 Nested LHd in low-dimensional space

First, space-filling properties of the nested Latin hypercube design (nested LHd) are compared with those of (i) a uniform design (obtained through i.i.d uniform sampling) and (ii) a Latin hypercube design (LHd).

Figure 4 shows the results obtained with the three criteria for each design. The input space dimension d equals 5. The sizes n of each design equal $(2^3, 2^4, \ldots, 2^{10})$. For the nested LHd, these sizes correspond to those of design $\mathscr{P}_\ell$ augmented over 8 consecutive steps. Results for the LHd and the nested LHd are overall similar and both better than results for the uniform design. As such, the iterative replication procedure comes with no loss in terms of space-filling properties of the designs.

Remark   One other class of designs well suited for the estimation of first-order Sobol' indices are low discrepancy sequences. These sequences are points sets sampled so as to approximate as close as possible a uniform distribution and are known to achieve both uniformity and regularity properties. Such sequences could be used in the iterative replication procedure in place of nested Latin hypercube designs. This alternative has recently been studied in Gilquin et al. (2017).

5.1.2 Nested LHd in high-dimensional space

The comparison conducted in the last section is repeated anew with an input space of dimension 20. The size n of each design equals $(2^5, \ldots, 2^{10})$. Since the discrepancy criterion
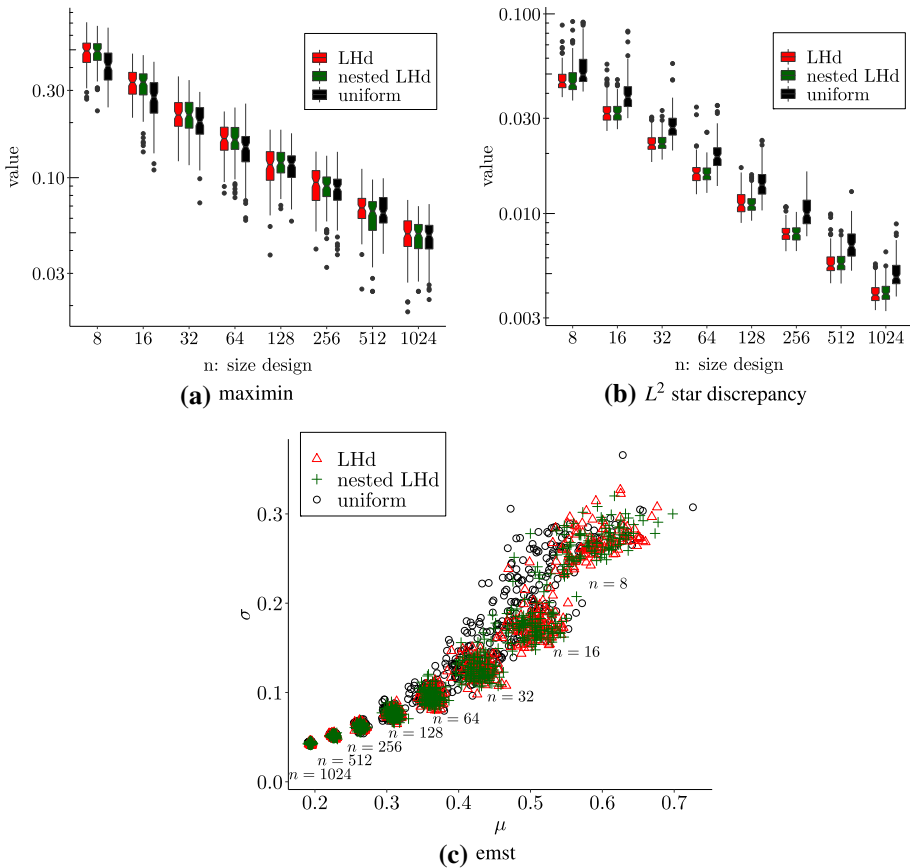
Fig. 4 (Color online) Results of maximin, star discrepancy and emst criteria over 100 repetitions for different sizes n of the designs used for the estimation of first-order indices. Logarithmic scales are used for the y-axis of graphs (**a**, **b**)

requires the number of points to be greater than the input space dimension, the first design size equals 2^5. Figure 5 shows the results obtained with the three criteria for each design.

The results are more mitigated. For the maximin and discrepancy criteria the three designs perform close to identical. The emst criterion shows that LHd and nested LHd perform better for small design sizes. Remark that the three criteria studied are subject to the "curse of dimensionality" as distances provide poorer contrasts between point neighbors in high-dimensional spaces. This phenomenon is a possible explanation as to why the results appear so similar.

5.1.3 Nested randomized OA of strength two

A third comparison is carried out between the following designs: (i) uniform design, (ii) "non-iterative" OA, (iii) *accept–reject* and (iv) *algebraic*. Design (ii) refers to the orthogonal array used in Tissot and Prieur (2015). Designs (iii) and (iv) refer to the design constructed with either the *accept–reject* or the *algebraic* method of Sect. 4.2. The input space dimension
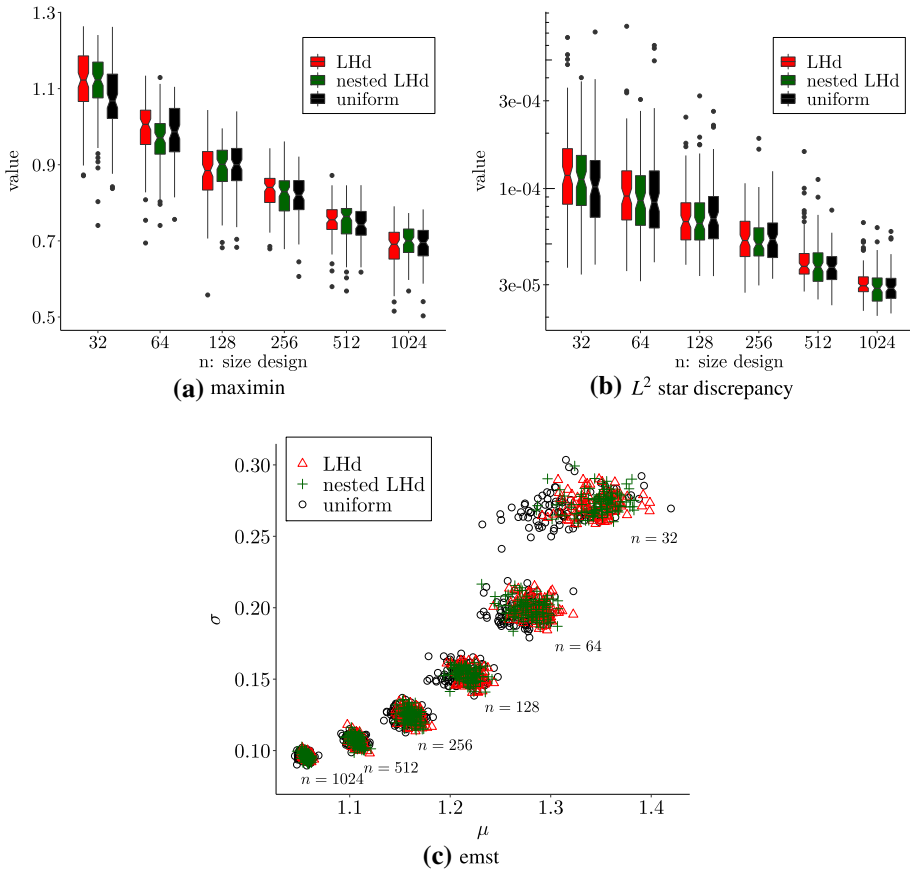
Fig. 5 (Color online) Results of maximin, star discrepancy and emst criteria over 100 repetitions for different sizes n of the designs used for the estimation of first-order indices. A logarithmic scale is used for the y-axis of graph (**b**)

still equals 5. The level q of the nested OA equals 8. Figure 6 shows the results obtained with each of the three criteria.

For the sake of visualization, results are represented only for the following sizes of the designs: $(2 \times 8^2, 3 \times 8^2, 5 \times 8^2, 8 \times 8^2, 11 \times 8^2, 16 \times 8^2)$. In terms of emst and discrepancy criteria, the "non-iterative" OA gives the best results while results for the *accept–reject* and *algebraic* designs are similar. The *algebraic* design gives better results for the maximin criterion than the *accept–reject* design but shows a greater disparity for the emst criterion.

The main conclusion is that the *algebraic* design possesses regularity and uniformity properties overall slightly better than those of the *accept–reject* design. These two designs possess slightly worse space-filling properties than their counterpart used in Tissot and Prieur (2015). This difference can be explained by the lack of progressive discretization of the inputs in both the *algebraic* and the *accept–reject* method. However, that is largely offset by the possibility to perform an iterative estimation of the indices.
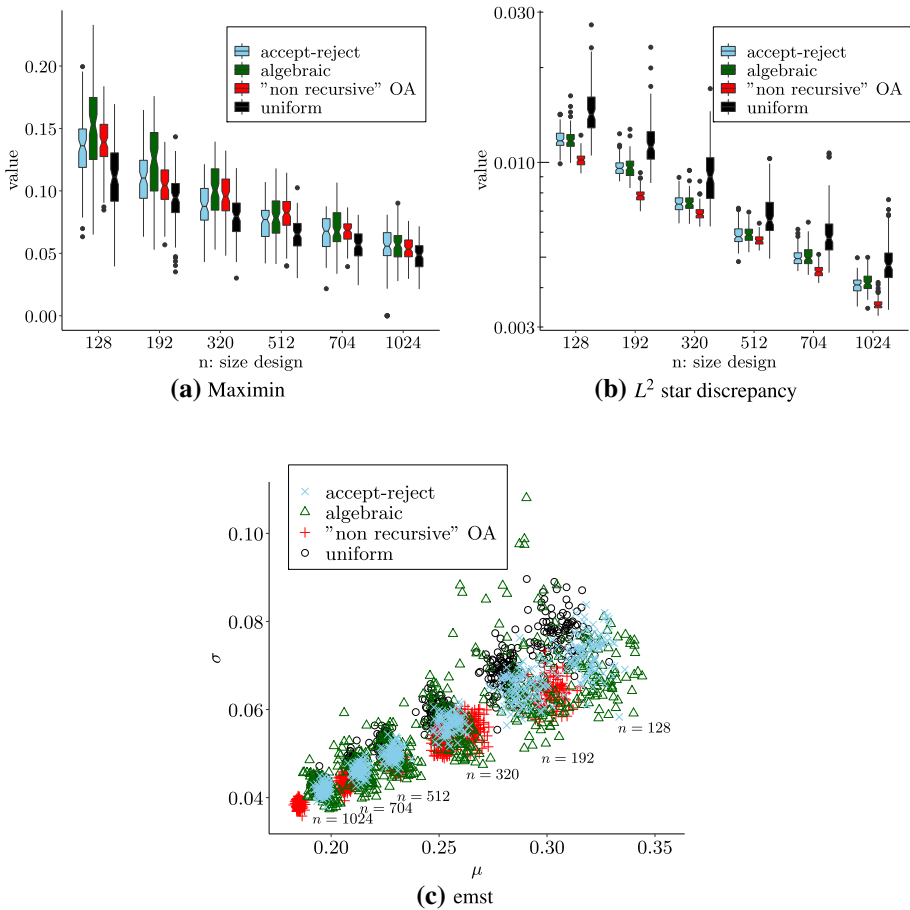
(a) Maximin

(b) L^2 star discrepancy

(c) emst

Fig. 6 (Color online) Results of maximin, emst and star discrepancy criteria over 100 repetitions for different sizes n of the designs used for the estimation of closed second-order indices. A logarithmic scale is used for the y-axis of graph (**b**)

5.2 Application to a toy example

The iterative replication procedure is tested and compared to the classical replication procedure using the Bratley et al. function (Bratley and Niederreiter 1992),

$$f(X_1, \ldots, X_d) = \sum_{i=1}^{d} (-1)^i \prod_{k=1}^{i} X_k,$$

where $X_1, \ldots, X_d$ are independent random variables uniformly distributed on $[0, 1]$. Both first- and closed second-order Sobol' indices are estimated with each procedure. Both procedures are repeated $r = 100$ times to get samples of estimates. The input space dimension is $d = 6$. Since f has an analytical expression, theoretical values of the Sobol' indices can be precisely calculated through symbolic integrals evaluations.

Recall ℓ_0 and $\ell_{\max}$ the two parameters of the stopping criterion defined by Eq. (11). Let K be any integer such that $\ell_0 \leq K \leq \ell_{\max} - 1$. Let r_K denote the number of repetitions where

the estimation procedure stopped at step K, then $\sum_{K=\ell_0}^{\ell_{\max}-1} r_K = r = 100$. Let $\alpha \in (0,1)$ and $\widetilde{q}_\alpha$ denote the empirical quantile of order α defined by:

$$\widetilde{q}_\alpha = \inf\{\, v \in \mathbb{N},\ v \geq \ell_0,\ r_{\ell_0} + \cdots + r_v \geq \alpha r \,\}.$$

To reach a fair comparison, $\underline{S}_u$ is also estimated $r = 100$ times with the classical replication procedure where the size of the two replicated designs equals the size of the designs $\mathscr{P}_{\widetilde{q}_{1/2}}$ and $\mathscr{P}'_{\widetilde{q}_{1/2}}$ in the iterative procedure.

On the choice of the hyper-parameters   The proposed iterative procedure possesses three hyper-parameters: ε, ℓ_0 and $\ell_{\max}$. Consider a budget of $2N$ evaluations, half of it used for the estimation of the first-order indices and the other half used for the estimation of the closed second-order indices. The choice of these parameters is discussed hereafter from a practitioner's perspective.

$\ell_{\max}$ should be set according to the budget of evaluations. For the estimation of first-order indices, let $n_0 = 2^p$, $p \geq 1$, be the size of designs $\mathscr{P}_0$ and $\mathscr{P}'_0$. The formula reads,

$$\ell_{\max} = \lfloor \log_2(N) - p \rfloor,$$

where $\lfloor \cdot \rfloor$ is the floor function. For the estimation of closed second-order indices, let $n_0 = pq^2$, $p \geq 1$, be the size of designs $\mathscr{P}_0$ and $\mathscr{P}'_0$. The formula reads,

$$\ell_{\max} = \lfloor N/(2q^2) + 1 - p \rfloor.$$

In practice, $p = 1$ and $q \in \{8, 11\}$ offers a good compromise between the number of layers available and the number of points contained within each layer.

ℓ_0 is used to counteract plateau effects that may occur during iterations. From experiments, $\ell_0 \in \{2, 3\}$ seems to be a good empirical prescription. Depending on the number of layers at hand, the value of ℓ_0 can be set higher.

ε is the tolerance for differences between consecutive estimates. The choice ε is the least straightforward. On the one hand, ε must be small enough to avoid poor precision. On the other hand, choosing ε too small will most certainly exhaust the budget of evaluations. In the examples hereafter (except the first one), the order of magnitude of ε is 10^{-2} for the estimation of first-order indices. Since most of the time, the number of layers available for the estimation of closed second-order indices is bigger, the order of magnitude of ε can be reduced to 10^{-3}. If the practitioner has a large budget (say 10^6 to 10^7 points), the values for ε can be further decreased ($\propto 10^{-4}$).

5.2.1 Estimation of first-order indices

A small value for $\ell_{\max}$ is selected to highlight that the iterative replication procedure can perform as well as the classical one for a restricted budget of evaluation points. The parameters of the stopping criterion (Eq. 11) are set as follows: $\varepsilon = 0.15$, $\ell_0 = 2$ and $\ell_{\max} = 9$. The size of designs $\mathscr{P}_\ell$ and $\mathscr{P}'_\ell$ at the end of the procedure can range from 2^3 up to 2^9. Note that ε is deliberately set higher than prescribed given the small number of layers and the limited budget of points.

Figure 7a shows a barplot representation of the r_K obtained. Figure 7b shows boxplots of the estimates obtained with the two replication procedures: iterative (black boxplots) and classical (grey boxplots).

The two methods give overall similar results. Hence, there is no drawback to render the replication procedure iterative for the estimation of the first-order Sobol' indices. Further-
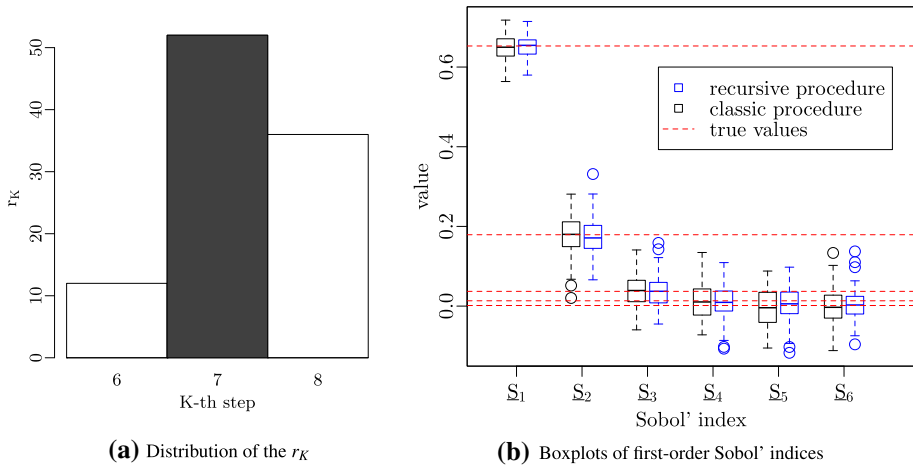
(a) Distribution of the r_K

(b) Boxplots of first-order Sobol' indices

Fig. 7 **a** Distribution of the r_K, $K = 2, \ldots, 8$ for the estimation of first-order indices. The black bar corresponds to $K = \widetilde{q}_{0.5}$. **b** (Color online) Boxplots of first-order Sobol' indices estimated $r = 100$ times with both the iterative replication procedure and the classical one. The dotted horizontal lines refer to the true values of the indices. True values of indices $\underline{S}_5$ and $\underline{S}_6$ are identical

more, Fig. 7a shows that the number of model evaluations can be decreased by adopting a sequential approach. One can calculate the number of model evaluations saved in Algorithm 1. This gain corresponds to the ratio of $\ell_{\max}$ to the iteration at which the procedure stopped. For this example, the median gain equals $9/8 = 1.125$ and the maximum gain equals $9/7 = 1.29$.

5.2.2 Estimation of closed second-order indices

The parameters of the stopping criterion are set as follows: $\varepsilon = 3 \times 10^{-3}$, $\ell_0 = 3$ and $\ell_{\max} = 100$. The initial orthogonal array A_0 used to augment designs $\mathscr{P}_\ell$ and $\mathscr{P}'_\ell$ is constructed by setting $q = 8$. The size of designs $\mathscr{P}_\ell$ and $\mathscr{P}'_\ell$ at the final step of Algorithm 1 range from 4×8^2 up to 100×8^2.

Figure 8 shows barplots representation of the r_K obtained when applying the iterative replication procedure with either the *algebraic* method or the *accept–reject* method. Results show that using the *accept–reject* method allow saving more iterations but the discrepancy is very thin.

Figure 9 gives the boxplots representation of the estimates obtained with the classical replication procedure (left boxplots) and with the iterative version, using either the *algebraic* method (middle boxplots) or the *accept–reject* method (right boxplots). Only the estimates of closed second-order indices higher than 0.1 are shown.

The main observation is that the three methods give overall similar results. Hence, the iterative replication procedure comes at no loss.

As for the case of first-order indices, one can calculate the number of model evaluations saved using the iterative replication procedure. Table 1 gives the gain at iterations $\widetilde{q}_{0.25}$, $\widetilde{q}_{0.5}$ and $\widetilde{q}_{0.75}$. Results show that the number of model evaluations can be further decreased by adopting a sequential approach for the estimation of closed second-order indices. When the input space dimension is small ($d \leq 4$), the number of blocks is rather limited. As such, the *algebraic* method should be preferred to the *accept–reject* one.
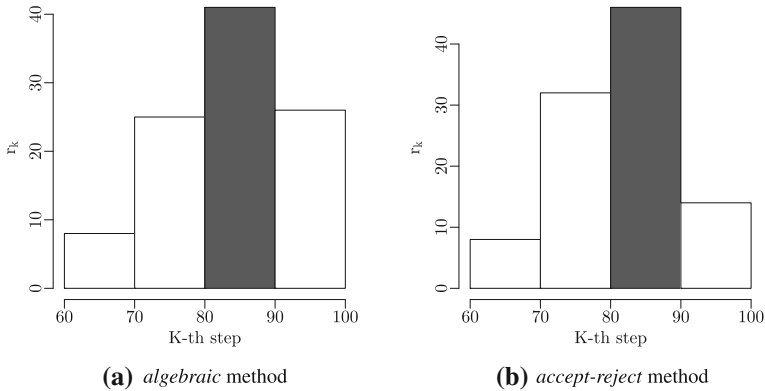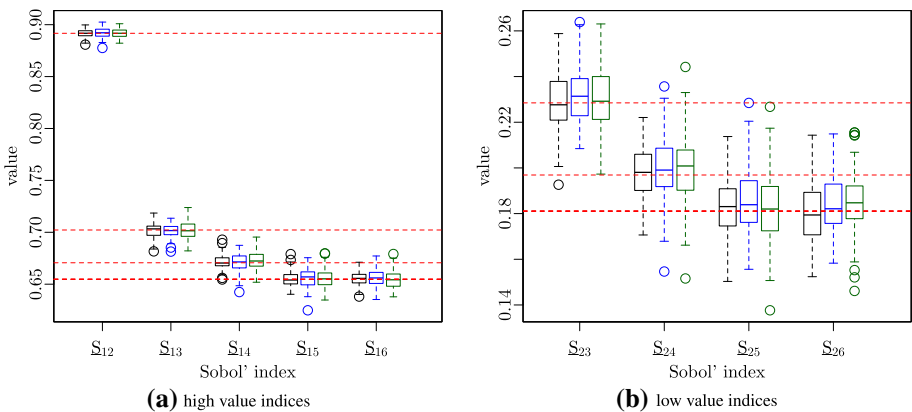
(a) *algebraic* method   **(b)** *accept-reject* method

Fig. 8 Distribution of the r_K, $K = 3, \ldots, 99$, when the iterative replication procedure is applied with either (**a**) the *algebraic* method or (**b**) the *accept–reject* method. For each graph, the black bar corresponds to $K = \widetilde{q}_{0.5}$

(a) high value indices   **(b)** low value indices

Fig. 9 (Color online) Boxplots of closed second-order Sobol' indices estimated $r = 100$ times with the iterative replication procedure and the classical one. For each index $\underline{S}_u$, the left boxplot refers to the classical procedure, the boxplot in the middle (resp. on the right) refers to the iterative procedure using the *algebraic* (resp. *accept–reject*) method. The horizontal dotted lines refer to the true values of the indices

Table 1 Gain of the iterative replication procedure using either the *algebraic* or the *accept–reject* construction, if it stops at step $\widetilde{q}_\alpha$ with $\alpha = 0.25$, 0.5 or 0.75

α	Construction	$\widetilde{q}_\alpha$	Gain$= \frac{\ell_{\max}}{q_\alpha}$
0.25	*Algebraic*	76	1.32
	Accept–reject	76	1.32
0.5	*Algebraic*	84	1.19
	Accept–reject	82	1.22
0.75	*Algebraic*	91	1.10
	Accept–reject	88	1.14

Table 2 Winter wheat dry matter model-inputs and associated distributions

Inputs	Distribution	Range of variation	Description
E_b	Uniform	[0.9, 2.8]	Radiation use efficiency
E_{imax}	Uniform	[0.9, 0.99]	Ratio max. of intercepted to incident radiation
K	Uniform	[0.6, 0.8]	Coefficient of extinction
$L_{\max}$	Uniform	[3, 12]	Maximal value of the leaf area index
A	Uniform	[0.0035, 0.01]	No physical interpretation
B	Uniform	[0.0011, 0.0025]	No physical interpretation
T_I	Uniform	[700, 1100]	Temperature threshold
C	Discrete uniform	$\{1, \ldots, 14\}$	Climate factor

5.3 Engineering application

A simple crop model will be used in this section to illustrate the iterative replication procedure. The crop model measures on a daily basis the above-ground winter wheat dry matter since sowing. The dry matter is calculated as a function of the cumulative daily temperature and the daily photosynthetically active radiation (a detailed description of the equations can be found in Monod et al. Monod et al. 2006, Section 2.2).

The quantity of interest (i.e., the output of the model) considered here is the dry matter at harvest. This output is subject to sources of uncertainties characterizing the daily climate and the ground. In the present study, the 8 input parameters listed in Table 2 are considered. The climate factor follows a discrete uniform distribution. 14 sets of annual climate data were gathered. The value of the climate factor allows to select one of the 14 climate data sets at random. The parameters A and B hold no physical interpretations.

The iterative replication procedure using the *algebraic* method is applied to estimate first-order and closed second-order Sobol' indices of the height input parameters.

For the first-order indices, The parameters of the stopping criterion (Eq. (11)) are: $\varepsilon = 0.05$, $\ell_0 = 2$ and $\ell_{\max} = 13$. The size of the two replicated designs equals 8 at first and then are doubled at each iteration up to $2^{15} = 32\,768$. Figure 10 shows values of the 8 first-order estimates along the iterations.

It can be observed that the iterative replication procedure stopped at the 10-th iteration, which implies a cost of $2 \times 4096 = 8192$ model evaluations. E_b is the most influential input trough its main effect, followed by (in order): A, B, $L_{\max}$ and E_{imax}. The remaining inputs, K and C have negligible main effects (lower than 0.01).

As the sum of the first-order estimates equals 0.88 it is interesting to estimate the closed second-order indices. To do so, the parameters of the stopping criterion (Eq. (11)) are fixed as follows: $\varepsilon = 5 \times 10^{-3}$, $\ell_0 = 3$ and $\ell_{\max} = 100$. The size of the two replicated designs equals $11^2 = 121$ at first and then, 121 points are added at each iteration up to 12 100 points. By subtracting the first second-order estimates to the closed second-order indices, a rough estimation of unclosed second-order indices can be obtained (see Eq. (5)). Figure 11 shows values of the five most influential unclosed second-order estimates along the iterations. These estimates correspond to those whose values are higher than 0.01 at the end of the procedure.

It can be observed that the iterative replication procedure stopped at the 43-th iteration, which implies a cost of $2 \times 5203 = 10\,406$ model evaluations. Five of the height inputs are influential through second-order interactions, namely: K, $L_{\max}$, A, B, C. In particular, A is influential through three different second-order interactions. The interaction effects of inputs
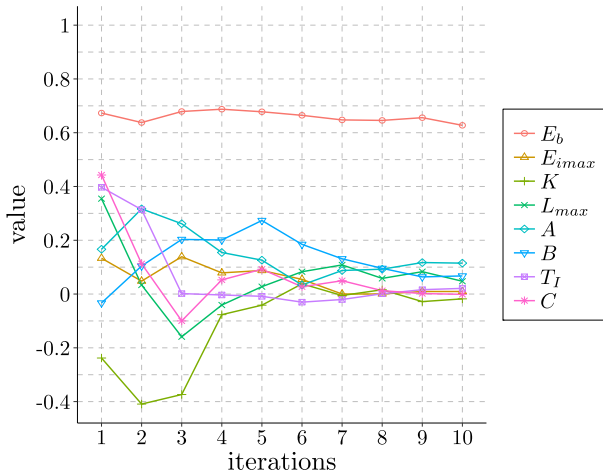
Fig. 10 (Color online) Values of the first-order Sobol' indices estimated with the iterative replication procedure, in function of the iterations
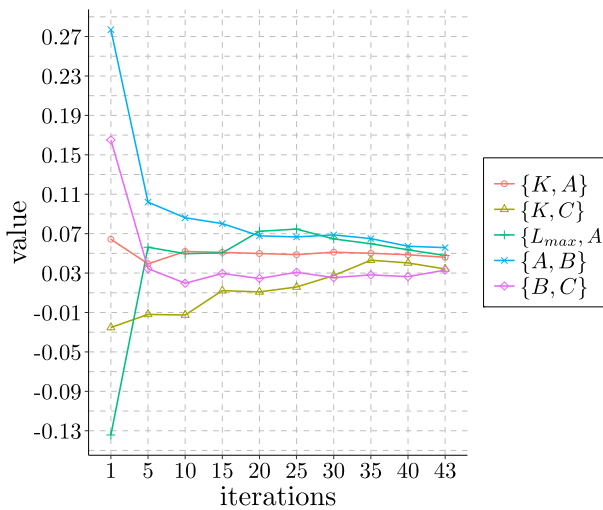

Fig. 11 (Color online) Values of the five most influential unclosed second-order Sobol' estimates obtained with the iterative replication procedure, in function of the iterations

A and B have also been highlighted in Monod et al. (2006) by means of the estimation of total effect Sobol' indices.

Overall, the main effects and second-order interactions captured with the iterative replication procedure match well (qualitatively speaking) the results of the study performed in Monod et al. (2006).

6 Conclusion

The present paper proposed a new approach rendering the replication procedure iterative to estimate first-order or closed second-order Sobol' indices. An iterative formula for the Sobol' index estimator was introduced. The iterative procedure presented consists in augmenting the two replicated designs with new sets of points through the construction of nested space-filling designs. For the case of closed second-order indices, two methods were proposed to construct a randomized nested orthogonal array of strength two: an *algebraic* method and an *accept–reject* method. The iterative replication procedure was compared to the classical replication procedure of Tissot and Prieur (2015). The comparison focused on the space-filling properties of the designs and on the precision of the Sobol' indices estimates.

The replication procedure proposed in Tissot and Prieur (2015) is known to be highly efficient in terms of number of simulations. Yet the results in this paper showed that it is possible to further decrease the number of simulations by adopting an iterative scheme. More precisely, the nested designs proposed here gave the same order of precision on Sobol' indices as the replicated designs used in Tissot and Prieur (2015) but with a random number of simulations of much smaller expectation. Furthermore, the space-filling properties of the nested designs constructed were overall as good as the one of the replicated designs used in Tissot and Prieur (2015).

For the case of first-order indices, considering Sobol' sequences could improve the nested designs (Gilquin et al. 2017). For the case of closed second-order indices, the proposed methodology could be further improved by working on the set C (Sect. 4.2, Algorithm 3). A more deterministic choice of the $g \in C$ could lead to a better exploration of the input space. A future perspective would be to define a more refined stopping criterion than the current one.

References

Bratley P, Niederreiter BL (1992) Implementation and tests of low-discrepancy sequences. ACM Trans Model Comput Simul 2(3):195–213. https://doi.org/10.1145/146382.146385

Devroye L (1986) Non-uniform random variate generation. Springer, New York

Dey A (2012) On the construction of nested orthogonal arrays. Australas J Combin 54:37–48

Franco J, Vasseur O, Corre N, Sergent M (2009) Minimum spanning tree: a new approach to assess the quality of the design of computer experiments. Chemom Intell Lab Syst 97(2):164–169. https://doi.org/10.1016/j.chemolab.2009.03.011

Gilquin L, Rugama LAJ, Arnaud E, Hickernell FJ, Monod H, Prieur C (2017) Iterative construction of replicated designs based on Sobol' sequences. C R Math 355(1):10–14

Hedayat AS, Sloane NJA, Stufken J (1999) Orthogonal arrays: theory and applications, Springer series in statistic. Springer, New York

Hoeffding W (1948) A class of statistics with asymptotically normal distributions. Ann Math Stat 19(3):293–325

Homma T, Saltelli A (1996) Importance measures in global sensitivity analysis of nonlinear models. Reliab Eng Syst Saf 52:1–17

Janon A, Klein T, Lagnoux A, Nodet M, Prieur C (2014) Asymptotic normality and efficiency of two Sobol' index estimators. ESAIM Probab Stat 18:342–364. https://doi.org/10.1051/ps/2013040

Jonshon ME, Moore LM, Ylvisaker D (1990) Minmax and maxi-min distance designs. J Stat Plan Inference 26(2):131s–148. https://doi.org/10.1016/0378-3758(90)90122-B

Mara TA, Joseph OR (2008) Comparison of some efficient met-hods to evaluate the main effect of computer model factors. J Stat Comput Simul 78(2):167–178. https://doi.org/10.1080/10629360600964454

McKay MD, Conover WJ, Beckman RJ (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21:239–245. https://doi.org/10.2307/1271432

Monod H, Naud C, Makowski D (2006) Uncertainty and sensitivity analysis for crop models. Working with dynamic crop models. Elsevier, New York, pp 55–100

Morokoff WJ, Caflisch RE (1994) Quasi-random sequences and their discrepancies. SIAM J Sci Comput 15(16):1251–1279

Prieur C, Tarantola S (2017) Variance-based sensitivity analysis: theory and estimation algorithms. Handbook of uncertainty quantification, pp 1217–1239

Qian PZG (2009) Nested Latin hypercube designs. Biometrika 96(4):957–970. https://doi.org/10.1093/biomet/asp045

Qian PZG, Ai M, Wu CFJ (2009a) Construction of nested space-filling designs. Ann Stat 37(6A):3616–3643. https://doi.org/10.1214/09-AOS690

Qian PZG, Tang B, Wu CFJ (2009b) Nested space-filling designs for computer experiments with two levels of accuracy. Stat Sin 19:287–300

Rennen G, Husslage B, Van Dam ER, Den Hertog D (2010) Nested maximin latin hypercube designs. Struct Multidiscip Optim 41(3):371–395

Sallaberry CJ, Helton JC, Hora SC (2008) Extension of latin hypercube samples with correlated variables. Reliab Eng Syst Saf 93(7):1047–1059

Saltelli A (2002) Making best use of models evaluations to compute sensitivity indices. Comput Phys Commun 145(2):280–297

Saltelli A, Chan K, Scott EM (2008) Sensitivity analysis. Wiley, New York

Sarrazin F, Pianosi F, Wagener T (2016) Global sensitivity analysis of environmental models: convergence and validation. Environ Modell Softw 79:135–152

Sheikholeslami R, Razavi S, Gupta HV, Becker W, Haghnegahdar A (2019) Global sensitivity analysis for high-dimensional problems: how to objectively group factors and measure robustness and convergence while reducing computational cost. Environ Model Softw 111:282–299

Sobol' IM (1993) Sensitivity indices for nonlinear mathematical models. Math Model Comput Exp 1:407–414

Stinson DR, Massey JL (1995) An infinite class of counterexamples to a conjecture concerning nonlinear resilient functions. J Cryptol 8:67–173. https://doi.org/10.1007/BF00202271

Terraz T, Ribes A, Fournier Y, Iooss B, Raffin B (2017) Melissa: large scale in transit sensitivity analysis avoiding intermediate files. In: Proceedings of the international conference for high performance computing, networking, storage and analysis, pp 1–14

Tissot JY, Prieur C (2015) A randomized orthogonal array-based procedure for the estimation of first- and second-order Sobol' indices. J Stat Comput Simul 85:1358–1381. https://doi.org/10.1080/00949655.2014.971799

Vorechovsky M, Novak D, Rusina R (2013) Sample size extension in stratified sampling: Theory and software implementation. In: Safety, reliability, risk and life-cycle performance of structures and infrastructures: proceedings of the 11th international conference on structural safety and reliability (ICOSSAR 2013), IASSAR. CRC Press, New York